การเรกูลาไรซ์ในตัวแบบถดถอยโลจิสติกที่ข้อมูลมีมิติสูงโดยใช้วิธีแลซโซแบบปรับได้

โดย

นายวสุรัตน์ ขำภาษี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต

สาขาวิชาสถิติประยุกต์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต

ภาควิชาสถิติ

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

ปีการศึกษา 2564

การเรกูลาไรซ์ในตัวแบบถดถอยโลจิสติกที่ข้อมูลมีมิติสูงโดยใช้วิธีแลซโซแบบปรับได้

โดย
นายวสุรัตน์ ขำภาษี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติประยุกต์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
ภาควิชาสถิติ
บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร
ปีการศึกษา 2564

REGULARIZATION IN HIGH DIMENSIONAL LOGISTIC REGRESSION MODEL BY USING ADAPTIVE LASSO METHOD

By

MR. Wasurat KHUMPASEE

A Thesis Submitted in Partial Fulfillment of the Requirements

for Master of Science (APPLIED STATISTICS)

Department of STATISTICS

Graduate School, Silpakorn University

Academic Year 2021

หัวข้อ	การเรกูลาไรซ์ในตัวแบบถดถอยโลจิสติกที่ข้อมูลมีมิติสูงโดยใช้วิธีแลซโซแบบปรับได้
โดย	นายวสุรัตน์ ขำภาษี
สาขาวิชา	สถิติประยุกต์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
อาจารย์ที่ปรึกษาหลัก	ดร. กรรณิกาณ์ หิรัญกสิ

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร ได้รับพิจารณาอนุมัติให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต

.... คณบดีบัณฑิตวิทยาลัย

(รองศาสตราจารย์ ดร.จุไรรัตน์ นันทานิช)

พิจารณาเห็นชอบโดย

.... ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. ประหยัด แสงงาม)

.... อาจารย์ที่ปรึกษาหลัก

(ดร. กรรณิกาณ์ หิรัญกสิ)

.... ผู้ทรงคุณวุฒิภายนอก

(ผู้ช่วยศาสตราจารย์ ดร. มานะชัย รอดชื่น)

61304202 : สถิติประยุกต์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
คำสำคัญ : ข้อมูลที่มีมิติสูง, ตัวแบบถดถอยโลจิสติกแบบบางเบา, การเรกูลาไรซ์, แลซโซแบบปรับได้, ค่าถ่วงน้ำหนัก

นาย วสุรัตน์ ขำภาษี: การเรกูลาไรซ์ในตัวแบบถดถอยโลจิสติกที่ข้อมูลมีมิติสูงโดยใช้วิธีแลซโซแบบปรับได้ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก : ดร. กรรณิกาณ์ หิรัญกสิ

การเรกูลาไรซ์หรือการวิเคราะห์การถดถอยโลจิสติกแบบพีนอลไลซ์เป็นวิธีที่ใช้กันอย่างแพร่หลายในการประมาณค่าพารามิเตอร์เมื่อข้อมูลมีมิติสูง งานวิจัยนี้มีวัตถุประสงค์เพื่อทำการศึกษาและเปรียบเทียบประสิทธิภาพการประมาณค่าสัมประสิทธิ์การถดถอยและการคัดเลือกตัวแปรของการเรกูลาไรซ์โดยใช้วิธีแลซโซแบบปรับได้ของตัวแบบถดถอยโลจิสติกในกรณีที่ข้อมูลมีมิติสูงแบบบางเบาทั้ง 3 วิธี ได้แก่ วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณริดจ์ (Adaptive LASSO using ridge initial weight), วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณแลซโซ (Adaptive LASSO using LASSO initial weight), และวิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณสไตน์-ริดจ์ (Adaptive LASSO using Stein-Ridge initial weight) รวมทั้งเปรียบเทียบกับวิธีแลซโซ (LASSO) ภายใต้สถานการณ์ที่มีขนาดตัวอย่างคือ 100 และ 200 มีจำนวนตัวแปรอธิบายแบบต่อเนื่องเป็นจำนวน 2, 3, และ 4 เท่าของขนาดตัวอย่าง รวมถึงจำนวนตัวแปรอธิบายแบบไม่ต่อเนื่องจำนวน 4 ตัวแปร รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแตกต่างกัน และจำนวนตัวแปรแบบต่อเนื่องที่สัมประสิทธิ์การถดถอยที่ไม่เท่ากับศูนย์ เท่ากับ 5, 10, และ 15 ตัว ในการจำลองแต่ละสถานการณ์ ทำซ้ำจำนวน 500 รอบ เกณฑ์ที่ใช้ในการทดสอบประสิทธิภาพ คือ ความถูกต้องของการทำนายจากค่าเฉลี่ยของค่าความไว ค่าความจำเพาะ และค่าพื้นที่ใต้โค้งของกราฟ Receiver Operating Characteristic (ROC) ความถูกต้องของการประมาณค่าจากค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย และความถูกต้องของการคัดเลือกตัวแปร ผลการวิจัยพบว่า วิธีที่มีประสิทธิภาพดีที่สุดทั้งการทำนาย การประมาณค่า และการคัดเลือกตัวแปร คือ วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีริดจ์ เมื่อมีตัวแปรอธิบายแบบต่อเนื่องที่สัมประสิทธิ์การถดถอยไม่เท่ากับ 0 จำนวน 5 ตัวในตัวแบบบางเบา และวิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีสไตน์-ริดจ์ เมื่อมีตัวแปรอธิบายแบบต่อเนื่องที่สัมประสิทธิ์การถดถอยไม่เท่ากับ 0 จำนวน 10 และ 15 ตัวอยู่ในตัวแบบบางเบา ในทุกกรณีของค่าพารามิเตอร์อื่น ๆ

61304202 : Major (APPLIED STATISTICS)
Keyword : High-dimensional data, Sparse Logistic Rgression Model, Regularization, Adaptive LASSO, Initial Weights

MR. WASURAT KHUMPASEE : REGULARIZATION IN HIGH DIMENSIONAL LOGISTIC REGRESSION MODEL BY USING ADAPTIVE LASSO METHOD THESIS ADVISOR : KANNIGARH HIRUNKASI, Ph.D.

Regularization or penalized logistic regression is widely used to estimate parameters for the high dimensional data. The purpose of this research was to compare the performance of three Adaptive LASSO (Least absolute shrinkage and selection operator)-types for logistic regression in high-dimensional sparse data; Adaptive LASSO using ridge initial weights, and Adaptive LASSO using LASSO initial weights and Adaptive LASSO using Stein-Ridge initial weight and also compared with LASSO under various conditions. The simulation study parameter setting was two cases of sample sizes as n=100, 200, number of quantitative predictors were 2n, 3n, and 4n and additioning 4 binary variables. There are two cases of the relationship structures between predictors and numbers of non-zero regression coefficients of quantitative predictors were 5, 10, and 15 predictors. For each condition, data was iteratively simulated 500 times. For the performance comparison, accuracy of prediction was measured by sensitivity, specificity, and area under ROC curve. The accuracy of parameter estimation was measured by of mean squared error of logistic regression coefficients estimate and variable selection performance was computed by the percentage of non-effected variables including in the model and percentage of effected predictors excluded from the model. The results showed that the Adaptive LASSO method using ridge initial weight had the best performances for all criterion when there are 5 nonzero coefficients of quantitative predictors in the sparse model and the Adaptive LASSO method using Stein-Ridge initial weight had the best performances when there are 10 and 15 nonzero coefficients of quantitative predictors in the sparse model in all cases of other parameters.

กิตติกรรมประกาศ

การดำเนินงานวิจัยและการเรียบเรียงวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้เป็นอย่างดีเนื่องจากได้รับความอนุเคราะห์จากอาจารย์กรรณิกาณ์ หิรัญกสิ ผู้เป็นอาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้กรุณาให้คำแนะนา คำปรึกษา แนวคิด องค์ความรู้ รวมถึงการตรวจทานและแก้ไขข้อบกพร่องต่าง ๆ ให้ข้าพเจ้าเป็นอย่างดี ผู้วิจัยจึงขอกราบขอบพระคุณอาจารย์เป็นอย่างสูงด้วยความซาบซึ้ง

ขอขอบคุณผู้ช่วยศาสตราจารย์ ดร.ประหยัด แสงงาม ที่กรุณาเป็นประธานกรรมการในการสอบวิทยานิพนธ์ และผู้ช่วยศาสตราจารย์ ดร.มานะชัย รอดชื่น ที่กรุณาเป็นผู้ทรงคุณวุฒิ สำหรับการให้คำแนะนำ การตรวจสอบความถูกต้อง และชี้แนะแนวทางทำให้วิทยานิพนธ์ฉบับนี้มีความสมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณคณาจารย์ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากรทุกท่านที่ได้มอบองค์ความรู้ ให้คำแนะนำ และให้กำลังใจตลอดระยะเวลาในการศึกษา รวมทั้งบุคลากรภาควิชาสถิติสายสนับสนุนทุกท่าน ไม่ว่าจะเป็นคุณนงลักษณ์ เอี๊ยวเจริญ คุณเดือนเพ็ญ เภาด้วง และคุณประจักษ์ ลำใยหวาน ที่ให้ความช่วยเหลือและอำนวยความสะดวกด้านงานเอกสารและการดำเนินงานวิจัยแก่ผู้วิจัย

ขอขอบพระคุณโครงการพัฒนากำลังคนด้านวิทยาศาสตร์ (ทุนเรียนดีวิทยาศาสตร์แห่งประเทศไทย) ที่ให้การสนับสนุนด้านทุนทรัพย์ในการศึกษาและการทำวิจัย

ขอบคุณรุ่นพี่และเพื่อน ๆ ภาควิชาสถิติที่ให้คำแนะนำ ให้ความช่วยเหลือ และเป็นกำลังใจให้การทำวิทยานิพนธ์นี้สำเร็จลุล่วงไปด้วยดี

สุดท้ายนี้ ขอกราบขอบพระคุณคุณพ่อ คุณแม่ที่เป็นผู้ปกครองของผู้ทำวิจัยเป็นอย่างสูง รวมถึงขอบคุณครอบครัวและคนสนิททุก ๆ ท่านที่สนับสนุนการศึกษาการทำวิจัยครั้งนี้ ตลอดจนให้ความรัก กำลังใจ แรงผลักดันและแรงสนับสนุนในทุกด้านแก่ผู้ทำวิจัย

นาย วสุรัตน์ ขำภาษี

สารบัญ

หน้า

สารบัญตาราง

หน้า

สารบัญภาพ

หน้า

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การวิเคราะห์การถดถอยโลจิสติกเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอธิบาย (explanatory variables) และตัวแปรตอบสนอง (response variables) ในลักษณะที่ตัวแปรตอบสนองนั้นเป็นตัวแปรแบบจำแนกประเภทหรือเป็นตัวแปรเชิงกลุ่ม โดยที่อาจจะแบ่งเป็นสองกลุ่มหรือมากกว่าสองกลุ่มก็ได้ ตัวแบบการถดถอยโลจิสติก (logistic regression model) นั้น เป็นตัวแบบที่มีการนำไปใช้อย่างแพร่หลายในสาขาวิชาต่าง ๆ เช่น แพทยศาสตร์ สาธารณสุขศาสตร์ พันธุศาสตร์ สังคมศาสตร์ และด้านอื่น ๆ (Agresti, 2007)

โดยทั่วไป การถดถอยโลจิสติก จะประมาณค่าพารามิเตอร์ในตัวแบบด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood: ML) ร่วมกับระเบียบวิธีเชิงตัวเลข (numerical method) โดยที่จำนวนขนาดตัวอย่าง (n) ต้องมากกว่าจำนวนตัวแปรอธิบาย (p) หรือ $n > p$ และตัวแปรอธิบายในตัวแบบจะต้องไม่มีความสัมพันธ์เชิงเส้นพหุ (multicollinearity) เพื่อให้ได้ตัวแบบที่มีประสิทธิภาพและมีความแม่นยำในการทำนายสูงสุด อย่างไรก็ตาม ในงานวิจัยบางสาขาวิชา เช่น งานวิจัยทางด้านชีววิทยาเกี่ยวกับอัตราการผลิต riboflavin หรือวิตามิน B2 ของแบคทีเรีย *Bacillus subtilis* ซึ่งมีจำนวนตัวแปรอธิบายที่เป็นตัวแปรของระดับยีนต่าง ๆ ที่พบในแบคทีเรียเป็นจำนวนมาก ในขณะที่หน่วยตัวอย่างที่สามารถเก็บรวบรวมข้อมูลของระดับยีนของทุกตัวแปรมีจำนวนน้อยกว่า (Bühlmann, 2014) ซึ่งงานวิจัยในกรณีนี้จะพบว่าข้อมูลที่ใช้ในการวิจัย มีลักษณะเป็นข้อมูลที่มีมิติสูง (high dimensional data) ซึ่งเป็นข้อมูลที่มีขนาดตัวอย่าง (n) น้อยกว่าจำนวนตัวแปรอธิบาย (p) หรือ $n < p$ นอกจากนี้ การมีจำนวนตัวแปรอธิบายเป็นจำนวนมากนั้น มีโอกาสที่จะทำให้ตัวแปรอธิบายหลาย ๆ ตัว อาจไม่มีความสัมพันธ์กับตัวแปรตอบสนอง และค่าสัมประสิทธิ์การถดถอยของตัวแปรอธิบายส่วนใหญ่เป็นศูนย์ ซึ่งในกรณีที่ตัวแบบถดถอยโลจิสติกที่มีตัวแปรอธิบายที่ไม่มีอิทธิพลต่อตัวแปรตอบสนองเป็นจำนวนมาก จะเรียกตัวแบบนี้ว่า ตัวแบบบางเบา (sparse model) เนื่องจากการวิเคราะห์การถดถอยโลจิสติกในกรณีที่ข้อมูลมีมิติสูง ไม่สามารถหาตัวประมาณพารามิเตอร์ด้วยวิธี Maximum Likelihood ด้วยวิธีการดั้งเดิมได้ จึงได้มีการศึกษาวิจัยเกี่ยวกับวิธีการประมาณค่าพารามิเตอร์ของตัวแบบโลจิสติกในกรณีที่ข้อมูลมีมิติสูงด้วยการเรกูลาไรซ์ (regularization) หรือเรียกว่า การถดถอยโลจิสติกแบบพีนอลไลซ์ (penalized logistic regression) ซึ่งเป็นวิธีการหาตัว

ประมาณที่มีความเอนเอียง (bias) ที่เพิ่มเข้าไปเพื่อลดสัดส่วนของความแปรปรวนลง และทำให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ (Mean Square Error: MSE) มีค่าลดลง โดยการเพิ่มฟังก์ชันพีนอลตี้ (penalty function) ซึ่งเป็นฟังก์ชันของพารามิเตอร์เข้าไปที่ฟังก์ชันภาวะความน่าจะเป็น (likelihood function) ของพารามิเตอร์ แล้วจึงหาตัวประมาณพารามิเตอร์ที่ทำให้ฟังก์ชันดังกล่าวมีค่าสูงสุด

Hoerl (1970) ได้เสนอการถดถอยริดจ์ (ridge regression) ซึ่งเป็นการถดถอยพีนอลไลซ์ที่มีฟังก์ชันพีนอลตี้คือ ผลรวมของค่ากำลังสองของพารามิเตอร์ โดยตัวประมาณพารามิเตอร์ด้วยวิธีริดจ์จะเป็นค่าประมาณสัมประสิทธิ์การถดถอยของทุกตัวแปรอธิบายถูกทำให้มีค่าเล็กลง (shrinkage) แต่ไม่เท่ากับ 0 ดังนั้น ตัวประมาณจากวิธีริดจ์จึงจะมีตัวแปรอธิบายทุกตัวอยู่ในตัวแบบ โดยไม่มีการคัดเลือกตัวแปรออกจากตัวแบบ ซึ่งอาจจะทำให้แปลผลได้ยากในกรณีที่ข้อมูลมีมิติสูง ซึ่งก่อนหน้านี้ Stein (1956) และ James (1961) ได้เสนอตัวประมาณสไตน์ (Stein estimator) ที่เป็นตัวประมาณที่ปรับค่าพารามิเตอร์ให้มีค่าเล็กลง ตามค่าความแปรปรวนของตัวประมาณ Maximum Likelihood โดยตัวประมาณพารามิเตอร์ทั้งสองวิธีมีค่าความคลาดเคลื่อนเฉลี่ยกำลังสองของค่าประมาณสัมประสิทธิ์ถดถอยต่ำกว่าตัวประมาณวิธี Maximum Likelihood ในกรณีที่ข้อมูลมีปัญหาความสัมพันธ์เชิงเส้นพหุ

Tibshirani (1996) ได้เสนอวิธีการเรกูลาไรซ์ของการถดถอยเชิงเส้นพหุคูณ เรียกว่าแลซโซ (LASSO) ซึ่งเป็นคำย่อมาจาก Least Absolute Shrinkage and Selection Operator เป็นวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นที่นำลักษณะที่ดีของวิธีริดจ์ และวิธีการคัดเลือกตัวแปร (Subset selection) มารวมกัน โดยใช้ฟังก์ชันพีนอลตี้คือ ผลรวมของค่าสัมบูรณ์ของค่าพารามิเตอร์ ต่อมา Van de Geer (2008) ได้นำวิธีการ LASSO นี้ไปใช้ในตัวแบบเชิงเส้นทั่วไปที่ข้อมูลมีมิติสูง และ Zou (2006) ได้ทำการพัฒนาวิธีการ LASSO ให้ดีขึ้น เนื่องจากเห็นว่าวิธีการคัดเลือกตัวแปรอธิบายของวิธี LASSO นั้นไม่คงเส้นคงวา และได้เสนอวิธีการประมาณค่าใหม่คือวิธีแลซโซแบบปรับได้ (Adaptive LASSO) ในตัวแบบถดถอยเชิงเส้นพหุคูณ โดยการเพิ่มเวกเตอร์ค่าถ่วงน้ำหนัก (weights) เข้าไปในส่วนของฟังก์ชันพีนอลตี้ของวิธี LASSO ซึ่งทำให้ตัวแปรอธิบายที่ถูกคัดเลือกเข้าสู่ตัวแบบมีค่าถ่วงน้ำหนักที่แตกต่างกันตามความสำคัญของตัวแปรอธิบาย และ Zou ได้แสดงให้เห็นว่าวิธีการ Adaptive LASSO นี้มีความคงเส้นคงวาในการคัดเลือกตัวแปรอธิบาย การประยุกต์ใช้วิธีการ Adaptive LASSO สำหรับการวิเคราะห์การถดถอยแบบโลจิสติกที่ข้อมูลมีมิติสูงได้นำไปใช้ในงานวิจัยต่าง ๆ เช่น งานวิจัยของ Algamal และ Lee (2015) ที่ได้นำวิธี Adaptive LASSO ไปใช้ในการ

คัดเลือกยีนของการคัดแยกโรคมะเร็งในข้อมูลที่มีมิติสูง, งานวิจัยของ Hashemian และคณะ (2017) ที่ได้นำวิธี Adaptive LASSO ไปใช้ในการวิเคราะห์การแสดงออกของยีนของผู้ป่วยมะเร็งต่อมลูกหมากด้วยวิธีการถดถอยแบบโลจิสติกที่ข้อมูลมีมิติสูง เป็นต้น

การประมาณค่าสัมประสิทธิ์ถดถอยด้วยวิธี Adaptive LASSO ข้างต้น สำหรับตัวแบบถดถอยเชิงเส้นพหุคูณ โดยทั่วไปจะใช้ฟังก์ชันของตัวประมาณจากวิธีกำลังสองน้อยที่สุด (Least Square Method) เป็นเวกเตอร์ถ่วงน้ำหนัก ส่วนตัวแบบถดถอยโลจิสติกพหุคูณ จะใช้ฟังก์ชันของตัวประมาณภาวะน่าจะเป็นสูงสุด (maximum likelihood estimator: MLE) เป็นเวกเตอร์ถ่วงน้ำหนัก แต่เนื่องจากในกรณีที่ข้อมูลมีมิติสูง จะไม่สามารถหาตัวประมาณกำลังสองน้อยที่สุดหรือตัวประมาณ MLE ได้ จึงต้องใช้ตัวประมาณจากวิธีการอื่น ๆ เพื่อนำมาใช้ในการคำนวณเวกเตอร์ค่าถ่วงน้ำหนักสำหรับวิธี Adaptive LASSO ต่อมา Guo และคณะ (Guo et al., 2015) ได้นำเสนอวิธีการที่พัฒนามาจาก LASSO และ Adaptive LASSO คือวิธีการแลซโซแบบผสมสองขั้นตอน (Two-stage Hybrid) ซึ่งมีวิธีการคือ ในขั้นตอนแรก จะใช้วิธี LASSO ในการหาตัวประมาณสัมประสิทธิ์ถดถอยขั้นต้น (initial estimator) และลดจำนวนมิติของตัวแบบให้มีขนาดเล็กลงด้วยการพิจารณาจากค่าสัมประสิทธิ์นั้นว่าเป็น 0 หรือไม่ จากนั้นจะทำการคำนวณค่าถ่วงน้ำหนักของตัวแปรที่มีนัยสำคัญและแทนค่าลงในฟังก์ชันพีนอลตี้ จากนั้นก็ใช้วิธี Adaptive LASSO เพื่อประมาณค่าสัมประสิทธิ์ถดถอยและพิจารณาเลือกตัวแปรอธิบายที่สำคัญเข้าสู่ตัวแบบอีกครั้ง จึงได้เป็นตัวแบบที่มีตัวแปรที่สำคัญที่มีผลต่อตัวแปรตอบสนองเท่านั้น และได้ทำการเปรียบเทียบประสิทธิภาพของวิธีดังกล่าวกับวิธีต่าง ๆ หลายวิธีรวมทั้ง LASSO และ Adaptive LASSO พบว่า วิธี Two-stage Hybrid นี้มีประสิทธิภาพโดยรวมที่ดีกว่าทุก ๆ วิธี

อย่างไรก็ตาม ทางผู้วิจัยมีความสนใจที่จะนำตัวประมาณจากวิธีอื่น ๆ มาใช้ในการคำนวณเวกเตอร์ค่าถ่วงน้ำหนักสำหรับตัวประมาณวิธี Adaptive LASSO นอกจากวิธี LASSO จึงได้เลือกตัวประมาณจากวิธีริดจ์ และตัวประมาณจากวิธีสไตน์มาใช้ในการคำนวณเวกเตอร์ค่าถ่วงน้ำหนักด้วย แต่เนื่องจากตัวประมาณสไตน์ เป็นตัวประมาณที่ใช้ตัวประมาณวิธี Maximum Likelihood ในการคำนวณ ซึ่งไม่สามารถหาได้ในกรณีที่ข้อมูลมีมิติสูง ผู้วิจัยจึงประยุกต์ใช้วิธีของตัวประมาณสไตน์มาใช้กับตัวประมาณจากวิธีริดจ์แทนตัวประมาณภาวะน่าจะเป็นสูงสุด และใช้ชื่อเรียกตัวประมาณนี้ว่าตัวประมาณวิธีสไตน์-ริดจ์ (Stein-Ridge estimator) ซึ่งตัวประมาณทั้งสองตัวก็จะเป็นตัวเลือกที่จะนำมาใช้กับฟังก์ชันถ่วงน้ำหนักสำหรับการประมาณค่าพารามิเตอร์และการคัดเลือกตัวแปรด้วยวิธี Adaptive LASSO ข้างต้น

ดังนั้นงานวิจัยนี้จึงสนใจที่จะศึกษาและเปรียบเทียบประสิทธิภาพการประมาณค่าสัมประสิทธิ์การถดถอยและการคัดเลือกตัวแปรของการถดถอยพีนอลไลซ์โดยใช้วิธีแลซโซแบบปรับได้ของตัวแบบถดถอยโลจิสติกในกรณีที่ข้อมูลมีมิติสูงแบบบางเบาทั้ง 3 วิธี ได้แก่ วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณริดจ์ (Adaptive LASSO using ridge initial weight), วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณแลซโซ (Adaptive LASSO using LASSO initial weight), และวิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณสไตน์-ริดจ์ (Adaptive LASSO using Stein-Ridge initial weight) รวมทั้งเปรียบเทียบกับวิธีแลซโซ ภายใต้สถานการณ์ของเงื่อนไขที่แตกต่างกันครอบคลุมข้อจำกัดคุณสมบัติของตัวประมาณแต่ละวิธี

1.2 วัตถุประสงค์การทำวิจัย

1. เพื่อศึกษาวิธีการประมาณค่าสัมประสิทธิ์ถดถอยและการคัดเลือกตัวแปรด้วยวิธีแลซโซแบบปรับได้โดยใช้ตัวประมาณที่แตกต่างกันในการคำนวณค่าถ่วงน้ำหนัก ในตัวแบบถดถอยโลจิสติกในกรณีที่ข้อมูลมีมิติสูงแบบบางเบา

2. เพื่อเปรียบเทียบประสิทธิภาพการประมาณค่าสัมประสิทธิ์การถดถอยและการคัดเลือกตัวแปรของการเรกูลาไรซ์โดยใช้วิธีแลซโซและวิธีแลซโซแบบปรับได้ของตัวแบบถดถอยโลจิสติกในกรณีที่ข้อมูลมีมิติสูงทั้ง 4 วิธี ได้แก่ วิธีแลซโซ, วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณริดจ์, วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณแลซโซ, และวิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณสไตน์-ริดจ์ โดยพิจารณาจากความถูกต้องในการทำนาย ความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอย และความถูกต้องในการคัดเลือกตัวแปรอธิบายเข้าสู่ตัวแบบ

1.3 ขอบเขตของงานวิจัย

1. ขนาดตัวอย่าง n = 100 และ 200

2. จำนวนตัวแปรอธิบายแบบต่อเนื่อง p ตัวแปรมีจำนวนเป็น 2, 3, และ 4 เท่าของขนาดตัวอย่าง n และจำนวนตัวแปรอธิบายแบบไม่ต่อเนื่องมีจำนวน 4 ตัวแปรทุกกรณี รวมจำนวนตัวแปรอธิบายทั้งสิ้น $p+4$ ตัวแปร

3. เมทริกซ์ของตัวแปรอธิบาย $(\boldsymbol{X}) = (\boldsymbol{X}^*_{n\times p},\ \boldsymbol{D}_{n\times 4})$ มีขนาด $n \times (p+4)$ โดยที่ $\boldsymbol{X}^* = (X_1, X_2, \ldots, X_p)$ เป็นเมทริกซ์ของตัวแปรอธิบายที่เป็นตัวแปรแบบต่อเนื่องจำนวน p ตัวแปร และ $\boldsymbol{D} = (D_1, D_2, D_3, D_4)$ เป็นเมทริกซ์ของตัวแปรแบบไม่ต่อเนื่องจำนวน 4 ตัวแปรที่มีการแจกแจงแบบเบอร์นูลลีที่มีพารามิเตอร์เป็น 0.8, 0.6, 0.4, และ 0.2 ตามลำดับ ผู้วิจัยจำลองข้อมูล $\boldsymbol{X}^*$ ให้

มีการแจกแจงแบบปรกติหลายตัวแปรที่เวกเตอร์ค่าเฉลี่ยเป็นเวกเตอร์ $\mathbf{0}$ และเมทริกซ์ความแปรปรวนร่วม $\mathbf{\Sigma}$ ที่แบ่งเป็น 2 กรณี คือ

กรณีที่ 1 ตัวแปรอธิบายจะถูกแบ่งเป็น 2 กลุ่มที่เป็นอิสระต่อกัน แต่มีความสัมพันธ์กันภายในกลุ่ม โดยที่ในกลุ่มที่ 1 จะมีจำนวน 15 ตัวแปรที่สัมพันธ์กันเอง และกลุ่มที่ 2 คือจำนวนตัวแปรที่เหลือมีความสัมพันธ์กันเอง นั่นคือ เมทริกซ์ของตัวแปรอธิบายแบบต่อเนื่องกรณีที่ 1 ใช้สัญลักษณ์แทนด้วย $\boldsymbol{X}^*_{\boldsymbol{ind}} = \left[\boldsymbol{X}_{\mathbf{1}_{(\boldsymbol{n}\times\mathbf{15})}} \vdots \boldsymbol{X}_{\mathbf{2}_{(\boldsymbol{n}\times(\boldsymbol{p}-\mathbf{15}))}}\right]$ โดยที่ $\boldsymbol{X}^*_{\boldsymbol{ind}} \sim N_p(\mathbf{0}, \mathbf{\Sigma}_{\boldsymbol{a}})$ ซึ่งเมทริกซ์ความแปรปรวนร่วมของตัวแปรอธิบาย $\boldsymbol{X}^*_{\boldsymbol{ind}}$ สำหรับกรณีที่ 1 คือ

$$\mathbf{\Sigma}_{\boldsymbol{a}} = \begin{bmatrix} \mathbf{\Sigma_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma_2} \end{bmatrix}_{p\times p}$$

โดยที่ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรตัวที่ X_k และ X_j ของเมทริกซ์ $\boldsymbol{X_1}$ คือ $\rho^{|k-j|}$; $j,k = 1,2,\dots,15$ และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรตัวที่ X_k และ X_j ของเมทริกซ์ $\boldsymbol{X_2}$ คือ $\rho^{|k-j|}$; $j,k = 1,2,\dots,(p-15)$ ดังนี้

$$\mathbf{\Sigma_1} = \mathbf{Cov}(\boldsymbol{X_1}) = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{14} \\ \rho & 1 & \rho & \cdots & \rho^{13} \\ \rho^2 & \rho & 1 & \cdots & \rho^{12} \\ \vdots & \vdots & \vdots & \ddots & \rho \\ \rho^{14} & \rho^{13} & \cdots & \rho & 1 \end{bmatrix}_{15\times 15} \quad \text{และ}$$

$$\mathbf{\Sigma_2} = \mathbf{Cov}(\boldsymbol{X_2}) = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{p-14} \\ \rho & 1 & \rho & \cdots & \rho^{p-13} \\ \rho^2 & \rho & 1 & \cdots & \rho^{p-12} \\ \vdots & \vdots & \vdots & \ddots & \rho \\ \rho^{p-14} & \rho^{p-13} & \cdots & \rho & 1 \end{bmatrix}_{(p-15)\times(p-15)}$$

กรณีที่ 2 ตัวแปรอธิบาย $X_1,\dots,X_p$ มีความสัมพันธ์กัน โดยที่เมทริกซ์ของตัวแปรอธิบายแบบต่อเนื่องกรณีที่ 2 จะใช้สัญลักษณ์แทนด้วย $\boldsymbol{X}^*_{\boldsymbol{cor}(p\times p)}$ โดยที่ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง X_k และ X_j คือ $\rho^{|k-j|}$; $j,k = 1,2,\dots,p$ ดังนั้น $\boldsymbol{X}^*_{\boldsymbol{cor}} \sim N_p(\mathbf{0}, \mathbf{\Sigma}_{\boldsymbol{b}})$ เมื่อเมทริกซ์ความแปรปรวนร่วม คือ

$$\mathbf{\Sigma}_{\boldsymbol{b}} = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{p-1} \\ \rho & 1 & \rho & \cdots & \rho^{p-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{p-3} \\ \vdots & \vdots & \vdots & \ddots & \rho \\ \rho^{p-1} & \rho^{p-2} & \cdots & \rho & 1 \end{bmatrix}_{p\times p}$$

4. กำหนดเวกเตอร์ของสัมประสิทธิ์การถดถอย $\boldsymbol{\beta}_{(p+4)\times 1} = (\boldsymbol{\beta}_{\boldsymbol{X}^*}, \boldsymbol{\beta}_{\boldsymbol{D}})$ จากตัวแบบการถดถอยโลจิสติกพหุคูณที่เป็นตัวแบบบางเบา ในแต่ละรูปแบบดังนี้

รูปแบบที่ 1 กำหนดให้มีตัวแปรจำนวน 9 ตัวแปรที่มีค่าพารามิเตอร์ที่ไม่เป็น 0 แบ่งเป็นตัวแปรแบบต่อเนื่อง 5 ตัวแปร คือ $X_1, X_2, \dots, X_5$ และตัวแปรแบบไม่ต่อเนื่อง 4 ตัวแปร คือ D_1, D_2, D_3, D_4 โดยที่ค่าพารามิเตอร์ที่ไม่เป็น 0 ของแต่ละตัวแปรเป็นดังนี้

$\boldsymbol{\beta}_{\boldsymbol{X}^*}^{(1)} = [2.5, 2.5, 2.5, 2.5, 2.5, 0, 0, \dots, 0]$ และ

$\boldsymbol{\beta}_{\boldsymbol{D}}^{(1)} = [-0.1, -0.1, -0.1, -0.1]$ ตามลำดับ

รูปแบบที่ 2 กำหนดให้มีตัวแปรจำนวน 14 ตัวแปรที่มีค่าพารามิเตอร์ที่ไม่เป็น 0 แบ่งเป็นตัวแปรแบบต่อเนื่อง 10 ตัวแปร คือ $X_1, X_2, \dots, X_{10}$ และตัวแปรแบบไม่ต่อเนื่อง 4 ตัวแปร คือ D_1, D_2, D_3, D_4 โดยที่ค่าพารามิเตอร์ที่ไม่เป็น 0 ของแต่ละตัวแปรเป็นดังนี้

$\boldsymbol{\beta}_{\boldsymbol{X}^*}^{(2)} = [2.5, 2.5, 2.5, 2.5, 2.5, 2, 2, 2, 2, 2, 0, 0, \dots, 0]$ และ

$\boldsymbol{\beta}_{\boldsymbol{D}}^{(2)} = [-0.1, -0.1, -0.1, -0.1]$ ตามลำดับ

รูปแบบที่ 3 กำหนดให้มีตัวแปรจำนวน 19 ตัวแปรที่มีค่าพารามิเตอร์ที่ไม่เป็น 0 แบ่งเป็นตัวแปรแบบต่อเนื่องอีก 15 ตัวแปร คือ $X_1, X_2, \dots, X_{15}$ และตัวแปรแบบไม่ต่อเนื่อง 4 ตัวแปร คือ D_1, D_2, D_3, D_4 โดยที่ค่าพารามิเตอร์ที่ไม่เป็น 0 ของแต่ละตัวแปรเป็นดังนี้

$\boldsymbol{\beta}_{\boldsymbol{X}^*}^{(3)} = [2.5, 2.5, 2.5, 2.5, 2.5, 2, 2, 2, 2, 2, 1.5, 1.5, 1.5, 1.5, 1.5, 0, 0, \dots, 0]$ และ

$\boldsymbol{\beta}_{\boldsymbol{D}}^{(3)} = [-0.1, -0.1, -0.1, -0.1]$ ตามลำดับ

5. เวกเตอร์ตัวแปรตามที่เป็นตัวแปรทวิภาค ซึ่งมี 2 กลุ่มคือ 0 และ 1 จะสร้างจากการคำนวณความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจจากตัวแบบการถดถอยโลจิสติกพหุคูณตามเมทริกซ์ของตัวแปรอธิบายและเวกเตอร์ของค่าสัมประสิทธิ์ถดถอยที่ได้กำหนดไว้ข้างต้น

$$\pi_i(\boldsymbol{x}) = P(y_i = 1|\boldsymbol{X_i}) = \frac{\exp(\beta_0 + \boldsymbol{X}_i'\boldsymbol{\beta})}{\mathbf{1} + \exp(\beta_0 + \boldsymbol{X}_i'\boldsymbol{\beta})} \;;\; i = 1, 2, \dots, n$$

โดยที่กำหนดให้ $\beta_0 =$ -0.5 และเมื่อได้ค่าความน่าจะเป็นมาแล้วนั้น จะกำหนดให้ค่านั้นเป็น 1 เมื่อค่าความน่าจะเป็นที่คำนวณได้มีค่ามากกว่า 0.5 และเป็น 0 เมื่อค่าความน่าจะเป็นที่คำนวณได้มีค่าน้อยกว่า 0.5

1.4 นิยามศัพท์เฉพาะ

ประสิทธิภาพของตัวประมาณ (Performance of estimator) คือ การวัดสมบัติของตัวประมาณสัมประสิทธิ์การถดถอยใน 3 ด้าน คือ ความถูกต้องในการทำนาย ความถูกต้องของการประมาณค่าพารามิเตอร์สัมประสิทธิ์การถดถอย และความถูกต้องในการคัดเลือกตัวแปรอธิบายเข้าในตัวแบบ

ข้อมูลมิติสูง (High-Dimensional data) คือ ข้อมูลที่มีจำนวนตัวแปรอธิบายเป็นจำนวนมาก และมีจำนวนมากกว่าขนาดตัวอย่าง

การเรกูลาไรซ์ (Regularization) คือ การแก้ปัญหาการประมาณค่าพารามิเตอร์ในตัวแบบถดถอยที่มีปัญหาให้มีความปกติ และทำให้ผลรวมกำลังสองของความคลาดเคลื่อนสุ่มมีค่าต่ำสุดภายใต้ข้อจำกัดของตัวประมาณพารามิเตอร์

ตัวแบบถดถอยบางเบา (Sparse Model) คือ ตัวแบบการถดถอยที่มีตัวแปรอธิบายเป็นจำนวนมาก ซึ่งตัวแปรอธิบายดังกล่าวจะประกอบด้วยตัวแปรที่มีค่าสัมประสิทธิ์การถดถอยเป็นศูนย์เป็นส่วนมาก และมีตัวแปรเพียงบางส่วนที่มีค่าสัมประสิทธิ์การถดถอยที่ไม่เป็นศูนย์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

เพื่อเป็นแนวทางในการตัดสินใจเลือกวิธีการประมาณค่าสัมประสิทธิ์การถดถอยโลจิสติกให้เหมาะสม และสามารถนำไปประยุกต์ใช้กับการวิเคราะห์การถดถอยโลจิสติกกับข้อมูลจริงที่มีลักษณะเป็นข้อมูลมิติสูงแบบบางเบาได้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพการประมาณค่าสัมประสิทธิ์การถดถอยและการคัดเลือกตัวแปรของการเรกูลาไรซ์โดยใช้วิธีแลซโซและวิธีแลซโซแบบปรับได้ของตัวแบบถดถอยโลจิสติกในกรณีที่ข้อมูลมีมิติสูงทั้ง 4 วิธี ได้แก่ วิธีแลซโซ, วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณริดจ์, วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณแลซโซ, และวิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณสไตน์-ริดจ์ โดยในการเปรียบเทียบประสิทธิภาพของแต่ละวิธีนั้น จะพิจารณาจากเกณฑ์ต่าง ๆ คือ ความถูกต้องในการทำนาย ความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอย และความถูกต้องในการคัดเลือกตัวแปรอธิบายเข้าในตัวแบบ ในบทนี้จะแบ่งออกเป็น 2 ส่วน คือ ส่วนแรก คือ ทฤษฎีที่เกี่ยวข้อง ได้แก่ การวิเคราะห์การถดถอยโลจิสติก, การประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด, วิธีการวิเคราะห์การถดถอยแบบริดจ์, วิธีการวิเคราะห์การถดถอยแบบ LASSO, วิธีการวิเคราะห์การถดถอยแบบ Adaptive LASSO, วิธีการวิเคราะห์การถดถอยแบบ Adaptive LASSO ที่มีค่าถ่วงน้ำหนักคำนวณจากตัวประมาณแต่ละวิธี และส่วนที่สอง คือ งานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 การถดถอยโลจิสติกทวิภาค (Binary Logistic Regression)

การถดถอยโลจิสติกทวิภาค (Binary Logistic Regression) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอธิบาย ซึ่งในที่นี้จะแทนด้วย $\boldsymbol{X}$ และตัวแปรตอบสนอง แทนด้วย Y โดยที่ตัวแปรตอบสนองนั้นจะเป็นตัวแปรแบบจำแนกประเภทหรือเป็นตัวแปรเชิงกลุ่มที่แบ่งเป็นสองกลุ่ม คือ กลุ่มที่สนใจ $(Y = 1)$ และ กลุ่มที่ไม่สนใจ $(Y = 0)$ โดยที่ค่าสังเกตของตัวแปร Y ขนาด n หน่วย ที่เป็นอิสระต่อกัน เนื่องจากตัวแปรตอบสนอง Y มีค่าที่เป็นไปได้ 2 ค่า ดังนั้นตัวแปรตอบสนองดังกล่าวจึงมีฟังก์ชันการแจกแจงแบบเบอร์นูลลี (Bernoulli Distribution) โดยกำหนดให้ความน่าจะเป็นที่ตัวแปรตอบสนองจะอยู่ในกลุ่มที่สนใจ หรือ $P(Y = 1) = \pi_i$ และความน่าจะเป็นที่ตัวแปรตอบสนองจะอยู่ในกลุ่มที่ไม่สนใจ หรือ $P(Y = 0) = 1 - \pi_i$ โดยที่ $i = 1,2,\dots,n$ จะได้รูปแบบของฟังก์ชันมวลความน่าจะเป็น (Probability mass function) คือ $f(y_i;\pi_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i}$ เมื่อ $y_i = 0, 1$

ตัวแบบการถดถอยโลจิสติกทวิภาคพหุคูณ คือ ตัวแบบการถดถอยโลจิสติกที่มีตัวแปรอธิบาย X จำนวน m ตัวแปร แสดงได้ดังฟังก์ชันต่อไปนี้

$$\pi_i = \pi(x_{i1}, x_{i2}, \dots, x_{im}) = P(Y = 1|\boldsymbol{X} = \boldsymbol{X_i}) = \frac{e^{(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_m X_m)}}{1+e^{(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_m X_m)}} \quad (1)$$

$$= \frac{e^{\left(\beta_0+\sum_{j=1}^{m}\beta_j X_j\right)}}{1+e^{\left(\beta_0+\sum_{j=1}^{m}\beta_j X_j\right)}}$$

โดยที่ $\beta_0, \beta_1, \beta_2, \dots, \beta_m$ คือสัมประสิทธิ์การถดถอยของตัวแบบและ $X_1, X_2, \dots, X_m$ คือตัวแปรอธิบาย ซึ่งตัวแบบการถดถอยโลจิสติกทวิภาคพหุคูณสามารถเขียนอยู่ในรูปเวกเตอร์ได้เป็น

$$\pi_i = \frac{e^{(\beta_0+\boldsymbol{X_i'\beta})}}{1+e^{(\beta_0+\boldsymbol{X_i'\beta})}} \quad ; \quad i = 1,2,\dots,n \qquad (2)$$

เพื่อให้มีลักษณะใกล้เคียงกันกับตัวแบบการถดถอยแบบเชิงเส้น (Linear Regression Model) จึงได้มีการแปลงสมการที่ (1) โดยใช้การแปลงโลจิต (Logit transformation) ซึ่งมีขั้นตอนดังนี้

$$\pi(x) = P(Y = 1|X = x) = \frac{e^{\left(\beta_0+\sum_{j=1}^{m}\beta_j X_j\right)}}{1+e^{\left(\beta_0+\sum_{j=1}^{m}\beta_j X_j\right)}}$$

$$\pi(x)[1 + e^{\left(\beta_0+\sum_{j=1}^{m}\beta_j X_j\right)}] = e^{\left(\beta_0+\sum_{j=1}^{m}\beta_j X_j\right)}$$

$$\pi(x) + \pi(x)e^{\left(\beta_0+\sum_{j=1}^{m}\beta_j X_j\right)} = e^{\left(\beta_0+\sum_{j=1}^{m}\beta_j X_j\right)}$$

$$\pi(x) = e^{\left(\beta_0+\sum_{j=1}^{m}\beta_j X_j\right)} - \pi(x)e^{\left(\beta_0+\sum_{j=1}^{m}\beta_j X_j\right)}$$

$$\pi(x) = e^{\left(\beta_0+\sum_{j=1}^{m}\beta_j X_j\right)}[1 - \pi(x)]$$

$$\frac{\pi(x)}{1 - \pi(x)} = e^{\left(\beta_0+\sum_{j=1}^{m}\beta_j X_j\right)}$$

$$e^{\ln\left(\frac{\pi(x)}{1-\pi(x)}\right)} = e^{\left(\beta_0+\sum_{j=1}^{m}\beta_j X_j\right)}$$

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m \qquad (3)$$

ดังนั้น ตัวแบบการถดถอยโลจิสติกที่แปลงแล้ว จะมีรูปแบบเป็นดังสมการที่ (3)

วิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยโลจิสติก

การวิเคราะห์การถดถอยโลจิสติก จะมีวิธีการประมาณค่าพารามิเตอร์ในตัวแบบโดยใช้วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood: ML) ที่ให้ค่าประมาณของพารามิเตอร์ที่มีค่าความน่าจะเป็นสูงสุด สำหรับเซตข้อมูลหรือค่าสังเกตที่ใช้ในการประมาณ ร่วมกับระเบียบวิธีเชิงตัวเลข

(Numerical Method) ที่มีการคำนวณซ้ำ (Iterative Algorithm) เช่นวิธี Fisher's Scoring Method เป็นต้น โดยจะหาค่าประมาณของ $\boldsymbol{\beta}_{(m+1)\times 1} = [\beta_0, \beta_1, \dots, \beta_m]'$ ที่ทำให้ฟังก์ชันภาวะน่าจะเป็น (Likelihood function) มีค่าสูงสุด ซึ่งมีขั้นตอนดังนี้ (จารุวรรณ, 2559)

1. หาฟังก์ชันภาวะน่าจะเป็น Likelihood function หรือ $\boldsymbol{L} = f(Y_1, Y_2, \dots, Y_n)$

$$\begin{aligned}\boldsymbol{L} &= f(Y_1, Y_2, \dots, Y_n)\\ &= \prod_{i=1}^{n} f(Y_i)\\ &= f(Y_1)\cdot f(Y_2)\cdot \dots \cdot f(Y_n)\end{aligned}$$

ฟังก์ชันภาวะน่าจะเป็นนี้สามารถเขียนในรูปของฟังก์ชันลอการิทึมฐานธรรมชาติ ได้ดังนี้

$$\begin{aligned}\ln \boldsymbol{L} &= \ln f(Y_1) + \ln f(Y_2) + \cdots + \ln f(Y_n)\\ &= \textstyle\sum_{i=1}^{n} \ln f(Y_i)\end{aligned}$$

เนื่องจาก Y เป็นตัวแปรสุ่มที่มีการแจกแจงแบบเบอร์นูลลีที่มีค่าเฉลี่ย $E(Y) = \pi$ และความแปรปรวน $V(Y) = \pi(1-\pi)$ ดังนั้น ฟังก์ชันมวลความน่าจะเป็นของ Y_i คือ

$$f(Y_i) = \pi^{y_i}(1-\pi)^{1-y_i}$$

นำ $f(Y_i)$ แทนลงในฟังก์ชันภาวะน่าจะเป็นที่อยู่ในรูปของฟังก์ชันลอการิทึมฐานธรรมชาติจะได้

$$\begin{aligned}\ln \boldsymbol{L} &= \textstyle\sum_{i=1}^{n} \ln f(Y_i) = \sum_{i=1}^{n} \ln\left[\pi_i^{y_i}(1-\pi_i)^{1-y_i}\right] && (4)\\ &= \textstyle\sum_{i=1}^{n}[y_i \ln \pi_i + (1-y_i)\ln(1-\pi_i)]\\ &= \textstyle\sum_{i=1}^{n}[y_i\{\boldsymbol{X}_i'\boldsymbol{\beta} + \ln(1-\pi_i)\} + (1-y_i)\ln(1-\pi_i)]\\ &= \textstyle\sum_{i=1}^{n}[y_i(\boldsymbol{X}_i'\boldsymbol{\beta}) + y_i\ln(1-\pi_i) + \ln(1-\pi_i) - y_i\ln(1-\pi_i)]\\ &= \textstyle\sum_{i=1}^{n}[y_i(\boldsymbol{X}_i'\boldsymbol{\beta}) + \ln(1-\pi_i)]\\ &= \textstyle\sum_{i=1}^{n}\left[y_i(\boldsymbol{X}_i'\boldsymbol{\beta}) + \ln\left(\frac{1}{1+e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}}\right)\right]\\ &= \textstyle\sum_{i=1}^{n}\left[y_i(\boldsymbol{X}_i'\boldsymbol{\beta}) - \ln\left(1+e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}\right)\right] && (5)\end{aligned}$$

เมื่อ $\boldsymbol{X}_{i[1\times(p+1)]} = [X_{i0}, X_{i1}, X_{i2}, \dots, X_{ip}]$

2. หาตัวประมาณภาวะน่าจะเป็นสูงสุดโดยการหาอนุพันธ์อันดับที่ 1 ของลอการิทึมของฟังก์ชันภาวะน่าจะเป็น หรือ $\ln \boldsymbol{L} = \ell(\boldsymbol{\beta})$ เทียบกับ $\boldsymbol{\beta}$ และกำหนดให้เท่ากับศูนย์ เพื่อตรวจสอบหาจุดที่มีค่าความชันเท่ากับ 0 ของฟังก์ชัน ดังนี้

$$U(\boldsymbol{\beta})_{[1\times(p+1)]} = \frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \left(\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_0}, \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_1}, \dots, \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_p}\right)$$
$$= \sum_{i=1}^{n} y_i \boldsymbol{X_i} - \sum_{i=1}^{n} \boldsymbol{X_i} \left(\frac{e^{\left(X_i'\beta\right)}}{1+e^{\left(X_i'\beta\right)}}\right) = 0$$

3. ตรวจสอบว่าฟังก์ชัน $\boldsymbol{\ell(\beta)}$ มีค่าสูงสุดหรือไม่ โดยใช้การหาอนุพันธ์อันดับที่ 2 ของลอการิทึมของฟังก์ชันภาวะน่าจะเป็น หรือ $\boldsymbol{\ell(\beta)}$ เทียบกับ $\boldsymbol{\beta}$ เพื่อตรวจสอบดูอัตราการเปลี่ยนแปลงของค่าความชันว่ามีค่าเป็นบวกหรือเป็นลบและจะได้ทราบว่าค่าประมาณ $\boldsymbol{\beta}$ ดังกล่าวทำให้ฟังก์ชันมีค่าสูงสุด ดังนี้

$$I(\boldsymbol{\beta})_{[(p+1)\times(p+1)]} = \frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = -\sum_{i=1}^{n} \left(\frac{\boldsymbol{X_i' X_i} e^{\left(\boldsymbol{X_i'\beta}\right)}}{\left(1+e^{\left(\boldsymbol{X_i'\beta}\right)}\right)^2}\right)$$

4. กำหนดค่าประมาณเริ่มต้น $\boldsymbol{\beta}^{(0)}$

5. ประมาณค่า $\boldsymbol{\beta}$ จากสมการ $\boldsymbol{\beta}^{new}_{[1\times(p+1)]} = \boldsymbol{\beta}^{old} - U(\boldsymbol{\beta}^{old})E[I^{-1}(\boldsymbol{\beta}^{old})]$ โดยที่ $E[I^{-1}(\boldsymbol{\beta}^{old})]$ คือ ค่าคาดหวังของเมทริกซ์ผกผัน (Inverse matrix) ของ $\boldsymbol{I(\beta)}$

โดยที่จะเริ่มจากการใช้ $\boldsymbol{\beta}^{(0)}$ แทน $\boldsymbol{\beta}^{old}$ ในการทำซ้ำครั้งแรก จนกระทั่งได้ $\boldsymbol{\beta}^{new}$ หรือ $\boldsymbol{\beta}^{(1)}$จากนั้นนำ $\boldsymbol{\beta}^{(1)}$แทน $\boldsymbol{\beta}^{old}$ ในการทำซ้ำครั้งที่ 2 จนกระทั่งได้ $\boldsymbol{\beta}^{(2)}$ จากนั้นทำซ้ำไปเรื่อย ๆ จนกว่าจะได้ค่าประมาณสัมประสิทธิ์การถดถอย $\boldsymbol{\beta}$ ที่มีค่าลู่เข้า นั่นคือ มีผลต่างน้อยมากระหว่าง $\boldsymbol{\beta}^{new}$ และ $\boldsymbol{\beta}^{old}$ ซึ่งน้อยกว่าค่าที่สามารถยอมรับได้ (error tolerance) ก็จะหยุดการทำซ้ำ (Garrett, 2015)

อย่างไรก็ตาม ในกรณีข้อมูลมีมิติสูง จะไม่สามารถหาตัวประมาณพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุดได้ เนื่องจากเมทริกซ์ของตัวแปรอธิบายในกรณีที่ข้อมูลมิติสูงหรือ $\boldsymbol{X}$ จะเป็น non-full ranked matrix ตัวแปรอธิบายจึงมักจะเกิดปัญหาความสัมพันธ์เชิงเส้นแบบพหุ นั่นคือ ตัวแปรอธิบายมีความสัมพันธ์เชิงเส้นกันระหว่างตัวแปร เมื่อทำการประมาณค่าพารามิเตอร์ของสัมประสิทธิ์การถดถอยโลจิสติก แล้วจะได้ว่าค่าประมาณของสัมประสิทธิ์การถดถอยโดยวิธีภาวะน่าจะเป็นสูงสุดนั้นมีความแปรปรวนสูง นั่นคือ ค่าสัมประสิทธิ์การถดถอยที่ได้จากการประมาณค่านี้ไม่ใกล้เคียงกับค่าพารามิเตอร์ ซึ่งจะส่งผลให้ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของค่าประมาณสัมประสิทธิ์การถดถอยที่ได้มีค่ามาก (Sur & Candes, 2019)

ดังนั้น จึงได้มีผู้วิจัยหลายท่านได้เสนอวิธีการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีการอื่น ๆ เพื่อที่จะแก้ปัญหาของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีภาวะน่าจะเป็นสูงสุดของการถดถอยโลจิสติกสำหรับกรณีนี้ ซึ่งวิธีการหนึ่งที่เป็นวิธีที่ใช้กันค่อนข้างแพร่หลาย คือ การเรกู-

ลาไรซ์หรือการถดถอยโลจิสติกแบบพีนอลไลซ์ คือการหาตัวประมาณที่มีความเอนเอียง (bias) ที่เพิ่มความเอนเอียงเข้าไปเพื่อลดสัดส่วนของความแปรปรวนลง และทำให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ (Mean Square Error: MSE) มีค่าลดลง

สำหรับวิธีการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีการประมาณค่าจากวิธีภาวะน่าจะเป็นสูงสุด การถดถอยโลจิสติกพหุคูณแบบพีนอลไลซ์มีหลักการในการหาตัวประมาณซึ่งเรียกว่า วิธีการประมาณค่าจากภาวะน่าจะเป็นสูงสุดแบบพีนอลไลซ์ (Penalized maximum likelihood estimation) โดยการเพิ่มฟังก์ชันพีนอลตี้ในฟังก์ชันลบลอการิทึมของฟังก์ชันภาวะน่าจะเป็น (Negative log-likelihood function) ดังนี้

จากสมการที่ (5) คือ $\ln \boldsymbol{L} = \sum_{i=1}^{n}\left[y_i(\boldsymbol{X}_i'\boldsymbol{\beta}) - \ln\left(1 + e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}\right)\right]$

จากเดิม ในการประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุด เราต้องการประมาณค่าพารามิเตอร์เพื่อให้ฟังก์ชันภาวะน่าจะเป็นนั้นมีค่าสูงสุด เพื่อให้สอดคล้องกับวิธีการของวิธีกำลังสองน้อยที่สุดของตัวแบบการถดถอยเชิงเส้น จึงเปลี่ยนจากการหาค่าประมาณพารามิเตอร์ที่ทำให้ฟังก์ชันนั้นมีค่าสูงสุด เป็นการหาค่าประมาณพารามิเตอร์ที่ทำให้ฟังก์ชันนั้นมีค่าน้อยสุด โดยการเปลี่ยนฟังก์ชันภาวะน่าจะเป็นนั้นเป็นค่าลบ นั่นคือ ฟังก์ชันลบลอการิทึมของฟังก์ชันภาวะน่าจะเป็น คือ

$$-\ln \boldsymbol{L} = \sum_{i=1}^{n}\left[-y_i(\boldsymbol{X}_i'\boldsymbol{\beta}) + \ln\left(1 + e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}\right)\right] \quad (6)$$

ทำการเพิ่มฟังก์ชันพีนอลตี้ในฟังก์ชันลอการิทึมของฟังก์ชันภาวะน่าจะเป็นสูงสุด จะได้เป็น

$$M(\boldsymbol{Y}) = -\ln \boldsymbol{L} + \lambda P\left((\|\boldsymbol{\beta}\|)_q^2\right)$$

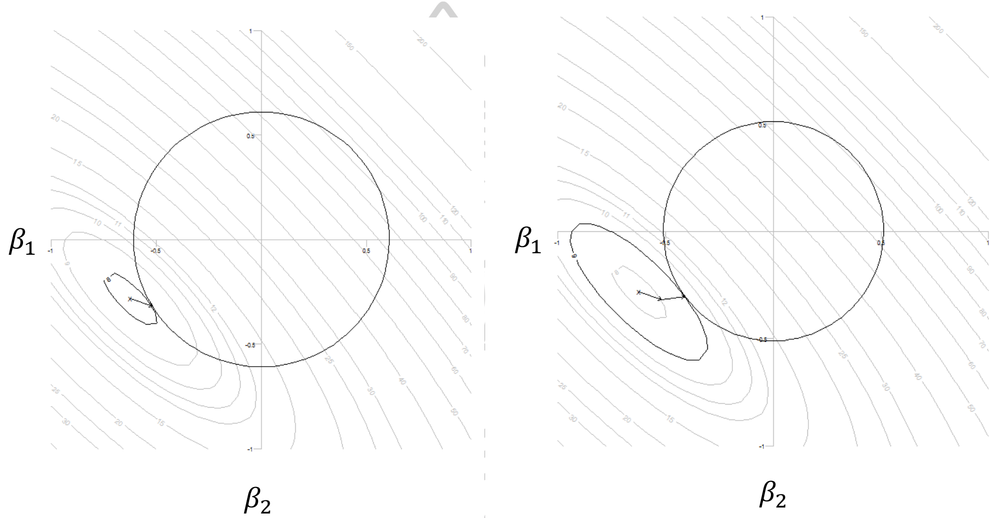
แล้วจึงทำการหาตัวประมาณของ $\boldsymbol{\beta}$ ที่ทำให้ฟังก์ชัน $M(\boldsymbol{Y})$ มีค่าน้อยที่สุด

ฟังก์ชันพีนอลตี้ที่ใช้ในวิธีการถดถอยโลจิสติกพหุคูณแบบพีนอลไลซ์มีหลายรูปแบบต่างกัน เช่น ฟังก์ชัน L_2 นอร์ม $P(\|\beta\|_2^2) = \sum_{j=1}^{p}\beta_j^2$ ที่ใช้ในการถดถอยวิธีริดจ์ หรือฟังก์ชัน L_1 นอร์ม $P(\|\beta\|_1^2) = \sum_{j=1}^{p}|\beta_j|$ ที่ใช้ในการถดถอยวิธีแลซโซ เป็นต้น โดยจะกล่าวถึงในหัวข้อถัดไป

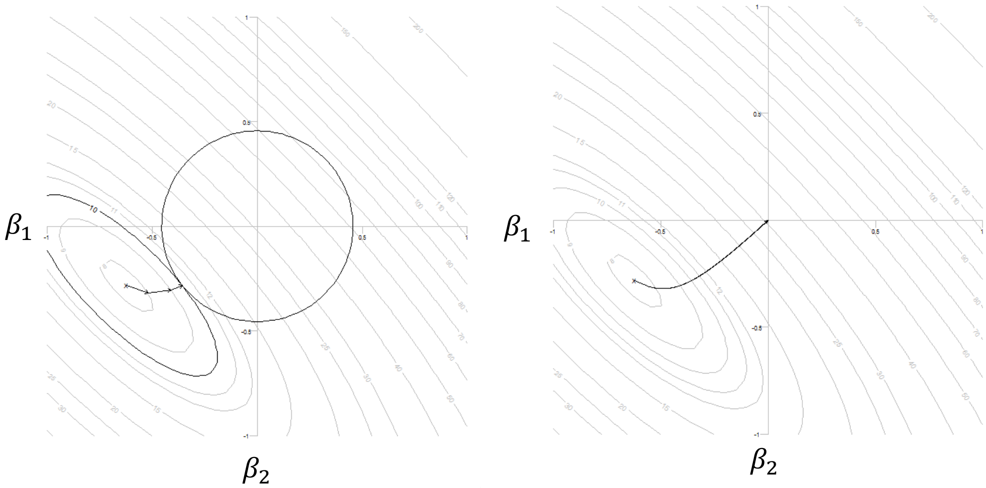
2.1.2 การถดถอยโลจิสติกพหุคูณด้วยวิธีริดจ์

Hoerl และ Kennard (1970) ได้เสนอวิธีการประมาณค่าตัวประมาณสัมประสิทธิ์การถดถอยที่เรียกว่า การถดถอยแบบริดจ์ ซึ่งเป็นวิธีการหนึ่งที่ใช้ในการแก้ปัญหาภาวะร่วมเชิงเส้นแบบพหุระหว่างตัวแปรอธิบายภายในตัวแบบถดถอยเชิงเส้น โดยการหาตัวค่าประมาณของพารามิเตอร์ที่ทำให้ผลรวมกำลังสองของส่วนเหลือมีค่าน้อยสุด โดยอยู่ภายใต้เงื่อนไขบังคับ $\sum_{j=1}^{p}\beta_j^2 = r^2 < t$ โดยที่ $0 < r^2 < t$ เมื่อ t เป็นค่าคงที่ใด ๆ ทั้งนี้ ตัวประมาณที่ได้จากการประมาณค่าด้วยการถดถอยแบบริดจ์จะเป็นตัวประมาณที่ปรับค่าเมทริกซ์ $\boldsymbol{X}'\boldsymbol{X}$ ของตัวประมาณแบบกำลังสองน้อยที่สุด

ทำให้ตัวประมาณที่ได้มีความเอนเอียง (Biased estimator) แต่ความแปรปรวนของตัวประมาณจะมีค่าน้อยกว่าตัวประมาณด้วยวิธีกำลังสองน้อยที่สุด และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) น้อยกว่าวิธีกำลังสองน้อยที่สุด หากพบว่าตัวแปรอธิบายภายในตัวแบบมีความสัมพันธ์เชิงเส้นระหว่างตัวแปรแล้ว เราสามารถเลือกใช้วิธีการถดถอยแบบริดจ์มาใช้ในการวิเคราะห์ โดยจะเป็นการดึงค่าของตัวประมาณแบบวิธีกำลังสองน้อยที่สุดลดลงไปด้านศูนย์ (Shrink) ทำให้ได้ค่า $\hat{\boldsymbol{\beta}}^{ridge}$ ทุกตัวมีขนาดเล็กแต่ไม่เท่ากับศูนย์ ดังนั้นวิธีริดจ์จึงมีความเหมาะสมในการวิเคราะห์ข้อมูลที่ตัวแบบถดถอยมีสัมประสิทธิ์ขนาดเล็กที่ไม่เท่ากับ 0 จำนวนมาก



(ก) (ข)



(ค) (ง)

ภาพที่ 1 กราฟแสดงริดจ์ต่ำสุดและเส้นทางของพิกัดของริดจ์ต่ำสุดของผลบวกกำลังสองของความคลาดเคลื่อนสุ่มภายใต้ข้อจำกัด

$\boldsymbol{\beta}'\boldsymbol{\beta} = \sum_{j=1}^{2} \beta_j^2 = r^2$ (Simon, 2016)

ภาพที่ 1 (ก) แสดงริดจ์ต่ำสุดเมื่อกำหนดให้ผลบวกกำลังสองของความคลาดเคลื่อนสุ่มมีค่าเท่ากับ 8 สัมผัสกับพื้นที่เงื่อนไขบังคับที่จุด $\hat{\beta}_1$ = -0.52 และ $\hat{\beta}_2$ = -0.32 และเมื่อกำหนดให้ผลบวกกำลังสองของความคลาดเคลื่อนสุ่มมีค่าเท่ากับ 9 และ 10 แล้วหารัศมีของวงกลมที่สัมผัสกับเส้นคอนทัวร์จะลดลงตามลำดับ ซึ่งจะทำให้พิกัดของริดจ์ต่ำสุด $(\hat{\beta}_1, \hat{\beta}_2)$ จะเข้าใกล้จุดกำเนิด (0,0) มากขึ้น ดังภาพที่ 1 (ข) และ (ค) ตามลำดับ และภาพที่ 1 (ง) จะแสดงเส้นทางของพิกัดของริดจ์ต่ำสุดไปยังจุดกำเนิด (0,0)

ในกรณีของการประมาณค่าพารามิเตอร์ของตัวแบบถดถอยเชิงเส้นพหุคูณด้วยวิธีกำลังสองน้อยที่สุดนั้น คือจะประมาณค่า $\boldsymbol{\beta}$ ที่ทำให้เส้นถดถอยของตัวอย่างใกล้เคียงกับเส้นถดถอยของประชากรมากที่สุด หรือทำให้เส้นถดถอยของตัวอย่างลากผ่านหรือใกล้เคียงกับค่าสังเกตมากที่สุด โดยพิจารณาจากค่าส่วนเหลือ (residual) ที่เป็นผลต่างระหว่างค่าสังเกตกับค่าประมาณคือ $\boldsymbol{\varepsilon_i} = y_i - \boldsymbol{X_i\beta}$ มีค่าน้อยที่สุด ตัวประมาณสัมประสิทธิ์การถดถอยที่ได้จากวิธีนี้ คือ ตัวประมาณที่ทำให้ผลรวมกำลังสองของส่วนเหลือ (residual sum of squares : RSS) มีค่าน้อยที่สุด หรือ $\sum_{i=1}^{n}(y_i - \boldsymbol{X_i\beta})^2$ ซึ่งสามารถเขียนอยู่ในรูปผลคูณของเวกเตอร์ของส่วนเหลือ ดังนี้

$$\begin{aligned}
\text{RSS} &= \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}' = (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) \\
&= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X\beta} - \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta X}'\boldsymbol{X\beta} \\
&= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta}
\end{aligned}$$

การหาค่าประมาณ $\boldsymbol{\beta}$ ที่ทำให้ $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$ มีค่าน้อยที่สุด ซึ่งสามารถทำได้โดยการหาอนุพันธ์ของ $\boldsymbol{e}'\boldsymbol{e}$ เทียบกับ $\boldsymbol{\beta}$ แล้วกำหนดให้เท่ากับศูนย์ ดังนี้

$$
\begin{aligned}
\frac{\partial \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta}} &= 0 - 2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = 0 \\
\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} &= \boldsymbol{X}'\boldsymbol{y} \\
(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\
\widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}
\end{aligned}
$$

ดังนั้น ตัวประมาณของสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุด คือ $\widehat{\boldsymbol{\beta}}^{ols} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$

ส่วนการประมาณค่าพารามิเตอร์ของตัวแบบถดถอยเชิงเส้นพหุคูณด้วยวิธีริดจ์นั้น จะมีลักษณะที่ใกล้เคียงกัน นั่นคือ จะประมาณค่า $\boldsymbol{\beta}$ ที่ทำให้ผลรวมกำลังสองของส่วนเหลือมีค่าน้อยที่สุด แต่จะถูกจำกัดอยู่ภายใต้ข้อจำกัดคือ $\boldsymbol{\beta}'\boldsymbol{\beta} = r^2$ โดยที่ $0 < r^2 < t$ เมื่อ t เป็นค่าคงที่ใด ๆ ซึ่งมีขั้นตอนดังนี้ (พัชราภรณ์, 2560)

การหาค่าเหมาะสมที่สุดภายใต้ข้อจำกัด (optimization with equality constraint) ซึ่งสามารถทำได้โดยใช้วิธีการของลากรานจ์ โดยกำหนด λ เป็นตัวคูณลากรานจ์ (Lagrange multiplier) และฟังก์ชันลากรานจ์ (Lagrangian expression) คือ

$$L(y, \boldsymbol{\beta}, \lambda) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) - \lambda(\boldsymbol{\beta}'\boldsymbol{\beta} - r^2)$$

หาค่าอนุพันธ์บางส่วนของ $L(y, \boldsymbol{\beta}, \lambda)$ เทียบกับ $\boldsymbol{\beta}$ คือ

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}} L(y, \boldsymbol{\beta}, \lambda) &= \frac{\partial}{\partial \boldsymbol{\beta}}[(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) - \lambda(\boldsymbol{\beta}'\boldsymbol{\beta} - r^2)] \\
&= \frac{\partial}{\partial \boldsymbol{\beta}}[\boldsymbol{y}'\boldsymbol{y} - \mathbf{2}\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} - \lambda(\boldsymbol{\beta}'\boldsymbol{\beta} - r^2)] \\
&= -\mathbf{2}\boldsymbol{X}'\boldsymbol{y} + \mathbf{2}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} - \mathbf{2}\lambda\boldsymbol{\beta}
\end{aligned}
$$

หาค่าอนุพันธ์บางส่วนของ $L(y, \boldsymbol{\beta}, \lambda)$ เทียบกับ λ คือ

$$
\begin{aligned}
\frac{\partial}{\partial \lambda} L(y, \boldsymbol{\beta}, \lambda) &= \frac{\partial}{\partial \lambda}[(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) - \lambda(\boldsymbol{\beta}'\boldsymbol{\beta} - r^2)] \\
&= -(\boldsymbol{\beta}'\boldsymbol{\beta} - r^2)
\end{aligned}
$$

กำหนดค่าอนุพันธ์เท่ากับ $\mathbf{0}$ และแก้สมการหาค่าของ $\widehat{\boldsymbol{\beta}}$

$$-\mathbf{2}\boldsymbol{X}'\boldsymbol{y} + \mathbf{2}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} - \mathbf{2}\lambda\boldsymbol{\beta} = \mathbf{0}$$

$$(\boldsymbol{X}'\boldsymbol{X} - \lambda\mathbf{I}_p)\boldsymbol{\beta} = \boldsymbol{X}'\boldsymbol{y}$$

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X} - \lambda\mathbf{I}_p)^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{y}$$

ดังนั้น ตัวประมาณแบบริดจ์จะนิยมเขียนอยู่รูปแบบของค่าคงที่ $k \geq 0$ โดยที่ $k = -\lambda$ ดังนี้

$$\widehat{\boldsymbol{\beta}}^{ridge} = (\boldsymbol{X}'\boldsymbol{X} + k\mathbf{I}_p)^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{y} \tag{7}$$

ในส่วนของการประมาณค่าพารามิเตอร์ของตัวแบบถดถอยโลจิสติกพหุคูณด้วยวิธีริดจ์นั้น จากการประมาณค่าพารามิเตอร์ของตัวแบบถดถอยโลจิสติกด้วยวิธีภาวะน่าจะเป็นสูงสุด เราต้องการประมาณค่า $\boldsymbol{\beta}$ ที่ทำให้ฟังก์ชันภาวะน่าจะเป็น (Likelihood function) มีค่าสูงสุด วิธีริดจ์นั้นก็จะทำในลักษณะเดียวกัน จากฟังก์ชันลบลอการิทึมของฟังก์ชันภาวะน่าจะเป็นของตัวแบบถดถอยโลจิสติกพหุคูณดังสมการที่ (6) คือ $-\ln \boldsymbol{L} = \sum_{i=1}^{n}\left[-y_i(\boldsymbol{X}_i'\boldsymbol{\beta}) + \ln\left(1 + e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}\right)\right]$ จากนั้นทำการประมาณค่าพารามิเตอร์ $\boldsymbol{\beta}$ ที่ทำให้ฟังก์ชันลบลอการิทึมดังกล่าวมีค่าต่ำสุด ซึ่งสามารถเขียนได้เป็น

$$\widehat{\boldsymbol{\beta}}^{ridge} = argmin_{\boldsymbol{\beta}}\left\{\sum_{i=1}^{n}\left[-y_i(\boldsymbol{X}_i'\boldsymbol{\beta}) + \ln\left(1 + e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}\right)\right] + \lambda\sum_{j=1}^{p}\beta_j^2\right\} \tag{8}$$

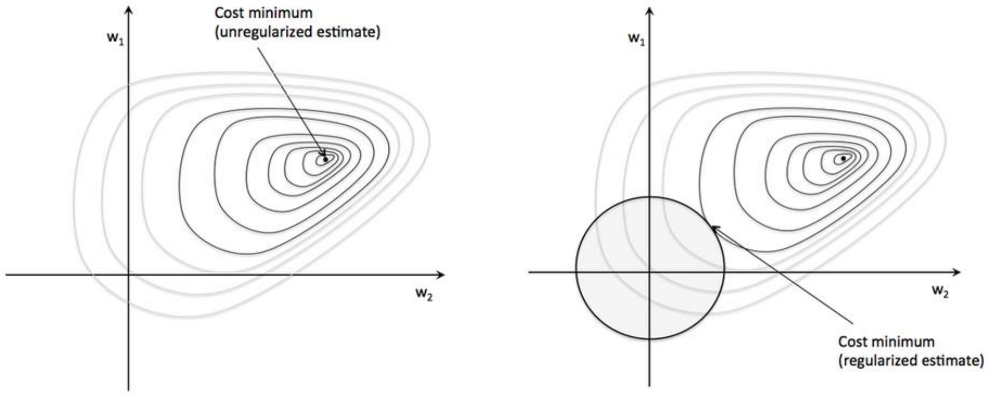
ตัวประมาณพารามิเตอร์จากวิธีริดจ์สำหรับตัวแบบถดถอยโลจิสติกพหุคูณนั้น สามารถเขียนให้อยู่ในรูปของผลคูณของตัวประมาณจากวิธี ML ได้ (Schaefer et al., 1984) ดังนี้

$$\widehat{\boldsymbol{\beta}}^{ridge} = (\boldsymbol{X}'\widehat{\boldsymbol{W}}\boldsymbol{X} + k\boldsymbol{I})^{-1}(\boldsymbol{X}'\widehat{\boldsymbol{W}}\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{ML}}) \tag{9}$$

โดยที่ $\widehat{\boldsymbol{W}}$ คือเมทริกซ์ที่มีค่าบนเส้นทแยงมุมเป็น $\widehat{w}_i = \widehat{\pi}_i(1 - \widehat{\pi}_i)$ เมื่อ $\widehat{\pi}_i$ คือค่าความน่าจะเป็นที่คำนวณได้จากการประมาณค่าด้วยวิธี ML

k คือพารามิเตอร์ชริงค์เกจ (Shrinkage parameter)

$\widehat{\boldsymbol{\beta}}_{\boldsymbol{ML}}$ คือค่าประมาณค่าสัมประสิทธิ์ถดถอยโลจิสติกจากวิธี ML



ภาพที่ 2 แสดงการใช้วิธีริดจ์ในการประมาณค่าพารามิเตอร์ในตัวแบบถดถอยโลจิสติก ประกอบด้วยการประมาณค่าพารามิเตอร์โดยไม่มีการเรกูลาไรซ์ และการใช้การเรกูลาไรซ์โดยวิธีริดจ์ในการประมาณค่า (Raschka, 2016)

อย่างไรก็ตาม ในกรณีที่ตัวแปรอธิบายทุกตัวมีความสัมพันธ์กับตัวแปรตาม เราสามารถเลือกใช้การถดถอยวิธีริดจ์ ในการประมาณค่าพารามิเตอร์ได้ ซึ่งจะเป็นการดึงค่าประมาณลดลงไปด้านศูนย์ ทำให้ได้ค่า $\hat{\boldsymbol{\beta}}^{ridge}$ ทุกตัวมีขนาดเล็กแต่ไม่เท่ากับศูนย์ ดังนั้น วิธีริดจ์จึงมีความเหมาะสมสำหรับวิเคราะห์ข้อมูลที่มีสัมประสิทธิ์การถดถอยขนาดเล็กที่ไม่เท่ากับ 0 จำนวนมาก ถึงแม้ว่าตัวประมาณนี้จะเป็นตัวประมาณที่เอนเอียง สำหรับพารามิเตอร์ $\boldsymbol{\beta}$ แต่ก็สามารถลดความแปรปรวนของตัวประมาณให้น้อยลง รวมถึงยังมีคุณสมบัติเด่น คือ ตัวประมาณที่ได้จะมีความเสถียร จึงเป็นอีกวิธีที่ใช้กันมากเพื่อแก้ไขปัญหาตัวแปรอธิบายมีความสัมพันธ์กันสูง แม้ว่าข้อมูลจะไม่มีมิติสูงก็ตาม แต่สำหรับข้อเสียของวิธีริดจ์นี้ คือ การขาดคุณสมบัติในการคัดเลือกตัวแปรเข้าสู่ตัวแบบ จึงทำให้การแปลผลตัวแปรในตัวแบบนั้นทำได้ยากในข้อมูลที่มีมิติสูง

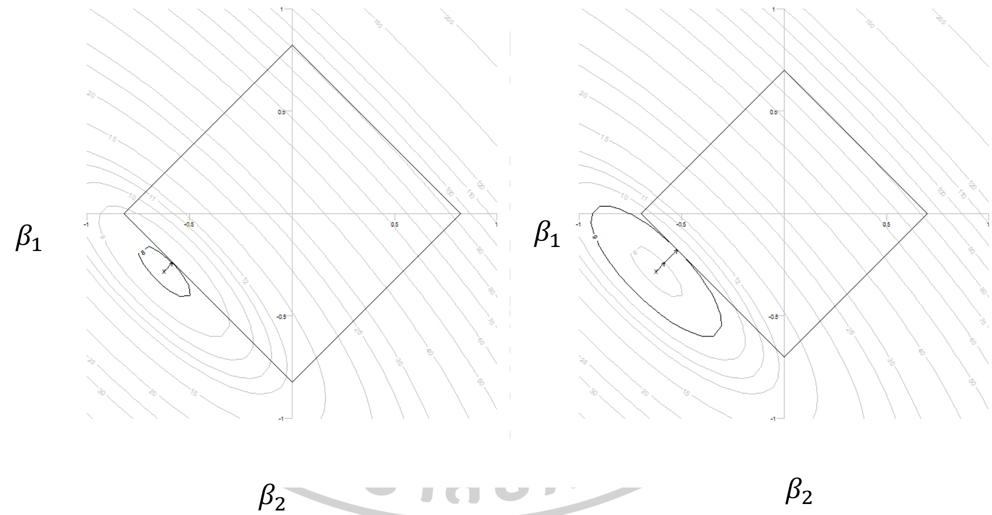
2.1.3 การถดถอยโลจิสติกพหุคูณด้วยวิธีแลซโซ

ตัวประมาณด้วยวิธีแลซโซ เป็นการนำคุณลักษณะที่ดีของวิธีริดจ์และวิธี best subset selection มารวมกัน โดยจะทำการดึงค่าประมาณสัมประสิทธิ์การถดถอยบางค่าให้ลดลง และให้ค่าตัวประมาณบางตัวเท่ากับศูนย์ ซึ่งเป็นการแก้ไขข้อบกพร่องของทั้งสองวิธี ได้แก่ ถ้าหากใช้วิธีการ subset selection ในการคัดเลือกตัวแปรเพียงอย่างเดียว อาจจะให้ผลลัพธ์ของสมการที่ได้ยังมีจำนวนตัวแปรอธิบายมาก เนื่องจากเป็นกระบวนการแบบไม่ต่อเนื่อง (discrete process) ซึ่งตัวแปรอธิบายอาจจะยังอยู่หรือถูกคัดออกจากสมการก็ได้ การเปลี่ยนแปลงข้อมูลเล็กน้อย ก็สามารถให้ผลลัพธ์ของการคัดเลือกตัวแปรที่แตกต่างกันได้ และทำให้ความถูกต้องในการทำนายลดลง ส่วนวิธีริดจ์เป็นกระบวนการต่อเนื่อง (continuous process) โดยมีการทำให้ดึงค่าของสัมประสิทธิ์การ

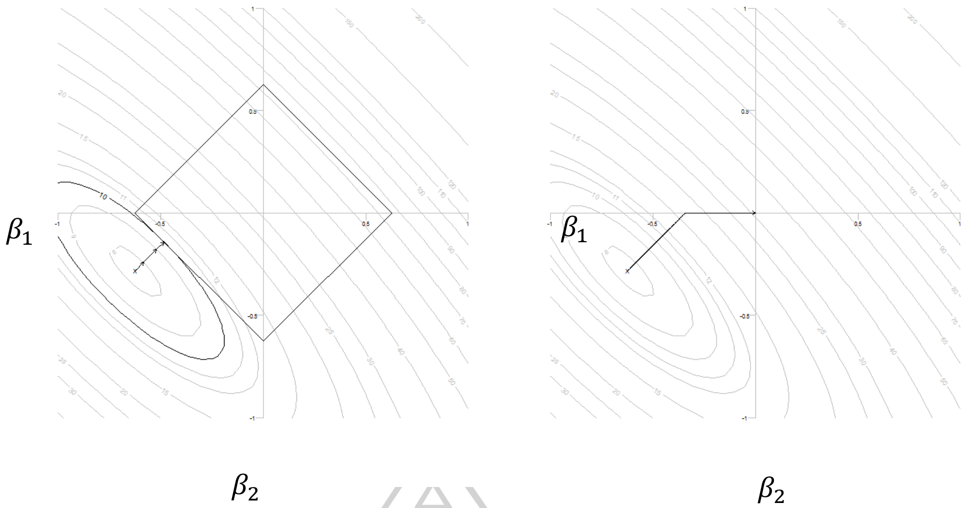
ถดถอยให้ลดลง ซึ่งทำให้มีความถูกต้องในการทำนายมากขึ้น แต่ก็ไม่สามารถกำหนดให้ตัวประมาณมีค่าเป็นศูนย์ได้ จึงทำให้การแปลผลตัวแปรในตัวแบบนั้นทำได้ยาก

Tibshirani (1996) ได้เสนอวิธีการประมาณพารามิเตอร์ที่ชื่อว่า LASSO ซึ่งเป็นหนึ่งในวิธีการเรกูลาไรซ์ของตัวแบบการถดถอยที่มีคุณลักษณะเช่นเดียวกันกับตัวประมาณแบบริดจ์ นั่นคือ การเพิ่มความเอนเอียงของตัวประมาณสัมประสิทธิ์การถดถอยเล็กน้อย เพื่อให้ความแปรปรวนของตัวประมาณมีค่าลดลง ซึ่งจะทำให้การทำนายมีความถูกต้องมากขึ้น และยังรวมเทคนิคการคัดเลือกตัวแปร ที่เป็นประโยชน์สำหรับการแปลความหมายของสมการถดถอย โดยเฉพาะในกรณีที่ตัวแปรอธิบายมีจำนวนมาก

หลักการของวิธี LASSO สำหรับการประมาณค่าพารามิเตอร์สำหรับตัวแบบถดถอยเชิงเส้นพหุคูณ คือการหาตัวประมาณของเวกเตอร์พารามิเตอร์ $\boldsymbol{\beta}$ ที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อนสุ่มมีค่าต่ำสุด ภายใต้ข้อจำกัดของ L_1 นอร์ม ของ $\boldsymbol{\beta}$ คือ $\|\beta\| = \sum_{j=1}^{p}|\beta_j| \leq t$ เมื่อ t เป็นค่าคงที่ ซึ่งเป็นลักษณะเดียวกันกับวิธีกำลังสองน้อยที่สุด



(ก) (ข)



(ค) (ง)

ภาพที่ 3 กราฟแสดงวิธีการหาตัวประมาณโดยวิธี LASSO และเส้นทางของพิกัดของสัมประสิทธิ์การถดถอยของตัวประมาณ LASSO ภายใต้ข้อจำกัด

$\sum_{j=1}^{2}|\boldsymbol{\beta_j}| \leq \boldsymbol{t}$ (Simon, 2016)

ภาพที่ 3 แสดงค่าประมาณสัมประสิทธิ์การถดถอยและเส้นทางของพิกัดของค่าประมาณสัมประสิทธิ์การถดถอยวิธี LASSO ภายใต้ข้อจำกัด $\sum_{j=1}^{p}|\beta_j| = t$ เมื่อกำหนดค่าผลบวกกำลังสองของความคลาดเคลื่อนสุ่มเป็นค่าตรึง แกนตั้งแสดงพิกัดของ β_1 และแกนนอนแสดงพิกัดของ β_2 ซึ่งพิกัดจุดของค่าประมาณสัมประสิทธิ์การถดถอยวิธีกำลังสองน้อยที่สุดเป็นค่าที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อนสุ่มมีค่าต่ำที่สุดมีค่าเท่ากับ 7.81 สำหรับค่าประมาณสัมประสิทธิ์การถดถอยวิธี LASSO กรณีตัวแปรอธิบาย 2 ตัว (p = 2) จะได้เงื่อนไขบังคับ คือ $|\beta_1| + |\beta_2| \leq t$ ซึ่งเป็นรูปสี่เหลี่ยมเอียง โดยจุดพิกัดที่เส้นเส้นคอนทัวร์สัมผัสกับพื้นที่เงื่อนไขบังคับจะเป็นค่าประมาณสัมประสิทธิ์การถดถอยวิธี LASSO ภาพที่ 3 (ก) แสดงพิกัดของค่าประมาณสัมประสิทธิ์การถดถอยเมื่อกำหนดให้ผลบวกกำลังสองของความคลาดเคลื่อนสุ่มมีค่าเท่ากับ 8 และเมื่อกำหนดให้ผลบวกกำลังสองของความคลาดเคลื่อนสุ่มมีค่าเท่ากับ 9 และ 10 แสดงค่าประมาณสัมประสิทธิ์การถดถอยได้ดังภาพที่ 3 (ข) และ (ค) ตามลำดับ ภาพที่ 3 (ง) แสดงเส้นทางของพิกัดของค่าประมาณสัมประสิทธิ์การถดถอยวิธี LASSO โดยค่าประมาณสัมประสิทธิ์การถดถอยจะมีค่าเข้าสู่แกนนอนทำมุม 45 องศา ตัวประมาณสัมประสิทธิ์การถดถอย β_1 มีค่าเป็นศูนย์ และค่าประมาณสัมประสิทธิ์ถดถอย β_2 มีค่าลู่เข้าสู่เป็นศูนย์

การประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี LASSO คือการหาตัวประมาณของเวกเตอร์พารามิเตอร์ $\boldsymbol{\beta}$ ที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อนสุ่มมีค่าต่ำสุด ภายใต้ข้อจำกัดของ L_1 นอร์ม ของ $\boldsymbol{\beta}$ คือ $\|\beta\| = \sum_{j=1}^{p}|\beta_j| \leq t$ เมื่อ t เป็นค่าคงที่ ดังนี้

ตัวประมาณสัมประสิทธิ์การถดถอยแบบ LASSO สามารถนิยามโดย

minimize $\quad (\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})$

ภายใต้ข้อจำกัด $\quad \sum_{j=1}^{p}|\beta_j| \leq t$

เมื่อ $t \geq 0$ เรียกว่า พารามิเตอร์การปรับแต่ง (tuning parameter) ซึ่งควบคุมขนาดของการหดตัว(shrinkage) ของตัวประมาณ $\widehat{\boldsymbol{\beta}}$ ถ้า $t_0 = \sum|\widehat{\boldsymbol{\beta}}^{ols}|$ เมื่อกำหนด ค่า $t < t_0$ จะทำให้ตัวประมาณสัมประสิทธิ์ถูกดึงค่าลดลงเข้าหาศูนย์ และตัวประมาณบางตัวจะมีค่าเท่ากับศูนย์ ซึ่งสามารถเขียนอยู่ในรูปสมการที่สอดคล้องกัน คือ

$$\widehat{\boldsymbol{\beta}}^{lasso} = argmin_{\boldsymbol{\beta}}(\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta}) + \lambda\sum_{j=1}^{p}|\beta_j| \qquad (10)$$

แต่เนื่องจากตัวประมาณ LASSO เป็นตัวประมาณที่เป็นคำตอบของฟังก์ชันที่ไม่เป็นเชิงเส้น และไม่สามารถหาอนุพันธ์เทียบกับ $\boldsymbol{\beta}$ ได้ สำหรับค่า $\boldsymbol{\lambda}$ ใดๆ ที่กำหนด เราจึงไม่สามารถเขียนตัวประมาณ LASSO อยู่ในสูตรที่ชัดเจนได้ในกรณีทั่วไป

ในส่วนของการประมาณค่าพารามิเตอร์ของตัวแบบถดถอยโลจิสติกพหุคูณด้วยวิธีแลซโซนั้น จากการประมาณค่าพารามิเตอร์ของตัวแบบถดถอยโลจิสติกด้วยวิธีภาวะน่าจะเป็นสูงสุด เราต้องการประมาณค่า $\boldsymbol{\beta}$ ที่ทำให้ฟังก์ชันภาวะน่าจะเป็น (Likelihood function) มีค่าสูงสุด วิธีแลซโซนั้นก็จะทำในลักษณะเดียวกัน โดยใช้ฟังก์ชันลบลอการิทึมของฟังก์ชันภาวะน่าจะเป็นของตัวแบบถดถอยโลจิสติกพหุคูณดังสมการที่ (6) คือ $-\ln \boldsymbol{L} = \sum_{i=1}^{n}\left[-y_i(\boldsymbol{X}_i'\boldsymbol{\beta}) + \ln\left(1 + e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}\right)\right]$ จากนั้นทำการประมาณค่าพารามิเตอร์ $\boldsymbol{\beta}$ ที่ทำให้ฟังก์ชันลบลอการิทึมดังกล่าวมีค่าต่ำสุด ซึ่งสามารถเขียนได้เป็น

$$\widehat{\boldsymbol{\beta}}^{\boldsymbol{lasso}} = argmin_{\boldsymbol{\beta}}\left\{-\sum_{i=1}^{n}\left[-y_i(\boldsymbol{X}_i'\boldsymbol{\beta}) + \ln\left(1 + e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}\right)\right] + \lambda\sum_{j=1}^{p}\left|\beta_j\right|\right\} \qquad (11)$$

ซึ่ง $\sum_{j=1}^{p}|\beta_j|$ คือ ฟังก์ชันพีนอลตี้

เมื่อเปรียบเทียบตัวประมาณแบบ LASSO กับตัวประมาณจาก subset selection และวิธีริดจ์ในกรณีของเมทริกซ์ $\boldsymbol{X'X}$ เป็นเมทริกซ์เชิงตั้งฉากปกติ (orthonormal design matrix)

วิธี subset selection: สำหรับค่า γ ใดๆ ที่กำหนดเป็นเกณฑ์สำหรับการคัดเลือก และตัวประมาณสัมประสิทธิ์ถดถอยที่ไม่ถูกเลือก มีค่าเท่ากับ 0 นั่นคือ

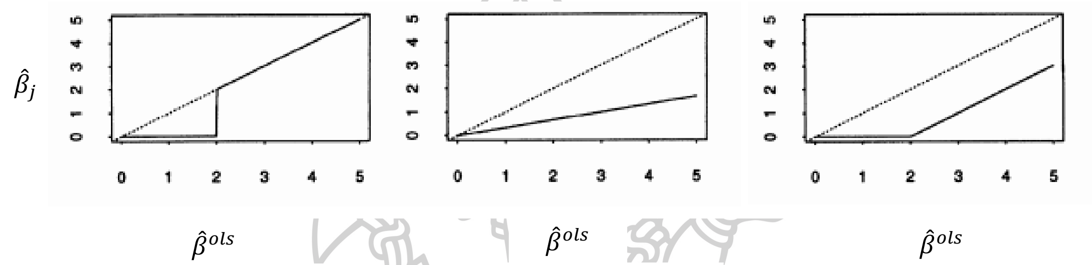
$$\hat{\beta}_{\boldsymbol{j}} = \begin{cases} \widehat{\beta}_j^{ols}\ ; & |\hat{\beta}_j^{ols}| > \gamma \\ 0\ ; & \text{อื่นๆ} \end{cases}$$

วิธีริดจ์: ตัวประมาณของสัมประสิทธิ์การถดถอยแบบริดจ์คือ $\widehat{\boldsymbol{\beta}}^{ridge} = \left[\boldsymbol{I_p} + k(\boldsymbol{X'X})^{-1}\right]^{-1}\widehat{\boldsymbol{\beta}}_j^{ols}$ ซึ่งในกรณีของเมทริกซ์ตั้งฉากปกติ จะได้ว่า

$$\widehat{\boldsymbol{\beta}}^{ridge} = \left[I_p + k\boldsymbol{I}_p\right]^{-1}\widehat{\boldsymbol{\beta}}_j^{ols}$$
$$= (1+k)^{-1}\widehat{\boldsymbol{\beta}}_j^{ols}$$

ดังนั้น $\hat{\beta}_j^{ridge} = (\frac{1}{1+k})\hat{\beta}_j^{ols}$ เมื่อ $j = 1,2,\dots,p$ และ k ขึ้นอยู่กับค่า t

จากฟังก์ชันตัวประมาณของแต่ละวิธีข้างต้น สามารถเขียนกราฟแสดงความสัมพันธ์ระหว่างตัวประมาณของแต่ละวิธีกับตัวประมาณ $\widehat{\boldsymbol{\beta}}_j^{ols}$ ได้ดังนี้

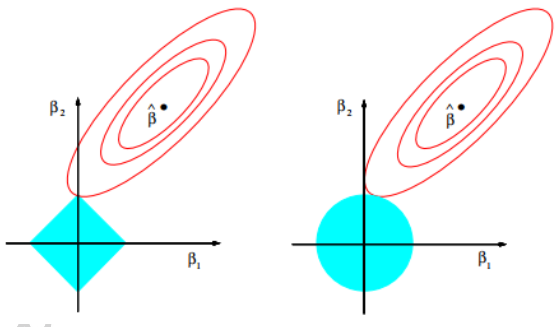


(ก) Subset regression (ข) Ridge regression (ค) LASSO

ภาพที่ 4 แสดงความสัมพันธ์ของการสัมประสิทธิ์การถดถอยโดยวิธีต่าง ๆ กับสัมประสิทธิ์การถดถอยวิธีกำลังสองน้อยที่สุด ในกรณีของเมทริกซ์ตั้งฉากปกติ (orthonormal design matrix) เมื่อ $\boldsymbol{\lambda}$ =2 (Tibshirani, 1996)

ภาพที่ 4 แสดงความสัมพันธ์ระหว่างตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยสุด และตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธี Subset selection, ตัวประมาณแบบริดจ์ และตัวประมาณแบบ LASSO แสดงในภาพที่ 4 (ก), (ข), (ค) ตามลำดับ แกนนอนคือค่าสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุดและแกนตั้งคือค่าสัมประสิทธิ์การถดถอยของแต่ละวิธี เส้นประทำมุม 45° กับแกนนอน คือ เส้นตรงที่ค่าสัมประสิทธิ์การถดถอยโดยวิธีต่าง ๆ เท่ากับค่าสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุด ซึ่งแสดงเฉพาะจตุภาค (quadrant) ที่ 1 จากภาพที่ 4 (ก) แสดงความสัมพันธ์ระหว่างตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยสุด และตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธี Subset selection โดย $\hat{\beta}_j = \hat{\beta}_j^{ols}$ เมื่อ $\hat{\beta}^{ols} > 2$ และ $\hat{\beta}_j = 0$ เมื่อ $\hat{\beta}^{ols} < 2$ แสดงว่า ตัวประมาณสัมประสิทธิ์การถดถอยที่มีค่ามากกว่า 2 จะถูกคัดเลือกเข้าสมการ จากภาพที่ 4 (ข) แสดงความสัมพันธ์ระหว่างตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยสุด และตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธีริดจ์ จะเห็นว่า $\hat{\beta}_j = (\frac{1}{1+2})\hat{\beta}_j^{ols} = (\frac{1}{3})\hat{\beta}_j^{ols}$ นั่นคือ ตัวประมาณแบบริดจ์ จะมีค่าลดลงเป็น 1 ใน 3 ของตัวประมาณโดยวิธีกำลังสองน้อยที่สุด จากภาพที่ 4 (ค) แสดงความสัมพันธ์ระหว่างตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธี

กำลังสองน้อยสุด และตัวประมาณสัมประสิทธิ์การถดถอยโดยวิธี LASSO จะเห็นว่า $\hat{\beta}_j = \hat{\beta}_j^{ols} - 2$ เมื่อ $\hat{\beta}_j^{ols} > 2$ และ $\hat{\beta}_j = 0$ เมื่อ $\hat{\beta}_j^{ols} \leq 2$ ดังนั้น ตัวประมาณสัมประสิทธิ์การถดถอยที่มีค่ามากกว่า 2 จะถูกคัดเลือกในสมการ และทำให้มีขนาดลดลงโดยมีค่าน้อยกว่า $\widehat{\beta}_j^{ols}$ เท่ากับ 2 หน่วยทุกค่า



ภาพที่ 5 แสดงการประมาณค่าพารามิเตอร์สัมประสิทธิ์การถดถอยแบบ LASSO และแบบริดจ์ในกรณี p = 2 (Tibshirani, 1996)

ภาพที่ 5 แสดงการเมื่อเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยวิธี LASSO และวิธีริดจ์ในกรณี p=2 เส้นวงรีแสดงเส้นคอนทัวร์ (Contour Line) ของ ฟังก์ชันภาวะน่าจะเป็นสูงสุดโดยมีค่าประมาณโดยวิธีกำลังสองน้อยที่สุด ($\hat{\beta}^{ols}$) เป็นจุดศูนย์กลางวงรี พื้นที่แรเงาคือ เงื่อนไขของพารามิเตอร์แบบ LASSO ได้แก่ $|\beta_1| + |\beta_2| \leq t$ ซึ่งเป็นรูปสี่เหลี่ยมเอียง และแบบริดจ์ได้แก่ $\beta_1^2 + \beta_2^2 \leq t^2$ ตามลำดับ และจุดที่เส้นความชันสูงที่น้อยที่สุดสัมผัสกับพื้นที่เงื่อนไขบังคับ จะเป็นตัวประมาณสัมประสิทธิ์การถดถอยแบบ LASSO และแบบริดจ์โดยจะพบว่า วิธี LASSO นั้นบางครั้งอาจจะมีโอกาสที่เส้นความชันสูงสัมผัสกับพื้นที่เงื่อนไขบังคับที่จุดมุมของรูปสีเหลี่ยม นั่นหมายความว่า สัมประสิทธิ์ $\beta_j = 0$ ส่งผลให้มีการคัดเลือกตัวแปรเกิดขึ้น โดยเส้นความชันสูงนี้จะต้องสัมผัสกับรูปสี่เหลี่ยมในจตุภาค (quadrant) เดียวกับ $\hat{\beta}^{ols}$ เนื่องจากข้างต้น $Q = -2\widehat{\beta'}_j^{ols}\beta_j + \beta_j'\beta_j + k|\beta_j|$ โดย Q จะมีค่าต่ำสุด เมื่อ $\widehat{\beta'}_j^{ols} \cdot \beta_j > 0$ ดังนั้น เมื่อ $\hat{\beta}_j^{ols} > 0$ β_j จะต้องมากกว่า 0 ด้วย และเมื่อ $\hat{\beta}_j^{ols} \leq 0$ นั้น $\beta_j \leq 0$ ด้วย จึงส่งผลให้ $\hat{\beta}^{lasso}$ อยู่จตุภาคเดียวกันกับ $\hat{\beta}^{ols}$ แต่การประมาณแบบริดจ์นั้นไม่มีโอกาสที่เส้นความชันสูงจะสัมผัสกับพื้นที่เงื่อนไขที่จุดมุม จึงไม่มีการคัดเลือกตัวแปรเกิดขึ้นนั่นเอง

วิธีการประมาณค่าพารามิเตอร์แบบ LASSO เป็นวิธีที่สามารถเลือกตัวแปรเข้าสู่ตัวแบบและประมาณค่าสัมประสิทธิ์การถดถอยได้ในคราวเดียวกัน ดังนั้นวิธี LASSO จึงเป็นหนึ่งในวิธีการเลือกตัวแปรได้โดยอัตโนมัติ ซึ่งวิธีนี้ถูกนำไปใช้งานอย่างแพร่หลาย แต่ถึงแม้ว่าวิธี LASSO จะมีลักษณะเด่น

ตรงที่มีการคัดเลือกตัวแปร แต่ก็ยังมีข้อจำกัดที่ว่า วิธี LASSO สามารถเลือกตัวแปรเข้าได้มากที่สุดจำนวน n ตัว ดังนั้นหากในกรณีที่ข้อมูลมีมิติสูง ตัวแบบ LASSO อาจจะไม่เหมาะสม รวมทั้งในกรณีที่ตัวแปรอธิบายมีความสัมพันธ์กันสูง วิธี LASSO มีแนวโน้มที่จะเลือกตัวแปรเพียงตัวแปรเดียวจากกลุ่มตัวแปรอธิบายที่มีความสัมพันธ์กันสูงนี้เข้าตัวแบบ โดยไม่สนใจว่าจะเป็นตัวแปรใดในกลุ่ม ดังนั้นจึงมีผู้คิดค้นดัดแปลงวิธีการประมาณค่าพารามิเตอร์แบบ LASSO ให้มีประสิทธิภาพมากยิ่งขึ้น

2.1.4 การถดถอยโลจิสติกพหุคูณด้วยวิธีแลซโซแบบปรับได้

Zou (2006) ได้ศึกษาคุณสมบัติในการคัดเลือกตัวแปรอธิบายและการประมาณค่าสัมประสิทธิ์การถดถอยของวิธี LASSO ว่าโดยทั่ว ๆ ไปแล้วการคัดเลือกตัวแปรอธิบายโดยวิธี LASSO ไม่คงเส้นคงวา และได้เสนอเงื่อนไขที่จำเป็นที่ทำให้ตัวประมาณวิธี LASSO มีความคงเส้นคงวา และได้เสนอวิธีการประมาณแบบแลซโซปรับได้ (Adaptive LASSO) และได้มีการปรับรูปแบบของตัวประมาณการถดถอยแบบ LASSO โดยจะมีการถ่วงน้ำหนักที่ L_1-นอร์ม จากวิธี LASSO เพื่อให้สอดคล้องกับสมบัติออราเคิล (Oracle properties) ดังนี้

การประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี Adaptive LASSO

ตัวประมาณการถดถอยโดยวิธี Adaptive LASSO คือ

minimize $\quad (\boldsymbol{Y}-\boldsymbol{X\beta})'(\boldsymbol{Y}-\boldsymbol{X\beta})$

ภายใต้ข้อจำกัด $\quad \sum_{j=1}^{p} w_j|\beta_j| \le t$

ซึ่งสามารถเขียนอยู่ในรูปสมการที่สอดคล้องกัน คือ

$$\widehat{\boldsymbol{\beta}}^{adp.lasso} = argmin_{\beta}(\boldsymbol{Y}-\boldsymbol{X\beta})'(\boldsymbol{Y}-\boldsymbol{X\beta}) + \lambda\sum_{j=1}^{p} w_j|\beta_j| \qquad (12)$$

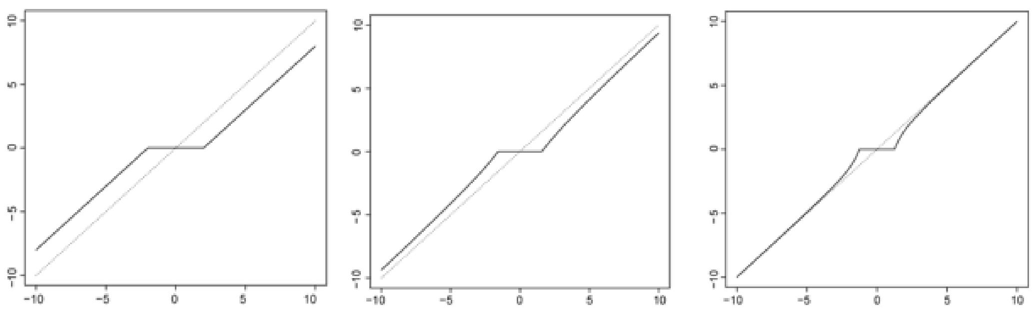
ซึ่ง $\sum_{j=1}^{p} w_j|\beta_j|$ คือ ฟังก์ชันพีนอลตี้ และ w_j คือ เวกเตอร์น้ำหนัก

ในส่วนของการประมาณค่าพารามิเตอร์ของตัวแบบถดถอยโลจิสติกพหุคูณด้วยวิธีแลซโซแบบปรับได้นั้น ก็จะใช้ฟังก์ชันลบลอการิทึมของฟังก์ชันภาวะน่าจะเป็นของตัวแบบถดถอยโลจิสติกพหุคูณเช่นกัน คือ $-\ln \boldsymbol{L} = \sum_{i=1}^{n}\left[-y_i(\boldsymbol{X}_i'\boldsymbol{\beta}) + \ln\left(1+e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}\right)\right]$ จากนั้นทำการประมาณค่าพารามิเตอร์ $\boldsymbol{\beta}$ ที่ทำให้ฟังก์ชันลบลอการิทึมดังกล่าวมีค่าต่ำสุด ซึ่งสามารถเขียนได้เป็น

$$\widehat{\boldsymbol{\beta}}^{\boldsymbol{adp.lasso}} = argmin_{\boldsymbol{\beta}}\left\{\sum_{i=1}^{n}\left[-y_i(\boldsymbol{X}_i'\boldsymbol{\beta}) + \ln\left(1+e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}\right)\right] + \lambda\sum_{j=1}^{p} w_j|\beta_j|\right\} \qquad (13)$$

ซึ่ง $\sum_{j=1}^{p} w_j|\beta_j|$ คือ ฟังก์ชันพีนอลตี้ และ w_j คือ เวกเตอร์น้ำหนัก

เมื่อเปรียบเทียบรูปแบบของสัมประสิทธิ์การถดถอยวิธี LASSO และ Adaptive LASSO ดังภาพที่ 6



$\hat{\beta}^{ols}$ $\hat{\beta}^{ols}$ $\hat{\boldsymbol{\beta}}^{ols}$

LASSO **Adaptive LASSO γ=0.5** **Adaptive LASSO γ=2**

ภาพที่ 6 แสดงรูปแบบของการสัมประสิทธิ์การถดถอยโดยวิธีต่าง ๆ ในกรณี orthonormal design เมื่อ $\boldsymbol{\lambda}$ =2 และ $\boldsymbol{w} = \frac{\mathbf{1}}{|\boldsymbol{\beta}^{ols}|^{\gamma}}$ (Zou, 2006)

ภาพที่ 6 แสดงความสัมพันธ์ระหว่างค่าประมาณสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุด ($\hat{\beta}^{ols}$) และค่าประมาณสัมประสิทธิ์การถดถอยโดยวิธี LASSO, Adaptive LASSO โดย γ=0.5 และแบบ Adaptive LASSO โดย γ=2 ตามลำดับ ในกรณี orthonormal design เมื่อ λ=2 โดยวิธี LASSO แบบดั้งเดิมนั้น $\hat{\beta}_j = \hat{\beta}_j^{ols} - \lambda$ เมื่อ $\hat{\beta}^{ols} > 2$, $\hat{\beta}_j = \hat{\beta}_j^{ols} + \lambda$ เมื่อ $\hat{\beta}^{ols} < -2$ และ $\hat{\beta}_j = 0$ เมื่อ $|\hat{\beta}^{ols}| < 2$ วิธีต่อมาคือการประมาณค่าพารามิเตอร์ Adaptive LASSO $\hat{\beta}_j = \hat{\beta}_j^{ols} - \lambda w_j = \hat{\beta}_j^{ols} - \frac{\lambda}{|\hat{\beta}^{ols}|^{\gamma}}$ เมื่อ $\hat{\beta}^{ols} > 2$ และ $\hat{\beta}_j = \hat{\beta}_j^{ols} + \lambda w_j = \hat{\beta}_j^{ols} + \frac{\lambda}{|\hat{\beta}^{ols}|^{\gamma}}$ เมื่อ $\hat{\beta}^{ols} < -2$ และ $\hat{\beta}_j = 0$ เมื่อ $|\hat{\beta}^{ols}| \leq 2$ เมื่อ γ=0.5 และ γ=2 ตามลำดับ เมื่อ γ มีขนาดใหญ่ จะทำให้ $\lambda w_j = \frac{\lambda}{|\hat{\beta}^{ols}|^{\gamma}}$ มีขนาดเล็ก นั่นคือ $\hat{\beta}_j$ มีค่าใกล้เคียงกับ $\hat{\beta}_j^{ols}$

อย่างไรก็ตาม ส่วนที่สำคัญที่สุดสำหรับวิธีการของ Adaptive LASSO คือเวกเตอร์ค่าถ่วงน้ำหนัก w_j ซึ่งเป็นส่วนที่ทำให้การคัดเลือกตัวแปรอธิบายเข้าสู่ตัวแบบมีความคงเส้นคงวาสอดคล้องกับคุณสมบัติของตัวประมาณ โดยปกติแล้วนั้น สำหรับการวิเคราะห์การถดถอยแบบโลจิสติกทั่วไปสามารถใช้ค่าถ่วงน้ำหนักที่คำนวณจากตัวประมาณวิธี ML ได้ แต่ในกรณีที่ข้อมูลมีมิติสูง จะไม่สามารถหาตัวประมาณของสัมประสิทธิ์ถดถอยจากวิธี ML ดังนั้น การเลือกใช้ค่าถ่วงน้ำหนักจากตัวประมาณอื่น ๆ ที่สามารถแก้ปัญหาที่พบจากการวิเคราะห์ข้อมูลมิติสูง จึงจะสามารถหาตัวประมาณสำหรับวิธี Adaptive LASSO นี้ได้ เช่น การนำตัวประมาณจากวิธี LASSO มาใช้เป็นค่าถ่วงน้ำหนักสำหรับวิธี Adaptive LASSO ซึ่งมีรายละเอียดดังนี้

2.1.5 การถดถอยโลจิสติกพหุคูณด้วยวิธี Two-stage Adaptive LASSO หรือวิธี Adaptive LASSO โดยใช้ตัวประมาณ LASSO เป็นค่าถ่วงน้ำหนัก

Guo และคณะ (2015) ได้เสนอวิธีการคัดเลือกตัวแปรโดยใช้วิธีการของ LASSO ในตัวแบบถดถอยโลจิสติกที่พัฒนามาจากวิธี Adaptive LASSO ที่เรียกว่าวิธี Two-stage Hybrid โดยวิธี Two-stage Hybrid นี้ มีลักษณะการคำนวณแบบเดียวกันกับวิธีการของ Adaptive LASSO เพียงแต่จะใช้เวกเตอร์น้ำหนัก w_j ที่คำนวณมาจากวิธีการของแลซโซ ซึ่งมีขั้นตอนดังนี้

1. เมื่อนำตัวแปรอธิบายและตัวแปรตามเข้าสู่ระบบแล้ว จะใช้ตัวประมาณ LASSO ในการหาตัวประมาณสัมประสิทธิ์ถดถอยขั้นต้น (initial estimator) คือ $\widehat{\boldsymbol{\beta}}^{\boldsymbol{lasso}} = argmin_{\boldsymbol{\beta}}\left\{\sum_{i=1}^{n}\left[-y_i(\boldsymbol{X}_{\boldsymbol{i}}'\boldsymbol{\beta}) + \ln\left(1 + e^{(X_i'\beta)}\right)\right] + \lambda_1\sum_{j=1}^{p}|\beta_j|\right\}$
2. ทำการแบ่งเซตของตัวแปรจากค่าสัมประสิทธิ์ถดถอยที่ประมาณได้จากข้อ 1 โดยกำหนดให้ $J_1 = \{j,\ \hat{\beta}_j \neq 0\}$ เป็นเซตของตัวแปรที่มีค่าสัมประสิทธิ์ถดถอยไม่เท่ากับ 0 และกำหนดให้ $J_1^* = \{j,\ \hat{\beta}_j = 0\}$ เป็นเซตของตัวแปรที่มีค่าสัมประสิทธิ์ถดถอยเท่ากับ 0
3. คำนวณเวกเตอร์น้ำหนัก $\widehat{\boldsymbol{w}} = 1/\left|\hat{\beta}(J_1)\right|^{\gamma}$ ที่จะนำไปใช้ในการคำนวณด้วยวิธีแลซโซแบบปรับได้
4. กำหนดค่าสัมประสิทธิ์ถดถอยของตัวแปรที่อยู่ในเซต $\boldsymbol{J_1^*}$ ให้เท่ากับ 0 ในขั้นตอนของวิธีแลซโซแบบปรับได้ $\hat{\beta}(J_1^*) = 0$
5. คำนวณค่าประมาณค่าสัมประสิทธิ์ถดถอยด้วยวิธีแลซโซแบบปรับได้ ดังนี้ $\widehat{\boldsymbol{\beta}}^{\boldsymbol{adp.lasso}} = argmin_{\boldsymbol{\beta}}\left\{\sum_{i=1}^{n}\left[-y_i(\boldsymbol{X}_{\boldsymbol{i}}'\boldsymbol{\beta}) + \ln\left(1 + e^{(X_i'\beta)}\right)\right] + \lambda_2\sum_{j=1}^{p}\widehat{\boldsymbol{w}}_j|\beta_j|\right\}$
6. หลังจากคำนวณค่า $\widehat{\boldsymbol{\beta}}$ ด้วยวิธีแลซโซแบบปรับได้แล้ว กำหนดให้ $J_2 = \{j,\ \hat{\beta}_j \neq 0\}$ เป็นเซตของตัวแปรที่มีค่าสัมประสิทธิ์ถดถอยไม่เท่ากับ 0
7. ทำการสร้างเซตใหม่ ซึ่งเป็นเซตของตัวแปรที่เข้าสู่ตัวแบบถดถอยที่ได้จากการคำนวณ คือ $S = X(J_2)$

แต่นอกจากการใช้ตัวประมาณจากวิธี LASSO มาเป็นค่าถ่วงน้ำหนักในวิธี Adaptive LASSO แล้ว ตัวประมาณอื่น ๆ ก็สามารถนำมาใช้ในการคำนวณเวกเตอร์ถ่วงน้ำหนักได้เช่นกัน ดังนั้น ทางผู้วิจัยจึงสนใจที่จะนำตัวประมาณจากวิธีอื่น ๆ มาใช้ในการคำนวณเวกเตอร์ค่าถ่วงน้ำหนักในวิธี Adaptive LASSO เพื่อเปรียบเทียบถึงประสิทธิภาพและความแตกต่างระหว่างแต่ละวิธี เช่น การใช้ตัวประมาณจากวิธีริดจ์มาคำนวณค่าเวกเตอร์ถ่วงน้ำหนัก หรือการใช้ตัวประมาณจากวิธีสไตน์มาคำนวณค่าเวกเตอร์ถ่วงน้ำหนักในวิธี Adaptive LASSO ซึ่งจะมีรายละเอียดแต่ละวิธีดังนี้

2.1.6 การถดถอยโลจิสติกพหุคูณด้วยวิธี Adaptive LASSO โดยใช้ตัวประมาณริดจ์ (Ridge estimator) เป็นค่าถ่วงน้ำหนัก

จากสมการที่ (8) ตัวประมาณพารามิเตอร์ด้วยวิธีริดจ์สำหรับตัวแบบถดถอยโลจิสติกพหุคูณสามารถนำมาใช้ในการคำนวณค่าถ่วงน้ำหนักสำหรับการประมาณค่าด้วยวิธี Adaptive LASSO ได้ ซึ่งมีขั้นตอนดังนี้

1. เมื่อนำตัวแปรอธิบายและตัวแปรตามเข้าสู่ระบบแล้ว ทำการหาตัวประมาณสัมประสิทธิ์ถดถอยจากวิธีริดจ์ คือ $\widehat{\boldsymbol{\beta}}^{\boldsymbol{ridge}} = argmin_{\boldsymbol{\beta}}\left\{\sum_{i=1}^{n}\left[-y_i(\boldsymbol{X}_{\boldsymbol{i}}'\boldsymbol{\beta}) + \ln\left(1 + e^{(\boldsymbol{X}_{\boldsymbol{i}}'\boldsymbol{\beta})}\right)\right] + \lambda\sum_{j=1}^{p}\beta_j^2\right\}$

2. คำนวณเวกเตอร์น้ำหนัก $\widehat{\boldsymbol{w}} = 1/\left|\widehat{\boldsymbol{\beta}}^{\boldsymbol{ridge}}\right|^{\gamma}$ ด้วยค่าประมาณพารามิเตอร์จากวิธีริดจ์หรือ $\widehat{\boldsymbol{\beta}}^{\boldsymbol{ridge}}$ ในวิธี Adaptive LASSO

3. คำนวณค่าประมาณค่าสัมประสิทธิ์ถดถอยด้วยวิธี Adaptive LASSO ดังนี้ $\widehat{\boldsymbol{\beta}}^{\boldsymbol{adp.lasso}} = argmin_{\boldsymbol{\beta}}\left\{\sum_{i=1}^{n}\left[-y_i(\boldsymbol{X}_{\boldsymbol{i}}'\boldsymbol{\beta}) + \ln\left(1 + e^{(\boldsymbol{X}_{\boldsymbol{i}}'\boldsymbol{\beta})}\right)\right] + \lambda_2\sum_{j=1}^{p}\widehat{\boldsymbol{w}}_j\left|\beta_j\right|\right\}$

ก็จะได้ว่า $\widehat{\boldsymbol{\beta}}^{\boldsymbol{adp.lasso}}$ ที่คำนวณได้เป็นค่าประมาณพารามิเตอร์ของตัวแบบถดถอยโลจิสติกจากวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณจากวิธีริดจ์

2.1.7 การถดถอยโลจิสติกพหุคูณด้วยวิธี Adaptive LASSO โดยใช้ตัวประมาณสไตน์-ริดจ์ (Stein-Ridge estimator) เป็นค่าถ่วงน้ำหนัก

James และ Stein (Stein, 1956) และ (W. James, 1961) ได้คิดค้นตัวประมาณสไตน์ขึ้นมา ซึ่งตัวประมาณสไตน์นั้นก็เป็นหนึ่งในตัวประมาณที่นิยมใช้กันในการเรกูลาไรซ์เช่นกันเพราะสามารถแก้ไขปัญหาความสัมพันธ์เชิงเส้นแบบพหุได้ โดยที่ตัวประมาณสไตน์สำหรับตัวแบบถดถอยโลจิสติกพหุคูณเป็นดังนี้ (Schaefer, 1986)

$$\widehat{\boldsymbol{\beta}}^{\boldsymbol{stein}} = c\widehat{\boldsymbol{\beta}}_{\boldsymbol{ML}} \tag{14}$$

โดยที่ $\boldsymbol{c}$ คือ พารามิเตอร์ชริงค์เกจ (Shrinkage parameter) สามารถคำนวณได้จาก

$$c = \frac{\widehat{\boldsymbol{\beta}}_{\boldsymbol{ML}}{}'\widehat{\boldsymbol{\beta}}_{\boldsymbol{ML}}}{\widehat{\boldsymbol{\beta}}_{\boldsymbol{ML}}{}'\widehat{\boldsymbol{\beta}}_{\boldsymbol{ML}} + tr(\boldsymbol{X}'\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}} \text{ และ } 0 < c < 1$$

โดยที่ $\widehat{\boldsymbol{W}}$ คือเมทริกซ์ที่มีค่าบนเส้นทแยงมุมเป็น $\widehat{w}_i = \widehat{\pi}_i(1 - \widehat{\pi}_i)$ เมื่อ $\widehat{\pi}_i$ คือค่าความน่าจะเป็นที่คำนวณได้จากการประมาณค่าด้วยวิธี ML

อย่างไรก็ตาม จะเห็นว่าตัวประมาณสไตน์นั้นเป็นตัวประมาณที่ขึ้นอยู่กับค่าประมาณพารามิเตอร์จากวิธี ML ซึ่งไม่สามารถหาค่าได้ในกรณีที่ข้อมูลมีมิติสูง จึงต้องมีการปรับรูปของตัวประมาณสไตน์ใหม่ โดยการใช้ตัวประมาณจากวิธีริดจ์แทนตัวประมาณจากวิธี ML จึงได้ตัวประมาณวิธีสไตน์-ริดจ์ที่จะนำไปใช้ในการคำนวณค่าถ่วงน้ำหนักในวิธี Adaptive LASSO สำหรับการวิเคราะห์การถดถอยโลจิสติกที่ข้อมูลมีมิติสูง ได้เป็นดังนี้

$$\widehat{\boldsymbol{\beta}}^{stein-ridge} = c\widehat{\boldsymbol{\beta}}^{ridge} \tag{15}$$

โดยที่ $\boldsymbol{c}$ คือ พารามิเตอร์ชริงค์เกจ (Shrinkage parameter) สามารถคำนวณได้จาก

$$c = \frac{\widehat{\boldsymbol{\beta}}^{ridge'}\widehat{\boldsymbol{\beta}}^{ridge}}{\widehat{\boldsymbol{\beta}}^{ridge'}\widehat{\boldsymbol{\beta}}^{ridge} + tr(\boldsymbol{X}'\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}} \text{ และ } 0 < c < 1$$

ตัวประมาณพารามิเตอร์ด้วยวิธีสไตน์-ริดจ์สำหรับตัวแบบถดถอยโลจิสติกพหุคูณสามารถนำมาใช้ในการคำนวณค่าถ่วงน้ำหนักสำหรับการประมาณค่าด้วยวิธี Adaptive LASSO ได้ ซึ่งมีขั้นตอนดังนี้

1. เมื่อนำตัวแปรอธิบายและตัวแปรตามเข้าสู่ระบบแล้ว ทำการหาตัวประมาณสัมประสิทธิ์ถดถอยจากวิธีริดจ์ คือ $\widehat{\boldsymbol{\beta}}^{ridge} = argmin_{\boldsymbol{\beta}}\left\{\sum_{i=1}^{n}\left[-y_i(\boldsymbol{X}_i'\boldsymbol{\beta}) + \ln\left(1 + e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}\right)\right] + \lambda\sum_{j=1}^{p}\beta_j^2\right\}$
2. ทำการคำนวณหาค่า c และคำนวณหาตัวประมาณสัมประสิทธิ์ถดถอยโลจิสติกด้วยวิธีสไตน์-ริดจ์จาก $\widehat{\boldsymbol{\beta}}^{stein-ridge} = c\widehat{\boldsymbol{\beta}}^{ridge}$
3. คำนวณเวกเตอร์น้ำหนัก $\widehat{\boldsymbol{w}} = 1/\left|\widehat{\boldsymbol{\beta}}^{stein}\right|^{\gamma}$ ด้วยค่าประมาณพารามิเตอร์จากวิธีสไตน์-ริดจ์หรือ $\widehat{\boldsymbol{\beta}}^{stein-ridge}$ ในวิธี Adaptive LASSO
4. คำนวณค่าประมาณค่าสัมประสิทธิ์ถดถอยด้วยวิธี Adaptive LASSO ดังนี้ $\widehat{\boldsymbol{\beta}}^{adp.lasso} = argmin_{\boldsymbol{\beta}}\left\{\sum_{i=1}^{n}\left[-y_i(\boldsymbol{X}_i'\boldsymbol{\beta}) + \ln\left(1 + e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}\right)\right] + \lambda_2\sum_{j=1}^{p}\widehat{\boldsymbol{w}}_j\left|\beta_j\right|\right\}$

ก็จะได้ว่า $\widehat{\boldsymbol{\beta}}^{adp.lasso}$ ที่คำนวณได้นั้นเป็นค่าประมาณพารามิเตอร์ของตัวแบบถดถอยโลจิสติกจากวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณจากวิธีสไตน์-ริดจ์

2.1.8 การหาพารามิเตอร์การปรับแต่ง (Tuning parameter) โดยวิธี Cross-validation

วิธีนี้เป็นวิธีที่นิยมในการทำวิจัยเพื่อใช้ในการวัดประสิทธิภาพของตัวแบบโดยใช้ค่าความถูกต้อง (Accuracy) เนื่องจากผลที่ได้มีความน่าเชื่อถือ โดยจะมีการแบ่งข้อมูลออกเป็น 2 ส่วน คือ

ชุดข้อมูลฝึกฝน (Training set) และชุดข้อมูลทดสอบ (Test set) โดย Cross-validation มีหลายวิธี แต่วิธีที่ใช้กันอย่างแพร่หลายคือ K-fold Cross validation ซึ่งมีขั้นตอนคือ เริ่มแรกจะนำชุดข้อมูล Training set มาแบ่งข้อมูลออกเป็น K ส่วนเท่า ๆ กัน ในการทดลองครั้งแรกจะให้ข้อมูลชุดที่ 1 เป็น Test set และชุดข้อมูล K-1 ที่เหลือเป็น Training set จากนั้นประมาณค่าพารามิเตอร์ที่สนใจจาก Training set แล้วนำค่าที่ได้ไปทำนายใน Test set แล้วคำนวณหาค่าความถูกต้องในการทำนายชุดที่ 1 ในการทดลองครั้งที่สองจะให้ข้อมูลชุดที่ 2 เป็น Test set และชุดข้อมูล K-1 ที่เหลือเป็น Training set จากนั้นประมาณค่าพารามิเตอร์ที่สนใจจาก Training set แล้วนำค่าที่ได้ไปทำนายใน Test set แล้วคำนวณหาค่าความถูกต้องในการทำนายชุดที่ 2 ทำแบบนี้จนกระทั่งข้อมูลทั้ง K ชุดถูกนำมาเป็น Test set แล้วหาค่าเฉลี่ยของความถูกต้องในการทำนายทุกชุด โดย Tuning parameter คือ พารามิเตอร์ ณ ตำแหน่งที่ทำให้ค่าความถูกต้องในการทำนายต่ำที่สุด ข้อดีของการเลือกสุ่มข้อมูลแบบ K-fold Cross validation คือ ข้อมูลทุกตัวจะถูกนำมาเป็น Training set และ Test set แต่ข้อเสียคือ ใช้เวลาในการทดลองมากเนื่องจากต้องทดลองข้อมูลทั้งหมด K ครั้ง โดยส่วนใหญ่นิยมกำหนดค่า K เป็น 5 หรือ 10

2.2 งานวิจัยที่เกี่ยวข้อง

ในปี 2008 Huang, Ma และ Zhang (2008) ได้เสนอวิธีการประมาณค่าสัมประสิทธิ์การถดถอยในตัวแบบถดถอยโลจิสติกส์ที่ข้อมูลมีมิติสูงที่ชื่อว่า Iterated LASSO ซึ่งเป็นวิธีการประมาณค่าสัมประสิทธิ์ถดถอยที่พัฒนามาจากวิธี Adaptive LASSO ที่มีวิธีการคือ ในขั้นแรก จะใช้วิธี LASSO ในการหาตัวประมาณสัมประสิทธิ์ถดถอยขั้นต้น (initial estimator) และลดจำนวนมิติของตัวแบบให้มีขนาดเล็กลงด้วยการพิจารณาจากค่าสัมประสิทธิ์นั้นว่าเป็น 0 หรือไม่ จากนั้นจะทำการคำนวณค่า weights ของตัวแปรที่มีนัยสำคัญและแทนค่าลงในฟังก์ชันพีนอลตี้ ℓ_1 จากนั้นก็ใช้วิธี Adaptive LASSO เพื่อประมาณค่าสัมประสิทธิ์ถดถอยและพิจารณาเลือกเซตของตัวแปรอธิบายที่สำคัญเข้าสู่ตัวแบบอีกครั้ง จึงได้เป็นตัวแบบที่มีตัวแปรที่สำคัญเท่านั้น จากนั้น ทางผู้วิจัยจะทำการเปรียบเทียบประสิทธิภาพของวิธีการ Iterated LASSO กับวิธีการประมาณค่าแบบ LASSO โดยกำหนดจำนวนตัวแปรอธิบายเท่ากับ 500 ขนาดตัวอย่างเท่ากับ 100 และ 200 มีสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอธิบายตัวที่ i และ j คือ $\rho^{|i-j|}$ เท่ากับ 0, 0.3 และ 0.5 พบว่า วิธี Iterated LASSO มีประสิทธิภาพในการทำนายสูงกว่าวิธี LASSO และมีจำนวนตัวแปรอธิบายเข้าสู่ตัวแบบใกล้เคียงกับที่กำหนดไว้มากกว่าวิธี LASSO และเมื่อนำวิธี Iterated LASSO และวิธี LASSO ไปเปรียบเทียบประสิทธิภาพในการวิเคราะห์ถดถอยโลจิสติกส์กับข้อมูลผู้ป่วยโรคมะเร็งเต้านม เพื่อวิเคราะห์การแพร่กระจายของ

เซลล์มะเร็งไปยังอวัยวะอื่น ๆ ภายในร่างกาย พบว่า วิธี Iterated LASSO สามารถเลือกจำนวนตัวแปรอธิบายเข้าสู่ตัวแบบน้อยกว่าวิธี LASSO และตัวแปรอธิบายที่เลือกด้วยวิธี Iterated LASSO นั้นยังมีความเหมาะสมในการวิเคราะห์ผลลัพธ์ทางการรักษาได้ดีกว่าวิธี LASSO ด้วย

ในปี 2015 Guo และคณะ (2015) ได้เสนอวิธีการคัดเลือกตัวแปรโดยใช้วิธีการของ LASSO ในตัวแบบถดถอยโลจิสติกส์สองวิธี ได้แก่ วิธี Two-stage Hybrid ซึ่งเป็นวิธีการประมาณค่าสัมประสิทธิ์ถดถอยที่พัฒนามาจากวิธี Adaptive LASSO ที่มีวิธีการคือ ในขั้นแรก จะใช้วิธี LASSO ในการหาตัวประมาณสัมประสิทธิ์ถดถอยขั้นต้น (initial estimator) และลดจำนวนมิติของตัวแบบให้มีขนาดเล็กลงด้วยการพิจารณาจากค่าสัมประสิทธิ์นั้นว่าเป็น 0 หรือไม่ จากนั้นจะทำการคำนวณค่า weights ของตัวแปรที่มีนัยสำคัญและแทนค่าลงในฟังก์ชันพีนอลตี้ ℓ_1 จากนั้นก็ใช้ Adaptive LASSO เพื่อประมาณค่าสัมประสิทธิ์ถดถอยและพิจารณาเลือกเซตของตัวแปรอธิบายที่สำคัญเข้าสู่ตัวแบบอีกครั้ง จึงได้เป็นตัวแบบที่มีตัวแปรที่สำคัญเท่านั้น และอีกวิธีคือ วิธี Bootstrap Ranking ซึ่งมีวิธีการคือ เมื่อทำการสร้างข้อมูลมาแล้วจะใช้วิธี LASSO มาประมาณค่าสัมประสิทธิ์ถดถอยขั้นต้นของแต่ละตัวแปรอธิบาย จากนั้นจะนำค่าสัมประสิทธิ์ที่ได้มาคำนวณเป็นคะแนนความสำคัญของแต่ละตัวแปรแล้วทำการเรียงอันดับ และประมาณค่าจุดตัดเพื่อที่จะใช้ในการเลือกตัวแปรที่มีอันดับสูงกว่าจุดตัดมาเป็นตัวแปรที่มีนัยสำคัญภายในตัวแบบนั้น จากนั้น ทางผู้วิจัยจะทำการเปรียบเทียบประสิทธิภาพของวิธีการทั้งสองวิธีกับวิธีการคัดเลือกตัวแปรต่าง ๆ ที่เคยมีการใช้มาก่อน ไม่ว่าจะเป็น Stepwise Selection, Stability Selection, LASSO, และ Bolasso โดยกำหนดจำนวนตัวแปรอธิบายเท่ากับ 100 และ 200 ขนาดตัวอย่างเท่ากับ 50, 100, 200, 300 และ 500 โดยจะเปรียบเทียบประสิทธิภาพของแต่ละวิธีจากค่า True Positive Rate (TPR), False Positive Rate (FPR), Area Under Curve (AUC) ของแต่ละวิธี พบว่า ทั้งสองวิธีที่ผู้วิจัยเสนอนั้น มีประสิทธิภาพโดยรวมดีกว่าหลาย ๆ วิธีที่นำมาเปรียบเทียบกัน และเมื่อนำทั้งสองวิธีนี้ไปใช้กับข้อมูลแบบสอบถามเกี่ยวกับปัจจัยการระบาดของโรคไวรัสตับอักเสบบีจากจำนวนผู้ทำแบบสอบถาม 5,357 คน และทำการเปรียบเทียบกับวิธีการก่อน ๆ หน้า พบว่า ทั้งสองวิธีที่ทางผู้วิจัยได้เสนอมีประสิทธิภาพในการคัดเลือกตัวแปรได้ดีกว่าวิธีการก่อนหน้าเช่นกัน

ในปี 2017 Farghali และ Abo-El-Hadid (2017) ได้ทำการประเมินประสิทธิภาพของตัวประมาณในตัวแบบถดถอยโลจิสติกที่มีปัญหา Multicollinearity ระหว่างตัวแปรอธิบายภายในตัวแบบ ได้แก่ ตัวประมาณ Maximum Likelihood (ML) เป็นตัวประมาณที่ใช้ในการประมาณค่าสัมประสิทธิ์ถดถอยของตัวแบบถดถอยโลจิสติก แต่จะเอนเอียงและมีความแปรปรวนสูงในกรณีที่มี

ปัญหา Multicollinearity เกิดขึ้น, ตัวประมาณ Stein เป็นตัวประมาณที่มีข้อดีคือสามารถลดค่าความคลาดเคลื่อนเฉลี่ยกำลังสองของค่าประมาณค่าสัมประสิทธิ์ถดถอยได้ แต่มีข้อเสียคือตัวประมาณ Stein เป็นตัวประมาณที่ใช้ ML ในการคำนวณ จึงทำให้ค่าที่ได้มีความเอนเอียงเช่นกัน, ตัวประมาณ Ridge เป็นตัวประมาณที่มีข้อดีคือค่าประมาณค่าสัมประสิทธิ์ถดถอยที่ได้มีค่าความคลาดเคลื่อนมาตรฐานต่ำกว่าตัวประมาณ ML แต่มีข้อเสียคือไม่มีความแน่นอนในการเลือกตัวแปร Ridge ในการคำนวณ และตัวประมาณ Liu ซึ่งเป็นตัวประมาณที่รวมข้อดีของทั้งตัวประมาณ Stein และตัวประมาณ Ridge เข้าด้วยกัน โดยทำการกำหนดขนาดตัวอย่างเป็น 50, 70, 100, 150 และ 200, กำหนดจำนวนตัวแปรคือ 2 และ 3 ตัวแปร โดยที่มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรที่แตกต่างกันคือ 0.75, 0.85 และ 0.95 มีเกณฑ์ที่ใช้ในการพิจารณาคือ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์, ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของค่าความน่าจะเป็นที่สังเกตได้กับค่าความน่าจะเป็นที่ทำนายได้ ซึ่งสามารถสรุปได้ว่า ในกรณีที่มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร 0.75 ทั้งสี่ตัวประมาณให้ค่าความคลาดเคลื่อนที่ไม่ต่างกัน ในขณะที่เมื่อเพิ่มค่าสัมประสิทธิ์สหสัมพันธ์ให้สูงขึ้น พบว่า ตัวประมาณ Liu มีค่าความคลาดเคลื่อนที่ต่ำกว่าตัวประมาณตัวอื่น

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยนี้เป็นการวิจัยเชิงทดลองโดยวิธีการจำลอง (Simulation study) ซึ่งมีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพการประมาณค่าสัมประสิทธิ์การถดถอยและการคัดเลือกตัวแปรของการเรกูลาไรซ์โดยใช้วิธีแลซโซและวิธีแลซโซแบบปรับได้ของตัวแบบถดถอยโลจิสติกในกรณีที่ข้อมูลมีมิติสูงทั้ง 4 วิธี ได้แก่ วิธีแลซโซ, วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณริดจ์, วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณแลซโซ, และวิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณสไตน์-ริดจ์ โดยในการเปรียบเทียบประสิทธิภาพของแต่ละวิธีนั้น จะพิจารณาจากเกณฑ์ต่าง ๆ คือ ความถูกต้องในการทำนาย ความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอย และความถูกต้องในการคัดเลือกตัวแปรอธิบายเข้าในตัวแบบ การวิจัยครั้งนี้ได้จำลอง ข้อมูลด้วยโปรแกรม R เพื่อสร้างข้อมูลตัวแปรอธิบายหรือตัวแปรอธิบาย และตัวแปรตามหรือตัวแปรตอบสนอง และกำหนดสถานการณ์ต่าง ๆ ตามขอบเขตของงานวิจัย โดยในแต่ละสถานการณ์มีการทำซ้ำ 500 ครั้ง ซึ่งมีขอบเขตของงานวิจัยและขั้นตอนในการวิจัย ดังนี้

3.1 ขอบเขตของการวิจัยและการทดลอง

1. ขนาดตัวอย่าง n = 100 และ 200

2. จำนวนตัวแปรอธิบายแบบต่อเนื่อง p ตัวแปรมีจำนวนเป็น 2, 3, และ 4 เท่าของขนาดตัวอย่าง $\boldsymbol{n}$ และจำนวนตัวแปรอธิบายแบบไม่ต่อเนื่องมีจำนวน 4 ตัวแปรทุกกรณี รวมจำนวนตัวแปรอธิบายทั้งสิ้น $p+4$ ตัวแปร

3. เมทริกซ์ของตัวแปรอธิบาย $(\boldsymbol{X}) = (\boldsymbol{X}^*_{n\times p},\ \boldsymbol{D}_{n\times 4})$ มีขนาด $n \times (p+4)$ โดยที่ $\boldsymbol{X}^* = (X_1, X_2, \dots, X_p)$ เป็นเมทริกซ์ของตัวแปรอธิบายที่เป็นตัวแปรแบบต่อเนื่องจำนวน p ตัวแปร และ $\boldsymbol{D} = (D_1, D_2, D_3, D_4)$ เป็นเมทริกซ์ของตัวแปรแบบไม่ต่อเนื่องจำนวน 4 ตัวแปรที่มีการแจกแจงแบบเบอร์นูลลีที่มีพารามิเตอร์เป็น 0.8, 0.6, 0.4, และ 0.2 ตามลำดับ ผู้วิจัยจำลองข้อมูล $\boldsymbol{X}^*$ ให้มีการแจกแจงแบบปรกติหลายตัวแปรที่เวกเตอร์ค่าเฉลี่ยเป็นเวกเตอร์ $\boldsymbol{0}$ และเมทริกซ์ความแปรปรวนร่วม $\boldsymbol{\Sigma}$ ที่แบ่งเป็น 2 กรณี คือ

กรณีที่ 1 ตัวแปรอธิบายจะถูกแบ่งเป็น 2 กลุ่มที่เป็นอิสระต่อกัน แต่มีความสัมพันธ์กันภายในกลุ่ม โดยที่ในกลุ่มที่ 1 จะมีจำนวน 15 ตัวแปรที่สัมพันธ์กันเอง และกลุ่มที่ 2 คือจำนวนตัวแปร

ที่เหลือมีความสัมพันธ์กันเอง นั่นคือ เมทริกซ์ของตัวแปรอธิบายแบบต่อเนื่องกรณีที่ 1 ใช้สัญลักษณ์แทนด้วย $\boldsymbol{X}^*_{\boldsymbol{ind}} = \left[\boldsymbol{X}_{\boldsymbol{1}_{(n\times 15)}} \vdots \boldsymbol{X}_{\boldsymbol{2}_{(n\times(p-15))}}\right]$ โดยที่ $\boldsymbol{X}^*_{\boldsymbol{ind}} \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma_a})$ ซึ่งเมทริกซ์ความแปรปรวนร่วมของตัวแปรอธิบาย $\boldsymbol{X}^*_{\boldsymbol{ind}}$ สำหรับกรณีที่ 1 คือ

$$\boldsymbol{\Sigma_a} = \begin{bmatrix} \boldsymbol{\Sigma_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma_2} \end{bmatrix}_{p\times p}$$

โดยที่ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรตัวที่ X_k และ X_j ของเมทริกซ์ $\boldsymbol{X_1}$ คือ $\rho^{|k-j|}$; $j, k = 1,2, \ldots, 15$ และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรตัวที่ X_k และ X_j ของเมทริกซ์ $\boldsymbol{X_2}$ คือ $\rho^{|k-j|}$; $j, k = 1,2, \ldots, (p-15)$ ดังนี้

$$\boldsymbol{\Sigma_1} = \mathbf{Cov}(\boldsymbol{X_1}) = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{14} \\ \rho & 1 & \rho & \cdots & \rho^{13} \\ \rho^2 & \rho & 1 & \cdots & \rho^{12} \\ \vdots & \vdots & \vdots & \ddots & \rho \\ \rho^{14} & \rho^{13} & \cdots & \rho & 1 \end{bmatrix}_{15\times 15} \quad \text{และ}$$

$$\boldsymbol{\Sigma_2} = \mathbf{Cov}(\boldsymbol{X_2}) = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{p-14} \\ \rho & 1 & \rho & \cdots & \rho^{p-13} \\ \rho^2 & \rho & 1 & \cdots & \rho^{p-12} \\ \vdots & \vdots & \vdots & \ddots & \rho \\ \rho^{p-14} & \rho^{p-13} & \cdots & \rho & 1 \end{bmatrix}_{(p-15)\times(p-15)}$$

กรณีที่ 2 ตัวแปรอธิบาย $X_1, \ldots, X_p$ มีความสัมพันธ์กัน โดยที่เมทริกซ์ของตัวแปรอธิบายแบบต่อเนื่องกรณีที่ 2 จะใช้สัญลักษณ์แทนด้วย $\boldsymbol{X}^*_{\boldsymbol{cor}_{(p\times p)}}$ โดยที่ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง X_k และ X_j คือ $\rho^{|k-j|}$; $j, k = 1,2, \ldots, p$ ดังนั้น $\boldsymbol{X}^*_{\boldsymbol{cor}} \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma_b})$ เมื่อเมทริกซ์ความแปรปรวนร่วม คือ

$$\boldsymbol{\Sigma_b} = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{p-1} \\ \rho & 1 & \rho & \cdots & \rho^{p-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{p-3} \\ \vdots & \vdots & \vdots & \ddots & \rho \\ \rho^{p-1} & \rho^{p-2} & \cdots & \rho & 1 \end{bmatrix}_{p\times p}$$

4. กำหนดเวกเตอร์ของสัมประสิทธิ์การถดถอย $\boldsymbol{\beta}_{(p+4)\times 1} = (\boldsymbol{\beta}_{\boldsymbol{X}^*}, \boldsymbol{\beta_D})$ จากตัวแบบการถดถอยโลจิสติกพหุคูณที่เป็นตัวแบบบางเบา ในแต่ละรูปแบบดังนี้

รูปแบบที่ 1 กำหนดให้มีตัวแปรจำนวน 9 ตัวแปรที่มีค่าพารามิเตอร์ที่ไม่เป็น 0 แบ่งเป็นตัวแปรแบบต่อเนื่อง 5 ตัวแปร คือ $X_1, X_2, \ldots, X_5$ และตัวแปรแบบไม่ต่อเนื่อง 4 ตัวแปร คือ D_1, D_2, D_3, D_4 โดยที่ค่าพารามิเตอร์ที่ไม่เป็น 0 ของแต่ละตัวแปรเป็นดังนี้

$\boldsymbol{\beta}_{\boldsymbol{X}^*}^{(1)} = [2.5, 2.5, 2.5, 2.5, 2.5,0,0, \ldots ,0]$ และ

$\boldsymbol{\beta}_{\boldsymbol{D}}^{(1)} = [-0.1, -0.1, -0.1, -0.1]$ ตามลำดับ

รูปแบบที่ 2 กำหนดให้มีตัวแปรจำนวน 14 ตัวแปรที่มีค่าพารามิเตอร์ที่ไม่เป็น 0 แบ่งเป็นตัวแปรแบบต่อเนื่อง 10 ตัวแปร คือ $X_1, X_2, \ldots , X_{10}$ และตัวแปรแบบไม่ต่อเนื่อง 4 ตัวแปร คือ D_1, D_2, D_3, D_4 โดยที่ค่าพารามิเตอร์ที่ไม่เป็น 0 ของแต่ละตัวแปรเป็นดังนี้

$\boldsymbol{\beta}_{\boldsymbol{X}^*}^{(2)} = [2.5, 2.5, 2.5, 2.5, 2.5,2,2,2,2,2,0,0, \ldots ,0]$ และ

$\boldsymbol{\beta}_{\boldsymbol{D}}^{(2)} = [-0.1, -0.1, -0.1, -0.1]$ ตามลำดับ

รูปแบบที่ 3 กำหนดให้มีตัวแปรจำนวน 19 ตัวแปรที่มีค่าพารามิเตอร์ที่ไม่เป็น 0 แบ่งเป็นตัวแปรแบบต่อเนื่องอีก 15 ตัวแปร คือ $X_1, X_2, \ldots , X_{15}$ และตัวแปรแบบไม่ต่อเนื่อง 4 ตัวแปร คือ D_1, D_2, D_3, D_4 โดยที่ค่าพารามิเตอร์ที่ไม่เป็น 0 ของแต่ละตัวแปรเป็นดังนี้

$\boldsymbol{\beta}_{\boldsymbol{X}^*}^{(3)} = [2.5,2.5,2.5,2.5,2.5,2,2,2,2,2,1.5,1.5,1.5,1.5,1.5,0,0, \ldots ,0]$ และ

$\boldsymbol{\beta}_{\boldsymbol{D}}^{(3)} = [-0.1, -0.1, -0.1, -0.1]$ ตามลำดับ

5. เวกเตอร์ตัวแปรตามที่เป็นตัวแปรทวิภาค ซึ่งมี 2 กลุ่มคือ 0 และ 1 จะสร้างจากการคำนวณความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจจากตัวแบบการถดถอยโลจิสติกพหุคูณตามเมทริกซ์ของตัวแปรอธิบายและเวกเตอร์ของค่าสัมประสิทธิ์ถดถอยที่ได้กำหนดไว้ข้างต้น

$$\pi_i(\boldsymbol{x}) = P(y_i = 1|\boldsymbol{X_i}) = \frac{\exp(\beta_0 + \boldsymbol{X}_i'\boldsymbol{\beta})}{1+\exp(\beta_0 + \boldsymbol{X}_i'\boldsymbol{\beta})} \; ; \; i = 1,2, \ldots , n$$

โดยที่กำหนดให้ $\beta_0 =$ -0.5 และเมื่อได้ค่าความน่าจะเป็นมาแล้วนั้น จะกำหนดให้ค่านั้นเป็น 1 เมื่อค่าความน่าจะเป็นที่คำนวณได้มีค่ามากกว่า 0.5 และเป็น 0 เมื่อค่าความน่าจะเป็นที่คำนวณได้มีค่าน้อยกว่า 0.5

3.2 เกณฑ์ที่ใช้ในการพิจารณา

เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพการประมาณค่าสัมประสิทธิ์และการคัดเลือกตัวแปรของการเรกูลาไรซ์โดยใช้วิธีแลซโซแบบปรับได้ของตัวแบบถดถอยโลจิสติกในกรณีที่ข้อมูลมีมิติสูงทั้ง 3 วิธี ได้แก่ วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณริดจ์, วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณแลซโซ, และวิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณสไตน์-ริดจ์ รวมทั้งเปรียบเทียบกับวิธีแลซโซด้วย มีดังนี้

3.2.1 ความถูกต้องของการทำนาย จากการทำนายโดยใช้ตารางการณ์จร

ตารางการณ์จร ขนาด 2×2 แสดงความถี่ในแต่ละกลุ่มจำแนกตามค่าสังเกตกับค่าที่ทำนายได้ของตัวแปรตาม เป็นดังนี้

ค่าสังเกต	ค่าทำนาย		รวม
	$\hat{Y} = 1$	$\hat{Y} = 0$	
$Y = 1$	a	b	a+b
$Y = 0$	c	d	c+d
รวม	a+c	b+d	a+b+c+d

จากตารางการณ์จรข้างต้น สามารถนำจำนวนต่าง ๆ ซึ่งเป็นผลลัพธ์ที่ได้จากการศึกษามาพิจารณาค่าสถิติที่ใช้ประเมินความถูกต้องในการทำนาย อันประกอบด้วย

- ความไว (Sensitivity) เป็นค่าสัดส่วนของจำนวนครั้งที่ทำนายถูกว่าจะเกิดเหตุการณ์ที่เราสนใจเทียบกับจำนวนครั้งทั้งหมดที่ทำนายว่าจะเกิดเหตุการณ์ที่สนใจ เช่น ผลการตรวจเป็นบวกของกลุ่มคนที่เกิดโรคจริง และประเมินค่าออกมาเป็นร้อยละ ที่มีค่าอยู่ระหว่าง 0 ถึง 100 โดยที่สามารถนำมาใช้เป็นตัวชี้วัดได้ว่า หากมีค่าความไวมาก ก็ถือว่ามีความถูกต้องในการจำแนกกลุ่มที่จะเกิดเหตุการณ์ที่สนใจได้ เช่น มีความสามารถตรวจพบหรือค้นหาผู้ป่วยได้ดี ซึ่งมีสูตรการคำนวณคือ

$$Sensitivity = \frac{a}{a+c}$$

- ความจำเพาะ (Specificity) เป็นค่าสัดส่วนของจำนวนครั้งที่ทำนายถูกว่าจะเกิดเหตุการณ์ที่เราไม่สนใจเทียบกับจำนวนครั้งทั้งหมดที่ทำนายว่าจะเกิดเหตุการณ์ที่เราไม่สนใจ เช่น ผลการตรวจเป็นลบของกลุ่มคนที่ไม่ได้เกิดโรคจริง และประเมินค่าออกมาเป็นร้อยละ ที่มีค่าอยู่ระหว่าง 0 ถึง 100 โดยที่สามารถนำมาใช้เป็นตัวชี้วัดได้ว่า หากมีค่าความจำเพาะมาก ก็ถือว่ามีความถูกต้องในการจำแนกกลุ่มที่จะเกิดเหตุการณ์ที่ไม่สนใจได้ เช่น มีความสามารถตรวจพบหรือค้นหาผู้ที่ไม่ป่วยได้ดี ซึ่งมีสูตรการคำนวณคือ

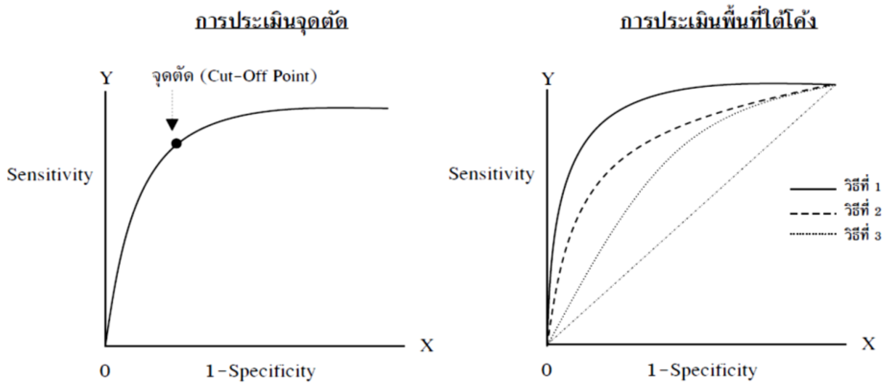
$$Specificity = \frac{d}{b+d}$$

- ความถูกต้อง (Accuracy) เป็นค่าสัดส่วนของจำนวนที่มีการจำแนกถูกต้องในการจำแนกความน่าจะเป็นของเหตุการณ์ที่สนใจและไม่สนใจได้เทียบกับจำนวนครั้งในการจำแนกทั้งหมด และประเมินค่าออกมาเป็นร้อยละ มีค่าอยู่ระหว่าง 0 ถึง 100 และสามารถนำมาใช้เป็นตัวชี้วัดได้ว่า หากมีค่าความถูกต้องมาก แสดงว่า วิธีการทดสอบมีความถูกต้องในการจำแนกเหตุการณ์ที่สนใจและ

ไม่สนใจได้ดี เช่น มีความถูกต้องในการตรวจพบผลบวกและลบของผู้ป่วยได้ถูกต้อง ซึ่งมีสูตรการคำนวณคือ

$$Accuracy = \frac{a + d}{a + b + c + d}$$

- ค่าพื้นที่ใต้โค้งของเส้นกราฟ ROC curve (Receiver Operating Characteristic curve) ซึ่ง เส้นกราฟ ROC curve เป็นเครื่องมือหนึ่งที่ช่วยกำหนดจุดตัด (Cut-off point) ในการตัดสินใจของวิธีการทดสอบวิธีการวินิจฉัยโรคหรือสมการทำนายโอกาสการเกิดผลลัพธ์ต่าง ๆ โดยเส้นกราฟ ROC เป็นการนำเอาค่าความไว (Sensitivity) และค่าความจำเพาะ (Specificity) จากตารางการณ์จรข้างต้นของวิธีการมาใช้ในการกำหนดจุดความสัมพันธ์บนเส้นกราฟ ซึ่งประกอบด้วยแกน X คือ ค่า 1- Specificity หรืออัตราผลบวกเท็จ (False positive rate) มีค่าอยู่ในช่วงร้อยละ 0 – 100 และแกน Y คือ ค่า Sensitivity หรืออัตราผลบวกจริง (True positive rate) มีค่าอยู่ในช่วงร้อยละ 0 – 100 จากกราฟของ ROC จะได้ว่า การประเมินผลจุดตัดที่ดีที่สุด ได้แก่ จุดในตำแหน่งที่มี Sensitivity เข้าใกล้ 100 และ 1-Specificity เข้าใกล้ 0 มากที่สุด นอกจากนี้ยังสามารถใช้เส้นกราฟ ROC มาประเมินประสิทธิภาพของวิธีการทดสอบได้ ด้วยการพิจารณาพื้นที่ใต้โค้งของวิธีการทดสอบแต่ละวิธี ซึ่ง วิธีการทดสอบที่มีพื้นที่มากกว่า แสดงว่า มีประสิทธิภาพมากกว่า



ภาพที่ 7 แนวคิดการวิเคราะห์ประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ด้วยเส้นกราฟ ROC (พงษ์เดช, 2560)

3.2.2 ความถูกต้องของการประมาณค่า โดยวัดจากค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์การถดถอย (Mean of Mean Square Error : Mean_MSE) ในการจำลองข้อมูล 500 ครั้ง มีค่าน้อยที่สุด คำนวณได้ดังนี้

Mean_MSE = Mean$(MSE_1, \ldots, MSE_{500})$

โดยที่ $MSE_r = \frac{\sum_{j=1}^{p+4}(\hat{\beta}_j - \beta_j)^2}{p+4}$ เมื่อ $p+4$ คือ จำนวนตัวแปรอธิบาย และ r คือรอบของการทำซ้ำ

3.2.3 ความผิดพลาดในการคัดเลือกตัวแปรเข้าสู่ตัวแบบ เมื่อกำหนดค่าพารามิเตอร์ β ตามสถานการณ์ต่าง ๆ ซึ่งส่วนใหญ่มีค่าเป็นศูนย์และบางส่วนไม่ได้มีค่าเป็นศูนย์ โดยทำการนับจำนวนตัวแปรที่ถูกคัดเลือกแตกต่างไปจากตัวแบบจริงที่ได้กำหนดไว้ข้างต้น โดยความผิดพลาดในการคัดเลือกตัวแปรมี 2 แบบ คือ กรณีที่ค่าพารามิเตอร์ไม่เท่ากับ 0 แต่ตัวแปรอธิบายไม่ถูกคัดเลือกเข้าสู่ตัวแบบ (IC1) หมายถึง จำนวนตัวแปรที่ควรอยู่ในตัวแบบ แต่ไม่ถูกคัดเลือก และกรณีที่ค่าพารามิเตอร์เท่ากับ 0 แต่ตัวแปรอธิบายถูกคัดเลือกเข้าสู่ตัวแบบ (IC2) หมายถึง จำนวนตัวแปรที่ไม่ควรอยู่ในตัวแบบ แต่ถูกคัดเลือก แล้วจึงพิจารณาค่าเฉลี่ยของจำนวนของตัวแปรอธิบายที่เกิดความผิดพลาดในการคัดเลือกตัวแปรอธิบาย (Mean of IC : Mean_IC) จากการทำซ้ำจำนวน 500 รอบ, และค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรอธิบายเมื่อเทียบกับจำนวนตัวแปรอธิบายในตัวแบบ (Percent of IC : %IC) จากการจำลอง 500 รอบ โดยมีวิธีการคำนวณ (พัชราภรณ์, 2560) ดังนี้

- คำนวณค่าเฉลี่ยของจำนวนของตัวแปรอธิบายที่เกิดความผิดพลาดในการคัดเลือกตัวแปรอธิบาย จากการจำลอง 500 รอบ คือ

$$\text{Mean\_IC1} = \frac{\sum_{r=1}^{500} \text{IC1}_r}{500} \quad \text{และ} \quad \text{Mean\_IC2} = \frac{\sum_{r=1}^{500} \text{IC2}_r}{500}$$

โดยที่ จำนวนของตัวแปรอธิบายที่เกิดความผิดพลาดในการคัดเลือกตัวแปรอธิบาย คือ $IC1_r = \#\{j : \beta_j \neq 0, \hat{\beta} = 0\}$ และ $IC2_r = \#\{j : \beta_j = 0, \hat{\beta} \neq 0\}$

- แล้วจึงมาคำนวณค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรอธิบายเมื่อเทียบกับจำนวนตัวแปรอธิบายในตัวแบบ คือ

$$\%\text{IC1} = \frac{\text{Mean\_IC1}}{15} \times 100 \quad \text{และ} \quad \%\text{IC2} = \frac{\text{Mean\_IC2}}{(p+4)-15} \times 100$$

3.3 ขั้นตอนการวิจัย

1. ทำการสร้างข้อมูลค่าสังเกต $(\boldsymbol{Y}, \boldsymbol{X})$ แล้วแบ่งออกเป็นสองชุด คือ ชุดข้อมูลฝึกฝน (Training set) จำนวน 80% และชุดข้อมูลทดสอบ (Test set) อีกจำนวน 20% โดยกำหนดให้ T และ T^C คือ ชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบ ตามลำดับ โดยมีขั้นตอนดังนี้

1.1 สร้างเมทริกซ์ของตัวแปรอธิบาย $(\boldsymbol{X})$ ขนาด $n \times (p+4)$ โดย $\boldsymbol{X}$ นั้นจะประกอบด้วยตัวแปรอธิบายแบบต่อเนื่องและตัวแปรอธิบายแบบไม่ต่อเนื่อง ซึ่งลักษณะของตัวแปรอธิบายมีลักษณะเป็นดังที่ได้กำหนดไว้ในขอบเขตของการศึกษา

1.2 สร้างเวกเตอร์ของค่าสัมประสิทธิ์ถดถอย $\boldsymbol{\beta}$ ที่มีรูปแบบต่าง ๆ ดังขอบเขตของการศึกษา

1.3 สร้างเวกเตอร์ของตัวแปรตาม ($\boldsymbol{Y}$) จากตัวแบบของการถดถอยโลจิสติกพหุคูณ คือ $\pi_i(\boldsymbol{x}) = P(y_i = 1|\boldsymbol{x_i}) = \frac{\exp(\beta_0+\boldsymbol{X_i'\beta})}{\mathbf{1}+\exp(\beta_0+\boldsymbol{X_i'\beta})}$ จากเมทริกซ์ของตัวแปรอธิบาย $\boldsymbol{X}$ ที่สร้างจากข้อ 1.1 และเวกเตอร์ของค่าสัมประสิทธิ์ถดถอย $\boldsymbol{\beta}$ ที่สร้างจากข้อ 1.2

2. นำ $\boldsymbol{X}$ และ $\boldsymbol{Y}$ จากชุดข้อมูลฝึกฝน T มาหาตัวประมาณสัมประสิทธิ์การถดถอย $\widehat{\boldsymbol{\beta}}$ ของแต่ละวิธีการประมาณค่า ได้แก่

ตัวประมาณวิธีริดจ์

$$\widehat{\boldsymbol{\beta}}^{\boldsymbol{ridge}} = argmin_{\boldsymbol{\beta}}\left\{\sum_{i=1}^{n}\left[-y_i(\boldsymbol{X_i'\beta}) + \ln\left(1 + e^{(\boldsymbol{X_i'\beta})}\right)\right] + \lambda\sum_{j=1}^{p+4}\beta_j^2\right\}$$

ตัวประมาณวิธีแลซโซ

$$\widehat{\boldsymbol{\beta}}^{\boldsymbol{L}} = argmin_{\boldsymbol{\beta}}\left\{-\sum_{i=1}^{n}\left[-y_i(\boldsymbol{X_i'\beta}) + \ln\left(1 + e^{(\boldsymbol{X_i'\beta})}\right)\right] + \lambda\sum_{j=1}^{p+4}\left|\beta_j\right|\right\}$$

ซึ่งสามารถหาพารามิเตอร์ปรับแต่ง (tuning parameter: t) ของแต่ละวิธีโดยวิธี 5-fold cross-validation โดยมีขั้นตอน ดังนี้

2.1 แบ่งข้อมูล $\boldsymbol{X_i}$ และ Y_i , $i = 1, \ldots, n$ ใน T ออกเป็น 5 ส่วน จะได้ T_1, T_2, T_3, T_4, T_5 ให้หนึ่งส่วนเป็นชุดข้อมูลทดสอบย่อย T_v และอีก 4 ส่วนที่เหลือเป็นชุดข้อมูลฝึกฝนย่อย $T - T_v$

2.2 หาค่าประมาณพารามิเตอร์ $\widehat{\boldsymbol{\beta}}_{\boldsymbol{v}}$ ของแต่ละวิธีข้างต้นที่สอดคล้องกับเงื่อนไขบังคับ และทำให้ $-\ln \boldsymbol{L} = \sum_{i=1}^{n}\left[-y_i(\boldsymbol{X_i'\beta}) + \ln\left(1 + e^{(\boldsymbol{X_i'\beta})}\right)\right]$ มีค่าต่ำที่สุด ทำเช่นนี้ในทุก ๆ ชุดข้อมูลฝึกฝนย่อย $T - T_v$ เมื่อ v = 1,...,5

2.3 จากนั้นนำค่าประมาณ $\widehat{\boldsymbol{\beta}}_{\boldsymbol{v}}$ ที่ได้จากแต่ละชุดข้อมูลฝึกฝนย่อยไปหาค่าคลาดเคลื่อนกำลังสองเฉลี่ย ในชุดข้อมูลทดสอบย่อย T_v เมื่อ v = 1, ..., 5 เพื่อหาค่าประมาณพารามิเตอร์ $\widehat{\boldsymbol{\beta}}_{\boldsymbol{v}}$ ที่มีค่าคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด

2.4 เลือกค่า $\boldsymbol{\lambda}$ ที่ทำให้ $\widehat{\boldsymbol{\beta}}_{\boldsymbol{v}}$ มีค่าคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด ซึ่งค่า $\boldsymbol{\lambda}$ ค่านั้นคือค่าพารามิเตอร์ปรับแต่ง (tuning parameter) ของแต่ละวิธี

3. นำตัวประมาณแบบริดจ์ที่คำนวณได้ไปหาตัวประมาณพารามิเตอร์ของวิธีสไตน์-ริดจ์ $\widehat{\boldsymbol{\beta}}^{\boldsymbol{stein-ridge}} = c\widehat{\boldsymbol{\beta}}^{\boldsymbol{ridge}}$ โดยที่ $c = \frac{\widehat{\boldsymbol{\beta}}^{\boldsymbol{ridge}'}\widehat{\boldsymbol{\beta}}^{\boldsymbol{ridge}}}{\widehat{\boldsymbol{\beta}}^{\boldsymbol{ridge}'}\widehat{\boldsymbol{\beta}}^{\boldsymbol{ridge}}+tr(\boldsymbol{X'\widehat{W}X})^{-1}}$ คือค่าพารามิเตอร์ปรับแต่งของวิธีสไตน์-ริดจ์

4. คำนวณเวกเตอร์น้ำหนัก $\widehat{\boldsymbol{w}}_j = 1/\left|\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}\right|^{\gamma}$ ของแต่ละวิธีที่จะนำไปใช้ในการประมาณค่าพารามิเตอร์ด้วยวิธี Adaptive LASSO โดยที่

4.1 Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีริดจ์ คำนวณค่าเวกเตอร์ถ่วงน้ำหนัก $\widehat{\boldsymbol{w}}_{AD\_R} = 1/\left|\widehat{\boldsymbol{\beta}}^{ridge}(\lambda)\right|^{\gamma}$

4.2 Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธี LASSO คำนวณเวกเตอร์น้ำหนัก $\widehat{\boldsymbol{w}}_{AD\_L} = 1/\left|\widehat{\boldsymbol{\beta}}^{L}(\lambda)\right|^{\gamma}$

4.3 Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีสไตน์-ริดจ์ คำนวณเวกเตอร์น้ำหนัก $\widehat{\boldsymbol{w}}_{AD\_SR} = 1/\left|\widehat{\boldsymbol{\beta}}^{stein-ridge}(\lambda)\right|^{\gamma}$

5. คำนวณค่าประมาณค่าสัมประสิทธิ์ถดถอยด้วยวิธี Adaptive LASSO ดังนี้ $\widehat{\boldsymbol{\beta}}^{adp.lasso} = argmin_{\boldsymbol{\beta}}\left\{\sum_{i=1}^{n}\left[-y_i\left(\boldsymbol{X}_i'\boldsymbol{\beta}\right) + \ln\left(1 + e^{(\boldsymbol{X}_i'\boldsymbol{\beta})}\right)\right] + \lambda_2\sum_{j=1}^{p+4}\widehat{\boldsymbol{w}}_j\left|\beta_j\right|\right\}$ โดยใช้เวกเตอร์ค่าถ่วงน้ำหนักที่คำนวณได้ในข้อ 4. ไป scale ที่เมทริกซ์ของตัวแปรอธิบาย $\boldsymbol{X}$ จนได้เป็น $\boldsymbol{X}_{\boldsymbol{w}}$ แล้วนำ $\boldsymbol{X}_{\boldsymbol{w}}$ และ $\boldsymbol{Y}$ จากชุดข้อมูลฝึกฝนไปใช้ในการประมาณค่าด้วยวิธี Adaptive LASSO ซึ่งวิธีการหา tuning parameter $\boldsymbol{\lambda}_2$ สามารถหาได้โดยใช้ขั้นตอนเดียวกันกับข้อ 2.

6. หลังจากการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี Adaptive LASSO แล้ว จะได้ค่าสัมประสิทธิ์การถดถอยที่ยังมีการถ่วงน้ำหนักอยู่ จะต้องทำการหารค่าถ่วงน้ำหนักออกจากค่าประมาณค่าสัมประสิทธิ์การถดถอยก่อน จึงจะได้ค่าประมาณค่าสัมประสิทธิ์การถดถอยที่แท้จริงจากวิธี Adaptive LASSO คือ $\widehat{\boldsymbol{\beta}}^{adp.lasso}$ ทั้งสามวิธีที่มีค่าถ่วงน้ำหนักต่างกัน ซึ่งทำให้ได้ค่าประมาณของค่าสัมประสิทธิ์การถดถอยของทั้ง 4 วิธีคือ $\widehat{\boldsymbol{\beta}}^{L}, \widehat{\boldsymbol{\beta}}^{AD\_R}, \widehat{\boldsymbol{\beta}}^{AD\_L}, \widehat{\boldsymbol{\beta}}^{AD\_SR}$

7. นำค่า $\widehat{\boldsymbol{\beta}}^{L}, \widehat{\boldsymbol{\beta}}^{AD\_R}, \widehat{\boldsymbol{\beta}}^{AD\_L}, \widehat{\boldsymbol{\beta}}^{AD\_SR}$ ที่ประมาณค่าได้ไปทำนายโดยใช้ข้อมูลจากชุดข้อมูลทดสอบ $\boldsymbol{T}^{C}$ แล้วจำแนกลงตารางการณ์จร ขนาด 2×2 ที่แสดงความถี่ในแต่ละกลุ่มจำแนกตามค่าสังเกตกับค่าที่ทำนายได้ของตัวแปรตาม

8. คำนวณค่าความไว ความจำเพาะ ความถูกต้องในการทำนาย และหาพื้นที่ใต้โค้งของกราฟ ROC Curve ที่ใช้ในการประเมินความถูกต้องของการทำนาย

9. คำนวณค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณค่าสัมประสิทธิ์ถดถอยเทียบกับค่าพารามิเตอร์ที่ได้สร้างไว้ในข้อ 1.2 ของแต่ละวิธีการประมาณค่า

10. นับจำนวนตัวประมาณสัมประสิทธิ์การถดถอยที่อยู่ในเซตของชุดข้อมูลทดสอบ ตามเกณฑ์ของ IC1 และ IC2 ที่ได้กำหนดไว้และตรวจสอบความผิดพลาดในการคัดเลือกตัวแปร คือ IC1, IC2 ในแต่ละรอบ

11. ทำซ้ำ ข้อ 1-10 จำนวน 500 ครั้ง

12. หาค่าเฉลี่ยของค่าความไว ค่าเฉลี่ยของความจำเพาะ ค่าเฉลี่ยความถูกต้องของการทำนาย ค่าเฉลี่ยของค่าพื้นที่ใต้เส้นกราฟของ ROC Curve มาใช้ในการพิจารณาประสิทธิภาพของแต่

ละวิธี, ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย MSE, ค่าเฉลี่ยของจำนวนของตัวแปรอธิบายที่เกิดความผิดพลาดในการคัดเลือกตัวแปรอธิบาย Mean_IC1, Mean_IC2, และค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรอธิบายเมื่อเทียบกับจำนวนตัวแปรอธิบายในตัวแบบ %IC1, และ %IC2 จากผลการทดลองทั้งหมด 500 ครั้ง

13. จากนั้นผู้วิจัยจะทำการเปรียบเทียบค่าเฉลี่ยของค่าความไว ค่าเฉลี่ยของความจำเพาะ ค่าเฉลี่ยความถูกต้องของการทำนาย ค่าเฉลี่ยของค่าพื้นที่ใต้เส้นกราฟของ ROC Curve มาใช้ในการพิจารณาประสิทธิภาพของแต่ละวิธี, ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย MSE, ค่าเฉลี่ยของจำนวนของตัวแปรอธิบายที่เกิดความผิดพลาดในการคัดเลือกตัวแปรอธิบาย Mean_IC1, Mean_IC2, และค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรอธิบายเมื่อเทียบกับจำนวนตัวแปรอธิบายในตัวแบบ %IC1, และ %IC2 เพื่อให้เห็นถึงความแตกต่างของความถูกต้องและประสิทธิภาพในแต่ละวิธี

บทที่ 4
ผลการวิเคราะห์ข้อมูล

งานวิจัยนี้มีวัตถุประสงค์เพื่อทำการศึกษาและเปรียบเทียบประสิทธิภาพการประมาณค่าสัมประสิทธิ์และการคัดเลือกตัวแปรของการเรกูลาไรซ์โดยใช้วิธีแลซโซและแลซโซแบบปรับได้ของตัวแบบถดถอยโลจิสติกในกรณีที่ข้อมูลมีมิติสูงแบบบางเบาทั้ง 4 วิธี ได้แก่ วิธีแลซโซ, วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีริดจ์, วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีแลซโซ, และวิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีสไตน์-ริดจ์ โดยพิจารณาจากความถูกต้องในการทำนาย ความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอย และความถูกต้องในการคัดเลือกตัวแปรอธิบายเข้าสู่ตัวแบบ ซึ่งจะมีการกำหนดสัญลักษณ์ที่ใช้ในงานวิจัยดังนี้

n คือ ขนาดตัวอย่าง

p คือ จำนวนตัวแปรอธิบาย

ρ คือ ค่าสัมประสิทธิ์สหสัมพันธ์

$\boldsymbol{D}$ คือ เมทริกซ์ของตัวแปรอธิบายแบบไม่ต่อเนื่องจำนวน 4 ตัวแปร

$\boldsymbol{X}^*_{ind}$ คือ เมทริกซ์ของตัวแปรอธิบายแบบต่อเนื่องในกรณีที่ 1 ตัวแปรอธิบายจะถูกแบ่งเป็น 2 กลุ่มที่เป็นอิสระต่อกัน แต่ตัวแปรอธิบายภายในแต่ละกลุ่มมีความสัมพันธ์กันเอง โดยที่ $\boldsymbol{X}^*_{ind} = \left[\boldsymbol{X}_{\boldsymbol{1}_{(n\times 15)}} \vdots \boldsymbol{X}_{\boldsymbol{2}_{(n\times(p-15))}}\right]$

$\boldsymbol{X}^*_{cor}$ คือ เมทริกซ์ของตัวแปรอธิบายแบบต่อเนื่องในกรณีที่ 2 ตัวแปรอธิบายจะมีเพียง 1 กลุ่ม จำนวน p ตัวแปรที่สัมพันธ์กันทั้งหมด

$\boldsymbol{\Sigma_1}$ คือ เมทริกซ์ของความแปรปรวนร่วมของเมทริกซ์ $\boldsymbol{X}^*_{ind}$

$\boldsymbol{\Sigma_2}$ คือ เมทริกซ์ของความแปรปรวนร่วมของเมทริกซ์ $\boldsymbol{X}^*_{cor}$

$\boldsymbol{\beta_{(1)}}$ คือ เวกเตอร์ของสัมประสิทธิ์การถดถอยรูปแบบที่ 1 ที่มีจำนวนตัวแปรอธิบายที่มีค่าพารามิเตอร์ไม่เป็น 0 จำนวน 9 ตัวแปร

$\boldsymbol{\beta_{(2)}}$ คือ เวกเตอร์ของสัมประสิทธิ์การถดถอยรูปแบบที่ 2 ที่มีจำนวนตัวแปรอธิบายที่มีค่าพารามิเตอร์ไม่เป็น 0 จำนวน 14 ตัวแปร

$\boldsymbol{\beta_{(3)}}$ คือ เวกเตอร์ของสัมประสิทธิ์การถดถอยรูปแบบที่ 3 ที่มีจำนวนตัวแปรอธิบายที่มีค่าพารามิเตอร์ไม่เป็น 0 จำนวน 19 ตัวแปร

L คือ วิธีแลซโซ

AD_R คือ วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณริดจ์

AD_L คือ วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณแลซโซ

AD_SR คือ วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณสไตน์-ริดจ์

ค่าความไว คือ ค่าสัดส่วนของจำนวนครั้งที่ทำนายถูกว่าจะเกิดเหตุการณ์ที่สนใจเทียบกับจำนวนครั้งที่ทำนายว่าจะเกิดเหตุการณ์ที่สนใจทั้งหมด

ค่าความจำเพาะ คือ ค่าสัดส่วนของจำนวนครั้งที่ทำนายถูกว่าจะเกิดเหตุการณ์ที่ไม่สนใจเทียบกับจำนวนครั้งที่ทำนายว่าจะเกิดเหตุการณ์ที่ไม่สนใจทั้งหมด

ROC curve คือ เส้นกราฟที่สร้างมาจากค่าความไวและ 1-ค่าความจำเพาะของการทำนาย ซึ่งจะใช้ในการประเมินประสิทธิภาพของการทำนายได้ด้วยการพิจารณาค่าพื้นที่ใต้โค้งของเส้นกราฟ ROC นี้

AUC คือ ค่าพื้นที่ใต้โค้งของเส้นกราฟ ROC ที่จะบ่งบอกถึงประสิทธิภาพของการทำนายโดยรวม

MSE คือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์การถดถอย

%IC1 คือ ค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรแบบที่ 1 หมายถึง จำนวนตัวแปรอธิบายที่ควรอยู่ในตัวแบบ แต่ไม่ถูกคัดเลือกเข้าสู่ตัวแบบ

และ %IC2 คือ ค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรแบบที่ 2 หมายถึง จำนวนตัวแปรอธิบายที่ไม่ควรอยู่ในตัวแบบ แต่ถูกคัดเลือกเข้าสู่ตัวแบบ

ในการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยได้ทำการศึกษา 2 กรณี คือ กรณีที่ 1 คือ ตัวแปรอธิบายจะถูกแบ่งออกเป็นสองกลุ่มที่เป็นอิสระกัน แต่ตัวแปรอธิบายภายในแต่ละกลุ่มมีความสัมพันธ์กันเอง โดยที่ในกลุ่มที่ 1 จะมีจำนวน 15 ตัวแปรที่สัมพันธ์กัน และกลุ่มที่ 2 คือจำนวนตัวแปรที่เหลือมีความสัมพันธ์กัน ซึ่งมีเมทริกซ์ของความแปรปรวนร่วมดัง $\boldsymbol{\Sigma_a}$ และกรณีที่ 2 คือ ตัวแปรอธิบายทั้งหมดจะมีความสัมพันธ์กันเองทุกตัวแปร ซึ่งมีเมทริกซ์ของความแปรปรวนร่วมดัง $\boldsymbol{\Sigma_b}$ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง X_k และ X_j คำนวณจาก $\rho^{|k-j|}$ เมื่อกำหนดค่าสัมประสิทธิ์สหสัมพันธ์ 3 กรณี คือ ρ = 0.3, 0.6 และ 0.9 รวมถึงมีตัวแปรอธิบายแบบไม่ต่อเนื่องจำนวน 4 ตัวแปรรวมอยู่ด้วย และกำหนดค่าพารามิเตอร์ให้สอดคล้องกับลักษณะของตัวแบบการถดถอยแบบบางเบาในหลากหลายแบบดังขอบเขตของการศึกษา โดยแต่ละกรณีจะทำการจำลองข้อมูลที่มีขนาดตัวอย่าง (n) เท่ากับ 100 และ 200 และจำนวนตัวแปรอธิบายแบบต่อเนื่อง p มีจำนวนเป็น 2, 3, และ 4 เท่าของขนาดตัวอย่าง n กับจำนวนตัวแปรอธิบายแบบไม่ต่อเนื่องมีจำนวน 4 ตัวแปรทุกกรณี รวมจำนวนตัวแปรทั้งสิ้น $p + 4$ ตัวแปร

โดยเกณฑ์ที่ใช้ในการพิจารณาเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยโลจิสติกทั้ง 4 วิธี ได้แก่

1. ประสิทธิภาพของการทำนายโดยวัดจากค่าเฉลี่ยของค่าความไว (Mean_sensitivity) ค่าเฉลี่ยของค่าความจำเพาะ (Mean_specificity) และค่าเฉลี่ยของพื้นที่ใต้โค้งของเส้นกราฟ Receiver Operating Characteristic curve หรือ ROC curve (Mean_AUC) ที่ได้จากการการทำนาย ในการจำลองข้อมูล 500 รอบ

2. ประสิทธิภาพของการประมาณโดยวัดจากค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์การถดถอย (Med_MSE) ในการจำลองข้อมูล 500 รอบ

3. ความผิดพลาดของการคัดเลือกตัวประมาณโดยวัดจากค่าร้อยละของความผิดพลาดของการคัดเลือกตัวแปรอธิบายเมื่อเทียบกับจำนวนตัวแปรอธิบายในตัวแบบ (%IC) จากการจำลองข้อมูล 500 รอบ

งานวิจัยนี้ทำการจำลองข้อมูลโดยโปรแกรม R เวอร์ชั่น 4.1.1 โดยแต่ละสถานการณ์มีการทำซ้ำ 500 รอบ ซึ่งผลการวิจัยมีรายละเอียด ดังนี้

4.1 ความถูกต้องของการทำนายด้วยสมการถดถอยโลจิสติกด้วยวิธีเรกูลาไรซ์ 4 วิธี

ความถูกต้องของการทำนายของตัวประมาณแต่ละวิธี จะวัดจากค่าเฉลี่ยของค่าความไว (Sensitivity), ค่าเฉลี่ยของค่าความจำเพาะ (Specificity), และค่าเฉลี่ยของค่าพื้นที่ใต้เส้นโค้งของกราฟ ROC curve (AUC) จากการจำลองข้อมูล 500 รอบ ซึ่งผลการวิจัยแบ่งออกเป็น 2 กรณี มีรายละเอียดดังนี้

ในกรณีที่ 1 ตัวแปรอธิบายจะถูกแบ่งเป็น 2 กลุ่มที่เป็นอิสระต่อกัน แต่มีความสัมพันธ์กันภายในกลุ่ม โดยที่ในกลุ่มที่ 1 จะมีจำนวน 15 ตัวแปรที่สัมพันธ์กันเอง และกลุ่มที่ 2 คือจำนวนตัวแปรที่เหลือมีความสัมพันธ์กันเอง นั่นคือ เมทริกซ์ของตัวแปรอธิบายแบบต่อเนื่องกรณีที่ 1 ใช้สัญลักษณ์แทนด้วย $\boldsymbol{X}_{ind}^{*} = \left[\boldsymbol{X}_{\boldsymbol{1}_{(n\times 15)}} \vdots \boldsymbol{X}_{\boldsymbol{2}_{(n\times(p-15))}}\right]$ โดยที่ $\boldsymbol{X}_{ind}^{*} \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_a)$ ซึ่งเมทริกซ์ความแปรปรวนร่วมของตัวแปรอธิบาย $\boldsymbol{X}_{ind}^{*}$ สำหรับกรณีที่ 1 คือ

$$\boldsymbol{\Sigma}_a = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 \end{bmatrix}_{p\times p}$$

โดยที่ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรตัวที่ X_k และ X_j ของเมทริกซ์ $\boldsymbol{X}_{\boldsymbol{1}}$ คือ $\rho^{|k-j|}$; $j,k = 1,2,\ldots,15$ และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรตัวที่ X_k และ X_j ของเมทริกซ์ $\boldsymbol{X}_{\boldsymbol{2}}$ คือ $\rho^{|k-j|}$; $j,k = 1,2,\ldots,(p-15)$ ดังนี้

$$\mathbf{\Sigma_1} = \mathbf{Cov}(\boldsymbol{X_1}) = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{14} \\ \rho & 1 & \rho & \cdots & \rho^{13} \\ \rho^2 & \rho & 1 & \cdots & \rho^{12} \\ \vdots & \vdots & \vdots & \ddots & \rho \\ \rho^{14} & \rho^{13} & \cdots & \rho & 1 \end{bmatrix}_{15\times 15} \quad \text{และ}$$

$$\mathbf{\Sigma_2} = \mathbf{Cov}(\boldsymbol{X_2}) = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{p-14} \\ \rho & 1 & \rho & \cdots & \rho^{p-13} \\ \rho^2 & \rho & 1 & \cdots & \rho^{p-12} \\ \vdots & \vdots & \vdots & \ddots & \rho \\ \rho^{p-14} & \rho^{p-13} & \cdots & \rho & 1 \end{bmatrix}_{(p-15)\times(p-15)}$$

ตารางที่ 1 ค่าความไวของการทำนายด้วยสมการถดถอยโลจิสติกเรกูลาไรซ์ทั้ง 4 วิธี ในกรณีที่ $X^*_{\text{ind}} \sim N_p(0, \mathbf{\Sigma_a})$

n	p+4	ρ	$\boldsymbol{\beta}_{(1)}$				$\boldsymbol{\beta}_{(2)}$				$\boldsymbol{\beta}_{(3)}$			
			L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR
100	204	0.3	0.9074	0.9554	0.9441	**0.9640**	0.8464	**0.8945**	0.8618	0.8925	0.8063	**0.8384**	0.8013	0.8352
		0.6	0.9255	**0.9756**	0.9437	0.9740	0.8904	0.9311	0.8926	**0.9312**	0.8661	**0.8972**	0.8599	0.8961
		0.9	0.9450	0.9814	0.9415	**0.9835**	0.9433	**0.9707**	0.9314	0.9689	0.9310	0.9523	0.9233	**0.9534**
	304	0.3	0.9076	0.9523	0.9455	**0.9623**	0.8460	0.8913	0.8602	**0.8953**	0.8048	0.8336	0.7933	**0.8363**
		0.6	0.9281	0.9712	0.9541	**0.9729**	0.8916	**0.9359**	0.8824	0.9352	0.8677	0.8948	0.8659	**0.8982**
		0.9	0.9498	0.9815	0.9442	**0.9821**	0.9334	**0.9633**	0.9334	0.9620	0.9296	**0.9557**	0.9152	0.9536
	404	0.3	0.9013	0.9472	0.9363	**0.9630**	0.8388	0.8800	0.8523	**0.8841**	0.7727	0.8141	0.7796	**0.8168**
		0.6	0.9170	0.9673	0.9350	**0.9700**	0.8841	0.9165	0.8883	**0.9240**	0.8607	0.8897	0.8550	**0.8909**
		0.9	0.9543	**0.9828**	0.9505	0.9818	0.9312	0.9619	0.9258	**0.9638**	0.9290	**0.9558**	0.9233	0.9530
200	404	0.3	0.9491	**0.9850**	0.9757	0.9844	0.9145	0.9505	0.9349	**0.9542**	0.8806	0.9182	0.8948	**0.9193**
		0.6	0.9554	**0.9920**	0.9814	0.9910	0.9329	0.9626	0.9367	**0.9644**	0.9200	0.9414	0.9155	**0.9422**
		0.9	0.9629	0.9934	0.9577	**0.9955**	0.9561	0.9819	0.9523	**0.9833**	0.9536	0.9716	0.9446	**0.9719**
	604	0.3	0.9445	**0.9866**	0.9780	0.9854	0.9120	0.9458	0.9355	**0.9476**	0.8811	**0.9112**	0.8829	0.9080
		0.6	0.9551	**0.9911**	0.9803	0.9908	0.9338	0.9634	0.9417	**0.9649**	0.9168	0.9378	0.9159	**0.9403**
		0.9	0.9637	0.9932	0.9600	**0.9934**	0.9518	0.9825	0.9541	**0.9836**	0.9496	0.9729	0.9442	**0.9735**
	804	0.3	0.9350	**0.9924**	0.9864	0.9914	0.9039	**0.9424**	0.9277	0.9414	0.8709	0.8957	0.8667	**0.8966**
		0.6	0.9501	0.9909	0.9805	**0.9914**	0.9274	**0.9631**	0.9410	0.9618	0.9154	**0.9363**	0.9117	0.9350
		0.9	0.9647	**0.9938**	0.9604	0.9937	0.9582	0.9824	0.9540	**0.9844**	0.9518	**0.9761**	0.9458	0.9749

จากตารางที่ 1 จะเห็นได้ว่าค่าความไวของวิธี AD_R และวิธี AD_SR มีค่ามากกว่าวิธีอื่น ๆ ซึ่งถ้าเปรียบเทียบระหว่างสองวิธีนี้ ในกรณีที่ตัวแบบบางเบามาก ๆ หรือ $\boldsymbol{\beta}_{(1)}$ วิธี AD_R จะมีค่าเฉลี่ยของค่าความไวสูงกว่า แต่เมื่อจำนวนตัวแปรที่มีค่าพารามิเตอร์ไม่เป็น 0 เพิ่มขึ้น คือ กรณี $\boldsymbol{\beta}_{(2)}$ และ $\boldsymbol{\beta}_{(3)}$ วิธี AD_SR จะมีค่าความไวมากกว่า

ตารางที่ 2 ค่าความจำเพาะของการทำนายด้วยสมการถดถอยโลจิสติกเรกูลาไรซ์ทั้ง 4 วิธี ในกรณีที่ $X^{*}_{\text{ind}} \sim N_{\text{p}}(\mathbf{0}, \mathbf{\Sigma_a})$

n	p+4	ρ	$\boldsymbol{\beta}_{(1)}$				$\boldsymbol{\beta}_{(2)}$				$\boldsymbol{\beta}_{(3)}$			
			L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR
100	204	0.3	**0.9202**	0.8754	0.9140	0.8806	**0.8475**	0.8352	0.8185	0.8273	**0.8001**	0.7818	0.7560	0.7794
		0.6	**0.9467**	0.9161	0.9275	0.9085	**0.8945**	0.8783	0.8741	0.8776	**0.8738**	0.8657	0.8407	0.8668
		0.9	**0.9556**	0.9353	0.9420	0.9358	**0.9468**	0.9397	0.9328	0.9353	**0.9409**	0.9372	0.9249	0.9362
	304	0.3	**0.9157**	0.8626	0.9001	0.8675	**0.8536**	0.8275	0.8149	0.8262	**0.8026**	0.7731	0.7514	0.7722
		0.6	**0.9329**	0.9028	0.9272	0.9030	**0.8922**	0.8792	0.8530	0.8775	**0.8839**	0.8634	0.8435	0.8650
		0.9	**0.9505**	0.9296	0.9372	0.9275	**0.9437**	0.9339	0.9234	0.9334	**0.9335**	0.9283	0.9106	0.9260
	404	0.3	**0.9059**	0.8589	0.8881	0.8658	**0.8247**	0.8018	0.7887	0.8058	**0.7857**	0.7600	0.7388	0.7567
		0.6	**0.9261**	0.9001	0.9098	0.8992	**0.8995**	0.8762	0.8629	0.8802	**0.8823**	0.8591	0.8394	0.8594
		0.9	**0.9535**	0.9358	0.9360	0.9310	**0.9337**	0.9291	0.9212	0.9296	**0.9271**	0.9239	0.9105	0.9246
200	404	0.3	**0.9672**	0.9057	0.9646	0.9013	**0.9192**	0.8900	0.8957	0.8875	**0.8942**	0.8613	0.8625	0.8626
		0.6	0.9634	0.9318	**0.9664**	0.9243	**0.9328**	0.9165	0.9139	0.9147	**0.9203**	0.9029	0.8868	0.8989
		0.9	**0.9742**	0.9444	0.9675	0.9423	**0.9571**	0.9468	0.9427	0.9466	0.9523	**0.9533**	0.9356	0.9513
	604	0.3	**0.9594**	0.9048	0.9550	0.9003	**0.9183**	0.8906	0.9020	0.8883	**0.8847**	0.8464	0.8382	0.8424
		0.6	0.9651	0.9260	**0.9691**	0.9220	**0.9322**	0.9204	0.9146	0.9164	**0.9121**	0.8988	0.8901	0.8981
		0.9	**0.9710**	0.9448	0.9634	0.9415	**0.9624**	0.9519	0.9451	0.9514	**0.9573**	0.9514	0.9373	0.9519
	804	0.3	**0.9661**	0.9139	0.9601	0.9045	**0.9185**	0.8846	0.8855	0.8786	**0.8786**	0.8393	0.8286	0.8393
		0.6	**0.9604**	0.9254	0.9575	0.9214	**0.9433**	0.9207	0.9154	0.9178	**0.9170**	0.9002	0.8855	0.9001
		0.9	**0.9695**	0.9487	0.9587	0.9463	**0.9598**	0.9525	0.9423	0.9516	**0.9547**	0.9526	0.9358	0.9524

จากตารางที่ 2 จะเห็นได้ว่าค่าความจำเพาะของวิธี LASSO มีค่าสูงกว่าวิธีการประมาณค่าอื่น ๆ แทบทุกกรณี มีเพียงบางกรณีเท่านั้นที่มีค่าน้อยกว่า แต่ไม่ได้แตกต่างกันมากนัก

ตารางที่ 3 ค่า AUC ของเส้นกราฟ ROC curve ของการทำนายด้วยสมการถดถอยโลจิสติกเรกูลาไรซ์ทั้ง 4 วิธี ในกรณีที่ $X^*_{\text{ind}} \sim N_p(0, \mathbf{\Sigma_a})$

n	p+4	ρ	$\boldsymbol{\beta}_{(1)}$				$\boldsymbol{\beta}_{(2)}$				$\boldsymbol{\beta}_{(3)}$			
			L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR
		0.3	0.9083	0.9178	**0.9294**	0.9247	0.8421	**0.8664**	0.8407	0.8613	0.7985	**0.8113**	0.7790	0.8085
	204	0.6	0.9313	**0.9480**	0.9356	0.9432	0.8896	**0.9051**	0.8829	0.9045	0.8668	**0.8816**	0.8499	0.8815
		0.9	0.9469	**0.9606**	0.9390	0.9593	0.9435	**0.9555**	0.9313	0.9523	0.9337	0.9447	0.9231	**0.9447**
		0.3	0.9063	0.9126	**0.9247**	0.9187	0.8439	0.8612	0.8385	**0.8623**	0.7982	0.8045	0.7721	**0.8054**
100	304	0.6	0.9279	0.9394	**0.9409**	0.9403	0.8888	**0.9088**	0.8674	0.9074	0.8722	0.8795	0.8546	**0.8825**
		0.9	0.9475	**0.9574**	0.9386	0.9564	0.9355	**0.9491**	0.9281	0.9482	0.9297	**0.9421**	0.9124	0.9400
		0.3	0.8986	0.9083	0.9139	**0.9179**	0.8277	0.8435	0.8225	**0.8475**	0.7727	0.7876	0.7594	**0.7876**
	404	0.6	0.9175	**0.9367**	0.9219	0.9356	0.8883	0.8971	0.8755	**0.9028**	0.8671	0.8746	0.8469	**0.8756**
		0.9	0.9523	**0.9605**	0.9426	0.9579	0.9308	0.9458	0.9230	**0.9474**	0.9267	**0.9401**	0.9165	0.9393
		0.3	0.9548	0.9483	**0.9700**	0.9459	0.9145	0.9221	0.9167	**0.9226**	0.8842	0.8907	0.8792	**0.8919**
	404	0.6	0.9578	0.9640	**0.9742**	0.9598	0.9317	0.9407	0.9257	**0.9407**	0.9190	**0.9229**	0.9018	0.9214
		0.9	0.9676	**0.9699**	0.9614	0.9699	0.9556	0.9646	0.9471	**0.9652**	0.9520	**0.9628**	0.9399	0.9619
		0.3	0.9485	0.9493	**0.9677**	0.9465	0.9128	**0.9197**	0.9196	0.9194	0.8801	**0.8802**	0.8617	0.8766
200	604	0.6	0.9582	0.9603	**0.9749**	0.9580	0.9318	**0.9426**	0.9285	0.9412	0.9136	0.9189	0.9034	**0.9197**
		0.9	0.9659	**0.9706**	0.9605	0.9691	0.9558	0.9676	0.9491	**0.9680**	0.9521	0.9625	0.9402	**0.9629**
		0.3	0.9447	0.9568	**0.9747**	0.9516	0.9078	**0.9148**	0.9078	0.9114	0.8691	0.8686	0.8484	**0.8717**
	804	0.6	0.9532	0.9604	**0.9699**	0.9587	0.9333	**0.9424**	0.9286	0.9405	0.9149	**0.9190**	0.8990	0.9182
		0.9	0.9659	**0.9722**	0.9588	0.9709	0.9582	0.9679	0.9481	**0.9685**	0.9524	**0.9644**	0.9406	0.9637

จากตารางที่ 3 จะเห็นได้ว่า ในกรณีที่ตัวแบบมีความบางเบามาก ๆ หรือ $\boldsymbol{\beta}_{(1)}$ ค่า AUC ของกราฟ ROC curve ของวิธี AD_R และวิธี AD_L จะมีค่าสูงกว่าวิธีอื่น ๆ โดยเฉพาะอย่างยิ่งในกรณีที่ $n = 200$ เมื่อ $\rho = 0.3, 0.6$ วิธี AD_L จะมีค่ามากกว่า แต่เมื่อ $\rho = 0.9$ วิธี AD_R จะมีค่าสูงกว่า แต่เมื่อจำนวนตัวแปรที่มีค่าพารามิเตอร์ไม่เป็น 0 เพิ่มขึ้นเป็นตัวแบบที่มีความบางเบาน้อยลง หรือ $\boldsymbol{\beta}_{(2)}$, และ $\boldsymbol{\beta}_{(3)}$ จะเห็นได้ว่าวิธี AD_SR และวิธี AD_R จะมีค่า AUC มากกว่าวิธีอื่น ๆ อย่างใกล้เคียงกัน

ในกรณีที่ 2 ตัวแปรอธิบาย $X_1, \ldots, X_p$ มีความสัมพันธ์กัน โดยที่เมทริกซ์ของตัวแปรอธิบายแบบต่อเนื่องกรณีที่ 2 จะใช้สัญลักษณ์แทนด้วย $\boldsymbol{X}^*_{\boldsymbol{cor}(p\times p)}$ โดยที่ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง X_k และ X_j คือ $\rho^{|k-j|}$; $j, k = 1,2, \ldots, p$ ดังนั้น $\boldsymbol{X}^*_{\boldsymbol{cor}} \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma_b})$ เมื่อเมทริกซ์ความแปรปรวนร่วม คือ

$$\text{โดยที่ } \boldsymbol{\Sigma_b} = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{p-1} \\ \rho & 1 & \rho & \cdots & \rho^{p-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{p-3} \\ \vdots & \vdots & \vdots & \ddots & \rho \\ \rho^{p-1} & \rho^{p-2} & \cdots & \rho & 1 \end{bmatrix}_{p\times p}$$

ตารางที่ 4 ค่าความไวของการทำนายด้วยสมการถดถอยโลจิสติกเรกูลาไรซ์ทั้ง 4 วิธี ในกรณีที่ $X^*_{\text{cor}} \sim N_{\text{p}}(0, \boldsymbol{\Sigma_b})$

n	p+4	ρ	$\boldsymbol{\beta}_{(1)}$				$\boldsymbol{\beta}_{(2)}$				$\boldsymbol{\beta}_{(3)}$			
			L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR
100	204	0.3	0.9031	0.9560	0.9372	**0.9594**	0.8558	0.9019	0.8630	**0.9078**	0.8128	**0.8476**	0.8094	0.8427
		0.6	0.9265	**0.9726**	0.9476	0.9713	0.8914	**0.9354**	0.8933	0.9345	0.8724	0.9085	0.8654	**0.9094**
		0.9	0.9479	**0.9791**	0.9379	0.9779	0.9462	0.9642	0.9336	**0.9661**	0.9310	0.9524	0.9144	**0.9534**
	304	0.3	0.9067	0.9550	0.9525	**0.9647**	0.8325	0.8816	0.8454	**0.8858**	0.7930	0.8283	0.7907	**0.8314**
		0.6	0.9264	0.9680	0.9466	**0.9707**	0.8894	**0.9230**	0.8896	0.9210	0.8669	0.9011	0.8622	**0.9041**
		0.9	0.9415	0.9770	0.9368	**0.9785**	0.9333	0.9616	0.9264	**0.9651**	0.9379	**0.9576**	0.9269	0.9542
	404	0.3	0.9040	0.9537	0.9492	**0.9605**	0.8353	0.8832	0.8416	**0.8882**	0.7876	0.8235	0.7884	**0.8258**
		0.6	0.9326	0.9703	0.9459	**0.9719**	0.9047	**0.9291**	0.8996	0.9288	0.8656	**0.8955**	0.8488	0.8943
		0.9	0.9463	0.9785	0.9431	**0.9806**	0.9387	**0.9665**	0.9309	0.9648	0.9359	**0.9532**	0.9285	0.9523
200	404	0.3	0.9429	0.9834	0.9740	**0.9865**	0.9146	**0.9476**	0.9366	0.9461	0.8828	0.9124	0.8903	**0.9178**
		0.6	0.9610	**0.9924**	0.9818	0.9913	0.9345	0.9604	0.9390	**0.9621**	0.9214	**0.9387**	0.9180	0.9357
		0.9	0.9634	0.9925	0.9586	**0.9940**	0.9613	0.9816	0.9559	**0.9827**	0.9533	0.9723	0.9488	**0.9744**
	604	0.3	0.9456	**0.9863**	0.9766	0.9854	0.9102	0.9453	0.9344	**0.9477**	0.8792	0.9057	0.8828	**0.9113**
		0.6	0.9562	0.9904	0.9809	**0.9922**	0.9312	0.9649	0.9457	**0.9664**	0.9144	0.9379	0.9117	**0.9388**
		0.9	0.9649	0.9924	0.9606	**0.9927**	0.9604	0.9816	0.9562	**0.9824**	0.9548	0.9718	0.9465	**0.9728**
	804	0.3	0.9467	0.9842	0.9805	**0.9863**	0.8995	0.9409	0.9257	**0.9442**	0.8699	0.8965	0.8784	**0.9034**
		0.6	0.9524	0.9906	0.9812	**0.9911**	0.9288	0.9580	0.9436	**0.9606**	0.9167	0.9339	0.9070	**0.9360**
		0.9	0.9609	0.9931	0.9610	**0.9942**	0.9588	**0.9826**	0.9538	0.9820	0.9530	0.9739	0.9484	**0.9761**

จากตารางที่ 4 จะเห็นได้ว่าค่าความไวของวิธี AD_SR มีค่ามากกว่าวิธีอื่น ๆ ในหลายกรณี โดยเฉพาะอย่างยิ่งเมื่อ $n = 200$

ตารางที่ 5 ค่าความจำเพาะของการทำนายด้วยสมการถดถอยโลจิสติกเรกูลาไรซ์ทั้ง 4 วิธี ในกรณีที่ $X^{*}_{cor} \sim N_p(0, \mathbf{\Sigma_b})$

n	p+4	ρ	$\boldsymbol{\beta}_{(1)}$				$\boldsymbol{\beta}_{(2)}$				$\boldsymbol{\beta}_{(3)}$			
			L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR
100	204	0.3	**0.9248**	0.8818	0.9148	0.8809	**0.8565**	0.8340	0.8252	0.8365	**0.8176**	0.7865	0.7713	0.7837
		0.6	**0.9401**	0.9103	0.9291	0.9065	**0.9049**	0.8920	0.8705	0.8882	**0.8851**	0.8788	0.8499	0.8786
		0.9	**0.9471**	0.9334	0.9425	0.9320	**0.9404**	0.9297	0.9198	0.9323	**0.9255**	0.9240	0.9069	0.9249
	304	0.3	**0.9383**	0.8794	0.9220	0.8855	**0.8439**	0.8171	0.8056	0.8163	**0.7917**	0.7736	0.7486	0.7780
		0.6	**0.9400**	0.9045	0.9232	0.9047	**0.8964**	0.8836	0.8658	0.8846	**0.8695**	0.8613	0.8378	0.8622
		0.9	**0.9500**	0.9227	0.9380	0.9251	**0.9346**	0.9342	0.9182	0.9335	**0.9308**	0.9284	0.9090	0.9246
	404	0.3	**0.9088**	0.8629	0.9009	0.8699	**0.8353**	0.8163	0.7924	0.8190	**0.7917**	0.7686	0.7402	0.7706
		0.6	0.9280	0.9072	**0.9283**	0.9054	**0.8985**	0.8768	0.8670	0.8738	**0.8713**	0.8552	0.8196	0.8574
		0.9	**0.9441**	0.9245	0.9354	0.9231	0.9340	**0.9367**	0.9199	0.9365	0.9267	**0.9285**	0.9049	0.9213
200	404	0.3	**0.9610**	0.9048	0.9588	0.9012	**0.9243**	0.8883	0.9006	0.8822	**0.8980**	0.8571	0.8508	0.8588
		0.6	0.9611	0.9259	**0.9657**	0.9220	**0.9378**	0.9156	0.9161	0.9153	**0.9187**	0.8934	0.8864	0.8925
		0.9	**0.9658**	0.9428	0.9639	0.9420	**0.9596**	0.9550	0.9440	0.9526	**0.9531**	0.9494	0.9376	0.9503
	604	0.3	**0.9568**	0.9032	0.9534	0.8974	**0.9220**	0.8856	0.8926	0.8837	**0.8832**	0.8536	0.8416	0.8577
		0.6	0.9662	0.9258	**0.9681**	0.9258	**0.9372**	0.9195	0.9216	0.9166	**0.9141**	0.8992	0.8904	0.9009
		0.9	**0.9687**	0.9416	0.9614	0.9410	**0.9569**	0.9529	0.9449	0.9505	**0.9510**	0.9498	0.9365	0.9494
	804	0.3	**0.9606**	0.9057	0.9583	0.9047	**0.9165**	0.8830	0.8933	0.8819	**0.8772**	0.8381	0.8278	0.8410
		0.6	0.9637	0.9290	**0.9662**	0.9260	**0.9319**	0.9131	0.9112	0.9049	**0.9183**	0.8911	0.8776	0.8885
		0.9	**0.9728**	0.9481	0.9628	0.9442	**0.9614**	0.9577	0.9486	0.9557	**0.9500**	0.9497	0.9357	0.9488

จากตารางที่ 5 จะเห็นได้ว่าค่าความจำเพาะของวิธี LASSO มีค่าสูงกว่าวิธีการประมาณค่าอื่น ๆ แทบทุกกรณี มีเพียงบางกรณีเท่านั้นที่มีค่าน้อยกว่า แต่ไม่ได้แตกต่างกันมากนัก เช่น กรณีที่ $\boldsymbol{\beta}_{(1)}$ เมื่อ $n = 200, \rho = 0.6$ ในทุก ๆ ค่า p ที่วิธี AD_L จะมีค่าเฉลี่ยของค่าความจำเพาะสูงกว่าวิธี LASSO แต่ก็มีค่าที่ใกล้เคียงกัน

ตารางที่ 6 ค่า AUC ของเส้นกราฟ ROC curve ของการทำนายด้วยสมการถดถอยโลจิสติกเรกูลาไรซ์ทั้ง 4 วิธี ในกรณีที่ $X^*_{cor} \sim N_p(0, \mathbf{\Sigma_b})$

n	p+4	ρ	$\boldsymbol{\beta}_{(1)}$				$\boldsymbol{\beta}_{(2)}$				$\boldsymbol{\beta}_{(3)}$			
			L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR
100	204	0.3	0.9078	0.9216	**0.9262**	0.9227	0.8520	0.8699	0.8448	**0.8743**	0.8100	**0.8181**	0.7900	0.8142
		0.6	0.9300	**0.9430**	0.9381	0.9406	0.8950	**0.9143**	0.8817	0.9123	0.8756	0.8938	0.8577	**0.8942**
		0.9	0.9455	**0.9574**	0.9373	0.9560	0.9417	0.9471	0.9259	**0.9493**	0.9265	0.9387	0.9096	**0.9396**
	304	0.3	0.9151	0.9194	**0.9379**	0.9273	0.8330	0.8509	0.8256	**0.8525**	0.7878	0.8019	0.7702	**0.8055**
		0.6	0.9295	**0.9400**	0.9352	0.9384	0.8891	**0.9042**	0.8777	0.9038	0.8652	0.8820	0.8501	**0.8837**
		0.9	0.9436	**0.9532**	0.9350	0.9514	0.9317	0.9481	0.9208	**0.9498**	0.9330	**0.9436**	0.9176	0.9400
	404	0.3	0.9013	0.9118	**0.9274**	0.9188	0.8304	0.8514	0.8181	**0.8556**	0.7827	0.7968	0.7644	**0.7990**
		0.6	0.9279	**0.9409**	0.9361	0.9408	0.8994	**0.9038**	0.8836	0.9022	0.8656	0.8764	0.8349	**0.8771**
		0.9	0.9431	0.9525	0.9382	**0.9528**	0.9353	**0.9519**	0.9249	0.9511	0.9301	**0.9412**	0.9161	0.9372
200	404	0.3	0.9482	0.9473	**0.9665**	0.9471	0.9170	**0.9196**	0.9195	0.9157	0.8870	0.8860	0.8714	**0.8895**
		0.6	0.9598	0.9615	**0.9743**	0.9591	0.9349	0.9388	0.9279	**0.9395**	**0.9191**	0.9169	0.9028	0.9150
		0.9	0.9637	**0.9690**	0.9601	0.9688	0.9599	**0.9686**	0.9499	0.9679	0.9524	0.9611	0.9429	**0.9626**
	604	0.3	0.9481	0.9482	**0.9658**	0.9448	0.9136	0.9168	0.9145	**0.9170**	0.8791	0.8808	0.8630	**0.8857**
		0.6	0.9594	0.9602	**0.9747**	0.9611	0.9329	**0.9430**	0.9342	0.9423	0.9131	0.9192	0.9015	**0.9205**
		0.9	0.9660	**0.9684**	0.9603	0.9682	0.9583	**0.9678**	0.9505	0.9669	0.9524	0.9610	0.9415	**0.9613**
	804	0.3	0.9495	0.9484	**0.9712**	0.9489	0.9047	0.9135	0.9104	**0.9145**	0.8708	0.8684	0.8542	**0.8733**
		0.6	0.9574	0.9613	**0.9730**	0.9601	0.9291	**0.9365**	0.9282	0.9337	**0.9159**	0.9134	0.8928	0.9132
		0.9	0.9656	**0.9716**	0.9611	0.9701	0.9594	**0.9705**	0.9509	0.9692	0.9510	0.9620	0.9423	**0.9626**

จากตารางที่ 6 จะเห็นได้ว่า ในกรณีที่ตัวแบบมีความบางเบามาก ๆ หรือ $\boldsymbol{\beta}_{(1)}$ ค่า AUC ของกราฟ ROC curve ของวิธี AD_R และวิธี AD_L จะมีค่าสูงกว่าวิธีอื่น ๆ โดยเฉพาะอย่างยิ่งในกรณีที่ $n = 200$ เมื่อ $\rho = 0.3, 0.6$ วิธี AD_L จะมีค่ามากกว่า แต่เมื่อ $\rho = 0.9$ วิธี AD_R จะมีค่าสูงกว่า แต่เมื่อจำนวนตัวแปรที่มีค่าพารามิเตอร์ไม่เป็น 0 เพิ่มขึ้นเป็นตัวแบบที่มีความบางเบาน้อยลง หรือ $\boldsymbol{\beta}_{(2)}$, และ $\boldsymbol{\beta}_{(3)}$ จะเห็นได้ว่าจะเป็นวิธี AD_SR และวิธี AD_R ที่จะมีค่า AUC สูงกว่าวิธีอื่น ๆ อย่างใกล้เคียงกัน

4.2 ความถูกต้องของการประมาณค่าด้วยสมการถดถอยโลจิสติกด้วยวิธีเรกูลาไรซ์ 4 วิธี

ความถูกต้องของการประมาณค่าของตัวประมาณแต่ละวิธี จะวัดจากค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าสัมประสิทธิ์การถดถอย (Mean_MSE) จากการจำลองข้อมูล 500 รอบ ซึ่งผลการวิจัยแบ่งออกเป็น 2 กรณี มีรายละเอียดดังนี้

ในกรณีที่ 1 ตัวแปรอธิบายจะถูกแบ่งเป็น 2 กลุ่มที่เป็นอิสระต่อกัน แต่มีความสัมพันธ์กันภายในกลุ่ม หรือ $\boldsymbol{X}^*_{\boldsymbol{ind}} = \left[\boldsymbol{X}_{\mathbf{1}_{(\boldsymbol{n}\times\mathbf{15})}} \vdots \boldsymbol{X}_{\mathbf{2}_{(\boldsymbol{n}\times(\boldsymbol{p}-\mathbf{15}))}}\right]$ โดยที่ $\boldsymbol{X}^*_{\boldsymbol{ind}} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{a}})$

ตารางที่ 7 ค่า MSE ของการประมาณค่าด้วยสมการถดถอยโลจิสติกเรกูลาไรซ์ทั้ง 4 วิธี ในกรณีที่ $X^*_{\text{ind}} \sim N_{\text{p}}(0, \boldsymbol{\Sigma}_{\mathbf{a}})$

n	p+4	ρ	$\boldsymbol{\beta}_{(1)}$				$\boldsymbol{\beta}_{(2)}$				$\boldsymbol{\beta}_{(3)}$			
			L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR
100	204	0.3	0.0846	0.0384	**0.0227**	0.0331	0.1739	0.1183	**0.1121**	0.1198	0.2308	0.1854	**0.1833**	0.1842
		0.6	0.0884	0.0392	0.0471	**0.0389**	0.1816	0.1178	0.1349	**0.1170**	0.2319	0.1776	0.1960	**0.1735**
		0.9	0.0961	**0.0666**	0.0984	0.0669	0.1885	**0.1518**	0.1855	0.1569	0.2396	0.2058	0.2363	**0.2031**
	304	0.3	0.0608	0.0339	**0.0165**	0.0234	0.1215	0.0880	**0.0784**	0.0810	0.1608	0.1308	**0.1272**	0.1285
		0.6	0.0620	**0.0252**	0.0264	0.0268	0.1238	0.0866	0.0933	**0.0855**	0.1606	**0.1220**	0.1322	0.1252
		0.9	0.0686	**0.0451**	0.0697	0.0489	0.1305	**0.1053**	0.1243	0.1059	0.1656	**0.1440**	0.1627	0.1450
	404	0.3	0.0452	0.0328	**0.0125**	0.0185	0.0950	0.0696	**0.0607**	0.0663	0.1240	0.1040	**0.0988**	0.1021
		0.6	0.0470	**0.0202**	0.0247	0.0204	0.0970	**0.0636**	0.0723	0.0657	0.1247	0.0959	0.1024	**0.0950**
		0.9	0.0511	**0.0323**	0.0508	0.0363	0.0988	**0.0790**	0.0937	0.0809	0.1254	0.1072	0.1209	**0.1064**
200	404	0.3	0.0351	**0.0072**	0.0077	0.0082	0.0780	0.0330	**0.0214**	0.0331	0.1045	0.0574	**0.0502**	0.0570
		0.6	0.0380	0.0086	0.0084	**0.0082**	0.0819	0.0372	0.0449	**0.0361**	0.1084	**0.0647**	0.0756	0.0651
		0.9	0.0391	0.0210	0.0363	**0.0196**	0.0865	0.0631	0.0804	**0.0608**	0.1137	**0.0860**	0.1098	0.0882
	604	0.3	0.0252	0.0055	**0.0049**	0.0050	0.0536	0.0235	**0.0134**	0.0230	0.0727	0.0440	**0.0340**	0.0414
		0.6	0.0249	0.0060	0.0054	**0.0053**	0.0568	0.0252	0.0305	**0.0245**	0.0751	0.0450	0.0504	**0.0439**
		0.9	0.0274	**0.0136**	0.0241	0.0146	0.0607	0.0425	0.0542	**0.0408**	0.0796	0.0601	0.0728	**0.0590**
	804	0.3	0.0207	0.0046	**0.0040**	0.0042	0.0426	0.0195	**0.0119**	0.0191	0.0568	0.0343	**0.0307**	0.0340
		0.6	0.0201	**0.0041**	0.0046	0.0045	0.0436	0.0209	0.0236	**0.0206**	0.0580	0.0357	0.0383	**0.0353**
		0.9	0.0219	0.0094	0.0185	**0.0087**	0.0457	**0.0310**	0.0414	0.0312	0.0596	0.0454	0.0551	**0.0447**

จากตารางที่ 7 จะเห็นได้ว่า ค่า MSE ของวิธี Adaptive LASSO ที่มีค่าถ่วงน้ำหนักที่แตกต่างกันทั้งสามวิธี มีค่าน้อยกว่าวิธี LASSO ทุกกรณี จึงจะเห็นได้ว่า การเลือกใช้วิธี Adaptive LASSO ในการประมาณค่าพารามิเตอร์สำหรับตัวแบบถดถอยโลจิสติกพหุคูณมีความถูกต้องในการประมาณค่ามากกว่าวิธี LASSO แบบปกติ

และในกรณีที่ 2 ตัวแปรอธิบายจะมีเพียง 1 กลุ่ม จำนวน p ตัวที่สัมพันธ์กันทั้งหมด หรือ $\boldsymbol{X}^*_{\boldsymbol{cor}\,(p\times p)}$ โดยที่ $\boldsymbol{X}^*_{\boldsymbol{cor}} \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma_b})$

ตารางที่ 8 ค่า MSE ของการประมาณค่าด้วยสมการถดถอยโลจิสติกเรกูลาไรซ์ทั้ง 4 วิธี ในกรณีที่ $X^*_{\text{cor}} \sim N_{\text{p}}(0, \boldsymbol{\Sigma_b})$

n	p+4	ρ	$\boldsymbol{\beta}_{(1)}$				$\boldsymbol{\beta}_{(2)}$				$\boldsymbol{\beta}_{(3)}$			
			L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR
100	204	0.3	0.0833	0.0307	**0.0234**	0.0348	0.1675	0.1165	**0.1067**	0.1113	0.2237	0.1849	**0.1805**	0.1822
		0.6	0.0878	**0.0367**	0.0412	0.0398	0.1829	0.1261	0.1355	**0.1179**	0.2325	0.1820	0.1899	**0.1799**
		0.9	0.0938	**0.0650**	0.0964	0.0697	0.1885	0.1570	0.1870	**0.1556**	0.2448	0.2092	0.2423	**0.2070**
	304	0.3	0.0608	0.0403	**0.0166**	0.0238	0.1196	0.0840	**0.0785**	0.0841	0.1617	0.1317	**0.1270**	0.1296
		0.6	0.0623	**0.0276**	0.0305	0.0299	0.1248	0.0851	0.0895	**0.0828**	0.1612	**0.1244**	0.1300	0.1246
		0.9	0.0646	0.0437	0.0651	**0.0430**	0.1288	0.1037	0.1260	**0.1018**	0.1660	**0.1411**	0.1609	0.1443
	404	0.3	0.0477	0.0271	**0.0118**	0.0187	0.0944	0.0648	**0.0622**	0.0665	0.1262	0.1029	0.1016	**0.1012**
		0.6	0.0450	0.0220	**0.0219**	0.0205	0.0944	**0.0654**	0.0687	0.0661	0.1210	**0.0958**	0.1007	0.0967
		0.9	0.0515	0.0357	0.0506	**0.0345**	0.0999	0.0781	0.0949	**0.0773**	0.1253	**0.1058**	0.1240	0.1088
200	404	0.3	0.0365	0.0084	**0.0077**	0.0079	0.0774	0.0298	**0.0208**	0.0278	0.1045	0.0571	**0.0500**	0.0576
		0.6	0.0372	0.0088	0.0078	**0.0074**	0.0814	0.0370	0.0454	**0.0359**	0.1092	0.0641	0.0740	**0.0617**
		0.9	0.0402	0.0215	0.0388	**0.0200**	0.0887	**0.0597**	0.0822	0.0620	0.1169	**0.0909**	0.1095	0.0910
	604	0.3	0.0255	0.0053	**0.0051**	0.0056	0.0553	0.0231	**0.0163**	0.0231	0.0715	0.0422	**0.0371**	0.0437
		0.6	0.0261	0.0061	**0.0056**	0.0060	0.0566	**0.0239**	0.0297	0.0261	0.0751	**0.0447**	0.0511	0.0448
		0.9	0.0275	0.0130	0.0236	**0.0123**	0.0606	**0.0417**	0.0545	0.0423	0.0785	**0.0595**	0.0716	0.0595
	804	0.3	0.0195	0.0044	**0.0039**	0.0051	0.0418	0.0193	**0.0124**	0.0201	0.0571	0.0347	**0.0288**	0.0341
		0.6	0.0199	0.0051	**0.0043**	0.0051	0.0436	0.0209	0.0236	**0.0201**	0.0570	0.0339	0.0379	**0.0337**
		0.9	0.0220	**0.0100**	0.0182	0.0102	0.0445	**0.0303**	0.0396	0.0321	0.0603	0.0443	0.0549	**0.0433**

จากตารางที่ 8 สามารถสรุปได้เช่นเดียวกันกับตารางที่ 7 คือ ค่า MSE ของวิธี Adaptive LASSO ที่มีค่าถ่วงน้ำหนักที่แตกต่างกันทั้งสามวิธี มีค่าน้อยกว่าวิธี LASSO ทุกกรณี จึงจะเห็นได้ว่า การเลือกใช้วิธี Adaptive LASSO ในการประมาณค่าพารามิเตอร์สำหรับตัวแบบถดถอยโลจิสติกพหุคูณมีความถูกต้องในการประมาณค่ามากกว่าวิธี LASSO แบบปกติ

4.3 ความถูกต้องของการคัดเลือกตัวแปรเข้าสู่ตัวแบบด้วยสมการถดถอยโลจิสติกด้วยวิธีเรกูลาไรซ์ 4 วิธี

ความถูกต้องของการคัดเลือกตัวแปรเข้าสู่ตัวแบบสำหรับตัวแบบถดถอยโลจิสติกของตัวประมาณแต่ละวิธี จะวัดจากค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรแบบที่ 1 เมื่อเทียบกับจำนวนตัวแปรอธิบายในตัวแบบ (%IC1), และค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรแบบที่ 2 เมื่อเทียบกับจำนวนตัวแปรอธิบายในตัวแบบ (%IC2) จากการจำลองข้อมูล 500 รอบ ซึ่งผลการวิจัยแบ่งออกเป็น 2 กรณี มีรายละเอียดดังนี้

ในกรณีที่ 1 ตัวแปรอธิบายจะถูกแบ่งเป็น 2 กลุ่มที่เป็นอิสระต่อกัน แต่มีความสัมพันธ์กันภายในกลุ่ม หรือ $\boldsymbol{X}^*_{ind} = \left[\boldsymbol{X}_{1_{(n\times 15)}} \vdots \boldsymbol{X}_{2_{(n\times(p-15))}}\right]$ โดยที่ $\boldsymbol{X}^*_{ind} \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_a)$

ตารางที่ 9 ค่า %IC1 ของการคัดเลือกตัวแปรอธิบายด้วยสมการถดถอยโลจิสติกเรกูลาไรซ์ทั้ง 4 วิธี ในกรณีที่ $X^*_{ind} \sim N_p(0, \boldsymbol{\Sigma}_a)$

n	p+4	ρ	$\boldsymbol{\beta}_{(1)}$				$\boldsymbol{\beta}_{(2)}$				$\boldsymbol{\beta}_{(3)}$			
			L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR
100	204	0.3	43.89	**42.44**	45.44	42.89	34.36	33.14	42.71	**32.71**	38.32	37.95	50.95	**37.89**
		0.6	44.67	44.56	50.33	**44.44**	40.43	**35.07**	54.00	35.14	42.21	38.68	59.21	**38.11**
		0.9	52.44	**51.11**	70.00	**51.11**	54.43	**50.79**	74.29	52.14	55.89	55.32	76.42	**54.89**
	304	0.3	**43.56**	**43.56**	46.11	43.89	36.14	32.86	43.43	**32.21**	42.58	39.53	53.47	**38.68**
		0.6	45.00	**44.56**	48.67	44.78	39.71	37.14	53.21	**36.57**	43.32	**37.95**	59.26	39.05
		0.9	54.67	**51.78**	71.22	53.33	57.64	54.21	74.07	**53.50**	58.00	57.47	78.37	**57.32**
	404	0.3	44.33	**42.56**	46.11	44.11	39.43	34.86	45.71	**34.57**	44.79	**41.42**	55.32	41.79
		0.6	45.56	**44.22**	52.33	44.78	41.86	**35.79**	55.86	**35.79**	47.05	41.16	61.37	**40.53**
		0.9	54.56	**51.56**	71.00	53.67	56.29	52.50	75.00	**52.43**	58.58	57.47	77.00	**56.16**
200	404	0.3	**43.22**	43.56	44.44	43.89	**27.43**	27.93	29.50	27.79	24.68	23.42	29.05	**22.79**
		0.6	44.22	**44.11**	44.78	44.22	30.79	29.00	38.21	**28.29**	27.74	**24.42**	41.37	24.68
		0.9	**44.67**	46.11	59.78	45.67	43.21	40.21	65.71	**39.14**	45.21	**41.26**	70.37	41.42
	604	0.3	44.00	44.11	44.44	**43.78**	28.29	28.43	29.29	**28.21**	24.95	24.05	30.58	**23.79**
		0.6	**43.89**	44.44	44.67	44.44	30.79	28.79	38.43	**28.57**	30.37	**25.21**	42.89	25.32
		0.9	47.11	**46.33**	60.11	**46.33**	44.29	40.50	65.79	**39.29**	46.53	42.68	69.58	**41.95**
	804	0.3	**43.92**	44.44	44.44	44.44	29.64	28.43	30.36	**28.36**	28.89	25.47	33.68	**25.21**
		0.6	**44.22**	**44.22**	45.11	**44.22**	31.64	**28.71**	39.07	28.79	31.63	**26.58**	43.68	**26.58**
		0.9	47.67	45.44	61.67	**44.78**	43.71	**38.86**	66.50	39.71	47.95	**41.84**	70.32	42.00

จากตารางที่ 9 จะเห็นได้ว่า ในกรณีที่ตัวแบบมีความบางเบามาก ๆ หรือ $\boldsymbol{\beta}_{(1)}$ ค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรแบบที่ 1 (%IC1) ที่มีความหมายคือ จำนวนตัวแปรที่ควรอยู่ใน

ตัวแบบ แต่ไม่ถูกคัดเลือกเข้าสู่ตัวแบบ ของวิธี AD_R และ AD_L จะมีค่าน้อยกว่าวิธีอื่น ๆ แต่เมื่อจำนวนตัวแปรที่มีค่าพารามิเตอร์ไม่เป็น 0 เพิ่มขึ้นเป็นตัวแบบที่มีความบางเบาน้อยลง หรือ $\boldsymbol{\beta}_{(2)}$, และ $\boldsymbol{\beta}_{(3)}$ จะเห็นได้ว่าจะเป็นวิธี AD_SR จะมี %IC1 น้อยกว่าวิธีอื่น ๆ

ตารางที่ 10 ค่า %IC2 ของการคัดเลือกตัวแปรอธิบายด้วยสมการถดถอยโลจิสติกเรกูลาไรซ์ทั้ง 4 วิธี ในกรณีที่ $X^{*}_{\text{ind}} \sim N_{\text{p}}(0, \boldsymbol{\Sigma}_{\mathbf{a}})$

n	p+4	ρ	$\boldsymbol{\beta}_{(1)}$				$\boldsymbol{\beta}_{(2)}$				$\boldsymbol{\beta}_{(3)}$			
			L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR
100	204	0.3	3.65	3.35	0.34	2.05	6.21	4.17	1.43	4.07	7.88	5.61	2.34	5.62
		0.6	2.24	0.94	0.24	0.99	3.01	1.88	0.51	1.88	4.46	2.22	0.90	2.33
		0.9	1.37	0.37	0.10	0.35	1.71	0.41	0.21	0.31	1.83	0.24	0.22	0.25
	304	0.3	2.38	3.82	0.26	2.00	4.61	4.55	1.12	3.81	5.55	4.25	1.68	4.54
		0.6	1.24	0.78	0.10	0.79	2.58	1.39	0.51	1.37	2.96	1.82	0.60	1.78
		0.9	0.57	0.28	0.08	0.28	1.07	0.30	0.11	0.28	1.18	0.15	0.14	0.16
	404	0.3	2.27	4.40	0.33	1.76	3.63	4.16	0.88	3.25	4.27	4.71	1.39	3.76
		0.6	1.47	0.78	0.16	0.80	1.53	1.12	0.24	1.05	2.10	1.54	0.46	1.61
		0.9	0.68	0.25	0.06	0.26	0.95	0.25	0.08	0.22	1.04	0.18	0.09	0.19
200	404	0.3	1.46	0.92	0.05	0.93	3.50	2.36	0.41	2.30	4.68	3.54	0.88	3.62
		0.6	1.10	0.31	0.03	0.42	2.27	1.00	0.29	1.01	2.93	1.25	0.43	1.27
		0.9	0.98	0.12	0.02	0.12	1.31	0.13	0.06	0.14	1.30	0.05	0.03	0.04
	604	0.3	1.15	0.66	0.03	0.81	2.57	1.98	0.34	2.00	3.44	2.85	0.81	3.28
		0.6	1.07	0.29	0.03	0.35	1.39	0.72	0.14	0.76	1.95	1.11	0.28	1.15
		0.9	0.69	0.10	0.02	0.10	0.62	0.08	0.03	0.09	0.57	0.05	0.03	0.07
	804	0.3	0.60	0.46	0.01	0.65	1.76	1.65	0.23	1.70	2.77	2.71	0.57	2.78
		0.6	0.67	0.29	0.01	0.28	0.98	0.54	0.11	0.56	1.40	0.85	0.22	0.95
		0.9	0.42	0.08	0.01	0.09	0.64	0.07	0.04	0.09	0.56	0.03	0.03	0.04

จากตารางที่ 10 จะเห็นได้ว่า ค่าร้อยละของความผิดในการคัดเลือกตัวแปรอธิบายแบบที่ 2 (%IC2) ที่มีความหมายคือ จำนวนตัวแปรที่ไม่ควรอยู่ในตัวแบบ แต่กลับถูกคัดเลือกเข้าสู่ตัวแบบของวิธี AD_L มีค่าน้อยกว่าวิธีการประมาณค่าวิธีอื่น ๆ ในทุกกรณี

และในกรณีที่ 2 ตัวแปรอธิบายจะมีเพียง 1 กลุ่ม จำนวน p ตัวที่สัมพันธ์กันทั้งหมด หรือ $\boldsymbol{X}^{*}_{\boldsymbol{cor}\,(p\times p)}$ โดยที่ $\boldsymbol{X}^{*}_{\boldsymbol{cor}} \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma_b})$

ตารางที่ 11 ค่า %IC1 ของการคัดเลือกตัวแปรอธิบายด้วยสมการถดถอยโลจิสติกเรกูลาไรซ์ทั้ง 4 วิธี ในกรณีที่ $X^{*}_{\mathrm{cor}} \sim N_{\mathrm{p}}(0, \boldsymbol{\Sigma}_{\mathbf{b}})$

n	p+4	ρ	$\boldsymbol{\beta}_{(1)}$				$\boldsymbol{\beta}_{(2)}$				$\boldsymbol{\beta}_{(3)}$			
			L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR
100	204	0.3	43.67	43.56	46.11	**43.22**	32.64	32.50	41.64	**31.57**	36.32	37.53	49.11	**37.32**
		0.6	44.56	**44.11**	49.00	44.33	39.21	36.43	51.79	**35.21**	40.63	38.16	56.84	**38.05**
		0.9	53.78	**51.78**	68.33	52.22	52.71	52.79	74.14	**51.79**	57.21	**55.63**	77.53	55.84
	304	0.3	44.44	**42.56**	45.56	44.00	35.36	**33.43**	44.21	33.57	41.84	**38.74**	53.05	39.16
		0.6	46.11	**45.00**	51.44	45.44	41.00	36.64	52.79	**36.21**	43.16	39.63	57.63	**38.63**
		0.9	52.56	**51.56**	69.22	**51.56**	55.57	52.14	74.71	**50.71**	60.42	**56.42**	100.00	57.89
	404	0.3	44.44	**42.56**	45.67	43.33	38.07	**33.86**	45.57	34.36	47.16	**39.32**	57.68	41.58
		0.6	45.11	44.67	49.44	**44.33**	40.79	**36.36**	53.29	36.43	44.26	**40.32**	59.79	40.42
		0.9	55.78	53.11	71.44	**52.67**	56.14	51.57	75.29	**51.36**	59.63	**55.11**	78.79	56.84
200	404	0.3	44.22	44.22	44.44	**44.11**	28.00	**27.79**	29.57	27.86	24.05	**22.84**	28.84	22.95
		0.6	**43.78**	44.33	44.78	44.11	30.36	29.14	37.36	**28.50**	28.74	**25.47**	41.79	25.53
		0.9	46.89	46.44	62.67	**45.78**	41.79	**37.29**	66.50	38.14	46.95	42.68	70.79	**42.16**
	604	0.3	**44.00**	44.11	44.44	44.33	28.50	**28.29**	29.64	28.43	25.05	24.26	30.89	**24.05**
		0.6	**43.89**	**43.89**	44.89	44.11	31.14	**28.79**	37.64	29.43	29.79	**25.21**	42.89	25.58
		0.9	47.00	**45.56**	59.78	46.11	45.57	**39.86**	65.93	40.21	47.11	**42.58**	69.47	42.89
	804	0.3	**44.00**	**44.00**	44.33	**44.00**	28.86	28.29	30.43	**28.21**	27.00	**23.58**	32.21	24.47
		0.6	**44.33**	**44.33**	45.00	44.44	31.07	29.14	37.86	**28.79**	30.05	25.74	42.63	**25.16**
		0.9	46.67	**45.78**	60.22	45.89	43.29	**38.86**	64.29	40.21	48.42	41.79	70.05	**41.16**

จากตารางที่ 11 จะเห็นได้ว่า ในกรณีที่ตัวแบบมีความบางเบามาก ๆ หรือ $\boldsymbol{\beta}_{(1)}$ ค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรแบบที่ 1 (%IC1) ที่มีความหมายคือ จำนวนตัวแปรที่ควรอยู่ในตัวแบบ แต่ไม่ถูกคัดเลือกเข้าสู่ตัวแบบ ของวิธี AD_R และ AD_L จะมีค่าน้อยกว่าวิธีอื่น ๆ แต่เมื่อจำนวนตัวแปรที่มีค่าพารามิเตอร์ไม่เป็น 0 เพิ่มขึ้นเป็นตัวแบบที่มีความบางเบาน้อยลง หรือ $\boldsymbol{\beta}_{(2)}$, และ $\boldsymbol{\beta}_{(3)}$ จะเห็นได้ว่าจะเป็นวิธี AD_SR และวิธี AD_R จะมี %IC1 น้อยกว่าวิธีอื่น ๆ และมีค่าที่ใกล้เคียงกันในระหว่างสองวิธี

ตารางที่ 12 ค่า %IC2 ของการคัดเลือกตัวแปรอธิบายด้วยสมการถดถอยโลจิสติกเรกูลาไรซ์ทั้ง 4 วิธี ในกรณีที่ $X^{*}_{cor} \sim N_{p}(\mathbf{0}, \mathbf{\Sigma_b})$

n	p+4	ρ	$\boldsymbol{\beta}_{(1)}$				$\boldsymbol{\beta}_{(2)}$				$\boldsymbol{\beta}_{(3)}$			
			L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR	L	AD_R	AD_L	AD_SR
100	204	0.3	3.49	2.36	**0.43**	2.07	7.49	4.25	**1.71**	4.55	9.07	5.77	**2.63**	6.01
		0.6	2.45	1.13	**0.26**	1.07	3.18	1.67	**0.65**	1.98	3.73	1.76	**0.92**	1.88
		0.9	1.73	0.54	**0.15**	0.47	1.50	0.36	**0.18**	0.35	1.76	0.42	**0.21**	0.36
	304	0.3	1.87	4.51	**0.20**	1.75	4.72	3.51	**1.20**	3.56	5.13	5.49	**1.49**	4.39
		0.6	1.41	0.81	**0.14**	0.65	2.44	1.44	**0.50**	1.61	3.17	1.86	**0.79**	1.94
		0.9	1.05	0.39	**0.09**	0.31	1.24	0.28	**0.14**	0.25	1.23	0.32	**0.15**	0.28
	404	0.3	1.65	3.55	**0.19**	1.61	3.51	3.27	**0.96**	3.12	3.68	5.73	**1.17**	3.54
		0.6	1.73	0.72	**0.19**	0.67	2.14	1.31	**0.35**	1.25	2.93	1.66	**0.61**	1.53
		0.9	0.71	0.20	**0.05**	0.26	0.82	0.20	**0.07**	0.23	1.07	0.22	**0.10**	0.21
200	404	0.3	1.50	0.92	**0.05**	0.86	3.80	2.65	**0.41**	2.98	5.09	3.87	**1.05**	3.82
		0.6	1.28	0.38	**0.05**	0.42	2.15	0.87	**0.22**	0.90	2.48	1.37	**0.32**	1.45
		0.9	0.96	0.15	**0.03**	0.17	0.93	0.14	**0.05**	0.14	1.03	0.14	**0.06**	0.12
	604	0.3	0.99	0.58	**0.02**	0.73	2.28	2.01	**0.31**	2.07	4.05	3.09	**0.84**	2.97
		0.6	0.74	0.39	**0.03**	0.34	1.52	0.83	**0.14**	0.70	2.05	1.11	**0.26**	1.09
		0.9	0.66	0.10	**0.02**	0.11	0.73	0.09	**0.04**	0.08	0.82	0.09	**0.05**	0.09
	804	0.3	0.84	0.55	**0.02**	0.53	2.04	1.74	**0.26**	1.64	2.52	3.64	**0.58**	2.59
		0.6	0.64	0.24	**0.02**	0.27	1.03	0.64	**0.11**	0.73	1.52	1.03	**0.26**	1.09
		0.9	0.45	0.08	**0.01**	0.08	0.79	0.08	**0.04**	0.08	0.55	0.09	**0.04**	0.10

จากตารางที่ 12 จะเห็นได้ว่า ค่าร้อยละของความผิดในการคัดเลือกตัวแปรอธิบายแบบที่ 2 (%IC2) ที่มีความหมายคือ จำนวนตัวแปรที่ไม่ควรอยู่ในตัวแบบ แต่กลับถูกคัดเลือกเข้าสู่ตัวแบบของวิธี AD_L มีค่าน้อยกว่าวิธีการประมาณค่าวิธีอื่น ๆ ในทุกกรณี

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

งานวิจัยนี้มีวัตถุประสงค์เพื่อทำการศึกษาและเปรียบเทียบประสิทธิภาพการประมาณค่าสัมประสิทธิ์และการคัดเลือกตัวแปรของการเรกูลาไรซ์โดยใช้วิธีแลซโซและวิธีแลซโซแบบปรับได้ของตัวแบบถดถอยโลจิสติกในกรณีที่ข้อมูลมีมิติสูงแบบบางเบาทั้ง 4 วิธี ได้แก่ วิธีแลซโซ, วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณริดจ์, วิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณแลซโซ, และวิธีแลซโซแบบปรับได้ที่ถ่วงน้ำหนักด้วยตัวประมาณสไตน์-ริดจ์ โดยพิจารณาจากความถูกต้องในการทำนาย ความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอย และความถูกต้องในการคัดเลือกตัวแปรอธิบายเข้าสู่ตัวแบบ ภายใต้ขอบเขตงานวิจัย โดยแบ่งเป็น 2 กรณี คือ คือ กรณีที่ 1 คือ ตัวแปรอธิบายจะถูกแบ่งออกเป็นสองกลุ่มที่เป็นอิสระกัน แต่ตัวแปรอธิบายภายในแต่ละกลุ่มมีความสัมพันธ์กันเอง ซึ่งมีเมทริกซ์ของค่าสัมประสิทธิ์สหสัมพันธ์ดัง $\boldsymbol{\Sigma_a}$ และกรณีที่ 2 คือ ตัวแปรอธิบายทั้งหมดจะมีความสัมพันธ์กันเองทุกตัวแปร ซึ่งมีเมทริกซ์ของค่าสัมประสิทธิ์สหสัมพันธ์ดัง $\boldsymbol{\Sigma_b}$ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง X_k และ X_j คำนวณจาก $\rho^{|k-j|}$ เมื่อกำหนดค่าสัมประสิทธิ์สหสัมพันธ์ 3 กรณี คือ ρ = 0.3, 0.6 และ 0.9 รวมถึงมีตัวแปรอธิบายแบบไม่ต่อเนื่องจำนวน 4 ตัวแปรรวมอยู่ด้วย และกำหนดค่าพารามิเตอร์ให้สอดคล้องกับลักษณะของตัวแบบการถดถอยแบบบางเบาในหลากหลายแบบดังขอบเขตของการศึกษา โดยแต่ละกรณีจะทำการจำลองข้อมูลที่มีขนาดตัวอย่าง (n) เท่ากับ 100 และ 200 และจำนวนตัวแปรอธิบายแบบต่อเนื่อง p มีจำนวนเป็น 2, 3, และ 4 เท่าของขนาดตัวอย่าง $\boldsymbol{n}$ กับจำนวนตัวแปรอธิบายแบบไม่ต่อเนื่องมีจำนวน 4 ตัวแปรทุกกรณี รวมจำนวนตัวแปรทั้งสิ้น $p + 4$ ตัวแปร

5.1 สรุปผลการวิจัย

จากผลการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยทั้ง 4 วิธี สามารถสรุปผลการวิจัยแบ่งเป็น 3 ประเด็น ดังนี้

1. ประสิทธิภาพของการทำนาย โดยวัดจากค่าเฉลี่ยของค่าความไว (Mean_sensitivity) ค่าเฉลี่ยของค่าความจำเพาะ (Mean_specificity) และค่าเฉลี่ยของพื้นที่ใต้โค้งของเส้นกราฟ Receiver Operating Characteristic curve หรือ ROC curve (Mean_AUC) ที่ได้จากการการทำนายเป็นเกณฑ์ในการเปรียบเทียบ สามารถสรุปแยกตามรูปแบบความสัมพันธ์ของตัวแปรอธิบายได้ดังนี้

สำหรับกรณีที่ตัวแปรอธิบายแบบต่อเนื่องแบ่งออกเป็นสองกลุ่มที่เป็นอิสระกัน แต่ตัวแปรอธิบายภายในแต่ละกลุ่มมีความสัมพันธ์กันเอง เมื่อพิจารณาถึงค่าความไว (Sensitivity) ที่มี

ความสำคัญคือ เป็นค่าที่บ่งบอกถึงความถูกต้องในการจำแนกกลุ่มที่จะเกิดเหตุการณ์ที่สนใจได้นั้น พบว่า วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีริดจ์ และวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีสไตน์-ริดจ์ มีค่าเฉลี่ยของค่าความไวมากกว่าวิธีอื่น ๆ ซึ่งถ้าเปรียบเทียบระหว่างสองวิธีนี้ ในกรณีที่ตัวแบบบางเบามาก ๆ วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีริดจ์จะมีค่าเฉลี่ยของค่าความไวสูงกว่า แต่เมื่อจำนวนตัวแปรที่มีค่าพารามิเตอร์ไม่เป็น 0 เพิ่มขึ้นทำให้ตัวแบบมีความบางเบาน้อยลง วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีสไตน์-ริดจ์จะมีค่าเฉลี่ยของค่าความไวมากกว่า จึงสรุปได้ว่า วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีริดจ์ และวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีสไตน์-ริดจ์ มีความถูกต้องในการจำแนกกลุ่มที่จะเกิดเหตุการณ์ที่สนใจได้มากกว่าวิธีอื่น

เมื่อพิจารณาถึงค่าความจำเพาะ (Specificity) ที่มีความสำคัญคือ เป็นค่าที่บ่งบอกถึงความถูกต้องในการจำแนกกลุ่มที่จะเกิดเหตุการณ์ที่ไม่สนใจได้นั้น พบว่า ค่าเฉลี่ยของค่าความจำเพาะของวิธี LASSO มีค่าที่สูงกว่าวิธีการประมาณค่าอื่น ๆ แทบทุกกรณี มีเพียงบางกรณีเท่านั้นที่มีค่าเฉลี่ยน้อยกว่า แต่ไม่ได้มีค่าแตกต่างกันมากนัก จึงสรุปได้ว่า วิธี LASSO มีความถูกต้องในการจำแนกกลุ่มที่จะเกิดเหตุการณ์ที่ไม่สนใจได้มากกว่าวิธีอื่น

และเมื่อพิจารณาถึงค่าพื้นที่ใต้เส้นโค้งของกราฟ ROC curve (Mean_AUC) ที่หมายถึงประสิทธิภาพในการทำนายโดยรวม พบว่า ในกรณีที่ตัวแบบมีความบางเบามาก ๆ วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์ และวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธี LASSO จะมีประสิทธิภาพในการทำนายโดยรวมดีกว่าวิธีอื่น ๆ โดยเฉพาะอย่างยิ่งในกรณีที่ข้อมูลมีขนาดตัวอย่างมาก และสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรมีค่าความสัมพันธ์ปานกลางไปจนถึงน้อย วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์จะทำนายได้ดีกว่า แต่เมื่อค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรมีค่าความสัมพันธ์มาก วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์จะทำนายได้ดีกว่า แต่เมื่อจำนวนตัวแปรที่มีค่าพารามิเตอร์ไม่เป็น 0 เพิ่มขึ้นเป็นตัวแบบที่มีความบางเบาน้อยลง จะเห็นได้ว่าวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีสไตน์-ริดจ์ และวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์ที่มีประสิทธิภาพในการทำนายที่ดีกว่าวิธีอื่น ๆ อย่างใกล้เคียงกันระหว่างสองวิธี

ในส่วนของกรณีที่ตัวแปรอธิบายแบบต่อเนื่องมีความสัมพันธ์กันเองทั้งหมด เมื่อพิจารณาถึงค่าความไว (Sensitivity) ที่มีความสำคัญคือ เป็นค่าที่บ่งบอกถึงความถูกต้องในการจำแนกกลุ่มที่จะเกิดเหตุการณ์ที่สนใจได้นั้น พบว่า วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีสไตน์-ริดจ์มีค่าเฉลี่ยของค่าความไวมากกว่าวิธีอื่น จึงสรุปได้ว่า วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีสไตน์-ริดจ์ มีความถูกต้องในการจำแนกกลุ่มที่จะเกิดเหตุการณ์ที่สนใจได้มากกว่าวิธีอื่น

เมื่อพิจารณาถึงค่าความจำเพาะ (Specificity) ที่มีความสำคัญคือ เป็นค่าที่บ่งบอกถึงความถูกต้องในการจำแนกกลุ่มที่จะเกิดเหตุการณ์ที่ไม่สนใจได้นั้น พบว่า ค่าเฉลี่ยของค่าความจำเพาะของ

วิธี LASSO มีค่าที่สูงกว่าวิธีการประมาณค่าอื่น ๆ แทบทุกกรณี มีเพียงบางกรณีเท่านั้นที่มีค่าเฉลี่ยน้อยกว่า แต่ไม่ได้มีค่าแตกต่างกันมากนัก จึงสรุปได้ว่า วิธี LASSO มีความถูกต้องในการจำแนกกลุ่มที่จะเกิดเหตุการณ์ที่ไม่สนใจได้มากกว่าวิธีอื่น

และเมื่อพิจารณาถึงค่าพื้นที่ใต้เส้นโค้งของกราฟ ROC curve (Mean_AUC) ที่หมายถึงประสิทธิภาพในการทำนายโดยรวม พบว่า ประสิทธิภาพในการทำนายมีผลไปในทิศทางเดียวกันกับกรณีที่ตัวแปรอธิบายถูกแบ่งออกเป็นสองกลุ่มที่เป็นอิสระกัน คือ ในกรณีที่ตัวแบบมีความบางเบามาก ๆ วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์ และวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธี LASSO จะมีประสิทธิภาพในการทำนายโดยรวมดีกว่าวิธีอื่น ๆ โดยเฉพาะอย่างยิ่งในกรณีที่ข้อมูลมีขนาดตัวอย่างมาก และสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรมีค่าความสัมพันธ์ปานกลางไปจนถึงน้อย วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธี LASSO จะทำนายได้ดีกว่า แต่เมื่อค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรมีค่าความสัมพันธ์มาก วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์จะทำนายได้ดีกว่า แต่เมื่อจำนวนตัวแปรที่มีค่าพารามิเตอร์ไม่เป็น 0 เพิ่มขึ้นเป็นตัวแบบที่มีความบางเบาน้อยลง จะเห็นได้ว่าวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีสไตน์-ริดจ์ และวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์ที่มีประสิทธิภาพในการทำนายที่ดีกว่าวิธีอื่น ๆ อย่างใกล้เคียงกันระหว่างสองวิธี

2. ประสิทธิภาพของการประมาณค่า โดยใช้ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าสัมประสิทธิ์การถดถอย (Mean_MSE) เป็นเกณฑ์ในการเปรียบเทียบ สามารถสรุป ได้ดังนี้

สำหรับทั้งในกรณีที่ตัวแปรอธิบายแบ่งออกเป็นสองกลุ่มที่เป็นอิสระกันแต่ตัวแปรอธิบายภายในแต่ละกลุ่มมีความสัมพันธ์กันเอง และกรณีที่ตัวแปรอธิบายมีความสัมพันธ์กันเองทั้งหมด พบว่า ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของวิธี Adaptive LASSO ที่มีค่าถ่วงน้ำหนักที่แตกต่างกันทั้งสามวิธี มีค่าเฉลี่ยของค่า MSE ที่น้อยกว่าวิธี LASSO ทุกกรณี หมายความว่า การเลือกใช้วิธี Adaptive LASSO ในการประมาณค่าพารามิเตอร์สำหรับตัวแบบถดถอยโลจิสติกพหุคูณนั้น สามารถให้ความถูกต้องในการประมาณค่าพารามิเตอร์สัมประสิทธิ์ถดถอยได้ดีกว่าวิธี LASSO แบบปกติ

3. ความผิดพลาดของการคัดเลือกตัวแปรอธิบายเข้าสู่ตัวแบบ โดยใช้ค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรแบบที่ 1 เมื่อเทียบกับจำนวนตัวแปรอธิบายในตัวแบบ (%IC1), และค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรแบบที่ 2 เมื่อเทียบกับจำนวนตัวแปรอธิบายในตัวแบบ (%IC2) เป็นเกณฑ์ในการเปรียบเทียบ สามารถสรุปแยกตามรูปแบบความสัมพันธ์ของตัวแปรอธิบาย ได้ดังนี้

ในกรณีที่ตัวแปรอธิบายแบ่งออกเป็นสองกลุ่มที่เป็นอิสระกันแต่ตัวแปรอธิบายภายในแต่ละกลุ่มมีความสัมพันธ์กันเอง สำหรับค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรแบบที่ 1 (%IC1) ซึ่งหมายถึงจำนวนตัวแปรที่ควรอยู่ในตัวแบบ แต่ไม่ถูกคัดเลือก พบว่า เมื่อตัวแบบมีความบาง

เบามาก ๆ วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์ และวิธี LASSO มีความถูกต้องในการคัดเลือกตัวแปรแบบที่ 1 มากกว่าวิธีอื่น แต่เมื่อตัวแบบมีความบางเบาน้อยลง วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีสไตน์-ริดจ์จะมีความถูกต้องในการคัดเลือกตัวแปรแบบที่ 1 ได้ดีกว่าวิธีอื่น

สำหรับกรณีที่ตัวแปรอธิบายมีความสัมพันธ์กันเองทั้งหมด เมื่อพิจารณาค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรแบบที่ 1 (%IC1) ถ้าตัวแบบมีความบางเบามาก ๆ พบว่า วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์ และวิธี LASSO มีความถูกต้องในการคัดเลือกตัวแปรแบบที่ 1 มากกว่าวิธีอื่น แต่ถ้าตัวแบบมีความบางเบาน้อยลง วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีสไตน์-ริดจ์ และวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์ จะมีความถูกต้องในการคัดเลือกตัวแปรแบบที่ 1 ได้ดีกว่าวิธีอื่น

ในส่วนของค่าร้อยละของความผิดพลาดในการคัดเลือกตัวแปรแบบที่ 2 (%IC2) ซึ่งหมายถึงจำนวนตัวแปรที่ควรไม่อยู่ในตัวแบบ แต่กลับถูกคัดเลือกเข้าสู่ตัวแบบ สำหรับทั้งกรณีที่ตัวแปรอธิบายถูกแบ่งออกเป็นสองกลุ่มที่เป็นอิสระกัน และกรณีที่ตัวแปรอธิบายมีความสัมพันธ์กันเองทั้งหมด พบว่า วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธี LASSO มีความถูกต้องในการคัดเลือกตัวแปรแบบที่ 2 มากกว่าวิธีอื่น ๆ ทุกกรณี

จากข้างต้น สามารถสรุปประสิทธิภาพของตัวประมาณสัมประสิทธิ์การถดถอยวิธีริดจ์, LASSO, Adaptive LASSO และ Relaxed LASSO ภายใต้ขอบเขตงานวิจัย ดังนี้

ตารางที่ 13 เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าแต่ละวิธีแยกตามกรณี

กรณี	ประสิทธิภาพของการทำนาย	ประสิทธิภาพของการประมาณค่า	ประสิทธิภาพของการคัดเลือกตัวแปรอธิบายของสู่ตัวแบบ
ตัวแปรอธิบายแบบต่อเนื่องแบ่งออกเป็นสองกลุ่มที่เป็นอิสระกัน แต่ตัวแปรอธิบายภายในแต่ละกลุ่มมีความสัมพันธ์กันเอง	-ค่าความไว ถ้าตัวแบบบางเบามาก วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์ดีกว่า แต่ถ้าตัวแบบบางเบาน้อย วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีสไตน์-ริดจ์ดีกว่า -ค่าความจำเพาะ วิธี LASSO ดีกว่าวิธีอื่น ๆ ทุกกรณี -ประสิทธิภาพโดยรวม ถ้าตัวแบบบางเบามาก วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์ และวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธี LASSO ดีกว่า ถ้าตัวแบบบางเบาน้อย วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์ และวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีสไตน์-ริดจ์ดีกว่า	วิธี Adaptive LASSO ทั้งสามวิธีที่มีค่าถ่วงน้ำหนักแตกต่างกันมีประสิทธิภาพในการประมาณค่าพารามิเตอร์สัมประสิทธิ์ถดถอยสำหรับตัวแบบถดถอยโลจิสติกพหุคูณดีกว่าวิธี LASSO แบบปกติ	-ค่า %IC1 ถ้าตัวแบบบางเบามาก วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์ และวิธี LASSO คัดเลือกตัวแปรได้ดีกว่า ถ้าตัวแบบบางเบาน้อย วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีสไตน์-ริดจ์คัดเลือกได้ดีกว่า -ค่า %IC2 วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธี LASSO คัดเลือกได้ดีกว่า
ตัวแปรอธิบายแบบต่อเนื่องมีความสัมพันธ์กันเองทั้งหมด	-ค่าความไว วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีสไตน์-ริดจ์ดีกว่า -ค่าความจำเพาะ วิธี LASSO ดีกว่าวิธีอื่นทุกกรณี -ประสิทธิภาพโดยรวม ถ้าตัวแบบบางเบามาก วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์ดีกว่าเมื่อตัวแปรสัมพันธ์กันปานกลางถึงน้อย และวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธี LASSO ดีกว่าเมื่อตัวแปรสัมพันธ์กันมาก ถ้าตัวแบบบางเบาน้อย วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์ และวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีสไตน์-ริดจ์ดีกว่า		-ค่า %IC1 ถ้าตัวแบบบางเบามาก วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์, วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีสไตน์-ริดจ์ และวิธี LASSO คัดเลือกตัวแปรได้ดีกว่า ถ้าตัวแบบบางเบาน้อย วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีสไตน์-ริดจ์และวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์คัดเลือกได้ดีกว่า -ค่า %IC2 วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธี LASSO คัดเลือกได้ดีกว่า

5.2 อภิปรายผลการวิจัย

ในกรณีข้อมูลมีมิติสูงแบบบางเบามาก ๆ พบว่า วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีริดจ์ และวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธี LASSO จะมีประสิทธิภาพโดยรวมทั้งสามด้านที่ดีกว่าวิธีอื่น โดยที่วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีริดจ์ จะมีความผิดพลาดในการคัดเลือกตัวแปรที่ควรอยู่ในตัวแบบ แต่ไม่ถูกคัดเลือกน้อยกว่า ในขณะที่วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธี LASSO จะมีความผิดพลาดในการคัดเลือกตัวแปรที่ไม่ควรอยู่ในตัวแบบ แต่ถูกคัดเลือกเข้าสู่ตัวแบบน้อยกว่า เพราะว่าวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธี LASSO นั้น เป็นกระบวนการหาตัวประมาณสัมประสิทธิ์การถดถอยที่มีการคัดเลือกตัวแปร 2 ขั้นตอน ซึ่งมีการคัดเลือกตัวแปรอธิบายจากกระบวนการแรกของ LASSO แล้วนำเซตของตัวแปรอธิบายนั้นมาทำการหาค่าถ่วงน้ำหนัก แล้วนำค่าถ่วงน้ำหนักไปใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยด้วย Adaptive LASSO อีกรอบหนึ่ง จึงทำให้มีการคัดเลือกตัวแปรอธิบายมากกว่าวิธีการอื่นที่มีการคัดเลือกตัวแปรอธิบายเพียงครั้งเดียว

ในกรณีข้อมูลมีมิติสูงแบบบางเบาน้อย พบว่า วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีริดจ์ และวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีสไตน์-ริดจ์ จะมีประสิทธิภาพโดยรวมทั้งสามด้านที่ดีกว่าวิธีอื่น ยกเว้นการคัดเลือกตัวแปรอธิบายแบบ IC2 ที่วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธี LASSO ยังคงทำได้ดี สาเหตุที่วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีริดจ์ และวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีสไตน์-ริดจ์นั้นมีประสิทธิภาพและมีความคงเส้นคงวา เนื่องจาก ตัวประมาณสัมประสิทธิ์การถดถอยวิธีริดจ์เป็นตัวประมาณที่ให้ค่าประมาณสัมประสิทธิ์การถดถอยทุกตัวมีค่าเข้าใกล้ศูนย์แต่ไม่เท่ากับศูนย์ จึงทำให้ค่าถ่วงน้ำหนักที่คำนวณจากตัวประมาณสัมประสิทธิ์การถดถอยของวิธีริดจ์รวมถึงตัวประมาณสไตน์-ริดจ์ที่คำนวณจากตัวประมาณของวิธีริดจ์มีค่าถ่วงน้ำหนักที่มีความคงเส้นคงวามากกว่าค่าถ่วงน้ำหนักที่คำนวณจากวิธี LASSO ที่มีการคัดเลือกตัวแปรที่ไม่คงเส้นคงวา และทำให้ตัวแปรที่นำมาคำนวณค่าถ่วงน้ำหนักไม่ได้นำมาใช้ครบทุกตัวแปร ทำให้ทั้งสองวิธีคือ วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีริดจ์ และวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยตัวประมาณวิธีสไตน์-ริดจ์ที่เป็นตัวประมาณแบบพีนอลไลซ์ที่มีฟังก์ชันพีนอลตี้เป็น L_2-norm มีประสิทธิภาพที่ดีมากกว่า ดังนั้นแล้ว การเลือกค่าถ่วงน้ำหนักที่เหมาะสม สำหรับวิธี Adaptive LASSO นั้น ก็จะทำให้ได้ผลลัพธ์และประสิทธิภาพที่แตกต่างกัน ขึ้นอยู่กับแต่ละสถานการณ์ในการวิเคราะห์ตัวแบบถดถอยโลจิสติกพหุคูณเมื่อข้อมูลมีมิติสูงแบบบางเบา

5.3 ข้อเสนอแนะ

1. ถ้าต้องการให้การทำนายมีค่าความไวสูง ควรเลือกใช้วิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีสไตน์-ริดจ์ และวิธี Adaptive LASSO ที่ถ่วงน้ำหนักด้วยวิธีริดจ์ในการประมาณค่าสัมประสิทธิ์การถดถอย

2. ถ้าต้องการให้การทำนายมีค่าความจำเพาะสูง ควรเลือกใช้วิธี LASSO ในการประมาณค่าสัมประสิทธิ์การถดถอย

3. ถ้าต้องการความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอย ควรเลือกใช้การประมาณค่าสัมประสิทธิ์ถดถอยด้วยวิธี Adaptive LASSO ดีกว่าวิธี LASSO ปกติ

4. ในงานวิจัยนี้ มีการกำหนดโครงสร้างของเมทริกซ์ความแปรปรวนร่วม (Covariance Structure) ให้มีความสัมพันธ์ระหว่างตัวแปรอธิบายที่อยู่ใกล้ชิดกันมีค่าสัมประสิทธิ์สหสัมพันธ์สูง หรือ $\rho = \rho^{|j-k|};\ j,k = 1,2,\ldots,p$ ซึ่งเป็นรูปแบบของเมทริกซ์ความแปรปรวนร่วมที่ค่อนข้างมีความเฉพาะเจาะจง ดังนั้น ก่อนที่จะใช้วิธีการประมาณค่าในงานวิจัยนี้ในการวิเคราะห์ ควรตรวจสอบโครงสร้างของเมทริกซ์ความแปรปรวนร่วมว่ามีความสอดคล้องกับโครงสร้างในงานวิจัยนี้หรือไม่

5.4 ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. ในงานวิจัยนี้ศึกษากรณีที่กำหนดตัวแปรอธิบายแบบไม่ต่อเนื่องอยู่ในตัวแบบจริงเพียงจำนวน 4 ตัวแปร ซึ่งไม่ค่อยส่งผลให้เห็นถึงความแตกต่างในแต่ละวิธีการประมาณค่าพารามิเตอร์ ซึ่งทางผู้วิจัยได้ทดลองเพิ่มจำนวนตัวแปรแบบไม่ต่อเนื่องให้มีจำนวนเท่า ๆ กันกับจำนวนตัวแปรแบบต่อเนื่อง แต่ผลลัพธ์ที่ได้ไม่เป็นที่น่าพึงพอใจ จึงไม่ได้ดำเนินการต่อ อาจจะต้องใช้เวลาในการศึกษาเกี่ยวกับการกำหนดค่าสัมประสิทธิ์การถดถอยเพิ่มเติม ดังนั้น ในงานวิจัยครั้งต่อไปควรกำหนดจำนวนตัวแปรอธิบายและจำนวนตัวแปรที่มีค่าสัมประสิทธิ์ถดถอยที่ไม่เป็นศูนย์ของทั้งตัวแปรแบบต่อเนื่องและตัวแปรแบบไม่ต่อเนื่องในตัวแบบให้มีความหลากหลายกว่านี้

2. ในงานวิจัยนี้ได้ศึกษาตัวประมาณที่นำมาใช้ในการคำนวณค่าถ่วงน้ำหนักเพียงค่าสามวิธี ยังมีตัวประมาณอีกหลายตัวที่สามารถนำมาใช้คำนวณค่าถ่วงน้ำหนักสำหรับวิธี Adaptive LASSO ในกรณีนี้ได้

รายการอ้างอิง

A.Farghali, R., & M. Abo-El-Hadid, S. (2017). Evaluating the Performance of Liu Logistic Regression Estimator. *Research Journal of Mathematics and Statistics*, *9*(2), 11-19. https://doi.org/10.19026/rjms.9.4715

Agresti, A. (2007). *An Introduction to Categorical Data Analysis* (Vol. 2). JohnWiley & Sons, Inc.

Algamal, Z. Y., & Lee, M. H. (2015). Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification. *Expert Systems with Applications*, *42*(23), 9326-9332. https://doi.org/10.1016/j.eswa.2015.08.016

Bühlmann, P., et al. (2014). High-Dimensional Statistics with a View Toward Applications in Biology. *Annual Review of Statistics and Its Application, 1*(1): 255-278.

Garrett, S. J. (2015). *Introduction to Actuarial and Financial Mathematical Methods: Chapter 13 - Introductory Numerical Methods* (S. J. Garrett, Ed.). Academic Press.

Guo, P., Zeng, F., Hu, X., Zhang, D., Zhu, S., Deng, Y., & Hao, Y. (2015). Improved Variable Selection Algorithm Using a LASSO-Type Penalty, with an Application to Assessing Hepatitis B Infection Relevant Factors in Community Residents. *PLoS One*, *10*(7), e0134151. https://doi.org/10.1371/journal.pone.0134151

Hashemian, A. H., Asl, M. G., & Shahsavari, S. (2017). Adaptive Lasso Logistic Regression Applied on Gene Expression of Prostate Cancer. *World Family Medicine Journal/Middle East Journal of Family Medicine*, *15*(7), 136-141. https://doi.org/10.5742/mewfm.2017.93028

Hoerl, A. E. a. R. W. K. (1970). Ridge Regression: Biased Estimation For Nonorthogonal Problems. *American Statistical Association*, *12*(1(February)), 55-67.

Jerome H. Friedman, T. H., Rob Tibshirani. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33(1)*, 1-22. https://doi.org/https://doi.org/10.18637/jss.v033.i01

Jian Huang, S. M., Cun-Hui Zhang. (2008). The Iterated Lasso for High-Dimensional

Logistic Regression.

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, *28(5)*, 1-26, Article 5. https://doi.org/http://dx.doi.org/10.18637/jss.v028.i05

Noah Simon, J. H. F., Trevor Hastie, Rob Tibshirani. (2011). Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *Journal of Statistical Software*, *39(5)*, 1-13. https://doi.org/https://doi.org/10.18637/jss.v039.i05

Raschka, S. (2016). *Regularization in Logistic Regression: Better Fit and Better Generalization?* https://www.kdnuggets.com/2016/06/regularization-logistic-regression.html

Ripley, W. N. V. a. B. D. (2002). *Modern Applied Statistics with S* (Fourth edition ed.). Springer. https://www.stats.ox.ac.uk/pub/MASS4/

Schaefer, R. L. (1986). Alternative estimators in logistic regression when the data are collinear. *Journal of Statistical Computation and Simulation*, *25*(1-2), 75-91. https://doi.org/10.1080/00949658608810925

Schaefer, R. L., Roi, L. D., & Wolfe, R. A. (1984). A ridge logistic estimator. *Communications in Statistics - Theory and Methods*, *13*(1), 99-113. https://doi.org/10.1080/03610928408828664

Simon, S. D. (2016). *Understanding how LASSO Regression works*. http://www.pmean.com/16/images/lasso.pdf

Stein, C. (1956). Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability.*, *1*, 197-206.

Sur, P., & Candes, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proc Natl Acad Sci U S A*, *116*(29), 14516-14525. https://doi.org/10.1073/pnas.1810420116

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society*, *58*(1 (January)), 267-288.

Van de Geer, S. A. (2008). High-dimensional generalized linear models and the lasso. *The Annals of Statistics*, *36*(2). https://doi.org/10.1214/009053607000000929

W. James, C. S. (1961). Estimation with quadratic loss. *Proceedings of the Fourth*

Berkeley Symposium on Mathematical Statistics and Probability, *1*, 361-379.

Xavier Robin, N. T., Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez and Markus Müller. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*, 77-84. https://doi.org/https://doi.org/10.1186/1471-2105-12-77

Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, *101*(476), 1418-1429. https://doi.org/10.1198/016214506000000735

จารุวรรณ, ม. (2559). การเปรียบเทียบประสิทธิภาพของวิธีการอนุมานในตัวแบบถดถอยลอจิสติก [วิทยานิพนธ์ปริญญามหาบัณฑิต, มหาวิทยาลัยศิลปากร]. มหาวิทยาลัยศิลปากร.

พงษ์เดช, ส. (2560). การวิเคราะห์ความเสี่ยงและสถิติในการวินิจฉัยโรค. เอกสารคำสอนรายวิชา *516 404* ชีวสถิติขั้นพื้นฐาน (*Fundamental of Biostatistics)* (pp. 243-299).

พัชราภรณ์, พ. (2560). การเปรียบเทียบประสิทธิภาพของการประมาณค่าพารามิเตอร์ด้วยวิธีแลซโซในการถดถอยเชิงเส้นที่มีมิติสูง [วิทยานิพนธ์ปริญญามหาบัณฑิต, มหาวิทยาลัยศิลปากร]. มหาวิทยาลัยศิลปากร.

ภาคผนวก

ก. แพ็กเกจและชุดคำสั่งของโปรแกรมที่ใช้ในงานวิจัย

ในงานวิจัยนี้ ใช้โปรแกรม R เวอร์ชั่น 4.1.1 ในการดำเนินงานวิจัย ซึ่งมีแพ็กเกจและชุดคำสั่งที่ได้ใช้งานภายในโปรแกรมดังนี้

1. แพ็กเกจ MASS (Ripley, 2002)

แพ็กเกจ MASS เป็นแพ็กเกจที่รวมคำสั่งที่ใช้ในด้านสถิติประยุกต์ร่วมสมัยที่สามารถใช้ได้ทั้งในโปรแกรม R และ โปรแกรม S-PLUS (Modern Applied Statistics with S) ซึ่งในงานวิจัยนี้จะใช้คำสั่ง mvrnorm() ของแพ็กเกจ MASS ในการสร้างเมทริกซ์ของตัวแปรสุ่มแบบต่อเนื่องที่มีการแจกแจงแบบปกติหลายตัวแปร

2. แพ็กเกจ glmnet (Jerome H. Friedman, 2010) และ (Noah Simon, 2011)

แพ็กเกจ glmnet เป็นแพ็กเกจที่รวมคำสั่งที่ใช้ในงานเกี่ยวกับการประมาณค่าสัมประสิทธิ์การถดถอยแบบพีนอลไลซ์ด้วยวิธีริดจ์, วิธีแลซโซ, รวมถึงวิธี Elastic-Net ของตัวแบบถดถอยต่าง ๆ ไม่ว่าจะเป็นตัวแบบถดถอยเชิงเส้น (Linear regression; Gaussian), Multi-task Gaussian process prediction, ตัวแบบถดถอยโลจิสติกทั้งแบบทวิภาคหรือแบบพหุภาค (Binary or Multinomial Logistic regression), ตัวแบบถดถอยแบบปัวซอง (Poisson regression) และตัวแบบอื่น ๆ

ในงานวิจัยนี้จะใช้คำสั่ง glmnet() ในการประมาณค่าสัมประสิทธิ์ถดถอยของตัวแบบถดถอยโลจิสติกด้วยวิธีริดจ์, วิธีแลซโซ, และวิธีแลซโซแบบปรับได้ โดยการกำหนด family = binomial สำหรับตัวแบบถดถอยโลจิสติก กำหนด alpha = 0 สำหรับการประมาณค่าด้วยวิธีริดจ์และ alpha = 1 สำหรับการประมาณค่าด้วยวิธีแลซโซ รวมทั้งใช้คำสั่ง cv.glmnet() ในการทำ Cross-validation ที่ใช้ในการหา Tuning parameter $\boldsymbol{\lambda}$ สำหรับวิธีการประมาณวิธีต่าง ๆ

3. แพ็กเกจ caret (Kuhn, 2008)

แพ็กเกจ caret เป็นแพ็กเกจที่รวมคำสั่งที่ใช้เกี่ยวกับการสร้างตัวแบบการจำแนกและตัวแบบการทำนายในตัวแบบถดถอยด้วยวิธีต่าง ๆ รวมถึงการตรวจสอบความถูกต้องการของการทำนายและการจำแนกของตัวแบบถดถอยต่าง ๆ ด้วย โดยที่ caret หมายถึง Classification And REgression Training

ในงานวิจัยนี้ จะใช้คำสั่ง sensitivity() และ specificity() ในการหาค่าความไวและค่าความจำเพาะของการทำนายในตัวแบบถดถอยโลจิสติกของแต่ละวิธีการประมาณค่า นอกจากนี้ ยัง

สามารถใช้คำสั่ง negPredValue() และ posPredValue() ในการพิจารณาถึงค่าทำนายผลลบ และค่าทำนายผลบวก คือสัดส่วนของจำนวนผลการตรวจที่เป็นผลบวกแท้หรือผลลบแท้ (การวินิจฉัยถูกต้อง) ต่อจำนวนผลการตรวจที่เป็นผลบวกหรือลบทั้งหมด โดยที่ค่าทำนายผลลบจะสอดคล้องกับค่าความไว และค่าทำนายผลบวกจะสอดคล้องกับค่าความจำเพาะ

4. แพ็กเกจ pROC (Xavier Robin, 2011)

แพ็กเกจ pROC เป็นแพ็กเกจที่รวมคำสั่งที่ใช้ในการสร้างเส้นกราฟ ROC curves ที่ใช้ในการพิจารณาถึงประสิทธิภาพของการทำนายของแต่ละวิธีการประมาณค่าได้ รวมถึงยังสามารถวิเคราะห์และเปรียบเทียบประสิทธิภาพของแต่ละวิธีการประมาณค่าโดยพิจารณาจากเส้นกราฟ ROC curves

ในงานวิจัยนี้ จะใช้คำสั่ง auc() ที่ใช้ในการหาค่าพื้นที่ใต้โค้งของเส้นกราฟ ROC curves ที่จะบ่งบอกถึงประสิทธิภาพของการทำนายได้

ข. โปรแกรมที่ใช้ในงานวิจัย

```
install.packages("MASS")
install.packages("glmnet")
install.packages("caret")
install.packages("pROC")

library(MASS)
library(glmnet)
library(caret)
library(pROC)

nrep=500
nsigma=2
r=0.3 #r=0.3,0.6,0.9
n=100 #n=100,200
p=200 #p=2n,3n,4n
p1=5 #p1=5,10,15
p13=p1/1 #if p1=5, use /1 /// if p1=10, use /2 /// if p=15, use /3
p2=p-p1
psigma1=15
psigma2=p-psigma1
gamma=1

tuning.ridge <- tuning.lasso <- rep(0,nrep)
ic1.lasso <- ic1.wridge <- ic1.wlasso <- ic1.wstein <- rep(0,nrep)
ic2.lasso <- ic2.wridge <- ic2.wlasso <- ic2.wstein <- rep(0,nrep)
bmse.lasso <- bmse.wridge <- bmse.wlasso <- bmse.wstein <- rep(0,nrep)
tuning.wridge <- tuning.wlasso <- tuning.wstein <- rep(0,nrep)
```

```
auc.lasso <- auc.wridge <- auc.wlasso <- auc.wstein <- rep(0,nrep)
accuracy.L <- accuracy.adwR <- accuracy.adwL <- accuracy.adwS <- rep(0,nrep)
sensitivity.L <-sensitivity.adwR <- sensitivity.adwL <- sensitivity.adwS <- rep(0,nrep)
specificity.L <-specificity.adwR <- specificity.adwL <- specificity.adwS <-rep(0,nrep)
negPredVal.L <- negPredVal.adwR <- negPredVal.adwL <- negPredVal.adwS <-
rep(0,nrep)
posPredVal.L <- posPredVal.adwR <- posPredVal.adwL <- posPredVal.adwS <-
rep(0,nrep)

#เมทริกซ์ความแปรปรวนร่วมกรณีที่ตัวแปรอธิบายแบบต่อเนื่องแบ่งออกเป็นสองกลุ่มที่เป็นอิสระกัน
sigma1=matrix(0,psigma1,psigma1)
for(a in 1:psigma1)
  for(b in 1:psigma1)
    sigma1[a,b]=r^(abs(a-b))
sigma0=matrix(0,psigma1,psigma2)
for(a in 1:psigma1)
  for(b in 1:psigma2)
    sigma0[a,b]=0
sigma00=matrix(0,psigma2,psigma1)
for(a in 1:psigma2)
  for(b in 1:psigma1)
    sigma00[a,b]=0
sigma2=matrix(0,psigma2,psigma2)
for(a in 1:psigma2)
  for(b in 1:psigma2)
    sigma2[a,b]=r^(abs(a-b))
sigmaa<-cbind(sigma1,sigma0)
sigmaaa<-cbind(sigma00,sigma2)
```

```
sigma<-rbind(sigmaa,sigmaaa)

beta <- c(-0.5,rep(2.5,p13),rep(0,p2),rep(0.1,4))

#เมทริกซ์ความแปรปรวนร่วมกรณีที่ตัวแปรอธิบายแบบต่อเนื่องมีความสัมพันธ์กันเองทั้งหมด
############################
sigma=matrix(0,p,p)
for(a in 1:p)
  for(b in 1:p)
    sigma[a,b]=r^(abs(a-b))
############################

# Calculate Time use
RunningTime <- function(t0,t1){
  dsec <- as.numeric(difftime(t1, t0, unit = "secs"))
  hours <- floor(dsec / 3600)
  minutes <- floor((dsec - 3600 * hours) / 60)
  seconds <- dsec - 3600*hours - 60*minutes
  secs <- format(seconds, digit = 4)
  Time <- c(hours,minutes,secs)
  return(Time)
}
t0 <- Sys.time()

set.seed(123456)
options(scipen = 999)

for(k in 1:nrep) {
```

```
cat1.train <- rbinom(n,1,0.8)
cat2.train <- rbinom(n,1,0.6)
cat3.train <- rbinom(n,1,0.4)
cat4.train <- rbinom(n,1,0.2)
cat1.test <- rbinom(n,1,0.8)
cat2.test <- rbinom(n,1,0.6)
cat3.test <- rbinom(n,1,0.4)
cat4.test <- rbinom(n,1,0.2)

xcat.train <- rbind(cat1.train,cat2.train,cat3.train,cat4.train)
xcat.test <- rbind(cat1.test,cat2.test,cat3.test,cat4.test)
xcat.train <- t(xcat.train)
xcat.test <- t(xcat.test)

x.train <- mvrnorm(n,mu = rep(0,p),sigma)
x.test <- mvrnorm(n,mu = rep(0,p),sigma)
x.train.center <- scale(x.train, center = TRUE, scale = FALSE)
x.test.center <- scale(x.test, center = TRUE, scale = FALSE)

bcat1 <- c(rep(0.1,4))
bcat <- as.matrix(bcat1)
bcat01 <- rep(-0.5,n)
bcat0 <- as.matrix(bcat01)

b1 <- c(rep(2.5,p13),rep(0,p2))
b <- as.matrix(b1)
b01 <- rep(-0.5,n)
```

```
  b0 <- as.matrix(b01)

  yprob.train <-
(exp(b0+(x.train.center%*%b))+(bcat0+(xcat.train%*%bcat)))/(1+(exp(b0+(x.train.center
%*%b))+(bcat0+(xcat.train%*%bcat))))
  yprob.test <-
(exp(b0+(x.test.center%*%b))+(bcat0+(xcat.test%*%bcat)))/(1+(exp(b0+(x.test.center%
*%b))+(bcat0+(xcat.test%*%bcat))))
  y.train <- ifelse(yprob.train>=0.5,1,0)
  y.test <- ifelse(yprob.test>=0.5,1,0)
  allx.train <- cbind(as.matrix(x.train.center),as.matrix(xcat.train))
  allx.test <- cbind(as.matrix(x.test.center),as.matrix(xcat.test))
  data.train <- data.frame(y=y.train,x=I(allx.train))
  data.test  <- data.frame(y=y.test,x=I(allx.test))

  #Finding tuning parameter(lambda)
  #ridge
  cridge.cv <- cv.glmnet(data.train$x, y=as.factor(data.train$y),
                       family="binomial", alpha = 0, nfolds = 5,
                       type.measure = "class")
  coef.ridge.cv <- coef(cridge.cv,cridge.cv$lambda.min)
  lamb.ridge <- cridge.cv$lambda.min
  coef.ridge.cv <- as.matrix(coef.ridge.cv[-1,])

  #lasso
  classo.cv <- cv.glmnet(data.train$x, y=as.factor(data.train$y),
                       family="binomial", alpha = 1, nfolds=5,
                       type.measure = "class")
```

```
coef.lasso.cv <- coef(classo.cv,classo.cv$lambda.min)
lamb.lasso <- classo.cv$lambda.min
coef.lasso.cv <- as.matrix(coef.lasso.cv[-1,])
var_lasso <- names(coef.lasso.cv[coef.lasso.cv[,1]!=0,])

#Finding coef(betahat)
ridge <- glmnet(data.train$x, y=as.factor(data.train$y),
                family="binomial", alpha = 0, lambda = lamb.ridge)

#find betahat ridge
coef.ridge <- as.matrix(coef(ridge,s=lamb.ridge))

#find pihat (probhat)
yprobhat.ridge <- predict(ridge, data.test$x, type="response")
tuning.ridge[k] <- lamb.ridge

#find W hat
What <- matrix(nrow=n, ncol=n)
for(i in 1:n) {
  What[i,i] <- yprobhat.ridge[i]*(1-yprobhat.ridge[i])
  What[is.na(What)]=0
}

#calculate X'WhatX
XtWX <- t(data.train$x)%*%What%*%data.train$x

#find eigenvalues of X'WhatX
eigenval <- eigen(XtWX)$values
```

```
#calculate trace from sum of 1/eigenvalues
eigenval1 <- eigenval[eigenval > 0.01]
trace <- sum(1/(eigenval1))

c <- as.numeric((t(coef.ridge)%*%coef.ridge)/((t(coef.ridge)%*%coef.ridge)+trace))

coef.stein <- c*coef.ridge

coef.stein.cv <- coef.stein[-1]

lasso <- glmnet(data.train$x, y=as.factor(data.train$y),
                family="binomial", alpha = 1, lambda = lamb.lasso)
coef.lasso <- as.matrix(coef(lasso,s=lamb.lasso))
yprobhat.lasso <- predict(lasso, data.test$x, type="response")
tuning.lasso[k] <- lamb.lasso
bmse.lasso[k] <- mean((coef.lasso-beta)^2)
ic1.lasso[k] <- sum(beta[-1] != 0 & coef.lasso[-1] == 0)
ic2.lasso[k] <- sum(beta[-1] == 0 & coef.lasso[-1] != 0)
yhat.lasso <- ifelse(yprobhat.lasso>=0.5,1,0)
accuracy.L[k] <- mean(data.test$y==yhat.lasso)
conf.matrix.L <- table(data.test$y,yhat.lasso)
sensitivity.L[k] <- sensitivity(conf.matrix.L)
specificity.L[k] <- specificity(conf.matrix.L)
negPredVal.L[k] <- negPredValue(conf.matrix.L)
posPredVal.L[k] <- posPredValue(conf.matrix.L)
auc.lasso[k] <- auc(data.test$y,as.numeric(yhat.lasso))
```

```
#Finding lambda for Adaptive LASSO
#ridge
cweights.ridge <- 1/(abs(coef.ridge.cv)^gamma)
xdata.train.w1 <- scale(data.train$x,center = FALSE, scale = cweights.ridge)
xdata.test.w1 <- scale(data.test$x,center = FALSE, scale = cweights.ridge)
wridge.cv <- cv.glmnet(x=xdata.train.w1,y=as.factor(data.train$y),
                  family="binomial",alpha=1,nfolds=5,
                  type.measure="class",standardize=FALSE)
lamb.wridge <- wridge.cv$lambda.min
coef.wridge <- coef(wridge.cv,wridge.cv$lambda.min)


#lasso
cweights.lasso <- 1/(abs(coef.lasso.cv)^gamma)
xdata.train.w2 <- scale(data.train$x,center = FALSE, scale = cweights.lasso)
xdata.test.w2 <- scale(data.test$x,center = FALSE, scale = cweights.lasso)
wlasso.cv <- cv.glmnet(x=xdata.train.w2,y=as.factor(data.train$y),
                  family="binomial",alpha=1,nfolds=5,
                  type.measure="class",standardize=FALSE)
lamb.wlasso <- wlasso.cv$lambda.min
coef.wlasso <- coef(wlasso.cv,wlasso.cv$lambda.min)

#stein-ridge
cweights.stein <- 1/(abs(coef.stein.cv)^gamma)
xdata.train.w3 <- scale(data.train$x,center = FALSE, scale = cweights.stein)
xdata.test.w3 <- scale(data.test$x,center = FALSE, scale = cweights.stein)
wstein.cv <- cv.glmnet(x=xdata.train.w3,y=as.factor(data.train$y),
                  family="binomial",alpha=1,nfolds=5,
```

```
                    type.measure="class",standardize=FALSE)
  lamb.wstein <- wstein.cv$lambda.min
  coef.wstein <- coef(wstein.cv,wstein.cv$lambda.min)

  #Adaptive LASSO
  #ridge
  weights.ridge1 <- 1/(abs(coef.ridge)^gamma)
  weights.ridge <- 1/(abs(coef.ridge[-1])^gamma)
  xdata.train.wridge <- scale(data.train$x,center = FALSE, scale = weights.ridge)
  xdata.test.wridge <- scale(data.test$x,center = FALSE, scale = weights.ridge)
  adwridge <- glmnet(x=xdata.train.wridge, y=as.factor(data.train$y),
                  family="binomial", alpha = 1, lambda = lamb.wridge, standardize =
FALSE)
  coef.adwridge <- coef(adwridge)
  coef.wridge <- (coef.adwridge)/(weights.ridge1)
  yprobhat.adwridge <- predict(adwridge, data.test$x, type="response")
  tuning.wridge[k] <- lamb.wridge
  bmse.wridge[k] <- mean((coef.wridge-beta)^2)
  ic1.wridge[k] <- sum(beta[-1] != 0 & coef.wridge[-1] == 0)
  ic2.wridge[k] <- sum(beta[-1] == 0 & coef.wridge[-1] != 0)
  yhat.wridge <- ifelse(yprobhat.adwridge>=0.5,1,0)
  accuracy.adwR[k] <- mean(data.test$y==yhat.wridge)
  conf.matrix.adwR <- table(data.test$y,yhat.wridge)
  sensitivity.adwR[k] <- sensitivity(conf.matrix.adwR)
  specificity.adwR[k] <- specificity(conf.matrix.adwR)
  negPredVal.adwR[k] <- negPredValue(conf.matrix.adwR)
  posPredVal.adwR[k] <- posPredValue(conf.matrix.adwR)
  auc.wridge[k] <- auc(data.test$y,as.numeric(yhat.wridge))
```

```
#####Finding criteria later#####

#lasso
weights.lasso1 <- 1/(abs(coef.lasso)^gamma)
weights.lasso <- 1/(abs(coef.lasso[-1])^gamma)
xdata.train.wlasso <- scale(data.train$x,center = FALSE, scale = weights.lasso)
xdata.test.wlasso <- scale(data.test$x,center = FALSE, scale = weights.lasso)
adwlasso <- glmnet(x=xdata.train.wlasso, y=as.factor(data.train$y),
          family="binomial", alpha = 1, lambda = lamb.wlasso, standardize =
FALSE)
coef.adwlasso <- coef(adwlasso)
coef.wlasso <- (coef.adwlasso)/(weights.lasso1)
yprobhat.adwlasso <- predict(adwlasso, data.test$x, type="response")
tuning.wlasso[k] <- lamb.wlasso
bmse.wlasso[k] <- mean((coef.wlasso-beta)^2)
ic1.wlasso[k] <- sum(beta[-1] != 0 & coef.wlasso[-1] == 0)
ic2.wlasso[k] <- sum(beta[-1] == 0 & coef.wlasso[-1] != 0)
yhat.wlasso <- ifelse(yprobhat.adwlasso>=0.5,1,0)
accuracy.adwL[k] <- mean(data.test$y==yhat.wlasso)
conf.matrix.adwL <- table(data.test$y,yhat.wlasso)
sensitivity.adwL[k] <- sensitivity(conf.matrix.adwL)
specificity.adwL[k] <- specificity(conf.matrix.adwL)
negPredVal.adwL[k] <- negPredValue(conf.matrix.adwL)
posPredVal.adwL[k] <- posPredValue(conf.matrix.adwL)
auc.wlasso[k] <- auc(data.test$y,as.numeric(yhat.wlasso))

#stein-ridge
weights.stein1 <- 1/(abs(coef.stein)^gamma)
```

```
  weights.stein <- 1/(abs(coef.stein[-1])^gamma)
  xdata.train.wstein <- scale(data.train$x,center = FALSE, scale = weights.stein)
  xdata.test.wstein <- scale(data.test$x,center = FALSE, scale = weights.stein)
  adwstein <- glmnet(x=xdata.train.wstein, y=as.factor(data.train$y),
                     family="binomial", alpha = 1, lambda = lamb.wstein, standardize =
FALSE)
  coef.adwstein <- coef(adwstein)
  coef.wstein <- (coef.adwstein)/(weights.stein1)
  yprobhat.adwstein <- predict(adwstein, data.test$x, type="response")
  tuning.wstein[k] <- lamb.wstein
  bmse.wstein[k] <- mean((coef.wstein-beta)^2)
  ic1.wstein[k] <- sum(beta[-1] != 0 & coef.wstein[-1] == 0)
  ic2.wstein[k] <- sum(beta[-1] == 0 & coef.wstein[-1] != 0)
  yhat.wstein <- ifelse(yprobhat.adwstein>=0.5,1,0)
  accuracy.adwS[k] <- mean(data.test$y==yhat.wstein)
  conf.matrix.adwS <- table(data.test$y,yhat.wstein)
  sensitivity.adwS[k] <- sensitivity(conf.matrix.adwS)
  specificity.adwS[k] <- specificity(conf.matrix.adwS)
  negPredVal.adwS[k] <- negPredValue(conf.matrix.adwS)
  posPredVal.adwS[k] <- posPredValue(conf.matrix.adwS)
  auc.wstein[k] <- auc(data.test$y,as.numeric(yhat.wstein))
}

tuning.out <- cbind(tuning.lasso,tuning.wridge,tuning.wlasso,tuning.wstein)
ic1.out <- cbind(ic1.lasso,ic1.wridge,ic1.wlasso,ic1.wstein)
ic2.out <- cbind(ic2.lasso,ic2.wridge,ic2.wlasso,ic2.wstein)
bmse.out <- cbind(bmse.lasso, bmse.wridge, bmse.wlasso,bmse.wstein)
sens.out <- cbind(sensitivity.L, sensitivity.adwR, sensitivity.adwL, sensitivity.adwS)
```

```
spec.out <- cbind(specificity.L, specificity.adwR, specificity.adwL, specificity.adwS)
negPredVal.out <-
cbind(negPredVal.L,negPredVal.adwR,negPredVal.adwL,negPredVal.adwS)
posPredVal.out <-
cbind(posPredVal.L,posPredVal.adwR,posPredVal.adwL,posPredVal.adwS)
auc.out <- cbind(auc.lasso,auc.wridge,auc.wlasso,auc.wstein)
accuracy.out <- cbind(accuracy.L,accuracy.adwR,accuracy.adwL,accuracy.adwS)


average.tuning <- apply(tuning.out,2,mean)
std.tuning <- apply(tuning.out,2,sd)
average.ic1 <- colMeans(ic1.out)
average.ic2 <- colMeans(ic2.out)
beta.mse <- apply(bmse.out,2,mean)
mean.sens <- apply(sens.out,2,mean)
mean.spec <- apply(spec.out,2,mean)
mean.negP <- apply(negPredVal.out,2,mean)
mean.posP <- apply(posPredVal.out,2,mean)
mean.auc <- apply(auc.out,2,mean)
mean.accuracy <- colMeans(accuracy.out)
summary <-
rbind(average.tuning,std.tuning,average.ic1,average.ic2,beta.mse,mean.sens,mean.spec
,mean.negP,mean.posP,mean.auc,mean.accuracy)
colsummary.names <- c("       LASSO       ",  "Adaptive LASSO (R)",  "Adaptive LASSO
(L)","Adaptive LASSO (S)")
rowsummary.names <-
c("tuning(average)","tuning(sd)","average.ic1","average.ic2","beta.mse","mean.sens","mean
.spec","mean.negP","mean.posP","mean.auc","mean.accuracy")
```

```
sink('D:/results1/new_Adlasso2(n=100,p=200,r=0.3,sigma2,beta5).txt')

t1 <- Sys.time()
RTime <- RunningTime(t0,t1)
print(paste("p =",p,",", "n =", n,",","rho =",r,",","sigma = ",nsigma,",","beta = ",p1))
cat("================================================================
==================================================\n")
dimnames(summary) <- list(rowsummary.names,colsummary.names)
print(summary)
cat("================================================================
==================================================\n")
print(paste("Start ",t0,"***","End ",t1,"***","Time use ",RTime[1],":", RTime[2],":",
RTime[3]))
cat("\n")
cat("\n")

sink()
file.show("D:/results1/new_Adlasso2(n=100,p=200,r=0.3,sigma2,beta5).txt")
```

ประวัติผู้เขียน

ชื่อ-สกุล	วสุรัตน์ ขำภาษี
วัน เดือน ปี เกิด	14 พฤษภาคม 2539
สถานที่เกิด	นครปฐม ประเทศไทย
วุฒิการศึกษา	วท.บ. (คณิตศาสตร์) มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
ที่อยู่ปัจจุบัน	441/5 ถนนสวนตะไคร้ ตำบลสนามจันทร์ อำเภอเมือง จังหวัดนครปฐม 73000
ผลงานตีพิมพ์	Khumpasee, W. and Hirunkasi, K. (2022). Weights Choices in Adaptive LASSO for High Dimensional Sparse Logistic Regression Model. The 13th National Science Research Conference, Phatthalung, Thailand.