การจำลองข้อมูลเพื่อเปรียบเทียบความแม่นยำในการพยากรณ์ระหว่างวิธีโครงข่ายประสาทเทียมกับวิธีซัพพอร์ตเวกเตอร์แมชชีน

นางสาวนันทณัฐ พันธุ์สีดา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2556

A SIMULATION STUDY TO COMPARE PREDICTION ACCURACY BETWEEN ARTIFICIAL NEURAL NETWORK AND SUPPORT VECTOR MACHINE

Miss Nantanat Pansida

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Statistics

Department of Statistics

Faculty of Commerce and Accountancy

Chulalongkorn University

Academic Year 2013

หัวข้อวิทยานิพนธ์	การจำลองข้อมูลเพื่อเปรียบเทียบความแม่นยำในการพยากรณ์ระหว่างวิธีโครงข่ายประสาทเทียมกับวิธีซัพพอร์ตเวกเตอร์แมชชีน
โดย	นางสาวนันทณัฐ พันธุ์สีดา
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	อาจารย์ ดร.นัท กุลวานิช

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะพาณิชยศาสตร์และการบัญชี
(รองศาสตราจารย์ ดร.พสุ เดชะรินทร์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ ดร.กัลยา วานิชย์บัญชา)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(อาจารย์ ดร.นัท กุลวานิช)

..... กรรมการ
(รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา)

..... กรรมการภายนอกมหาวิทยาลัย
(อาจารย์ ดร.อรุณี กำลัง)

นันทณัฐ พันธุ์สีดา : การจำลองข้อมูลเพื่อเปรียบเทียบความแม่นยำในการพยากรณ์ระหว่างวิธีโครงข่ายประสาทเทียมกับวิธีซัพพอร์ตเวกเตอร์แมชชีน. (A SIMULATION STUDY TO COMPARE PREDICTION ACCURACY BETWEEN ARTIFICIAL NEURAL NETWORK AND SUPPORT VECTOR MACHINE) อ. ที่ปรึกษาวิทยานิพนธ์หลัก : อ.ดร. นัท กุลวานิช , 79 หน้า.

การวิจัยในครั้งนี้ มีวัตถุประสงค์เพื่อเปรียบเทียบความแม่นยำในการพยากรณ์ระหว่างวิธีการโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับกับวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนล โดยใช้ Receiver Operating Characteristic (ROC) เป็นเครื่องมือวัดประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูล โดยใช้พื้นที่ใต้โค้ง ROC (Area Under ROC Curve : AUC) และใช้อัตราความผิดพลาดในการจำแนกประเภทข้อมูล (Misclassification Rate : MCR)

พิจารณาค่าเฉลี่ยของ AUC จะได้ว่า วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลได้ดีที่สุดในกรณีที่ข้อมูลมีการแจกแจงแบบชี้กำลังและข้อมูลที่มีการแจกแจงแบบปกติ ส่วนกรณีที่ข้อมูลมีการแจงแจกแบบปัวส์ซงนั้น วิธีโครงข่ายประสาทเทียมแบบย้อนกลับ ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุด

พิจารณาค่าของเฉลี่ยของ MCR จะได้ว่า ในทุกกรณีของการแจกแจงที่ศึกษาในงานวิจัยนี้นั้น วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุด และวิธีโครงข่ายประสาทเทียมแบบย้อนกลับ ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้สูงที่สุดในทุกกรณี

ภาควิชา สถิติ ลายมือชื่อนิสิต .....

สาขาวิชา สถิติ ลายมือ ชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก .....

ปีการศึกษา 2556

5481600626 : MAJOR STATISTICS

KEYWORDS: Artificial Neural Network , Support Vector Machine Classification , Kernel Function

NANTANAT PANSIDA : A SIMULATION STUDY TO COMPARE PREDICTION ACCURACY BETWEEN ARTIFICIAL NEURAL NETWORK AND SUPPORT VECTOR MACHINE. ADVISOR : NAT KULVANICH, Ph.D., 79 pp.

This thesis attempted to identify a simulation study to compare prediction accuracy between the backpropagation artificial neural network and support vector machine. The method was to compare the forecasting accuracy using area under ROC curve (AUC) and misclassification rate (MCR).

The average values of the AUC was the support vector machines with the laplacian kernel method provides better prediction performance than the backpropagation artificial neural network in most cases ,except in case independent variables was poisson distribution

The average values of the MCR was that in all cases the distributions was studied in this research . Method with support vector machines with the laplacian Kernel to the error rate in the classification have the lowest . The Backpropagation Artificial Neural Network have the error rate in the highest classification in all cases.

Department : Statistics Student's Signature……..……..…….

Field of Study : Statistics Advisor's Signature……..……..

Academic Year : 2013

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สามารถสำเร็จลุล่วงไปได้ด้วยดีด้วยความอนุเคราะห์เป็นอย่างสูง ขอกราบขอบพระคุณอาจารย์ ดร. นัท กุลวานิช อาจารย์ที่ปรึกษาวิทยานิพนธ์ของผู้วิจัย ซึ่งให้ความกรุณาแก่ผู้วิจัยเป็นอย่างมาก ทั้งสละเวลาให้คำปรึกษา คำแนะนำเพื่อปรับปรุงแก้ไขวิทยานิพนธ์ และให้กำลังใจในการทำงาน ตลอดจนอ่านตรวจทานแก้ไขจุดบกพร่องของวิทยานิพนธ์นี้ และขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา ที่เสียสละเวลาช่วยแนะนำและให้คำปรึกษาวิทยานิพนธ์ในเบื้องต้น จนกระทั้งวิทยานิพนธ์เล่มนี้สำเร็จสมบูรณ์

ขอกราบขอบพระคุณรองศาสตราจารย์ ดร.กัลยา วานิชย์บัญชา ประธานกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา และอาจารย์ ดร. อรุณี กำลัง กรรมการสอบวิทยานิพนธ์ ที่กรุณาให้คำแนะนำ ตรวจสอบ และแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น

ขอกราบขอบพระคุณคณาจารย์ประจำภาควิชาสถิติ ที่ให้โอกาสทางการศึกษาและประสิทธิประสาทความรู้ให้แก่ผู้วิจัยจนกระทั่งสำเร็จการศึกษาในครั้งนี้

สุดท้ายนี้ผู้วิจัยขอกราบขอบพระคุณครอบครัวที่ช่วยส่งเสริมสนับสนุน เป็นกำลังใจให้ผู้วิจัยในการศึกษาเล่าเรียนจนสำเร็จการศึกษาในครั้งนี้ และขอขอบคุณเพื่อน ๆ ทุกคน ที่คอยช่วยเหลือ ให้คำแนะนำไขข้อสงสัยต่าง ๆ และให้กำลังใจผู้วิจัยมาโดยตลอด

สารบัญ

หน้า

สารบัญตาราง

สารบัญภาพ

ภาพที่ หน้า

ภาพที่ หน้า

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันนี้มีการพัฒนาด้านเทคโนโลยีสารสนเทศต่าง ๆ ให้มีประสิทธิภาพและมีความสะดวกรวดเร็วเพิ่มมากขึ้น ทำให้ผู้ที่มีความสามารถในการพยากรณ์เหตุการณ์หรือคาดเดาสถานการณ์ที่กำลังจะเกิดขึ้นได้ในอนาคตนั้น สามารถกำหนดเป้าหมายและแนวทางแก้ไขปัญหาในสถานการณ์ต่าง ๆ ได้อย่างเหมาะสม โดยวิธีการพยากรณ์จำแนกประเภทกลุ่มทางสถิติที่นิยมใช้มีดังนี้

1.1.1 วิธีการพยากรณ์จำแนกประเภทของตัวแบบที่ใช้พารามิเตอร์ (Parametric)

เป็นตัวแบบที่สามารถอธิบายความสัมพันธ์ของตัวแปรอิสระที่มีผลต่อตัวแปรตามได้ เช่น

การวิเคราะห์ความถดถอยโลจิสติค (Logistic Regression Analysis : LRA) เป็นเทคนิคที่ใช้ในการศึกษาความสัมพันธ์ของตัวแปรตามที่เป็นตัวแปรเชิงคุณภาพกับตัวแปรอิสระที่เป็นตัวแปรเชิงปริมาณหรือตัวแปรเชิงคุณภาพก็ได้ โดยอธิบายความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระที่อยู่ในรูปสมการเชิงเส้น ซึ่งใช้ประโยชน์จากความสัมพันธ์เชิงเส้นที่ได้จากการวิเคราะห์ตัวแบบนั้นมาประมาณค่าหรือพยากรณ์ค่าของตัวแปรตามที่ใช้อธิบายลักษณะของเหตุการณ์ที่สนใจกับเหตุการณ์ที่ไม่สนใจ ซึ่งมีข้อกำหนดของค่าความคลาดเคลื่อนเป็นอิสระกันและตัวแปรอิสระต้องไม่มีความสัมพันธ์กัน

การวิเคราะห์จำแนกประเภทเชิงเส้น (Linear Discriminant Analysis : LDA) มีลักษณะและแนวความคิดคล้ายคลึงกับการวิเคราะห์ความถดถอยโลจิสติค แต่มีเงื่อนไขของการแจกแจงของตัวแปรอิสระ p ตัว ซึ่งต้องมีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) และเมทริกซ์ค่าความแปรปรวนร่วม (Multivariate Analysis of Variance) ของตัวแปรอิสระ p ตัวของทุกกลุ่มต้องเท่ากัน

และการวิเคราะห์จำแนกประเภทของสมการกำลังสอง (Quadratic Discriminant Analysis : QDA) มีลักษณะและความคิดคล้ายคลึงกับการวิเคราะห์จำแนกประเภทเชิงเส้น เพียงแต่ความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระจะอยู่ในรูปของสมการกำลังสอง

และเมทริกซ์ค่าความแปรปรวนร่วม (Multivariate Analysis of Variance) ของตัวแปรอิสระ p ตัว ของแต่ละกลุ่มไม่เท่ากัน

จากตัวอย่างวิธีการทางสถิติที่นิยมใช้กันนั้น จะมีประสิทธิภาพก็ต่อเมื่อชุดข้อมูลที่นำมาวิเคราะห์มีลักษณะการแจกแจงที่สอดคล้องกับข้อสมมติของตัวแบบทางสถิตินั้น ๆ แต่ในความเป็นจริง ข้อมูลโดยส่วนใหญ่มีลักษณะการแจกแจงที่เบ้ และไม่สอดคล้องกับข้อสมมติของตัวแบบทางสถิติที่กล่าวมาข้างต้น จึงทำให้ประสิทธิภาพในการพยากรณ์จำแนกประเภทลดลง และเมื่อพิจารณาวิธีการพยากรณ์จำแนกประเภทข้อมูลที่ไม่มีข้อสมมติดังนี้

1.1.2 **วิธีการพยากรณ์จำแนกประเภทข้อมูลของตัวแบบที่ไม่ใช้พารามิเตอร์ (Nonparametric)** ซึ่งเป็นวิธีการของทางปัญญาประดิษฐ์ โดยตัวแบบของการวิเคราะห์ที่ได้นั้นไม่สามารถอธิบายความสัมพันธ์ของตัวแปรอิสระที่มีผลต่อตัวแปรตามได้ เช่น

วิธีการวัดความใกล้เคียง (k – Nearest Neighbor : k – NN) มีแนวคิดมาจากการจำลองกระบวนการคิดของมนุษย์ ซึ่งเมื่อมีปัญหาเข้ามาจะทำการเปรียบเทียบกับเงื่อนไขและวิธีแก้ปัญหาจากประสบการณ์ที่มีอยู่ แล้วเลือกกรณีการแก้ปัญหาที่ใกล้เคียงที่สุด จากการแบ่งกลุ่มตามการตรวจสอบความใกล้เคียง มีจำนวนเต็มบวก k ซึ่งค่านี้เป็นตัวบอกจำนวนของกรณีในทุกมิติ วิธีการในการหาความใกล้เคียงที่นิยมใช้กัน คือ ระยะห่างยุคลิด (Euclidean distance) ซึ่งใช้ในการคำนวณน้ำหนักของความใกล้เคียง และไม่จำเป็นที่จะต้องทราบจำนวนกลุ่มของชุดข้อมูลมาก่อน

วิธีโครงข่ายประสาทเทียม (Artificial Neural Networks : ANN) มีแนวคิดมาจากการจำลองกระบวนการคิดของมนุษย์ โดยอาศัยการนำเข้าข้อมูล เพื่อสร้างตัวแบบการจำลอง เพื่อใช้ในการพยากรณ์ข้อมูลในอนาคต แล้วทำการปรับปรุงข้อมูลให้มีเหมาะสมกับเงื่อนไขของข้อมูลที่มีการเปลี่ยนแปลง โดยแนวความคิดนี้ จะพยายามลดจำนวนของการพยากรณ์เพื่อจำแนกประเภทให้ผิดพลาดต่ำที่สุด

และวิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) มีแนวความคิดที่มีการพัฒนาจากวิธีโครงข่ายประสาทเทียมแบบชั้นเดียว (Single – Layer Neural Networks) โดยวิธีซัพพอร์ตเวกเตอร์แมชชีน มีจุดประสงค์ของการใช้ประโยชน์จากระนาบหลายมิติเพื่อสร้างเส้นแบ่งแยกประเภทของข้อมูลที่ดีที่สุด (Optimal Separating Hyper plane) และมีการใช้แนวความคิดของการเกิดความผิดพลาดต่ำที่สุด (Structural Risk Minimization) เป็นต้น

ผู้วิจัยได้ทำการศึกษางานวิจัยของ Zan Huang และคณะ (2004) ได้ทำการศึกษางานวิจัยที่เกี่ยวข้องกับการพยากรณ์จำแนกประเภททางการเงินต่างๆ โดยงานวิจัยนี้ได้มีการรวมรวบเอกสารงานวิจัยที่เกี่ยวข้องกับการเปรียบเทียบการพยากรณ์จำแนกประเภทโดยวิธีการของตัวแบบที่ใช้พารามิเตอร์ (Parametric) และวิธีการของตัวแบบที่ไม่ใช้พารามิเตอร์ (Nonparametric) ผลจากการรวบรวมงานวิจัยดังกล่าว ได้ผลสรุปที่ว่าวิธีการของตัวแบบที่ไม่ใช้พารามิเตอร์ (Nonparametric) มีประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกปะเภทที่ดีกว่าวิธีการของตัวแบบที่ใช้พารามิเตอร์ (Parametric)

และผลการจากศึกษางานวิจัยดังกล่าว ทำให้ Zan Huang และคณะ (2004) ได้ทำการศึกษาเพิ่มเติมเกี่ยวกับการเปรียบเทียบการจัดอันดับความน่าเชื่อถือของบริษัทโดยศึกษาข้อมูลจริงที่เกี่ยวข้องกับการเงินของบริษัทจากประเทศไต้หวัน และประเทศอเมริกา เพื่อศึกษาประสิทธิภาพของวิธีการพยากรณ์จำแนกประเภทระหว่างวิธีความถดถอยโลจิสติค , วิธีโครงข่ายประสาทเทียม และวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยพิจารณาจากเปอร์เซ็นต์ของความแม่นยำในการพยากรณ์ (Percentage of prediction accuracy) ผลลัพธ์ที่ได้คือวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Gaussian Kernel และวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ จะให้ผลลัพธ์ของการพยากรณ์จำแนกประเภทที่มีความแม่นยำมากกว่า 80% ซึ่งวิธีทั้งสองให้ผลลัพธ์ที่ดีกว่าวิธีการวิเคราะห์ความถดถอยโลจิสติค

ผู้วิจัยได้ทำการศึกษางานวิจัยที่เกี่ยวข้องกับการเปรียบเทียบการจำแนกประเภทโดยวิธีการของตัวแบบที่ใช้พารามิเตอร์ (Parametric) และวิธีการของตัวแบบที่ไม่ใช้พารามิเตอร์ (Nonparametric) เพิ่มเติมดังนี้

ฐิติ อ่วมสวัสดิ์ (2002) ได้ทำการศึกษาเปรียบเทียบวิธีการพยากรณ์จำแนกประเภทข้อมูลด้วยวิธีการวิเคราะห์ความถดถอยโลจิสติค กับวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ โดยการวิจัยนั้นทำการจำลองข้อมูลของตัวแปรอิสระที่มีการแจกแจงแบบปกติ และกำหนดให้ตัวแปรอิสระบางตัวมีความสัมพันธ์กัน งานวิจัยนี้ใช้เกณฑ์ของค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Average Root Mean Square Error : ARMSE) ในการวัดประสิทธิภาพความถูกต้องแม่นยำในการพยากรณ์การจำแนกประเภท ผลลัพธ์ที่ได้คือวิธีการโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับจะให้ผลลัพธ์ของการพยากรณ์จำแนกประเภทที่มีความแม่นยำกว่าวิธีการวิเคราะห์ความถดถอยโลจิสติคเมื่อตัวอย่างมีขนาดเล็ก และเมื่อระดับความสัมพันธ์ของตัวแปรอิสระหรือระดับส่วนเบี่ยงเบนมาตรฐานของการแจกแจงของตัวแปรอิสระมีค่าสูง ซึ่งวิธีการ

วิเคราะห์ความถดถอยโลจิสติคจะให้ผลลัพธ์ของการพยากรณ์จำแนกประเภทที่ดี เมื่อขนาดของตัวอย่างมีขนาดใหญ่

Diego Alejandro และคณะ (2012) ได้ทำการศึกษาเปรียบเทียบวิธีการพยากรณ์ในการวิเคราะห์ความถดถอยโลจิสติค กับวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยการวิจัยนั้นทำการจำลองข้อมูลตัวแปรอิสระที่มีการแจกแจงหลายหลายรูปแบบ และกำหนดขนาดตัวอย่างให้มีความแตกต่างกัน โดยใช้อัตราความผิดพลาดในการจำแนกประเภทข้อมูล (Misclassification Rate : MCR) เป็นเกณฑ์ในการวัดประสิทธิภาพ จากผลการศึกษาวิธีซัพพอร์ตเวกเตอร์ที่ใช้ฟังก์ชันเคอร์เนลแบบเส้นตรงนั้นให้ผลลัพธ์ที่ไม่ดีที่สุดในทุกกรณี โดยวิธีซัพพอร์ตเวกเตอร์ด้วยฟังก์ชันเคอร์เนลอื่น ๆ ให้ผลลัพธ์ที่ดีกว่าวิธีการเคราะห์ความถดถอยโลจิสติคในทุกสถานการณ์ที่ทำการจำลองข้อมูล เมื่อข้อมูลมีลักษณะของการแจกแจงระหว่างสองกลุ่มใกล้เคียงกัน

Yuan – chin Ivan Chang ได้ทำการศึกษาเปรียบเทียบวิธีการพยากรณ์จำแนกประเภทของข้อมูลในกรณีที่ชุดข้อมูลมีจำนวนขนาดของกลุ่มตัวอย่างไม่สมดุลกัน โดยทำการศึกษาเปรียบเทียบวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนลต่าง ๆ กับวิธีซัพพอร์ตเวกเตอร์แมชชีนที่พัฒนาร่วมกับวิธีการเคราะห์ความถดถอยโลจิสติค โดยการวิจัยนั้นทำการจำลองข้อมูลของตัวแปรอิสระที่มีการแจกแจงแบบปกติ และกำหนดให้สร้างขนาดจำนวนตัวอย่างของกลุ่มที่ไม่สนใจมีจำนวนมากกว่าขนาดจำนวนตัวอย่างของกลุ่มที่สนใจ งานวิจัยนี้ใช้เกณฑ์ของค่าพื้นที่ใต้โค้ง ROC (Area Under ROC Curve : AUC) ซึ่งผลจากการศึกษาวิธีซัพพอร์ตเวกเตอร์แมชชีนที่พัฒนาร่วมกับวิธีการเคราะห์ความถดถอยโลจิสติค ให้ประสิทธิภาพของการวิธีการพยากรณ์จำแนกประเภทของข้อมูลได้สูงกว่า 80 % ของการพยากรณ์จำแนกประเภท และมีประสิทธิภาพที่ดีแม้ว่าลักษณะของการแจกแจงทั้งสองมีลักษณะที่คล้ายคลึงกัน

จากการศึกษางานวิจัยที่กล่าวมาทั้งหมด ทำให้ผู้วิจัยมีความสนใจที่จะทำการศึกษาเกี่ยวกับประสิทธิภาพของการพยากรณ์จำแนกประเภทของตัวแบบที่ไม่ใช้พารามิเตอร์ เนื่องมาจากงานวิจัยที่ทำการศึกษามานั้น ได้แสดงให้เห็นถึงประสิทธิภาพของการพยากรณ์จำแนกประเภทที่ดีกว่าตัวแบบที่ใช้พารามิเตอร์ และยังไม่มีผู้ใดที่ทำการศึกษาเปรียบเทียบประสิทธิภาพของการพยากรณ์จำแนกประเภทด้วยตัวแบบที่ไม่ใช้พารามิเตอร์เพียงอย่างเดียวมาก่อน ดังนั้นผู้วิจัยจึงทำการศึกษา โดยทำการจำลองข้อมูลเพื่อเปรียบเทียบความแม่นยำในการพยากรณ์จำแนกประเภทระหว่างวิธีโครงข่ายประสาทเทียมกับวิธีซัพพอร์ตเวกเตอร์แมชชีน ซึ่งกำหนดให้ตัวแปรตามเป็นตัวแปรเชิงคุณภาพ และตัวแปรอิสระเป็นตัวแปรเชิงปริมาณที่มีการแจกแจงหลากหลายมากขึ้นจากงานวิจัยที่ได้ทำการศึกษามา เพื่อหาแนวโน้มว่าวิธีการใด

มีประสิทธิภาพที่ดีกว่า โดยใช้เครื่องมือวัดความมีประสิทธิภาพของการพยากรณ์จำแนกประเภทเพื่อบอกความถูกต้องของการพยากรณ์ด้วยพื้นที่ใต้โค้ง ROC (Area Under ROC Curve : AUC) ซึ่งได้มาจากการพล็อตกราฟระหว่างค่า Sensitivity และ 1 – Specificity และใช้อัตราความผิดพลาดในการจำแนกประเภทข้อมูล (Misclassification Rate : MCR) เพื่อศึกษาว่าวิธีการใดมีความผิดพลาดในการจำแนกประเภทน้อยที่สุด โดยการศึกษาในครั้งนี้มีความสนใจที่จะศึกษาผลของเหตุการณ์เกิดขึ้นสองเหตุการณ์ (dichotomous)

1.2 วัตถุประสงค์ของการศึกษา

เพื่อเปรียบเทียบความแม่นยำในการพยากรณ์ระหว่างวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับกับวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนล โดยทำการจำลองข้อมูลที่มีสถานการณ์ของการเกิดเหตุการณ์สองเหตุการณ์ (dichotomous)

1.3 สมมติฐานในการศึกษา

การพยากรณ์จำแนกประเภทของการจำลองข้อมูลที่ศึกษานั้น จะแสดงแนวโน้มที่ว่าวิธีซัพพอร์ตเวกเตอร์แมชชีนที่มีประสิทธิภาพของความแม่นยำในการจำแนกประเภทดีกว่าวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ

1.4 ข้อตกลงเบื้องต้น

การศึกษาครั้งนี้มีข้อตกลงเบื้องต้นสำหรับการดำเนินงานวิจัยดังนี้

1. ทำการศึกษาเหตุการณ์เกิดขึ้นสองเหตุการณ์ (dichotomous) คือเหตุการณ์ที่สนใจกับเหตุการณ์ที่ไม่สนใจ เนื่องมาจากโดยหลักการพื้นฐานของวิธีโครงข่ายประสาทเทียมและวิธีซัพพอร์ตเวกเตอร์แมชชีนจะทำการพยากรณ์เพื่อจำแนกประเภทของสองเหตุการณ์ก่อนและค่อยทำตามกระบวนการเดิมซ้ำอีกครั้ง เพื่อเพิ่มการจำแนกประเภทของกลุ่มข้อมูลที่เพิ่มมากขึ้น ดังนั้น งานวิจัยนี้จึงนำเสนอการพยากรณ์จำแนกประเภทที่มีผลของเหตุการณ์เกิดขึ้นสองเหตุการณ์ (dichotomous) ซึ่งสามารถทำความเข้าใจได้ง่าย และมีวิธีการที่ไม่ซับซ้อน

2. ศึกษาตัวแบบที่มีตัวแปรอิสระที่มีการแจกแจงแบบชี้กำลัง (Exponential Distribution) นั่นคือ $x_i \sim Exp(\beta)$; $i = 1, 2, \ldots, n$

3. ศึกษาตัวแบบที่มีตัวแปรอิสระที่มีการแจกแจงแบบปัวส์ซง (Poisson Distribution) นั่นคือ $x_i \sim Poi(\lambda)$; $i = 1, 2, \ldots, n$

4. ศึกษาตัวแบบที่มีตัวแปรอิสระที่มีการแจกแจงแบบปกติ (Normal Distribution) นั่นคือ $x_i \sim N(\mu, \sigma^2 = I)$; $i = 1, 2, \ldots, n$

5. ศึกษาตัวแบบที่มีตัวแปรอิสระที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) นั่นคือ $x_i \sim N(\mu, \sigma^2 = I)$; $i = 1, 2, \ldots, n$

1.5 ขอบเขตของการศึกษา

ในการวิจัยครั้งนี้จะทำการศึกษาภายใต้ขอบเขตดังนี้

1. ศึกษาตัวแปรอิสระ (X) คือ 1 และ 2 ตัวแปร โดยมีการแจกแจงข้อตกลงเบื้องต้น และในการแจกแจงตัวแปรอิสระของกลุ่มตัวอย่างจะมีลักษณะการแจกแจงแบบเดียวกัน

2. ตัวแปรตาม (Y) เป็นข้อมูลเชิงกลุ่มที่อยู่ในระดับนามบัญญัติ (Nominal Scale) โดยกำหนดให้ตัวแปรตาม (Y) แบ่งเป็น 2 กลุ่ม คือ

$$Y = \begin{cases} 1 \; ; \; Group\,1 \quad \text{มีขนาด} \quad n_1 \\ 0 \; ; \; Group\,2 \quad \text{มีขนาด} \quad n_2 \end{cases}$$

3. ขนาดตัวอย่าง (n) ในการศึกษาครั้งนี้จะกำหนดให้มีอย่างน้อย 30 เท่าของจำนวนตัวแปรอิสระ ซึ่งกำหนดตัวอย่างของกลุ่มที่หนึ่งเป็นกลุ่มตัวอย่างที่สนใจและกลุ่มที่สองเป็นกลุ่มตัวอย่างที่ไม่สนใจ ซึ่งสามารถกำหนดขนาดของจำนวนตัวอย่าง ดังนี้ $n_s = 30, 60, 90, 120, 150$; $s = 1, 2$ โดยผู้วิจัยแบ่งกรณีศึกษาเป็น 3 กรณี คือ

3.1 กรณีที่ขนาดตัวอย่างของแต่ละกลุ่มเท่ากัน ดังนี้ $n_1 = n_2$ จะสามารถจัดกลุ่มของขนาดตัวอย่างได้เท่ากับ $\binom{5}{1} = 5$ กรณี

3.2 กรณีที่ขนาดตัวอย่างของกลุ่มตัวอย่างที่สนใจ**มากกว่า**กลุ่มตัวอย่างที่ไม่สนใจ ดังนี้ $n_1 > n_2$ จะสามารถจัดกลุ่มของขนาดตัวอย่างได้เท่ากับ $\binom{5}{2} = 10$ กรณี

3.3 กรณีที่ขนาดตัวอย่างของกลุ่มตัวอย่างที่สนใจ**น้อยกว่า**กลุ่มตัวอย่างที่ไม่สนใจ ดังนี้ $n_1 < n_2$ จะสามารถจัดกลุ่มของขนาดตัวอย่างได้เท่ากับ $\binom{5}{2} = 10$ กรณี

4. ศึกษาลักษณะของการแจกแจงข้อมูล โดยมีการศึกษาตัวอย่างที่มีการแจกแจง 2 แบบดังนี้

กรณีที่ 1 ตัวแปรอิสระ 1 ตัวแปร

การแจกแจง (Distribution)	**กลุ่มที่หนึ่ง** $Y=1\,;\,Group\,1=n_1$	**กลุ่มที่สอง** $Y=0\,;\,Group\,2=n_2$	**ค่าของพารามิเตอร์**
แบบชี้กำลัง	$\mathrm{Exp}(\beta=1)$	$\mathrm{Exp}(\beta=d)$	$d=\{3,5,7,9\}$
แบบปัวส์ซง	$\mathrm{Poi}(\lambda=1)$	$\mathrm{Poi}(\lambda=d)$	$d=\{3,5,7,9\}$
แบบปกติ	$N(\mu=0,\sigma^2=I)$	$N(\mu=d,\sigma^2=I)$	$d=\{0.5,1,1.5,2\}$

กรณีที่ 2 ตัวแปรอิสระ 2 ตัวแปร ที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution)

การแจกแจง (Distribution)	**กลุ่มที่หนึ่ง** $Y=1\,;\,Group\,1=n_1$	**กลุ่มที่สอง** $Y=0\,;\,Group\,2=n_2$	**ค่าของพารามิเตอร์**
แบบปกติหลายตัวแปร	$N(\mu=0,\sigma^2=I)$	$N(\mu=d,\sigma^2=I)$	$d=\{0.5,1,1.5,2\}$

กำหนดให้

กลุ่มที่สนใจ $(Y=1)$ เป็นกลุ่มฐาน ซึ่งเป็นกลุ่มที่มีเกณฑ์กำหนดที่ชัดเจน เพื่ออธิบายลักษณะของกลุ่มที่สนใจ โดยที่ค่าของพารามิเตอร์ของการแจกแจงข้อมูลเป็นค่าคงที่

กลุ่มที่ไม่สนใจ $(Y=0)$ ซึ่งเป็นกลุ่มที่มีการเปลี่ยนแปลงไปจากกลุ่มที่สนใจ หรือมีลักษณะการแจกแจงของข้อมูลแตกต่างกับกลุ่มที่สนใจ โดยที่ค่าของพารามิเตอร์ของการแจกแจงข้อมูลมีการเปลี่ยนแปลงตามค่าของ d

ทำการวิเคราะห์โดยสลับเปลี่ยนค่าของ d ไปเรื่อย ๆ ตามค่าที่กำหนดจนครบสำหรับการแจกแจงที่ทำการศึกษา ซึ่งสามารถอธิบายได้ว่า เมื่อค่าของ d มีค่าเพิ่มมากขึ้น จะแสดงว่ากลุ่มตัวอย่างของกลุ่มที่สนใจกับกลุ่มที่ไม่สนใจมีลักษณะการแจกแจงของข้อมูลแตกต่างกันเพิ่มมากขึ้นหรือสามารถอธิบายความแตกต่างระหว่างข้อมูลทั้งสองกลุ่มได้ชัดเจนเพิ่มมากขึ้น

5. กำหนดระดับความสัมพันธ์ของตัวแปรอิสระ ใช้สำหรับกรณีที่ตัวแปรอิสระมีการแจกแจงแบบปกติหลายตัวแปร โดยที่ค่าสหสัมพันธ์ระหว่างตัวแปรอิสระ (ρ) ต้องมีความสัมพันธ์เชิงเส้นเท่านั้น คือ $cor(x_1,x_2)$ = $0, 0.3, 0.5$ และ 0.9

6. ในการศึกษาครั้งนี้ทำการจำลองข้อมูลให้มีสถานการณ์ที่แตกต่างกันตามข้อกำหนดข้างต้นโดยใช้เทคนิคมอนติคาร์โล (Monte Carlo Simulation Technique) โดยทำการจำลองในแต่ละสถานการณ์จะกระทำซ้ำ 500 รอบ

1.6 คำจำกัดความที่ใช้ในการศึกษา

1. วิธีโครงข่ายประสาทเทียม (Artificial Neural Networks : NN) คือ ตัวแบบที่ทำการจำลองกระบวนการคิดของมนุษย์ ซึ่งมีกระบวนการนำเข้าข้อมูลระหว่างกัน โดยการสร้างเส้นเชื่อมโยงเป็นโครงข่ายทั่วถึงกัน และทำการรวบรวมข้อมูลเหล่านั้น มาแปลงเป็นผลลัพธ์เพื่อแสดงผลตามจุดประสงค์ที่ต้องการ ซึ่งวิธีการนี้มีความยืดหยุ่นในการทำงานสูง และสามารถปรับเปลี่ยนไปได้ตลอดการทำงานของกระบวนการวิเคราะห์

2. วิธีโครงข่ายประสาทเทียมแบบย้อนกลับ (The Backpropagation Artificial Neural Network : BP) เป็นวิธีโครงข่ายประสาทเทียมรูปแบบหนึ่งที่มีการพัฒนาจากรูปแบบเดิม โดยการนำค่าความผิดพลาดที่ได้มาใช้ในการปรับค่าน้ำหนัก ซึ่งค่าน้ำหนักที่เปลี่ยนแปลงไปมากหรือน้อยขึ้นอยู่กับค่าความผิดพลาดที่ได้รับ โดยกระบวนการของโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ จะมีกระบวนการวนซ้ำในขั้นตอนแรกจนกระทั่งค่าความผิดพลาดที่ได้มีค่าต่ำสุดที่กำหนดไว้หรือครบกระบวนการวนซ้ำตามที่กำหนดไว้

3. วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) เป็นวิธีโครงข่ายประสาทเทียมรูปแบบหนึ่งที่มีการพัฒนาจากวิธีโครงข่ายประสาทเทียมแบบชั้นเดียว (Single – Layer Neural Networks) โดยนำข้อมูลส่งเข้าชั้นรับข้อมูลแล้วส่งผ่านไปในปริภูมิลักษณะของมิติที่สูงขึ้น และเริ่มกระบวนการของโครงข่ายประสาทเทียม โดยซัพพอร์ตเวกเตอร์แมชชีนมีจุดประสงค์ของการใช้ประโยชน์จากระนาบหลายมิติเพื่อสร้างเส้นแบ่งแยกประเภทของข้อมูลที่ดีที่สุด (Optimal Separating Hyper plane)

4. ค่าพื้นที่ใต้โค้งอาร์โอซี (Area Under ROC Curve) คือ ค่าที่อธิบายความสามารถในการจำแนกประเภทของข้อมูลหรือความเชื่อถือได้ของตัวแบบกรณีที่มีเหตุการณ์เกิดขึ้น 2 เหตุการณ์

1.7 วิธีดำเนินการศึกษา

1. ศึกษาค้นคว้าเอกสารและข้อมูลที่เกี่ยวข้องกับงานวิจัย

2. กำหนดเงื่อนไขและขอบเขตของการวิจัย

- กำหนดค่าพารามิเตอร์ตามการแจกแจงที่กำหนดในขอบเขตของการศึกษา (d)
- กำหนดขนาดตัวอย่าง (n)
- ค่าสหสัมพันธ์ระหว่างตัวแปรอิสระ (ρ)

3. จำลองข้อมูลตามการแจกแจงและขอบเขตที่ต้องการศึกษา

- จำลองค่า x_i ตามการแจกแจงของข้อตกลงเบื้องต้น และจำลองค่า y เป็นข้อมูลเชิงกลุ่มที่อยู่ในระดับนามบัญญัติ (Nominal Scale) ตัวอย่างเช่น

$$y = \begin{cases} 1 & ; \quad x_i \sim Poi(\lambda = 1) \\ 0 & ; \quad x_i \sim Poi(\lambda = d) \end{cases}$$

4. ทำการประมาณค่าสัมประสิทธิ์การถดถอยของพารามิเตอร์ เพื่อสร้างตัวแบบสำหรับนำมาพยากรณ์

5. นำตัวแบบที่ใช้วิธีโครงข่ายประสาทเทียมแบบย้อนกลับและวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนลต่าง ๆ แล้วนำตัวแบบที่ได้ไปพยากรณ์ข้อมูลต่อไป เพื่อตรวจสอบผลของการพยากรณ์เทียบกับเหตุการณ์ที่เกิดขึ้นจริง และนำผลลัพธ์ของการพยากรณ์ที่ได้ไปสร้างตารางการแบ่งกลุ่ม

6. นำข้อมูลที่ได้ไปคำนวณหาค่าประมาณพื้นที่ใต้โค้ง ROC (Area Under ROC Curve : AUC) และคำนวณค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูล (Misclassification Rate : MCR)

7. วิเคราะห์และสรุปผลการเปรียบเทียบวิธีการที่ใช้ในการวิจัย

1.8 ประโยชน์ที่คาดว่าจะได้รับ

1. เป็นแนวทางในการเลือกใช้ระหว่างวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ หรือวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนลเพื่อให้เหมาะสมกับข้อมูลจริง

2. เพื่อวัดประสิทธิภาพของการพยากรณ์จำแนกประเภทด้วยการจำลองข้อมูลของการเกิดเหตุการณ์ที่เกิดขึ้นสองเหตุการณ์ ระหว่างวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับหรือวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนล

บทที่ 2

ทฤษฎีและตัวสถิติที่เกี่ยวข้อง

2.1 วิธีโครงข่ายประสาทเทียม (Artificial Neural Networks)

วิธีโครงข่ายประสาทเทียม (Artificial Neural Networks) มีแนวความคิดมาจากการจำลองกระบวนการคิดในสมองของมนุษย์ ซึ่งเป็นการนำเข้าข้อมูล โดยการเชื่อมโยงผลรวมของข้อมูลเหล่านั้นเข้าด้วยกัน เพื่อสร้างตัวแบบในการพยากรณ์ข้อมูลในอนาคต และในกระบวนการนำเข้าข้อมูลนั้น จะพยายามทำการปรับปรุงและพัฒนาข้อมูลให้มีความเหมาะสมกับเงื่อนไขที่กำหนด โดยวัตถุประสงค์ของโครงข่ายประสาทเทียม คือ พยายามลดความผิดพลาดในการพยากรณ์ข้อมูลให้ต่ำที่สุดแบ่งออกเป็น 2 ประเภทคือ

โครงสร้างของวิธีโครงข่ายประสาทเทียม ประกอบไปด้วย 3 ส่วนดังนี้

1. สถาปัตยกรรมของโครงข่าย (Network Architecture) ประกอบไปด้วย 3 ชั้นหลัก

1.1 ชั้นรับข้อมูล (Input Layer) คือ ข้อมูลที่จะนำเข้าสู่ระบบ หรือตัวแปรอิสระที่เก็บข้อมูลอยู่ในรูปของเวกเตอร์ และโครงข่ายประสาทเทียมสามารถทำการประมวลผลข้อมูลที่เป็นตัวเลขเท่านั้น

1.2 ชั้นแฝง (Hidden Layer) คือ ชั้นที่อยู่ระหว่างชั้นรับข้อมูลกับชั้นแสดงผล เป็นชั้นที่ช่วยในกระบวนการสร้างรูปแบบความสัมพันธ์และแปลงความสัมพันธ์ใหม่โดยการใช้ฟังก์ชันกระตุ้น

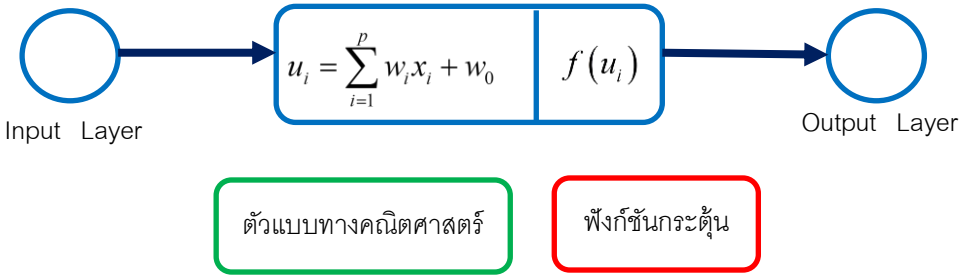
1.3 ชั้นแสดงผล (Output layer) คือ ผลลัพธ์ที่เกิดขึ้นจริง (Actual Outputs) จากกระบวนการสอนของโครงข่ายประสาทเทียม

2. การปรับค่าน้ำหนัก (Adjusting Weight) คือ สิ่งที่ได้จากการสอนของวิธีโครงข่ายประสาทเทียมหรือค่าความรู้ (Knowledge)

3. ฟังก์ชันกระตุ้น (Activation Function) เป็นฟังก์ชันการแปลงแสดงสมการทางคณิตศาสตร์

2.1.1 **วิธีโครงข่ายประสาทเทียมแบบชั้นเดียว** (Single – Layer Neural Networks)

วิธีโครงข่ายประสาทเทียมแบบชั้นเดียวประกอบด้วยชั้นรับข้อมูล (Input Layer) และชั้นแสดงผล (Output Layer) เท่านั้น ซึ่งในชั้นรับข้อมูลจะส่งข้อมูลเข้าสู่กระบวนการหาค่าน้ำหนักตามเส้นเชื่อมโยงต่าง แล้วนำมาคำนวณด้วยฟังก์ชันทางคณิตศาสตร์ จากนั้นส่งค่ามาที่ฟังก์ชันกระตุ้นเพื่อทำการพิจารณาตามเงื่อนไขของฟังก์ชันว่าชั้นแสดงผลควรให้ผลลัพธ์ใด ซึ่งวิธีโครงข่ายประสาทเทียมชั้นเดียวเหมาะสำหรับการแบ่งข้อมูลด้วยเส้นตรงเท่านั้น ตัวอย่างวิธีโครงข่ายแบบชั้นเดียว ลักษณะโครงสร้างของโครงข่ายแบบชั้นเดียวแสดงดังภาพที่ 2.1



ภาพที่ 2.1 แสดงโครงข่ายประสาทเทียมแบบชั้นเดียว (Single – Layer Neural Networks)

ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$u_i = \sum_{i=1}^{p} w_i x_i + w_0 \quad ; \quad i = 1, 2, \ldots, p \tag{2.1}$$

เมื่อ x_i คือ เวกเตอร์ของข้อมูลนำเข้า มีขนาด p มิติ

w_i คือ ค่าน้ำหนักที่เชื่อมต่อระหว่างชั้นรับข้อมูลกับชั้นแสดงผล

w_0 คือ ค่าน้ำหนักคงที่ ในชั้นรับข้อมูลกับชั้นแสดงผล

โดยค่าเริ่มต้นของ w_0 และ w_i เป็นค่าที่ได้จากการสุ่มข้อมูลในช่วง $[-1,1]$ ซึ่งค่าที่โปรแกรมของการคำนวณได้กำหนดเอาไว้

u_i คือ เวกเตอร์ผลรวมของค่าถ่วงน้ำหนักในชั้นรับข้อมูลกับชั้นแสดงผล

นำผลลัพธ์จากสมการ (2.1) มาทำการปรับค่าตามฟังก์ชันกระตุ้นที่ได้กำหนด คือ ฟังก์ชันซิกมอยด์ (Sigmoid Function) ซึ่งมีฟังก์ชันความหนาแน่นอยู่ในรูป

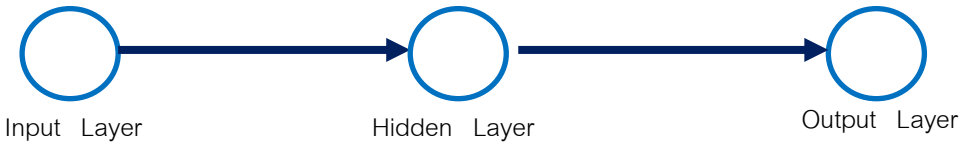
$$y = f(u_i) = \frac{1}{1+\exp^{-(u_i)}} \quad (2.2)$$

ซึ่งให้ผลลัพธ์มีค่าอยู่ในช่วง $[0,1]$

เมื่อ y คือ เวกเตอร์ของผลลัพธ์ที่ได้จากกระบวนการของโครงข่ายประสาทเทียม

2.1.2 **วิธีโครงข่ายประสาทเทียมแบบหลายชั้น** (Multi – Layer Neural Networks)

วิธีโครงข่ายประสาทเทียมแบบหลายชั้น เป็นวิธีโครงข่ายที่พัฒนาจากวิธีโครงข่ายประสาทเทียมแบบชั้นเดียว โดยการเพิ่มให้มีชั้นแฝง (Hidden Layer) ตั้งแต่ 1 ชั้นขึ้นไป โดยวิธีโครงข่ายแบบหลายชั้นจะใช้ในกรณีที่มีปัญหาซับซ้อนสูงซึ่งวิธีโครงข่ายแบบชั้นเดียวไม่สามารถใช้ในการแก้ไขปัญหาได้ ซึ่งปัญหาของข้อมูลโดยส่วนใหญ่ไม่สามารถทำการแบ่งประเภทได้ด้วยเส้นตรงที่ใช้วิธีโครงข่ายประสาทเทียมแบบชั้นเดียว ดังนั้นวิธีการนี้จึงเป็นที่นิยมใช้ในการแก้ไขปัญหาในการทำงาน ลักษณะโครงสร้างของวิธีโครงข่ายแบบหลายชั้นแสดงดังภาพที่ 2.2

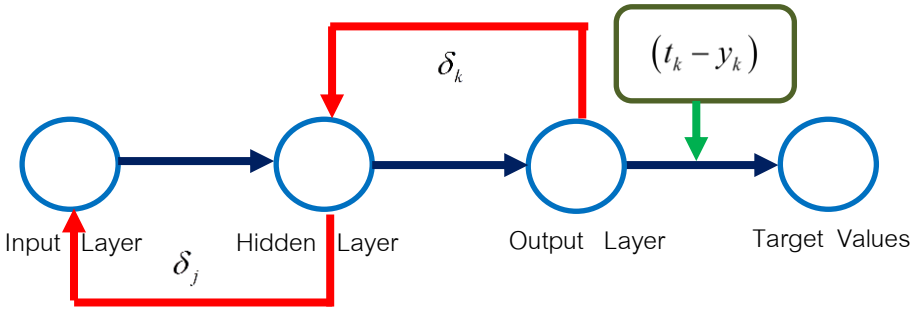


ภาพที่ 2.2 แสดงโครงข่ายประสาทเทียมแบบหลายชั้น (Multi – Layer Neural Networks)

2.1.2.1 **วิธีโครงข่ายประสาทเทียมแบบย้อนกลับ** (The Backpropagation Artificial Neural Network)

วิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ เป็นรูปแบบหนึ่งของวิธีโครงข่ายประสาทเทียมแบบหลายชั้นที่ได้รับความนิยม เนื่องจากสามารถแก้ปัญหาการแบ่งประเภทข้อมูลในลักษณะเชิงเส้น (Linear) และไม่เป็นเชิงเส้น (Nonlinear) ได้ และมีการพัฒนาจากรูปแบบเดิมโดยการนำค่าความผิดพลาดที่ได้มาใช้ในการปรับค่าน้ำหนัก ซึ่งค่าน้ำหนักที่เปลี่ยนแปลงไปมากหรือน้อยขึ้นอยู่กับค่าความผิดพลาดที่ได้รับ โดยกระบวนการของวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ จะมีกระบวนการวนซ้ำในขั้นตอนแรกจนกระทั่งค่าความผิดพลาดที่ได้มีค่าต่ำสุดที่กำหนดไว้ หรือครบกระบวนการวนซ้ำตามที่กำหนดไว้

โครงข่ายมีลักษณะ $p-q-m$ ในแต่ละชั้น ประกอบไปด้วย 3 ชั้นหลัก คือ ชั้นรับข้อมูล (Input layer) มี p โนด , ชั้นแฝง (Hidden Layer) หนึ่งชั้นหรืออาจมีมากกว่าหนึ่งชั้น มี q โนด และชั้นแสดงผล (Output Layer) จำนวนหนึ่งชั้น มี m โนด



ภาพที่ 2.3 แสดงลักษณะการทำงานของวิธีโครงข่ายประสาทเทียมแบบย้อนกลับ (The Backpropagation Artificial Neural Network)

โดยกำหนดให้มีสัญลักษณ์ในการคำนวณดังต่อไปนี้

เมื่อ x_i คือ เวกเตอร์ของข้อมูลนำเข้า มีขนาด p มิติ

v_{ij} คือ ค่าน้ำหนักที่เชื่อมต่อระหว่างชั้นรับข้อมูลกับชั้นแสดงผล

v_{0j} คือ ค่าของขอบเขต ในชั้นรับข้อมูลกับชั้นแสดงผล

u_j คือ เวกเตอร์ผลรวมของค่าถ่วงน้ำหนักในชั้นรับข้อมูลกับชั้นแสดงผล

z_j คือ ผลลัพธ์ที่ได้จากกระบวนการของโครงข่ายประสาทเทียมระหว่างชั้นรับข้อมูลกับชั้นแฝง

w_{jk} คือ ค่าน้ำหนักที่เชื่อมต่อระหว่างชั้นแฝงกับชั้นแสดงผล

w_{0k} คือ ค่าความเอนเอียงซึ่งเป็นค่าคงที่ ในชั้นแฝงกับชั้นแสดงผล

h_k คือ ผลรวมของค่าถ่วงน้ำหนักในชั้นแฝงกับชั้นแสดงผล

y_k คือ ผลลัพธ์ที่ได้จากกระบวนการของโครงข่ายประสาทเทียม

t_k คือ เวกเตอร์ของข้อมูลแสดงผลจริง มีขนาด k มิติ

δ_k คือ ค่าความผิดพลาดของผลลัพธ์ในชั้นแสดงผลกับชั้นแฝง

δ_j คือ ค่าความผิดพลาดของผลลัพธ์ในชั้นรับข้อมูลกับชั้นแฝง

$f'(S)$ คือ อนุพันธ์ของสมการผลรวมในตัวแบบโครงข่ายประสาทเทียม

e^q คือ ค่าความคลาดเคลื่อนในแต่ละแถวของข้อมูล

ขั้นตอนการคำนวณของวิธีโครงข่ายประสาทเทียมแบบย้อนกลับ

1. กำหนดจำนวนชั้นรับข้อมูล (p) , จำนวนชั้นแฝง (q) และจำนวนชั้นแสดงผล (m) ที่คาดว่าเหมาะสมกับปัญหาของข้อมูลตัวแปรอิสระและข้อมูลตัวแปรตาม

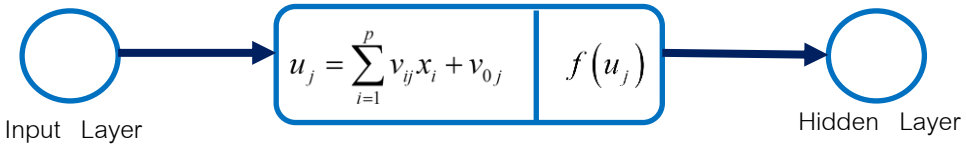
2. กำหนดค่าพารามิเตอร์ของอัตราการเรียนรู้ $(\eta : Learning\ rate)$ เป็น 0.5 และมีค่าเพิ่มขึ้นรอบละ 1.2

3. ทำการสุ่มค่าน้ำหนักเริ่มต้นให้กับทุกๆ เส้นเชื่อมโยงภายในโครงข่ายประสาทเทียมในชั้นของชั้นรับข้อมูลกับชั้นแฝง โดยให้มีค่าอยู่ระหว่าง $[-1,1]$

4. กำหนดจำนวนรอบสูงสุดที่จะใช้ในการเรียนรู้ $(R=100,000)$ และกำหนดค่าความผิดพลาดที่ยอมรับได้ $(e^q=0.05)$ เพื่อเป็นเกณฑ์ที่จะหยุดกระบวนการทำงาน ถ้าผลลัพธ์ไม่ผ่านเกณฑ์ที่กำหนดให้ทำตามกระบวนการที่ $5-13$ จนกว่าจะผ่านเกณฑ์ที่กำหนดไว้

5. นำเข้าเวกเตอร์ของข้อมูลส่งเข้าไปในชั้นรับข้อมูลชุดแรกหรือข้อมูลแถวแรก

6. คำนวณตามกระบวนการโครงข่ายประสาทเทียมแบบไปข้างหน้า (Feed – forward) จะทำการส่งข้อมูลจากชั้นรับข้อมูลไปสู่ชั้นแฝง

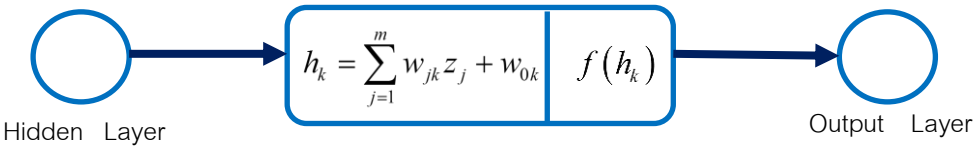


ภาพที่ 2.4 แสดงลักษณะการส่งข้อมูลจากชั้นรับข้อมูลไปสู่ชั้นแฝง

$$u_j = \sum_{i=1}^{p} v_{ij} x_i + v_{0j} \quad ; \quad i=1,2,\ldots,p \quad , \quad j=1,2,\ldots,q \tag{2.1}$$

$$z_j = f(u_j) = \frac{1}{1+\exp^{-(u_j)}} \tag{2.2}$$

7. คำนวณตามกระบวนการโครงข่ายประสาทเทียมแบบไปข้างหน้า (Feed – forward) จะทำการส่งข้อมูลจากชั้นแฝงไปสู่ชั้นแสดงผล

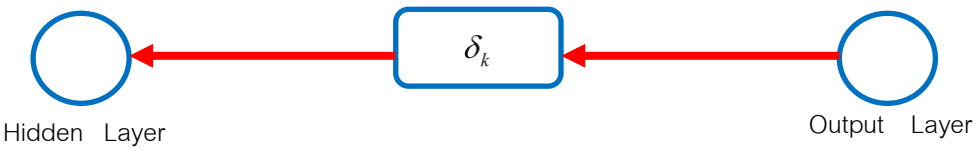


ภาพที่ 2.5 แสดงลักษณะการส่งข้อมูลจากชั้นแฝงไปสู่ชั้นแสดงผล

$$h_k = \sum_{j=1}^{m} w_{jk} z_j + w_{0k} \quad ; \quad j = 1,2,\ldots,q \quad , \quad k = 1,2,\ldots,m \tag{2.1}$$

$$y_k = f(h_k) = \frac{1}{1+\exp^{-(h_k)}} \tag{2.2}$$

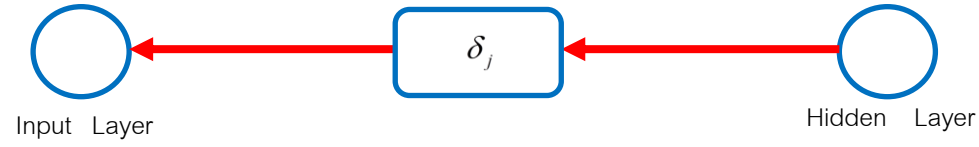
8. คำนวณตามกระบวนการโครงข่ายประสาทเทียมแบบย้อนกลับค่าความผิดพลาด (Error backpropagation) จะทำการส่งข้อมูลจากชั้นแสดงผลไปสู่ชั้นแฝง



ภาพที่ 2.6 แสดงลักษณะการส่งข้อมูลจากชั้นแสดงผลไปสู่ชั้นแฝง

$$\delta_k = (t_k - y_k) \bullet \left(f'(h_k) \right) \tag{2.3}$$

9. คำนวณตามกระบวนการโครงข่ายประสาทเทียมแบบย้อนกลับค่าความผิดพลาด (Error backpropagation) จะทำการส่งข้อมูลจากชั้นแฝงไปสู่ชั้นรับข้อมูล



ภาพที่ 2.7 แสดงลักษณะการส่งข้อมูลจากชั้นรับข้อมูลไปสู่ชั้นแฝง

$$\delta_j = \sum_{j=1}^{q} (\delta_k \bullet w_{jk}) \bullet \left(f'(u_j) \right) \tag{2.4}$$

10. การปรับค่าน้ำหนักระหว่างโนดในชั้นแฝงกับชั้นแสดงผล สามารถเขียนสมการได้ดังนี้ $w_{jk}^{(r+1)} = w_{jk}^{(r)} + \Delta w_{jk}$ เมื่อ $\Delta\left(w_{jk}\right) = \eta \bullet \delta_k \bullet z_j$ (2.5)

การปรับค่าน้ำหนักระหว่างโนดในชั้นรับข้อมูลกับชั้นแฝง สามารถเขียนสมการได้ดังนี้

$v_{ij}^{(r+1)} = v_{ij}^{(r)} + \Delta v_{ij}$ เมื่อ $\Delta v_{ij} = \eta \bullet \delta_j \bullet x_i$ (2.6)

11. คำนวณค่าความคลาดเคลื่อนเฉลี่ยในแต่ละแถวของข้อมูล สามารถเขียนสมการได้ดังนี้ $e^q = \frac{1}{2}\sum_{k=1}^{m}\left(t_k - y_k\right)^2$ (2.7)

12. คำนวณค่าความคลาดเคลื่อนเฉลี่ย (Mean Squared Error : MSE) สามารถเขียนสมการได้ดังนี้ $MSE = \frac{1}{Q}\sum_{q=1}^{Q} e^q$ (2.8)

การหาค่าความคลาดเคลื่อนเฉลี่ยนั้น เพื่อใช้ในการตรวจสอบว่าผลลัพธ์ของทุกๆ ข้อมูลในแต่ละรอบนั้นมีค่าน้อยกว่าค่าผิดพลาดที่ยอมรับได้ในทุกๆ แถวของข้อมูล

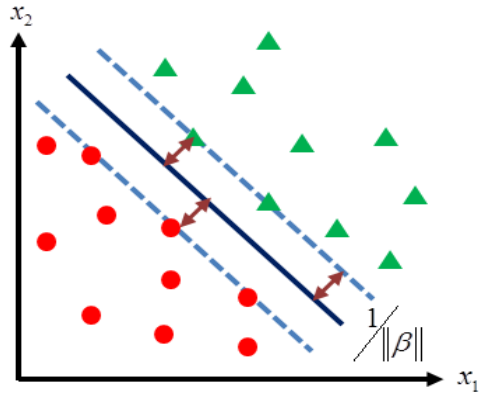
13. หยุดกระบวนการทำงานของโครงข่ายประสาทเทียมแบบย้อนกลับ

2.2 วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM)

วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) มีความคล้ายคลึงกับวิธีโครงข่ายประสาทเทียมแบบชั้นเดียว (Single – Layer Neural Networks) โดยวิธีซัพพอร์ตเวกเตอร์แมชชีนมีจุดประสงค์ของการใช้ประโยชน์จากระนาบหลายมิติเพื่อสร้างเส้นแบ่งแยกประเภทของข้อมูลที่ดีที่สุด (Optimal Separating Hyper plane) ซึ่งต้องมีคุณสมบัติของเงื่อนไขดังนี้

1. ค่าความผิดพลาดในการปฏิบัติเป็นศูนย์ (Zero Training Error)

2. ระยะระหว่างซัพพอร์ตเวกเตอร์ของทั้ง 2 ชนิดห่างกันมากที่สุด (Maximum Margin)



ภาพที่ 2.8 แสดงลักษณะหลักการหาระนาบเส้นแบ่งแยกประเภทของข้อมูลที่ดีที่สุดด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

กำหนดให้ลักษณะของข้อมูลเป็น $(\vec{x}_i, y_i)$ ซึ่ง $\vec{x}_i \in R^p$ โดยที่ $i = 1,2,\ldots,p$ และ $y_i \in \{-1,1\}$ จะทำการหาระนาบที่แบ่งตัวอย่างบวกและลบออกจากกัน และนำมาสร้างเส้นตรงบนไฮเปอร์เพลน (Hyper – plane) ซึ่งแบ่งกลุ่มข้อมูลที่มีลักษณะเป็นเชิงเส้นสองกลุ่มออกจากกัน

สามารถเขียนสมการที่เป็นตัวจำแนกประเภทข้อมูลได้ดังนี้

$$\beta x_i^T + \beta_o \geq +1 \quad for \quad y_i = 1 \tag{2.9}$$

และ $$\beta x_i^T + \beta_o \leq -1 \quad for \quad y_i = -1 \tag{2.10}$$

เมื่อ β คือ เวกเตอร์น้ำหนัก

x_i คือ เวกเตอร์ข้อมูล

β_0 คือ ค่าคงที่ (ค่าที่ตัดแกน y)

โดยสามารถนำสมการมาเขียนรวมกัน ได้ดังนี้

$$y_i\left(\beta x_i^T + \beta_0\right) \geq 1 \quad ; \quad \forall i \tag{2.11}$$

การหาระนาบสำหรับการแบ่งกลุ่มที่เหมาะสมที่สุด ทำได้โดยการหาค่าระยะ $\left(\text{Distance}\,; d\left(\beta, \beta_0; x\right)\right)$ จากตำแหน่งของซัพพอร์ตเวกเตอร์ x ถึงระนาบ (β, β_0) ได้ดังนี้

$$d\left(\beta, \beta_0; x\right) = \frac{\left|\langle \beta, x_i \rangle + \beta_0\right|}{\|\beta\|} \tag{2.12}$$

โดยวิเคราะห์จากค่าของระยะขอบที่มากที่สุด (Maximum margin) ซึ่งสามารถหาได้ดังนี้

$$\begin{aligned} p(\beta,\beta_0) &= \min_{x_i,y_i=-1} d(\beta,\beta_0;x_i) + \min_{x_i,y_i=+1} d(\beta,\beta_0;x_i) \\ &= \min_{x_i,y_i=-1} \frac{\left|\langle \beta, x_i\rangle + \beta_0\right|}{\|\beta\|} + \min_{x_i,y_i=+1} \frac{\left|\langle \beta, x_i\rangle + \beta_0\right|}{\|\beta\|} \\ &= \frac{1}{\|\beta\|}\left(\min_{x_i,y_i=-1} \left|\langle \beta, x_i\rangle + \beta_0\right| + \min_{x_i,y_i=+1} \left|\langle \beta, x_i\rangle + \beta_0\right| \right) \\ &= \frac{2}{\|\beta\|} \end{aligned}$$

สำหรับการหาสัมประสิทธิ์ที่ดีที่สุด ในการลดระยะทางดังกล่าวได้ดังนี้

$$\begin{aligned} \text{Minimize } \Phi(\beta,\beta_0) &= \frac{1}{2}\|\beta^2\| \\ \textit{Subject to} \quad & y_i\left(\beta x_i^T + \beta_0\right) \geq 1 \quad ; \quad \forall i \end{aligned} \tag{2.13}$$

กรณีที่ไม่สามารถแยกข้อมูลได้ด้วยไฮเปอร์เพลน (Hyper – plane)

เราสามารถหาได้เพียงไฮเปอร์เพลนที่สามารถแยกจุดตัวอย่างออกจากกันให้ได้มากที่สุด และยอมให้มีจุดตัวอย่างส่วนน้อยเพียงบางจุดที่ผิดพลาด โดยที่ข้อผิดพลาดที่เกิดขึ้นในกรณีที่ไม่สามารถทำการแบ่งกลุ่มของข้อมูลได้ กำหนดให้ $\xi = (\xi_1, \xi_2, \ldots, \xi_n)^T$ แทน เวกเตอร์ของตัวแปรที่ทำให้เกิดความผิดพลาดในการแบ่งข้อมูล (Slack Variables)

นำมาเขียนเป็นสมการตามเงื่อนไขใหม่ได้ดังนี้

$$\begin{aligned} \text{Minimize } \Phi(\beta,\beta_0) &= \frac{1}{2}\|\beta^2\| + \frac{1}{2}C\sum_{i=1}^{n}\xi_i \\ \textit{Subject to} \quad & y_i\left(\beta' x_i + \beta_0\right) \geq 1-\xi_i \quad ; \quad \xi_i \geq 0 \quad ; \quad \forall i \end{aligned} \tag{2.14}$$

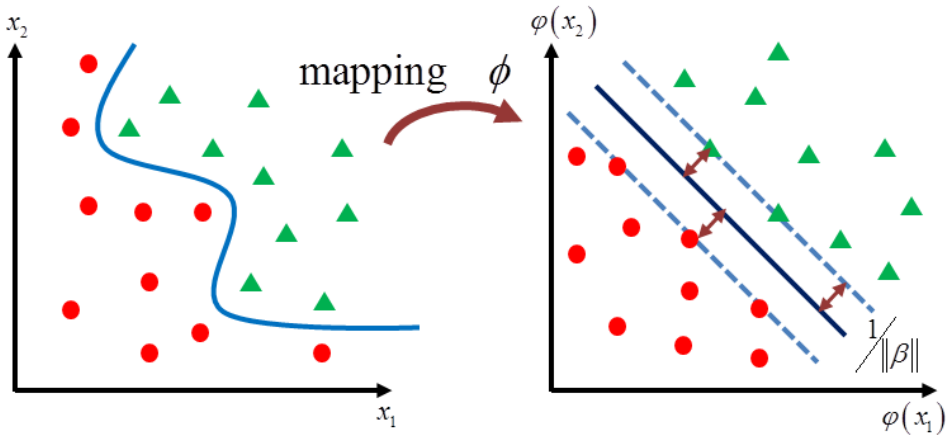
เมื่อ C คือ ค่าควบคุมการ Trade – off ระหว่างขอบเขตกับค่าผิดพลาด

นำมาเขียนสมการให้อยู่ในรูปของ Lagrangian [x] ได้ดังนี้

$$\begin{aligned} L_P(\beta,\beta_0,\xi,\alpha,\gamma) &= \sum_{i=1}^{n}\xi_i + \frac{1}{2}\|\beta\|^2 - \sum_{i=1}^{n}\alpha_i\left[1-y_i\left(\beta x_i^T+\beta_0\right)-\xi_i\right] - \sum_{i=1}^{n}\gamma_i\xi_i \\ \textit{Subject to} \quad & \alpha_i \geq 0 \text{ และ } \gamma_i \geq 0 \quad ; \quad \forall i \end{aligned} \tag{2.15}$$

กรณีการแบ่งกลุ่มโดยการใช้ระนาบแบบไม่เป็นเส้นตรง

ซัพพอร์ตเวกเตอร์แมชชีน จะอาศัยหลักการแปลงข้อมูลจากปริภูมิขาเข้า (Input space) ให้เป็นปริภูมิลักษณะ (Feature space) ที่มีมิติสูงขึ้น โดยใช้ฟังก์ชันเคอร์เนล (Kernel Function)



ภาพที่ 2.9 แสดงหลักการแปลงข้อมูลจากปริภูมิขาเข้าให้เป็นปริภูมิที่มีมิติสูงขึ้น

คุณสมบัติตามทฤษฎีของ Mercer (Mercer's Theorem) ดังนี้

$$\text{Kernel Function} \quad : \quad k(x_i, x_j) \quad = \quad \Phi(\vec{x}_i) \bullet \Phi(\vec{x}_j) \tag{2.16}$$

นำมาจัดให้อยู่ในรูปแบบของปัญหาคู่ (Dual Problem) โดยการแทนค่า β ในสมการของลากรานจ์ จะได้

$$\text{Maximize} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j}^{n} \alpha_i \alpha_j y_i y_j \Phi(x_i)^T \Phi(x_j) \tag{2.17}$$

$$\textit{With respect to} \quad \alpha \quad = \quad (\alpha_1, \alpha_2, \ldots \alpha_n)^T$$

$$\textit{Subject to} \quad 0 \le \alpha_i \le C \quad \textit{and} \quad \sum_{i=1}^{n} \alpha_i y_i = 0 \quad ; \quad \forall i$$

สมการที่แสดงการจำแนกข้อมูลบนไฮเปอร์เพลน (Hyper – plane) ได้ดังนี้

$$\textit{Hyperplane}\,; h(\vec{x}) \quad = \quad \text{Sgn}\left(\sum_{x_i=1}^{n} \alpha_i y_i \; K(x_i, x) + \beta_0 \right) \tag{2.18}$$

เมื่อ Sgn คือ Signum Function

กำหนดเคอร์เนลฟังก์ชันดังนี้

$$\text{Polynomial Kernel} \quad : \quad k(x_i, x_j) \quad = \quad (x_i \bullet x_j + 1)^d \tag{219}$$

$$\text{Gaussian Kernel} \quad : \quad k(x_i, x_j) \quad = \quad \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right) \tag{2.20}$$

$$\text{Laplacian Kernel} \quad : \quad k(x_i, x_j) \quad = \quad \exp\left(-\frac{\|x_i - x_j\|}{2\sigma^2} \right) \tag{2.21}$$

ขั้นตอนการคำนวณของซัพพอร์ตเวกเตอร์แมชชีน

1. นำเข้าเวกเตอร์ของข้อมูล และกำหนดขอบเขตของความผิดพลาดที่ยอมรับได้ $\left(C=\{1:10\}\right)$ และกำหนดพารามิเตอร์ของเคอร์เนลฟังก์ชัน $\left(\text{degree}=\{1:10\},\sigma^2=\left\{2^{-5}:2^5\right\}\right)$ ทำการคัดเลือกค่าพารามิเตอร์ต่างๆ ให้เหมาะสมกับข้อมูลในแต่ละชุด โดยการกระทำซ้ำขั้นตอนที่ 2 – 6 เพื่อปรับเปลี่ยนค่าพารามิเตอร์ที่กำหนดไว้ให้ครบตามที่กำหนด

2. คำนวณตามหลักการของการแปลงข้อมูลจากปริภูมิขาเข้า (Input space) ให้เป็นปริภูมิลักษณะ (Feature space) ที่มีมิติสูงขึ้น คุณสมบัติตามทฤษฎีของ Mercer (Mercer's Theorem) ดังนี้

$$\text{Kernel Function} \quad : \quad k\left(x_i,x_j\right) \;=\; \Phi\left(\vec{x}_i\right)\bullet\Phi\left(\vec{x}_j\right) \qquad (2.16)$$

3. เขียนสมการให้อยู่ในรูปของ Lagrangian [x] ได้ดังนี้

$$\text{Maximize} \quad \sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j}^{n}\alpha_i\alpha_j y_i y_j \Phi\left(x_i\right)^T\Phi\left(x_j\right) \qquad (2.17)$$

$$\textit{With respect to} \quad \alpha \;=\; \left(\alpha_1,\alpha_2,\ldots\alpha_n\right)^T$$

$$\textit{Subject to} \quad 0\le\alpha_i\le C \quad \textit{and} \quad \sum_{i=1}^{n}\alpha_i y_i=0 \quad ; \quad \forall i$$

และใช้วิธีการ Quadratic Programming Problem เพื่อหาผลลัพธ์ของตัวคูณลากรานจ์ (Lagrange Multipliers)

4. ทำการหาขนาดของขอบเขตที่ได้จากผลลัพธ์ของตัวคูณลากรานจ์ (Lagrange Multipliers) เพื่อหาตำแหน่งของข้อมูลที่เป็นซัพพอร์ตเวกเตอร์ของข้อมูลที่มีค่าตัวคูณลากรานจ์ไม่เท่ากับศูนย์

5. การหาค่าพารามิเตอร์ β,β_0 และ ξ_i โดยการ Differential ซึ่งกำหนดให้ผลลัพธ์มีค่าเป็นศูนย์ ตามสมการ (2.15) ดังนี้

$$L_P\left(\beta,\beta_0,\xi,\alpha,\gamma\right) \;=\; \sum_{i=1}^{n}\xi_i+\frac{1}{2}\|\beta\|^2-\sum_{i=1}^{n}\alpha_i\left[1-y_i\left(\beta x_i^T+\beta_0\right)-\xi_i\right]-\sum_{i=1}^{n}\gamma_i\xi_i$$

$$\textit{Subject to} \quad \alpha_i\ge 0 \quad \textit{and} \quad \gamma_i\ge 0 \quad ; \quad \forall i$$

$$\frac{\partial L_P}{\partial\beta} \;=\; \beta-\sum_{i=1}^{n}\alpha_i y_i x_i \;=\; 0 \qquad \text{จะได้} \qquad \beta \;=\; \sum_{i=1}^{n}\alpha_i y_i x_i \qquad (2.22)$$

$$\frac{\partial L_P}{\partial\beta_0} \;=\; \sum_{i=1}^{n}\alpha_i y_i \;=\; 0 \qquad \text{จะได้} \qquad \sum_{i=1}^{n}\alpha_i y_i \;=\; 0 \qquad (2.23)$$

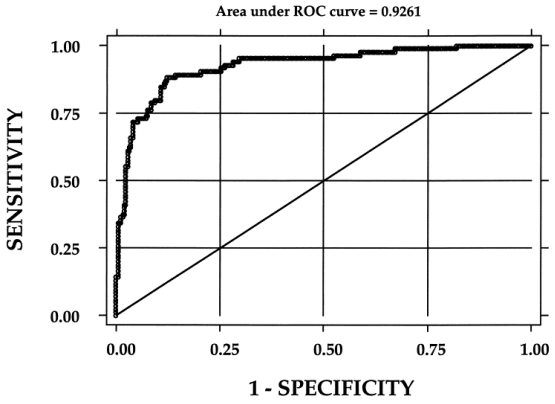
$$\frac{\partial L_P}{\partial\xi_i} \;=\; C-\alpha_i-\gamma_i \;=\; 0 \qquad \text{จะได้} \qquad \gamma_i \;=\; C-\alpha_i \qquad (2.24)$$

6. คำนวณหาสมการที่แสดงการแบ่งกลุ่มข้อมูลบนไฮเปอร์เพลน (Hyper – plane) ได้ดังนี้

$$Hyperplane\,;h\left(\vec{x}\right) \;=\; \mathrm{Sgn}\left(\sum_{x_i=1}^{n}\alpha_i y_i\, K\left(x_i,x\right)+\beta_0\right) \qquad (2.25)$$

2.3 เครื่องมือวัดความมีประสิทธิภาพของการพยากรณ์จำแนกประเภท

Receiver Operating Characteristic (ROC) ถูกนำมาใช้ในการประเมินความถูกต้องของการพยากรณ์เหตุการณ์ในระบบการจำแนกกลุ่มกรณีแบ่งข้อมูลเป็น 2 กลุ่ม ได้แก่ กลุ่มที่เกิดเหตุการณ์ที่สนใจและกลุ่มที่ไม่เกิดเหตุการณ์ที่สนใจ โดยอยู่ในรูปของกราฟที่พล็อตระหว่างค่า Sensitivity และค่า 1 – Specificity ซึ่งกราฟอยู่ในช่วง $[0,1]$ ดังนี้



ภาพที่ 2.10 กราฟแสดงพื้นที่ใต้โค้ง ROC

สำหรับการคำนวณค่า Sensitivity และ 1 – Specificity จะทำโดยการกำหนดค่าจุดตัด (Cutoff) ที่ระดับต่างๆ ระหว่าง 0 ถึง 1 เพื่อเปรียบเทียบกับความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจที่ได้จากการพยากรณ์ของแต่ละหน่วยตัวอย่างด้วยตัวแบบประมาณที่ได้จากการวิเคราะห์ $\left(p_i\right)$ แล้วทำการจำแนกกลุ่มของตัวแปรตามซึ่งเป็นตัวแปรที่ต้องการพยากรณ์ออกเป็น 2 กลุ่ม

โดยที่ ถ้า $p_i < cutoff \;\rightarrow\; Y_i = 0$ (ตัวอย่างจะถูกพยากรณ์ให้อยู่ในกลุ่มที่ไม่เกิดเหตุการณ์)

ถ้า $p_i \geq cutoff \;\rightarrow\; Y_i = 1$ (ตัวอย่างจะถูกพยากรณ์ให้อยู่ในกลุ่มที่เกิดเหตุการณ์)

จากนั้นจึงคำนวณหาสัดส่วนของการพยากรณ์เหตุการณ์ เพื่อนำค่าที่ได้ไปทำการพล็อตโค้ง ROC และคำนวณหาพื้นที่ใต้โค้ง ซึ่งสูตรการคำนวณ Sensitivity และ 1 – Specificity ได้ดังนี้

TP (True Positive) คือ จำนวนตัวอย่างที่พยากรณ์**ถูกต้อง**ของการ**เกิด**เหตุการณ์

FP (False Positive) คือ จำนวนตัวอย่างที่พยากรณ์**ผิด**ของการ**ไม่เกิด**เหตุการณ์

TN (True Negative) คือ จำนวนตัวอย่างที่พยากรณ์**ถูกต้อง**ของการ**ไม่เกิด**เหตุการณ์

FN (False Negative) คือ จำนวนตัวอย่างที่พยากรณ์**ผิด**ของการ**เกิด**เหตุการณ์

Sensitivity (True Positive Rate) เป็นความน่าจะเป็นหรืออัตราส่วนของการพยากรณ์เหตุการณ์ได้**ถูกต้อง**ของการ**เกิด**เหตุการณ์ที่สนใจ

Specificity (True Negative Rate) เป็นความน่าจะเป็นหรืออัตราส่วนของการพยากรณ์เหตุการณ์ได้**ถูกต้อง**ของการ**ไม่เกิด**เหตุการณ์ที่สนใจ

1 – Specificity (False – Positive Rate) เป็นความน่าจะเป็นหรืออัตราส่วนของการพยากรณ์เหตุการณ์ได้**ผิด**ของการ**ไม่เกิด**เหตุการณ์ที่สนใจ

ซึ่งจะได้ว่า

$$Sensitivity = \frac{TP}{Total\ acttual\ Positive} = \frac{TP}{TP + FN} \tag{2.26}$$

$$Specificity = \frac{TN}{Total\ acttual\ Negative} = \frac{TN}{FP + TN} \tag{2.27}$$

$$1 - Specificity = \frac{FP}{Total\ acttual\ Negative} = \frac{FP}{FP + TN} \tag{2.28}$$

จากกราฟที่ได้จะนำมาหาค่าประมาณพื้นที่ใต้โค้ง ROC (Area under the Curve หรือ $A\hat{U}C$) โดย $A\hat{U}C$ จะใช้เทคนิคการประมาณค่าเกี่ยวกับการคำนวณอินทริกรัลจำกัดเขต ซึ่งแสดงได้ดังนี้ $\int_a^b f(x)dx \quad ; \quad a \le x \le b$ (2.29)

อัตราความผิดพลาดในการจำแนกประเภทข้อมูล (Misclassification Rate : MCR)

อัตราความผิดพลาดในการจำแนกประเภทข้อมูล โดยการกำหนดจุดตัดเป็นค่ากึ่งกลางของการจำแนกข้อมูล $\left(p_i = 0.5\right)$ แล้วนำมาใช้ในการประเมินความความผิดพลาดของการพยากรณ์เหตุการณ์ในการจำแนกประเภทกรณีแบ่งข้อมูลเป็น 2 กลุ่ม

$$\text{MCR} = \frac{FP + FN}{n} \tag{2.30}$$

เมื่อ n คือ จำนวนข้อมูลทั้งหมด

บทที่ 3

วิธีการดำเนินการศึกษา

การศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความแม่นยำในวิธีการพยากรณ์จำแนกประเภทของวิธีการโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับกับวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนล โดยในการเปรียบเทียบความแม่นยำในการพยากรณ์จำแนกประเภทจะพิจารณาจากพื้นที่ใต้โค้งอาร์โอซี (Receiver Operating Characteristic : ROC) และใช้อัตราความผิดพลาดในการจำแนกประเภทข้อมูล (Misclassification Rate : MCR) เพื่อศึกษาว่าวิธีการใดมีความผิดพลาดในการจำแนกประเภทน้อยที่สุด

การศึกษาในครั้งนี้ทำการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล (Monte Carlo Method) โดยใช้โปรแกรม R เวอร์ชัน 2.15.3 ในการทำการศึกษาภายใต้ขอบเขตการณ์ศึกษาต่อไปนี้

3.1 ขอบเขตของการศึกษา

ในการวิจัยครั้งนี้จะทำการศึกษาภายใต้ขอบเขตดังนี้

1. กำหนดศึกษาตัวแปรอิสระ (X) ให้เป็นตัวแปรเชิงปริมาณ และตัวแปรตาม (Y) เป็นข้อมูลเชิงกลุ่มที่อยู่ในระดับนามบัญญัติ (Nominal Scale) มี 2 กรณี คือ

กรณีที่ 1 มีการศึกษาตัวแปรอิสระ (X) 1 ตัวแปร

- ศึกษาตัวแบบที่มีตัวแปรอิสระที่มีการแจกแจงแบบปัวส์ซง (Poisson Distribution) นั่นคือ $Y = \begin{cases} 1 & ; \quad x_1 \sim Poi(\lambda = 1) \\ 0 & ; \quad x_1 \sim Poi(\lambda = d) \end{cases}$
- ศึกษาตัวแบบที่มีตัวแปรอิสระที่มีการแจกแจงแบบชี้กำลัง (Exponential Distribution) นั่นคือ $Y = \begin{cases} 1 & ; \quad x_1 \sim Exp(\beta = 1) \\ 0 & ; \quad x_1 \sim Exp(\beta = d) \end{cases}$
- ศึกษาตัวแบบที่มีตัวแปรอิสระที่มีการแจกแจงแบบปกติ (Normal Distribution) นั่นคือ $Y = \begin{cases} 1 & ; \quad x_1 \sim Nor(\mu = 0, \sigma^2 = I) \\ 0 & ; \quad x_1 \sim Nor(\mu = d, \sigma^2 = I) \end{cases}$

กรณีที่ 2 มีการศึกษาตัวแปรอิสระ (X) 2 ตัวแปร

- ศึกษาตัวแบบที่มีตัวแปรอิสระที่มีการแจกแจงแบบปกติหลายตัวแปร (The Multivariate Normal Distribution) นั่นคือ

$$Y = \begin{cases} 1 \;\; ; \;\; \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim Nor\left(\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 = I \right) \\ 0 \;\; ; \;\; \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim Nor\left(\mu = \begin{bmatrix} d \\ d \end{bmatrix}, \sigma^2 = I \right) \end{cases}$$

เมื่อกำหนดให้ $d = 3,5,7,9$ สำหรับข้อมูลที่มีการแจกแจงแบบปัวส์ซงและการแจกแจงแบบซี้กำลัง และ $d = 0.5,1,1.5,2$ สำหรับข้อมูลที่มีการแจกแจงแบบแบบปกติและการแจกแจงแบบปกติหลายตัวแปร

กำหนดให้

กลุ่มที่สนใจ $(Y = 1)$ เป็นกลุ่มฐาน ซึ่งเป็นกลุ่มที่มีเกณฑ์กำหนดที่ชัดเจน เพื่ออธิบายลักษณะของกลุ่มที่สนใจ โดยที่ค่าของพารามิเตอร์ของการแจกแจงข้อมูลเป็นค่าคงที่

กลุ่มที่ไม่สนใจ $(Y = 0)$ ซึ่งเป็นกลุ่มที่มีการเปลี่ยนแปลงไปจากกลุ่มที่สนใจ หรือมีลักษณะการแจกแจงของข้อมูลแตกต่างกับกลุ่มที่สนใจ โดยที่ค่าของพารามิเตอร์ของการแจกแจงข้อมูลมีการเปลี่ยนแปลงตามค่าของ d

ทำการวิเคราะห์โดยสลับเปลี่ยนค่าของ d ไปเรื่อย ๆ ตามค่าที่กำหนดจนครบสำหรับการแจกแจงที่ทำการศึกษา ซึ่งสามารถอธิบายได้ว่า เมื่อค่าของ d มีค่าเพิ่มมากขึ้น จะแสดงว่ากลุ่มตัวอย่างของกลุ่มที่สนใจกับกลุ่มที่ไม่สนใจมีลักษณะการแจกแจงของข้อมูลแตกต่างกันเพิ่มมากขึ้นหรือสามารถอธิบายความแตกต่างระหว่างข้อมูลทั้งสองกลุ่มได้ชัดเจนเพิ่มมากขึ้น

2. ขนาดตัวอย่าง (n) ในการศึกษาครั้งนี้จะกำหนดให้มีอย่างน้อย 30 เท่าของจำนวนตัวแปรอิสระ ซึ่งกำหนดตัวอย่างของกลุ่มที่หนึ่งเป็นกลุ่มตัวอย่างที่สนใจและกลุ่มที่สองเป็นกลุ่มตัวอย่างที่ไม่สนใจ ซึ่งสามารถกำหนดขนาดของจำนวนตัวอย่าง ดังนี้ $n_s = 30,60,90,120,150 \;\; ; \;\; s = 1,2$ โดยผู้วิจัยแบ่งกรณีศึกษาเป็น 3 กรณี คือ

2.1 กรณีที่ขนาดตัวอย่างของแต่ละกลุ่มเท่ากัน ดังนี้ $n_1 = n_2$ จะสามารถจัดกลุ่มของขนาดตัวอย่างได้เท่ากับ $\binom{5}{1} = 5$ กรณี

2.2 กรณีที่ขนาดตัวอย่างของกลุ่มตัวอย่างที่สนใจ**มากกว่า**กลุ่มตัวอย่างที่ไม่สนใจ ดังนี้ $n_1 > n_2$ จะสามารถจัดกลุ่มของขนาดตัวอย่างได้เท่ากับ $\binom{5}{2} = 10$ กรณี

2.3 กรณีที่ขนาดตัวอย่างของกลุ่มตัวอย่างที่สนใจ**น้อยกว่า**กลุ่มตัวอย่างที่ไม่สนใจ ดังนี้ $n_1 < n_2$ จะสามารถจัดกลุ่มของขนาดตัวอย่างได้เท่ากับ $\binom{5}{2} = 10$ กรณี

3. กำหนดระดับความสัมพันธ์ของตัวแปรอิสระ กรณีที่ตัวแปรอิสระมีการแจกแจงแบบปกติหลายตัวแปร โดยที่ค่าสหสัมพันธ์ระหว่างตัวแปรอิสระ (ρ) ต้องมีความสัมพันธ์เชิงเส้นเท่านั้น คือ $cor(x_1, x_2) = 0, 0.3, 0.5$ และ 0.9

4. ในการศึกษาครั้งนี้ทำการจำลองข้อมูลให้มีสถานการณ์ที่แตกต่างกันตามข้อกำหนดข้างต้นโดยใช้เทคนิคมอนติคาร์โล (Monte Carlo Simulation Technique) โดยการจำลองในแต่ละสถานการณ์จะกระทำซ้ำ 500 รอบ

3.2 ขั้นตอนในการดำเนินการศึกษา

1. ศึกษาค้นคว้าเอกสารและข้อมูลที่เกี่ยวข้องกับงานวิจัย
   - วิธีการโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ
   - วิธีซัพพอร์ตเวกเตอร์แมชชีน
   - Receiver Operating Characteristic (ROC)
2. กำหนดเงื่อนไขและขอบเขตของการวิจัย
   - กำหนดค่าพารามิเตอร์ตามการแจกแจงที่กำหนดในขอบเขตของการศึกษา $(d = 3, 5, 7, 9 \;\; ; \;\; d = 0.5, 1, 1.5, 2)$
   - กำหนดขนาดตัวอย่าง $(n_s \;\; ; \;\; s = 1, 2)$
   - ค่าสหสัมพันธ์ระหว่างตัวแปรอิสระ $(\rho = 0, 0.3, 0.5, 0.9)$
3. จำลองข้อมูลตามการแจกแจงและขอบเขตที่ต้องการศึกษา
4. ทำการประมาณค่าสัมประสิทธิ์การถดถอยของพารามิเตอร์ เพื่อสร้างตัวแบบสำหรับนำมาพยากรณ์จำแนกประเภท

กำหนดค่าที่ใช้ในวิธีการโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับดังนี้

- ชั้นรับข้อมูล จำนวน 1 ชั้น
- ชั้นแฝง จำนวน 1 ชั้น

- ชั้นแสดงผล จำนวน 1 ชั้น
- กำหนดค่าพารามิเตอร์ของอัตราการเรียนรู้ (η) เป็น 0.5 และมีค่าเพิ่มขึ้นรอบละ 1.2
- สุ่มค่าของน้ำหนักเริ่มต้นให้มีค่าอยู่ระหว่าง $[-1,1]$
- กำหนดรอบสูงที่สุดที่ใช้ในการเรียนรู้ที่ $100,000$ รอบ

กำหนดค่าที่ใช้ในวิธีการซัพพอร์ตเวกเตอร์แมชชีน ดังนี้

- กำหนดช่วงของค่าควบคุมการ Trade – off ระหว่างขอบเขตกับค่าความผิดพลาดที่ดีที่สุด เป็นค่าตั้งแต่ 1 ถึง 50
- กำหนดช่วงของ degree ของ Polynomial Kernel เป็นค่าตั้งแต่ 1 ถึง 10
- กำหนดช่วงของ gamma ของ Gaussian Kernel และ Laplacian Kernel เป็นค่าตั้งแต่ 2^{-5} ถึง 2^{5}

ในการการประมาณค่าด้วยวิธีการซัพพอร์ตเวกเตอร์แมชชีนนั้น ผู้วิจัยได้มีการคัดเลือกค่าพารามิเตอร์ต่าง ๆ ให้เหมาะสมกับข้อมูลในแต่ละชุดก่อน โดยพิจารณาค่าตามที่กำหนดไว้ข้างต้น แล้วจึงทำการประมาณค่าสัมประสิทธิ์การถดถอยของพารามิเตอร์นั้นๆ

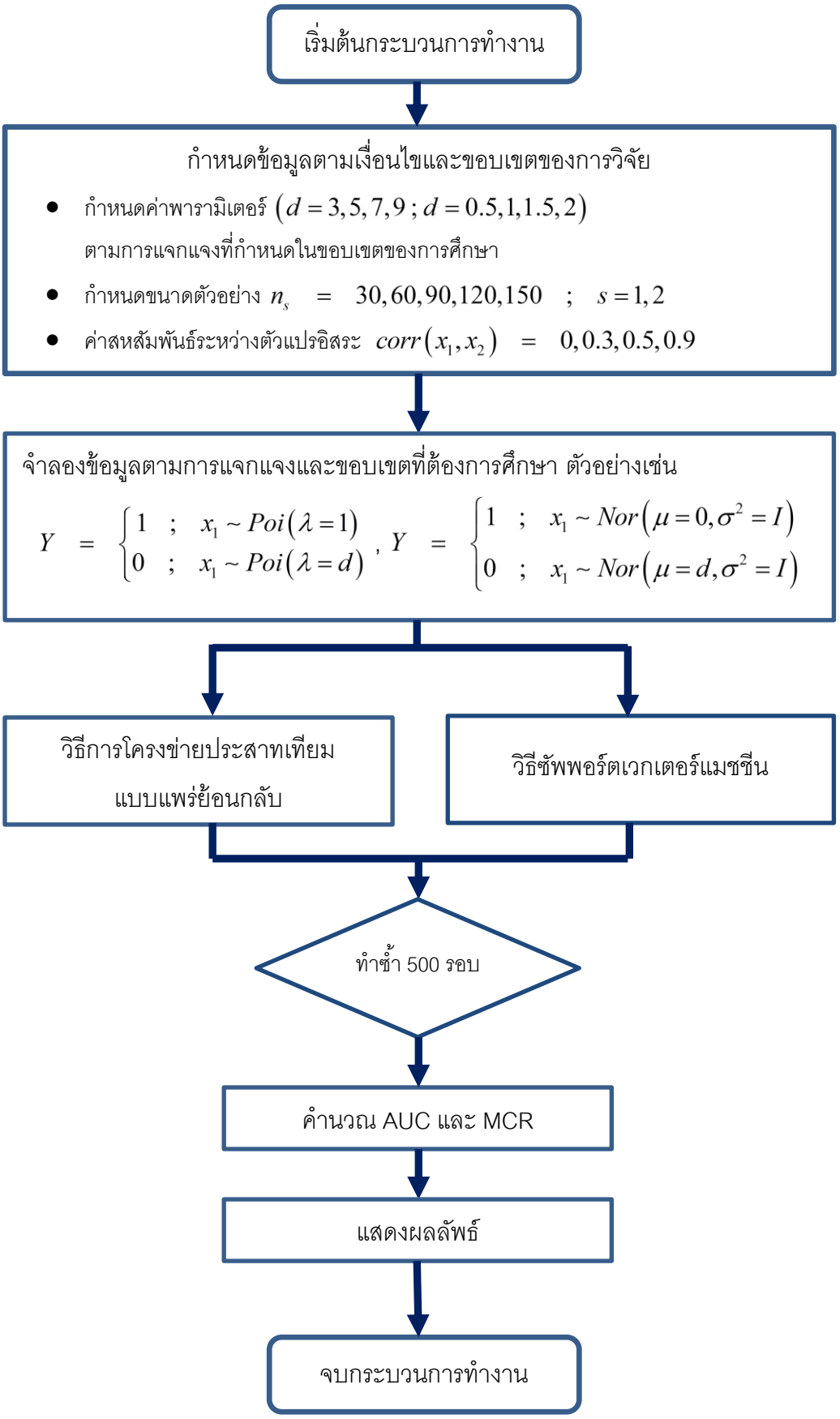
5. นำตัวแบบที่ใช้วิธีโครงข่ายประสาทเทียมและตัวแบบที่ใช้วิธีซัพพอร์ตเวกเตอร์แมชชีน แล้วนำตัวแบบที่ได้ไปพยากรณ์ข้อมูลต่อไป เพื่อตรวจสอบผลของการพยากรณ์เทียบกับเหตุการณ์ที่เกิดขึ้นจริง และนำผลลัพธ์ของการพยากรณ์ที่ได้ไปสร้างตารางการแบ่งกลุ่ม

6. นำข้อมูลที่ได้ไปสร้างตารางการแบ่งกลุ่มไปสร้างกราฟ ROC เพื่อคำนวณหาค่าประมาณพื้นที่ใต้โค้ง ROC และคำนวณค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูล (Misclassification Rate : MCR)

7. วิเคราะห์และสรุปผลการเปรียบเทียบวิธีการที่ใช้ในการวิจัย

3.3 ขั้นตอนการทำงานของโปรแกรม

โปรแกรมที่ใช้ในการศึกษาการวิจัยในครั้งนี้เขียนด้วยโปรแกรม R เวอร์ชัน 2.15.3 ซึ่งในแต่ละสถานการณ์ของการทดลองจะกระทำซ้ำ 500 รอบ สามารถแสดงขั้น ตอนการทำงานของโปรแกรมได้ดังนี้


เริ่มต้นกระบวนการทำงาน
กำหนดข้อมูลตามเงื่อนไขและขอบเขตของการวิจัย
กำหนดค่าพารามิเตอร์ $(d=3,5,7,9\,;d=0.5,1,1.5,2)$ ตามการแจกแจงที่กำหนดในขอบเขตของการศึกษา
กำหนดขนาดตัวอย่าง $n_s = 30,60,90,120,150 \;;\; s=1,2$
ค่าสหสัมพันธ์ระหว่างตัวแปรอิสระ $corr(x_1,x_2) = 0,0.3,0.5,0.9$
จำลองข้อมูลตามการแจกแจงและขอบเขตที่ต้องการศึกษา ตัวอย่างเช่น
$Y = \begin{cases} 1 & ; \; x_1 \sim Poi(\lambda=1) \\ 0 & ; \; x_1 \sim Poi(\lambda=d) \end{cases}$, $Y = \begin{cases} 1 & ; \; x_1 \sim Nor(\mu=0,\sigma^2=I) \\ 0 & ; \; x_1 \sim Nor(\mu=d,\sigma^2=I) \end{cases}$
วิธีการโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ
วิธีซัพพอร์ตเวกเตอร์แมชชีน
ทำซ้ำ 500 รอบ
คำนวณ AUC และ MCR
แสดงผลลัพธ์
จบกระบวนการทำงาน

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การศึกษางานวิจัยในครั้งนี้ มีวัตถุประสงค์เพื่อเปรียบเทียบความแม่นยำในการพยากรณ์จำแนกประเภทระหว่างวิธีโครงข่ายประสาทเทียมกับวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนล โดยทำการจำลองข้อมูลเพื่อศึกษาผลกระทบจากระดับค่าพารามิเตอร์ของการแจกแจงข้อมูล (d) , ค่าระดับความสัมพันธ์ของตัวแปรอิสระ (ρ) และขนาดของกลุ่มตัวอย่าง (n_1, n_2) ทำการพิจารณาผลการศึกษาด้วย Receiver Operating Characteristic (ROC) ใช้เป็นเครื่องมือวัดประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูล โดยใช้พื้นที่ใต้โค้ง ROC และใช้อัตราความผิดพลาดในการจำแนกประเภทข้อมูล (Misclassification Rate : MCR) เพื่อศึกษาว่าวิธีการใดมีความผิดพลาดในการจำแนกประเภท ซึ่งในงานวิจัยนี้ทำการศึกษาผลของเหตุการณ์เกิดขึ้นสองเหตุการณ์ (dichotomous)

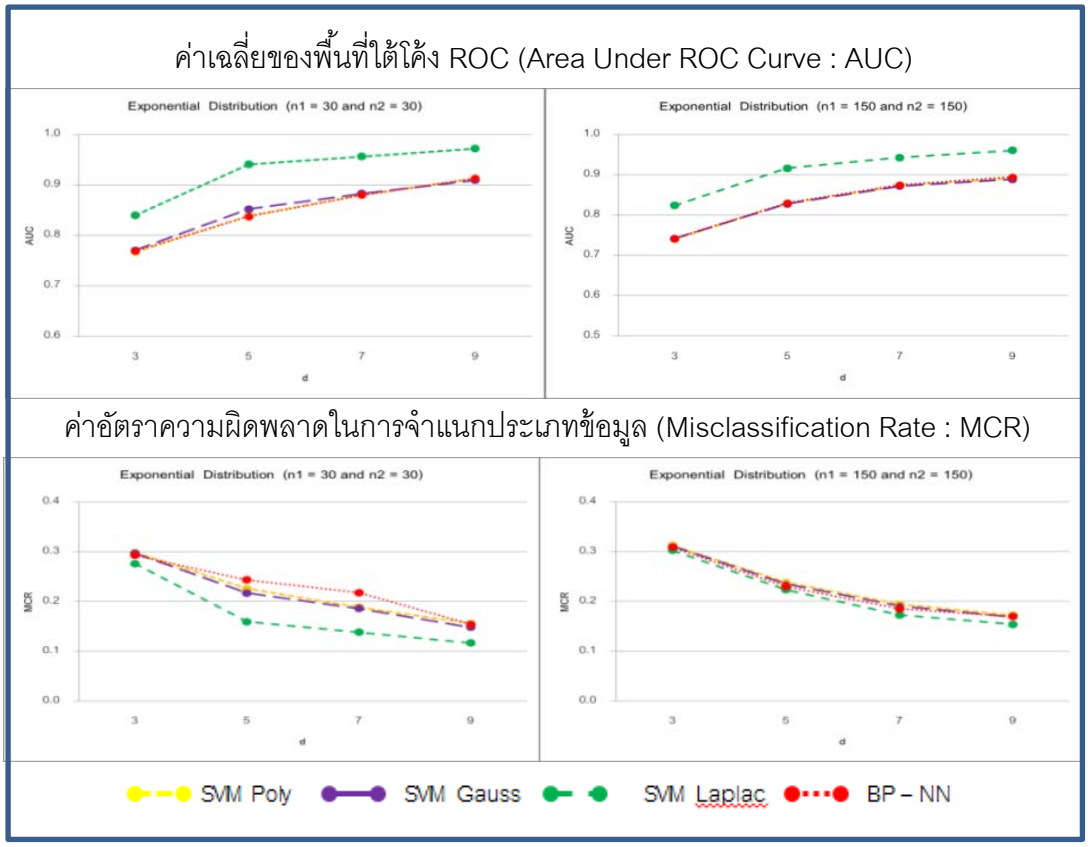
ในการนำเสนอผลการวิจัยจะแสดงในรูปแบบของกราฟ โดยมีสัญลักษณ์ที่ใช้แทนความหมายต่างๆ ดังนี้

n_1	แทน	ขนาดของตัวอย่างของกลุ่มที่ 1
n_2	แทน	ขนาดของตัวอย่างของกลุ่มที่ 2
d	แทน	ค่าของพารามิเตอร์ของการแจกแจงของข้อมูลในตัวอย่างกลุ่มที่ไม่สนใจที่มีการเปลี่ยนแปลงไปเรื่อย ๆ
AUC	แทน	ค่าของพื้นที่ใต้โค้ง ROC (Area Under ROC Curve)
MCR	แทน	ค่าของอัตราความผิดพลาดในการจำแนกประเภทข้อมูล (Misclassification Rate)

4.1 ตัวแปรอิสระ 1 ตัวแปร

4.1.1 การแจกแจงแบบชี้กำลัง (Exponential Distribution)

4.1.1.1. กรณีที่ขนาดตัวอย่างของแต่ละกลุ่มเท่ากัน



ภาพที่ 4.1 แสดงการเปรียบเทียบค่าเฉลี่ยของ AUC และค่าเฉลี่ยของ MCR ของข้อมูลที่มีการแจกแจงแบบชี้กำลัง กรณีข้อมูลมีขนาดตัวอย่างเท่ากัน

จากผลการศึกษาตามภาพที่ 4.1 เป็นการนำเสนอผลลัพธ์ของการจำลองข้อมูลที่มีขนาดตัวอย่างเล็กสุด ซึ่งเป็นภาพทางด้านซ้าย คือ $n_1 = 30$ กับ $n_2 = 30$ และการจำลองข้อมูลที่มีขนาดตัวอย่างใหญ่สุด ซึ่งเป็นภาพทางด้านขวา คือ $n_1 = 150$ กับ $n_2 = 150$ โดยทำการศึกษาจากค่าของพารามิเตอร์ของการแจกแจงของข้อมูล (d) ได้ผลลัพธ์ของการศึกษาดังนี้

ค่าเฉลี่ยของ AUC จากวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุดในทุกกรณีที่ขนาดของกลุ่มตัวอย่างที่มีขนาดเท่ากันและในทุกระดับของค่าพารามิเตอร์ d

เมื่อขนาดของกลุ่มตัวอย่างเพิ่มมากขึ้นวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับและวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนลอื่นๆ ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลได้ใกล้เคียงกันในทุกระดับของค่าพารามิเตอร์ d

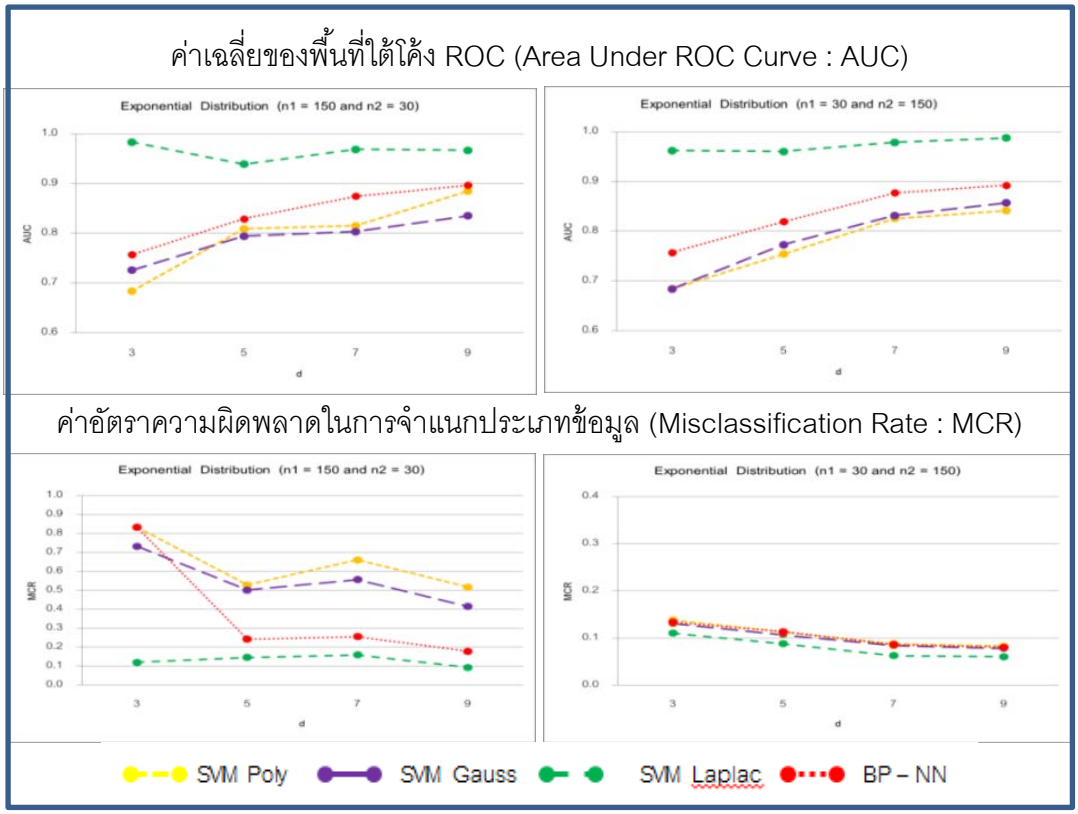
ค่าเฉลี่ยของ MCR จากวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุดในทุกกรณีที่ขนาดของกลุ่มตัวอย่างที่มีขนาดเท่ากันและในทุกระดับของค่าพารามิเตอร์ d

เมื่อขนาดของกลุ่มตัวอย่างเพิ่มมากขึ้นวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับและวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนลอื่นๆ ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ใกล้เคียงกันในทุกระดับของค่าพารามิเตอร์ d

4.1.1.2. กรณีที่ขนาดตัวอย่างของกลุ่มตัวอย่างที่สนใจ**มีความแตกต่างกับ**กลุ่มตัวอย่างที่ไม่สนใจ

ผลลัพธ์ของการศึกษาประสิทธิภาพของการจำแนกประเภทมีลักษณะความสามารถในการจำแนกประเภทไปในทิศทางเดียวกัน จึงนำเสนอผลลัพธ์ของข้อมูล ที่มีลักษณะขนาดของตัวอย่างที่มีความแตกต่างกันของข้อมูลมากที่สุดกับลักษณะขนาดของตัวอย่างที่มีความแตกต่างกันของข้อมูลน้อยที่สุด

เมื่อขนาดตัวอย่างของกลุ่มที่สนใจมีขนาดตัวอย่าง**แตกต่าง**กับกลุ่มที่ไม่สนใจ อยู่ 120 ตัวอย่าง



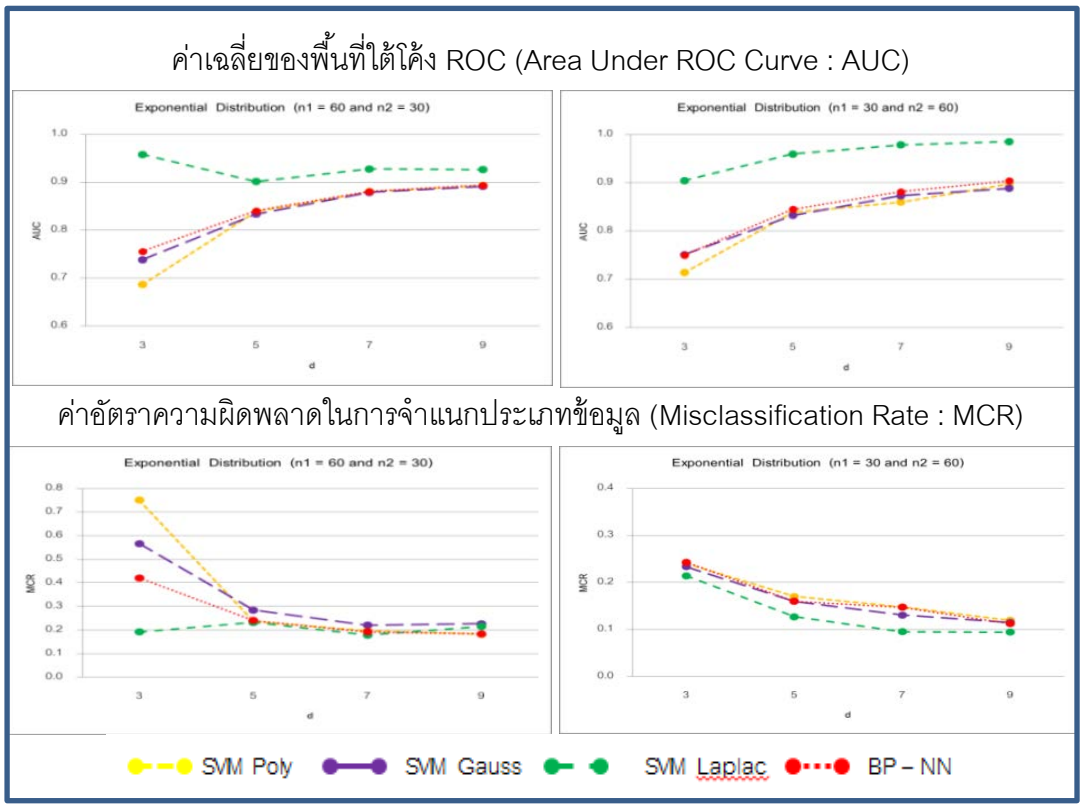
ภาพที่ 4.2 แสดงการเปรียบเทียบค่าเฉลี่ยของ AUC และค่าเฉลี่ยของ MCR ของข้อมูลที่มีการแจกแจงแบบชี้กำลัง กรณีขนาดตัวอย่างของกลุ่มตัวอย่างที่สนใจ**แตกต่าง**กับกลุ่มตัวอย่างที่ไม่สนใจ อยู่ 120 ตัวอย่าง

จากผลการศึกษาตามภาพที่ 4.2 เป็นการนำเสนอผลลัพธ์ของการจำลองข้อมูลที่มีผลต่างของขนาดตัวอย่างกลุ่มที่สนใจกับกลุ่มที่ไม่สนใจ ที่มีขนาดตัวอย่างมีความแตกต่างกันมากที่สุดตามที่ได้ทำการศึกษา ซึ่งภาพทางด้านซ้ายมือ คือ $n_1 = 150$ กับ $n_2 = 30$ และภาพทางด้านขวามือ คือ $n_1 = 30$ กับ $n_2 = 150$ โดยทำการศึกษาจากค่าของพารามิเตอร์ของการแจกแจงของข้อมูล (d) ได้ผลลัพธ์ของการศึกษาดังนี้

ค่าเฉลี่ยของ AUC จากวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุดในทุกระดับของค่าพารามิเตอร์ d โดยที่วิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ มีประสิทธิภาพเป็นอันดับสองในทุกระดับของค่าพารามิเตอร์ d

ค่าเฉลี่ยของ MCR ของกลุ่มตัวอย่างกลุ่มที่ไม่สนใจมีขนาดเล็ก จะได้ว่า วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุดในทุกระดับของค่าพารามิเตอร์ d และค่าเฉลี่ยของ MCR ของกลุ่มตัวอย่างกลุ่มที่ไม่สนใจมีขนาดใหญ่ จะได้ว่า วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุดในทุกระดับของค่าพารามิเตอร์ d และวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับและวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนลอื่น ๆ จะให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ใกล้เคียงกัน

เมื่อขนาดตัวอย่างของกลุ่มที่สนใจมีขนาดตัวอย่าง**แตกต่าง**กับกลุ่มที่ไม่สนใจ อยู่ 30 ตัวอย่าง



ภาพที่ 4.3 แสดงการเปรียบเทียบค่าเฉลี่ยของ AUC และค่าเฉลี่ยของ MCR ของข้อมูลที่มีการแจกแจงแบบชี้กำลัง กรณีขนาดตัวอย่างของกลุ่มตัวอย่างที่สนใจ**แตกต่าง**กับกลุ่มตัวอย่างที่ไม่สนใจอยู่ 30 ตัวอย่าง

จากผลการศึกษาตามภาพที่ 4.3 เป็นการนำเสนอผลลัพธ์ของการจำลองข้อมูลที่มีผลต่างของขนาดตัวอย่างกลุ่มที่สนใจกับกลุ่มที่ไม่สนใจ ที่มีขนาดตัวอย่างมีความแตกต่างกันน้อยที่สุด

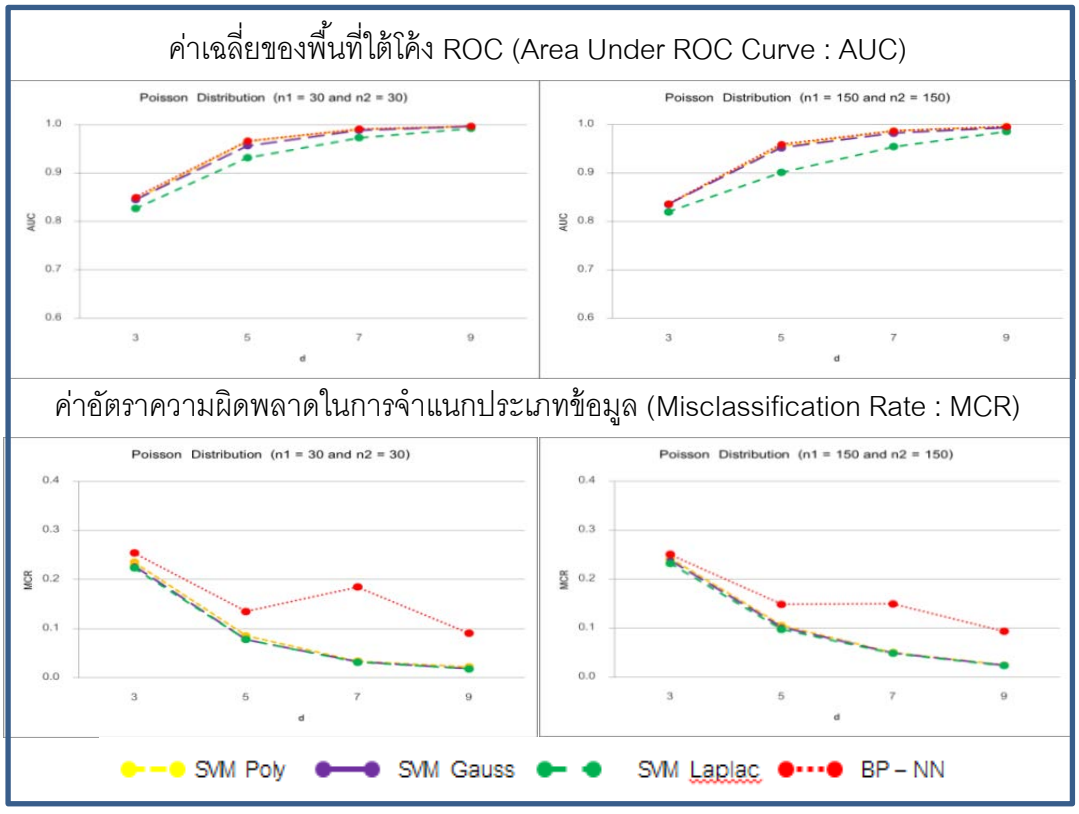
ตามที่ได้ทำการศึกษา ซึ่งภาพทางด้านซ้ายมือ คือ $n_1 = 60$ กับ $n_2 = 30$ และการจำลองข้อมูลที่มีขนาดตัวอย่างใหญ่สุด เป็นภาพทางด้านขวามือ คือ $n_1 = 30$ กับ $n_2 = 60$ โดยทำการศึกษาจากค่าของพารามิเตอร์ของการแจกแจงของข้อมูล (d) ได้ผลลัพธ์ของการศึกษาดังนี้

ค่าเฉลี่ยของ AUC จากวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุดในทุกระดับของค่าพารามิเตอร์

ค่าเฉลี่ยของ MCR จะได้ว่า วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุดในทุกระดับของค่าพารามิเตอร์

4.1.2 การแจกแจงแบบปัวส์ซง(Poisson Distribution)

4.1.2.1. กรณีที่ขนาดตัวอย่างของแต่ละกลุ่มเท่ากัน



ภาพที่ 4.4 แสดงการเปรียบเทียบค่าเฉลี่ยของ AUC และค่าเฉลี่ยของ MCR ของข้อมูลที่มีการแจกแจงแบบปัวส์ซง กรณีข้อมูลมีขนาดตัวอย่างเท่ากัน

จากผลการศึกษาตามภาพที่ 4.4 เป็นการนำเสนอผลลัพธ์ของการจำลองข้อมูลที่มีขนาดตัวอย่างเล็กสุด ซึ่งเป็นภาพทางด้านซ้าย คือ $n_1 = 30$ กับ $n_2 = 30$ และการจำลองข้อมูลที่มีขนาดตัวอย่างใหญ่สุด ซึ่งเป็นภาพทางด้านขวา คือ $n_1 = 150$ กับ $n_2 = 150$ โดยทำการศึกษาจากค่าของพารามิเตอร์ของการแจกแจงของข้อมูล (d) ได้ผลลัพธ์ของการศึกษาดังนี้

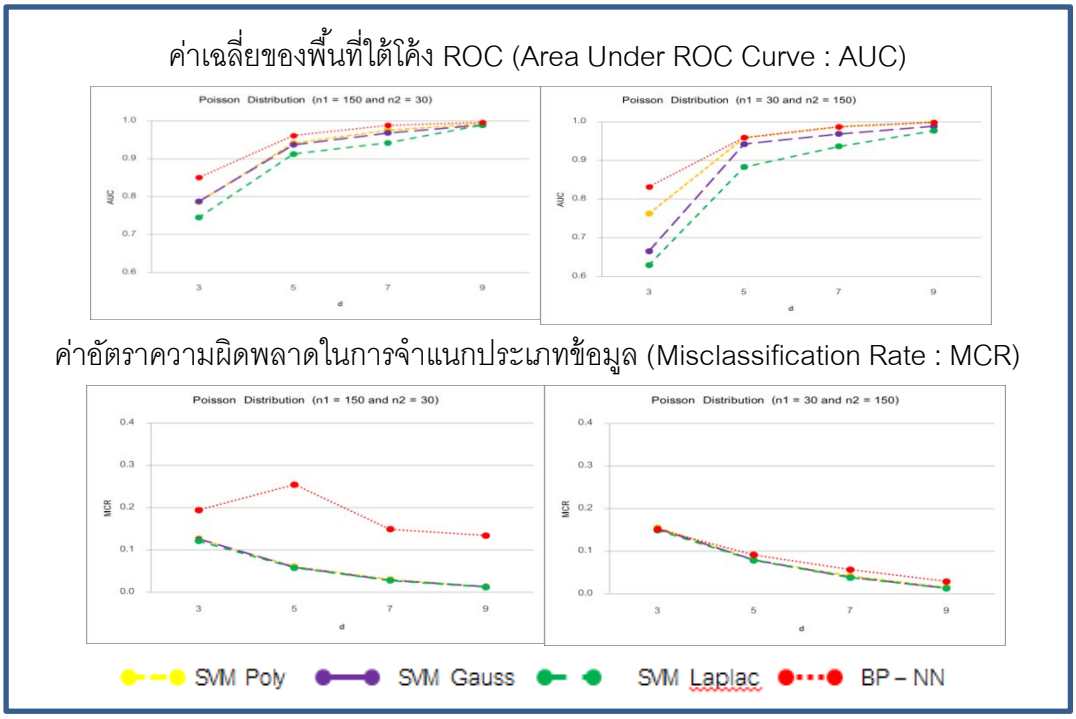
ค่าเฉลี่ยของ AUC จากวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับและวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Polynomial Kernel และ Gaussian Kernel ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลได้ดีใกล้เคียงกันในทุกกรณีที่ขนาดของกลุ่มตัวอย่างที่มีขนาดเท่ากันและในทุกระดับของค่าพารามิเตอร์ d

ค่าเฉลี่ยของ MCR จากวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนลทุกฟังก์ชันให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุดใกล้เคียงกัน ในทุกกรณีที่ขนาดของกลุ่มตัวอย่างที่มีขนาดเท่ากันและในทุกระดับของค่าพารามิเตอร์ d แต่วิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้แย่ที่สุด

4.1.2.2. กรณีที่ขนาดตัวอย่างของกลุ่มตัวอย่างที่สนใจ**มีความแตกต่างกับ**กลุ่มตัวอย่างที่ไม่สนใจ

ผลลัพธ์ของการศึกษาประสิทธิภาพของการจำแนกประเภทมีลักษณะความสามารถในการจำแนกประเภทไปในทิศทางเดียวกัน จึงนำเสนอผลลัพธ์ของข้อมูล ที่มีลักษณะขนาดของตัวอย่างที่มีความแตกต่างกันของข้อมูลมากที่สุดกับลักษณะขนาดของตัวอย่างที่มีความแตกต่างกันของข้อมูลน้อยที่สุด

เมื่อขนาดตัวอย่างของกลุ่มที่สนใจมีขนาดตัวอย่าง**แตกต่าง**กับกลุ่มที่ไม่สนใจ อยู่ 120 ตัวอย่าง



ภาพที่ 4.5 แสดงการเปรียบเทียบค่าเฉลี่ยของ AUC และค่าเฉลี่ยของ MCR ของข้อมูลที่มีการแจกแจงแบบปัวส์ซง กรณีขนาดตัวอย่างของกลุ่มตัวอย่างที่สนใจ**แตกต่าง**กับกลุ่มตัวอย่างที่ไม่สนใจ อยู่ 120 ตัวอย่าง

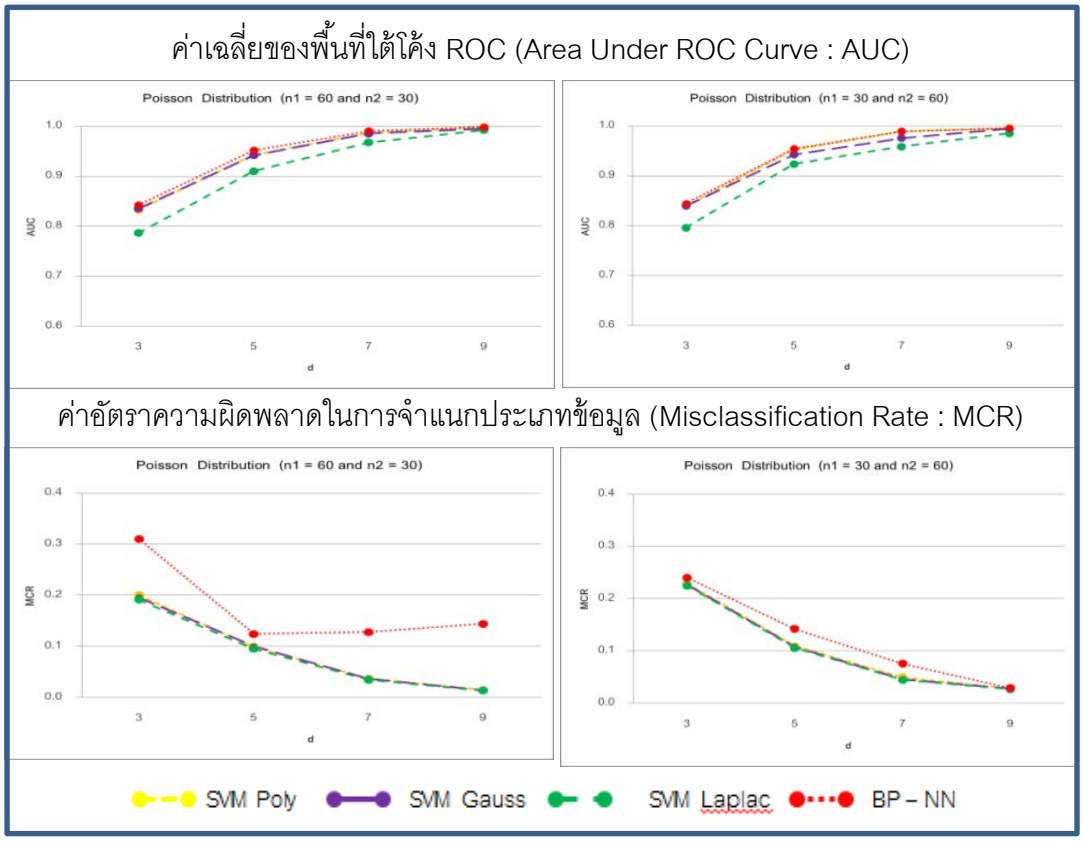
จากผลการศึกษาตามภาพที่ 4.5 เป็นการนำเสนอผลลัพธ์ของการจำลองข้อมูลที่มีผลต่างของขนาดตัวอย่างกลุ่มที่สนใจกับกลุ่มที่ไม่สนใจ ที่มีขนาดตัวอย่างมีความแตกต่างกันมากที่สุดตามที่ได้ทำการศึกษา ซึ่งภาพทางด้านซ้ายมือ คือ $n_1 = 150$ กับ $n_2 = 30$ และภาพทางด้านขวามือ คือ $n_1 = 30$ กับ $n_2 = 150$ โดยทำการศึกษาจากค่าของพารามิเตอร์ของการแจกแจงของข้อมูล (d) ได้ผลลัพธ์ของการศึกษาดังนี้

ค่าเฉลี่ยของ AUC จากวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุดในทุกระดับของค่าพารามิเตอร์ d และวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Polynomial Kernel และ Gaussian Kernel จะให้ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลได้ดีที่สุดใกล้เคียงกัน

ค่าเฉลี่ยของ MCR จากวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนลทุกฟังก์ชันให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุดใกล้เคียงกันในทุกระดับของ

ค่าพารามิเตอร์ d และวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้แย่ที่สุด

เมื่อขนาดตัวอย่างของกลุ่มที่สนใจมีขนาดตัวอย่าง**แตกต่าง**กับกลุ่มที่ไม่สนใจ อยู่ 30 ตัวอย่าง



ภาพที่ 4.6 แสดงการเปรียบเทียบค่าเฉลี่ยของ AUC และค่าเฉลี่ยของ MCR ของข้อมูลที่มีการแจกแจงแบบปัวส์ซง กรณีขนาดตัวอย่างของกลุ่มตัวอย่างที่สนใจ**แตกต่าง**กับกลุ่มตัวอย่างที่ไม่สนใจ อยู่ 30 ตัวอย่าง

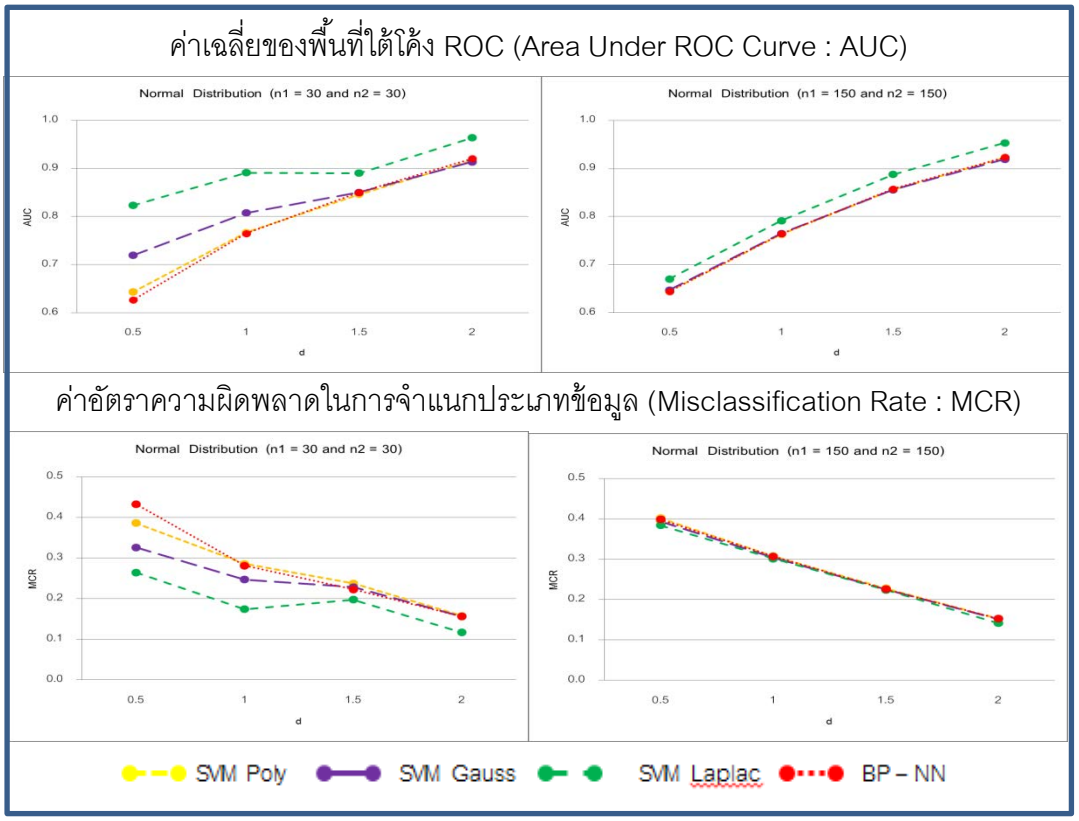
จากผลการศึกษาตามภาพที่ 4.6 เป็นการนำเสนอผลลัพธ์ของการจำลองข้อมูลที่มีผลต่างของขนาดตัวอย่างกลุ่มที่สนใจกับกลุ่มที่ไม่สนใจ ที่มีขนาดตัวอย่างมีความแตกต่างกันน้อยที่สุดตามที่ได้ทำการศึกษา ซึ่งภาพทางด้านซ้ายมือ คือ $n_1 = 60$ กับ $n_2 = 30$ และภาพทางด้านขวามือ คือ $n_1 = 30$ กับ $n_2 = 60$ โดยทำการศึกษาจากค่าของพารามิเตอร์ของการแจกแจงของข้อมูล (d) ได้ผลลัพธ์ของการศึกษาดังนี้

ค่าเฉลี่ยของ AUC จากวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุด และวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Polynomial Kernel และ Gaussian Kernel มีประสิทธิภาพใกล้เคียงกัน

ค่าเฉลี่ยของ MCR จากวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนลทุกฟังก์ชันให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุดใกล้เคียงกันในทุกระดับของค่าพารามิเตอร์ d และวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้แย่ที่สุด

4.1.3 การแจกแจงแบบปกติ (Normal Distribution)

4.1.3.1. กรณีที่ขนาดตัวอย่างของแต่ละกลุ่มเท่ากัน



ภาพที่ 4.7 แสดงการเปรียบเทียบค่าเฉลี่ยของ AUC ของข้อมูลที่มีการแจกแจงแบบปกติ กรณีข้อมูลมีขนาดตัวอย่างเท่ากัน

จากผลการศึกษาตามภาพที่ 4.7 เป็นการนำเสนอผลลัพธ์ของการจำลองข้อมูลที่มีขนาดตัวอย่างเล็กสุด เป็นภาพทางด้านซ้ายมือ คือ $n_1 = 30$ กับ $n_2 = 30$ และการจำลองข้อมูลที่มีขนาดตัวอย่างใหญ่สุด เป็นภาพทางด้านขวามือ คือ $n_1 = 150$ กับ $n_2 = 150$ โดยทำการศึกษาจากค่าของพารามิเตอร์ของการแจกแจงของข้อมูล (d) ได้ผลลัพธ์ของการศึกษาดังนี้

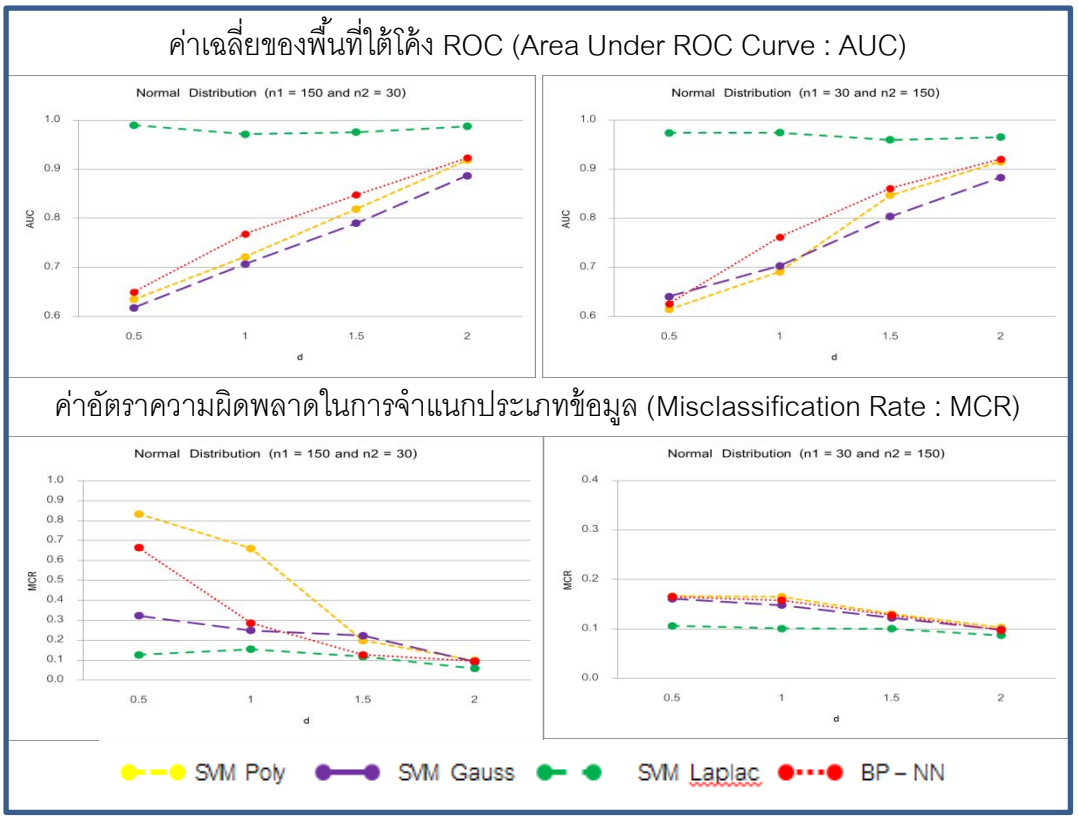
ค่าเฉลี่ยของ AUC จากวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุดในทุกกรณีที่ขนาดของกลุ่มตัวอย่างที่มีขนาดเท่ากันและในทุกระดับของ ค่าพารามิเตอร์ d

ค่าเฉลี่ยของ MCR จากวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุดในทุกกรณีที่ขนาดของกลุ่มตัวอย่างที่มีขนาดเท่ากันและในทุกระดับของค่าพารามิเตอร์ d และเมื่อขนาดของข้อมูลเพิ่มมากขึ้น วิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับและวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนลจะให้ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ใกล้เคียงกันในทุกระดับของค่าพารามิเตอร์ d

4.1.3.2. กรณีที่ขนาดตัวอย่างของกลุ่มตัวอย่างที่สนใจ**มีความแตกต่างกับ**กลุ่มตัวอย่างที่ไม่สนใจ

ผลลัพธ์ของการศึกษาประสิทธิภาพของการจำแนกประเภทมีลักษณะความสามารถในการจำแนกประเภทไปในทิศทางเดียวกัน จึงนำเสนอผลลัพธ์ของข้อมูล ที่มีลักษณะขนาดของตัวอย่างที่มีความแตกต่างกันของข้อมูลมากที่สุดกับลักษณะขนาดของตัวอย่างที่มีความแตกต่างกันของข้อมูลน้อยที่สุด

เมื่อขนาดตัวอย่างของกลุ่มที่สนใจมีขนาดตัวอย่าง**แตกต่าง**กับกลุ่มที่ไม่สนใจอยู่ 120 ตัวอย่าง



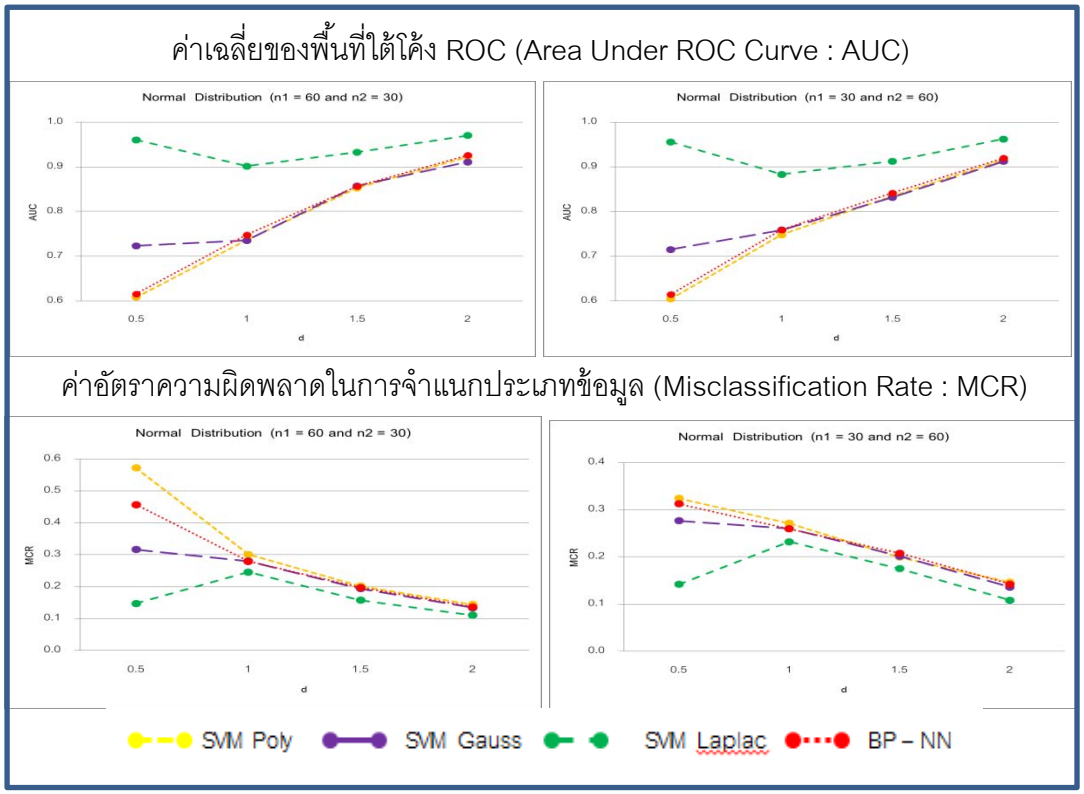
<u>ภาพที่ 4.8</u> แสดงการเปรียบเทียบค่าเฉลี่ยของ AUC และค่าเฉลี่ยของ MCR ของข้อมูลที่มีการแจกแจงแบบปกติ กรณีขนาดตัวอย่างของกลุ่มตัวอย่างที่สนใจ**แตกต่าง**กับกลุ่มตัวอย่างที่ไม่สนใจ อยู่ 120 ตัวอย่าง

จากผลการศึกษาตามรูปที่ 4.8 เป็นการนำเสนอผลลัพธ์ของการจำลองข้อมูลที่มีผลต่างของขนาดตัวอย่างกลุ่มที่สนใจกับกลุ่มที่ไม่สนใจ ที่มีขนาดตัวอย่างมีความแตกต่างกันมากที่สุดตามที่ได้ทำการศึกษา ซึ่งภาพทางด้านซ้ายมือ คือ $n_1 = 150$ กับ $n_2 = 30$ และภาพทางด้านขวามือ คือ $n_1 = 30$ กับ $n_2 = 150$ โดยทำการศึกษาจากค่าของพารามิเตอร์ของการแจกแจงของข้อมูล (d) ได้ผลลัพธ์ของการศึกษาดังนี้

ค่าเฉลี่ยของ AUC จากวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุดในทุกระดับของค่าพารามิเตอร์ d

ค่าเฉลี่ยของ MCR จากวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุดในทุกกรณีที่ขนาดของกลุ่มตัวอย่างที่มีขนาดเท่ากันและในทุกระดับของค่าพารามิเตอร์ d

เมื่อขนาดตัวอย่างของกลุ่มที่สนใจมีขนาดตัวอย่าง**แตกต่าง**กับกลุ่มที่ไม่สนใจ อยู่ 30 ตัวอย่าง



ภาพที่ 4.9 แสดงการเปรียบเทียบค่าเฉลี่ยของ AUC และค่าเฉลี่ยของ MCR ของข้อมูลที่มีการแจกแจงแบบปกติ กรณีขนาดตัวอย่างของกลุ่มตัวอย่างที่สนใจ**แตกต่าง**กับกลุ่มตัวอย่างที่สนใจ อยู่ 30 ตัวอย่าง

จากผลการศึกษาตามภาพที่ 4.9 เป็นการนำเสนอผลลัพธ์ของการจำลองข้อมูลที่มีผลต่างของขนาดตัวอย่างกลุ่มที่สนใจกับกลุ่มที่ไม่สนใจ ที่มีขนาดตัวอย่างมีความแตกต่างกันน้อยที่สุดตามที่ได้ทำการศึกษา ซึ่งภาพทางด้านซ้ายมือ คือ $n_1 = 60$ กับ $n_2 = 30$ และภาพทางด้านขวามือ คือ $n_1 = 30$ กับ $n_2 = 60$ โดยทำการศึกษาจากค่าของพารามิเตอร์ของการแจกแจงของข้อมูล (d) ได้ผลลัพธ์ของการศึกษาดังนี้

ค่าเฉลี่ยของ AUC จากวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุดในทุกระดับของค่าพารามิเตอร์ d

ค่าเฉลี่ยของ MCR ของกลุ่มตัวอย่างกลุ่มที่สนใจมีขนาดเล็ก และเมื่อระดับของค่าพารามิเตอร์ d มีค่าน้อย จะได้วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุด โดยเมื่อค่า d มีค่าเพิ่มมากขึ้นจะให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ใกล้เคียงกันทุกวิธีการ

ตารางที่ 4.1 แสดงค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของการจำลองข้อมูลกรณีที่มีตัวแปรอิสระ 1 ตัวแปร โดยจำแนกตามลักษณะของการแจกแจงของข้อมูล

	Exponential				Poisson				Normal			
d	3	5	7	9	3	5	7	9	0.5	1	1.5	2
SVM P	0.74	0.83	0.87	0.90	0.83	0.95	0.99	1.00	0.63	0.75	0.85	0.92
SVM G	0.75	0.82	0.87	0.89	0.82	0.95	0.98	0.99	0.66	0.76	0.85	0.91
SVM L	0.87	0.93	0.95	0.97	0.79	0.91	0.96	0.99	0.81	0.85	0.91	0.96
BP	0.75	0.83	0.88	0.90	0.84	0.96	0.99	1.00	0.64	0.76	0.86	0.92

ผลจากตารางจะได้ว่า ค่าเฉลี่ยของ AUC ในทุกการจำลองข้อมูลนั้น วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุด สำหรับกรณีที่ข้อมูลมีการแจกแจงแบบชี้กำลังและกรณีที่ข้อมูลที่มีการแจกแจงแบบปกติ โดยที่วิธีโครงข่ายประสาทเทียมแบบย้อนกลับให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุด ในกรณีที่ข้อมูลมีการแจกแจงแบบปัวส์ซง

ตารางที่ 4.2 แสดงค่าเฉลี่ยอัตราความผิดพลาดในการจำแนกประเภทข้อมูลของการจำลองข้อมูลกรณีที่มีตัวแปรอิสระ 1 ตัวแปร โดยจำแนกตามลักษณะของการแจกแจงของข้อมูล

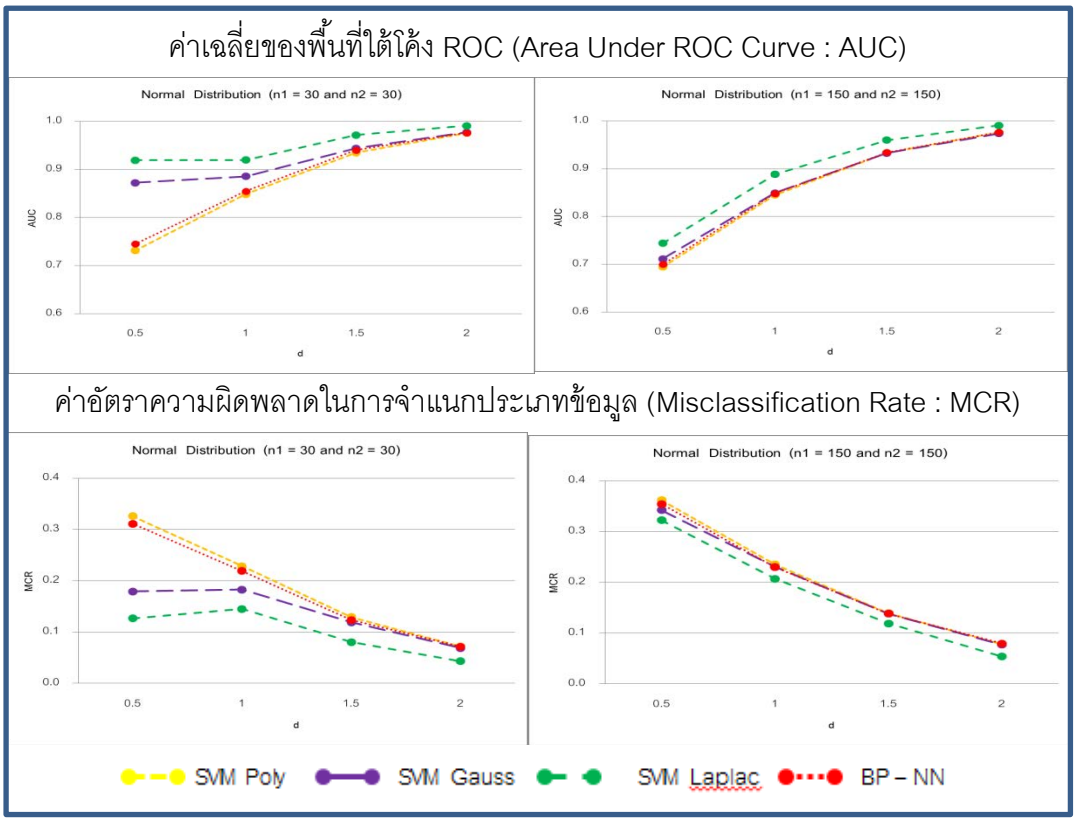
	Exponential				Poisson				Normal			
d	3	5	7	9	3	5	7	9	0.5	1	1.5	2
SVM P	0.35	0.24	0.20	0.17	0.22	0.10	0.04	0.02	0.42	0.30	0.21	0.15
SVM G	0.33	0.23	0.20	0.16	0.22	0.10	0.04	0.02	0.35	0.28	0.21	0.15
SVM L	0.25	0.19	0.16	0.13	0.21	0.09	0.04	0.02	0.29	0.25	0.19	0.13
BP	0.31	0.22	0.18	0.15	0.25	0.15	0.12	0.12	0.39	0.28	0.21	0.15

ผลจากตารางจะได้ว่า ค่าเฉลี่ยอัตราความผิดพลาดในการจำแนกประเภทข้อมูล ในทุกการจำลองข้อมูลนั้น วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุด

4.2 ตัวแปรอิสระ 2 ตัวแปร

4.2.1 การแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) ระดับความสัมพันธ์ของตัวแปรอิสระ คือ $cor(x_1, x_2) = 0$

4.2.1.1. กรณีที่ขนาดตัวอย่างของแต่ละกลุ่มเท่ากัน



ภาพที่ 4.10 แสดงการเปรียบเทียบค่าเฉลี่ยของ AUC และค่าเฉลี่ยของ MCR ของข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปร ที่มีระดับความสัมพันธ์ของตัวแปรอิสระเป็นศูนย์ กรณีข้อมูลมีขนาดตัวอย่างเท่ากัน

จากผลการศึกษาตามภาพที่ 4.10 เป็นการนำเสนอผลลัพธ์ของการจำลองข้อมูลที่มีขนาดตัวอย่างเล็กสุด เป็นภาพทางด้านซ้ายมือ คือ $n_1 = 30$ กับ $n_2 = 30$ และการจำลองข้อมูลที่มีขนาดตัวอย่างใหญ่สุด เป็นภาพทางด้านขวามือ คือ $n_1 = 150$ กับ $n_2 = 150$ โดยทำการศึกษาจากค่าของพารามิเตอร์ของการแจกแจงของข้อมูล (d) ได้ผลลัพธ์ของการศึกษาดังนี้

ค่าเฉลี่ยของ AUC จากวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุดในทุกกรณีที่ขนาดของกลุ่มตัวอย่างที่มีขนาดเท่ากันและในทุกระดับของค่าพารามิเตอร์ d เมื่อขนาดของกลุ่มตัวอย่างเพิ่มมากขึ้นวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับและวิธีซัพพอร์ตเวกเตอร์แมชชีน

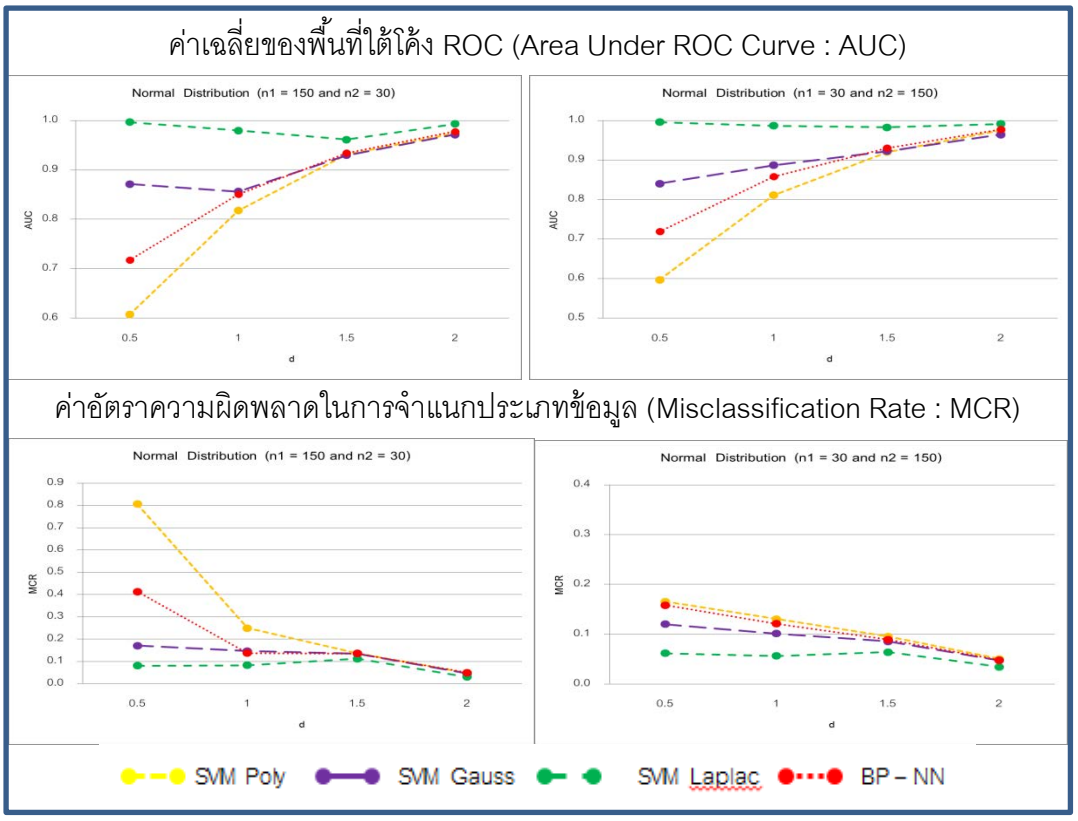
ด้วยฟังก์ชันเคอร์เนลอื่น ๆ ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลได้ใกล้เคียงกัน

ค่าเฉลี่ยของ MCR จากวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุดในทุกกรณีที่ขนาดของกลุ่มตัวอย่างที่มีขนาดเท่ากันและในทุกระดับของค่าพารามิเตอร์ d ในกรณีที่ขนาดของข้อมูลมีขนาดใหญ่ และในทุกระดับของค่าพารามิเตอร์ d วิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับและวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนลอื่น ๆ ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ใกล้เคียงกัน

4.2.1.2. กรณีที่ขนาดตัวอย่างของกลุ่มตัวอย่างที่สนใจ**มีความแตกต่างกับ**กลุ่มตัวอย่างที่ไม่สนใจ

ผลลัพธ์ของการศึกษาประสิทธิภาพของการจำแนกประเภทมีลักษณะความสามารถในการจำแนกประเภทไปในทิศทางเดียวกัน จึงนำเสนอผลลัพธ์ของข้อมูล ที่มีลักษณะขนาดของตัวอย่างที่มีความแตกต่างกันของข้อมูลมากที่สุดกับลักษณะขนาดของตัวอย่างที่มีความแตกต่างกันของข้อมูลน้อยที่สุด

เมื่อขนาดตัวอย่างของกลุ่มที่สนใจมีขนาดตัวอย่าง**แตกต่าง**กับกลุ่มที่ไม่สนใจ อยู่ 120 ตัวอย่าง



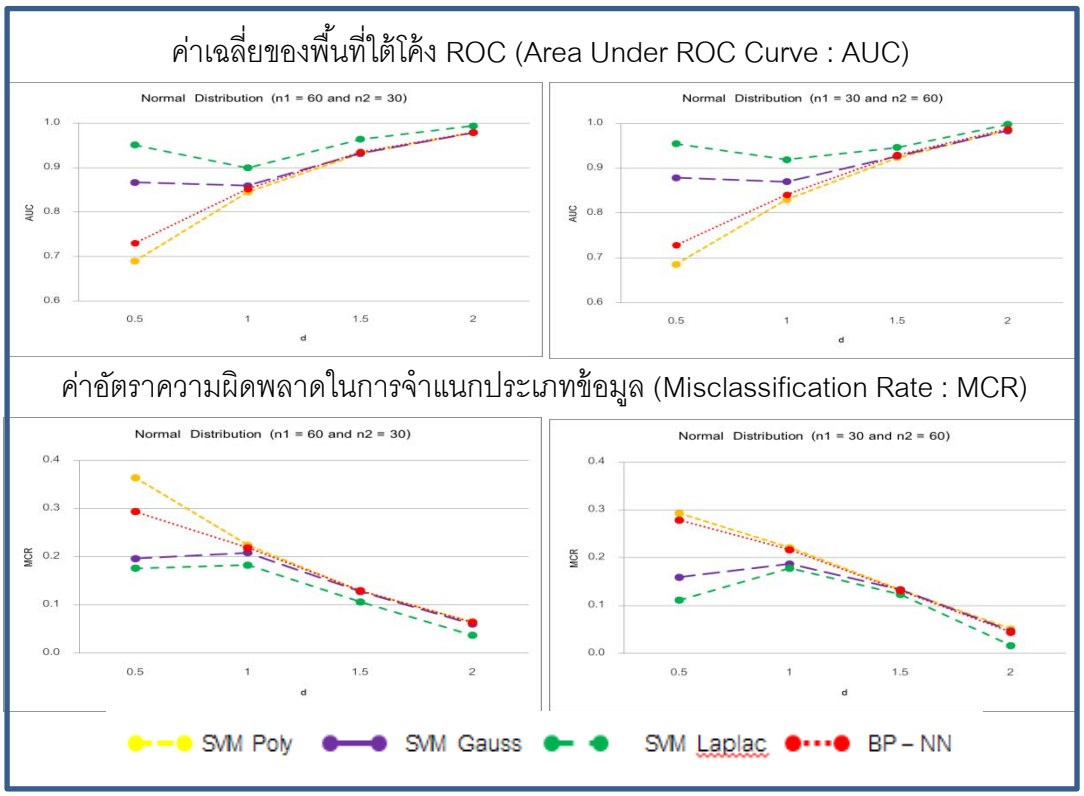
<u>ภาพที่ 4.11</u> แสดงการเปรียบเทียบค่าเฉลี่ยของ AUC และค่าเฉลี่ยของ MCR ของข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปร ที่มีระดับความสัมพันธ์ของตัวแปรอิสระเป็นศูนย์ กรณีขนาดตัวอย่างของกลุ่มตัวอย่างที่สนใจ**แตกต่าง**กับกลุ่มตัวอย่างที่ไม่สนใจ อยู่ 120 ตัวอย่าง

จากผลการศึกษาตามภาพที่ 4.11 เป็นการนำเสนอผลลัพธ์ของการจำลองข้อมูลที่มีผลต่างของขนาดตัวอย่างกลุ่มที่สนใจกับกลุ่มที่ไม่สนใจ ที่มีขนาดตัวอย่างมีความแตกต่างกันมากที่สุด ตามที่ได้ทำการศึกษา ซึ่งภาพทางด้านซ้ายมือ คือ $n_1 = 150$ กับ $n_2 = 30$ และภาพทางด้านขวามือ คือ $n_1 = 30$ กับ $n_2 = 150$ โดยทำการศึกษาจากค่าของพารามิเตอร์ของการแจกแจงของข้อมูล (d) ได้ผลลัพธ์ของการศึกษาดังนี้

ค่าเฉลี่ยของ AUC จากวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุดในทุกระดับของค่าพารามิเตอร์ d

ค่าเฉลี่ยของ MCR จากวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุดในทุกระดับของค่าพารามิเตอร์ d และวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ กับวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนลอื่นๆ ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ใกล้เคียงกันเมื่อระดับของค่าพารามิเตอร์ d มีค่าเพิ่มมากขึ้น

เมื่อขนาดตัวอย่างของกลุ่มที่สนใจมีขนาดตัวอย่าง**แตกต่าง**กับกลุ่มที่ไม่สนใจ อยู่ 30 ตัวอย่าง



ภาพที่ 4.12 แสดงการเปรียบเทียบค่าเฉลี่ยของ AUC และค่าเฉลี่ยของ MCR ของข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปร ที่มีระดับความสัมพันธ์ของตัวแปรอิสระเป็นศูนย์ กรณีขนาดตัวอย่างของกลุ่มตัวอย่างที่สนใจ**แตกต่าง**กับกลุ่มตัวอย่างที่ไม่สนจอยู่ 30 ตัวอย่าง

จากผลการศึกษาตามภาพที่ 4.12 เป็นการนำเสนอผลลัพธ์ของการจำลองข้อมูลที่มีผลต่างของขนาดตัวอย่างกลุ่มที่สนใจกับกลุ่มที่ไม่สนใจ ที่มีขนาดตัวอย่างมีความแตกต่างกันน้อยที่สุด ตามที่ได้ทำการศึกษา ซึ่งภาพทางด้านซ้ายมือ คือ $n_1 = 150$ กับ $n_2 = 30$ และภาพทางด้าน

ขวามือ คือ $n_1 = 30$ กับ $n_2 = 150$ โดยทำการศึกษาจากค่าของพารามิเตอร์ของการแจกแจงของข้อมูล (d) ได้ผลลัพธ์ของการศึกษาดังนี้

ค่าเฉลี่ยของ AUC จากวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุดในทุกระดับของค่าพารามิเตอร์ d

ค่าเฉลี่ยของ MCR จากวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุดในทุกระดับของค่าพารามิเตอร์ d และวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ กับวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนลอื่นๆ ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ใกล้เคียงกันเมื่อระดับของค่าพารามิเตอร์ d มีค่าเพิ่มมากขึ้น

จากการจำลองข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) โดยมีการจำลองข้อมูลให้มีการระดับความสัมพันธ์ของตัวแปรอิสระ $cor(x_1, x_2) = 0.3, 0.5$ และ 0.9 นั้นให้ผลลัพธ์ของการศึกษาให้ผลคล้ายคลึงกับการจำลองข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปรที่ไม่มีความสัมพันธ์ของตัวแปรอิสระ หรือ $cor(x_1, x_2) = 0$ แต่จะมีค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC และค่าเฉลี่ยอัตราความผิดพลาดในการจำแนกประเภทข้อมูลลดน้อยลงไปบางเล็กน้อย สามารถพิจารณาได้จากตารางที่ 4.3

ตารางที่ 4.3 แสดงค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของการจำลองข้อมูลกรณีที่มีการแจกแจงแบบปกติหลายตัวแปร โดยจำแนกตามระดับความสัมพันธ์ของตัวแปรอิสระ

ρ	$cor(x_1, x_2) = 0$				$cor(x_1, x_2) = 0.3$			
d	0.5	1	1.5	2	0.5	1	1.5	2
SVM Poly	0.687	0.842	0.932	0.977	0.672	0.808	0.907	0.961
SVM Gauss	0.799	0.857	0.933	0.975	0.798	0.834	0.911	0.960
SVM Laplace	0.866	0.906	0.966	0.992	0.866	0.890	0.951	0.984
BP_NN	0.716	0.850	0.936	0.978	0.699	0.820	0.911	0.963
ρ	$cor(x_1, x_2) = 0.5$				$cor(x_1, x_2) = 0.9$			
d	0.5	1	1.5	2	0.5	1	1.5	2
SVM Poly	0.670	0.795	0.897	0.948	0.654	0.772	0.876	0.927
SVM Gauss	0.800	0.822	0.898	0.946	0.753	0.795	0.880	0.925
SVM Laplace	0.852	0.871	0.934	0.972	0.827	0.853	0.919	0.957
BP_NN	0.689	0.805	0.901	0.949	0.669	0.782	0.879	0.929

ผลจากตารางจะได้ว่า ค่าเฉลี่ยของ AUC ในทุกการจำลองข้อมูลนั้น วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุด และเมื่อระดับความสัมพันธ์ของตัวแปรอิสระเพิ่มขึ้น ทุกวิธีการที่ทำการศึกษาจะมีค่าเฉลี่ยของ AUC ลดลงเล็กน้อยในทุกกรณีของ d ที่เปลี่ยนแปลงไป โดยเฉพาะในกรณีที่ $d = 0.5$ และ $d = 1$ จะแสดงผลลัพธ์ของประสิทธิภาพของการพยากรณ์จำแนกประเภทที่แตกต่างกันชัดเจน

ตารางที่ 4.4 แสดงค่าเฉลี่ยอัตราความผิดพลาดในการจำแนกประเภทของการจำลองข้อมูลกรณีที่มีการแจกแจงแบบปกติหลายตัวแปร โดยจำแนกตามระดับความสัมพันธ์ของตัวแปรอิสระ

ρ	$cor(x_1, x_2) = 0$				$cor(x_1, x_2) = 0.3$			
d	0.5	1	1.5	2	0.5	1	1.5	2
SVM Poly	0.360	0.223	0.130	0.070	0.377	0.247	0.157	0.095
SVM Gauss	0.250	0.206	0.126	0.066	0.250	0.220	0.148	0.091
SVM Laplace	0.205	0.177	0.100	0.043	0.203	0.195	0.122	0.067
BP_NN	0.319	0.213	0.128	0.069	0.334	0.234	0.153	0.092
ρ	$cor(x_1, x_2) = 0.5$				$cor(x_1, x_2) = 0.9$			
d	0.5	1	1.5	2	0.5	1	1.5	2
SVM Poly	0.391	0.270	0.171	0.113	0.402	0.290	0.195	0.138
SVM Gauss	0.238	0.230	0.166	0.111	0.279	0.258	0.186	0.134
SVM Laplace	0.179	0.202	0.142	0.087	0.211	0.221	0.165	0.114
BP_NN	0.344	0.248	0.167	0.111	0.361	0.267	0.187	0.135

ผลจากตารางจะได้ว่า ค่าเฉลี่ยของ MCR ในทุกการจำลองข้อมูลนั้น วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุด โดยที่วิธีการอื่น ๆ มีค่าเฉลี่ยของ MCR ใกล้เคียงกันในทุกระดับที่ d มีการเปลี่ยนแปลงไป และจะมีค่าความผิดพลาดเพิ่มมากขึ้น เมื่อระดับความสัมพันธ์ของตัวแปรอิสระมีค่าเพิ่มมากขึ้นในทุกวิธีการที่ศึกษา

แต่จากผลการศึกษาค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลของวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ในกรณีที่ $d = 0.5$ และ $d = 1$ ซึ่งเป็นกรณีที่ให้ผลลัพธ์ของค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลไม่สอดคล้องกับผลลัพธ์ของการศึกษาโดยส่วนใหญ่ ซึ่งจะมีค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลลดลงเมื่อระดับของค่าพารามิเตอร์เพิ่มมากขึ้น โดยอาจมีผลมาจากเกณฑ์ของการจำแนกประเภทที่กำหนดให้นั้นเป็นค่าความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจกับเหตุการณ์ที่ไม่สนใจเท่ากัน และสองกรณีดังกล่าวมีลักษณะการแจกแจงข้อมูลที่แบ่งแยกได้ยาก จึงเกิดความไม่เหมาะสมในกรณีดังกล่าว

ซึ่งทำให้ $d=1$ มีผลลัพธ์ของค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลสูงขึ้นจากกรณีที่ $d=0.5$

ตารางที่ 4.5 แสดงค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC และค่าเฉลี่ยอัตราความผิดพลาดในการจำแนกประเภทข้อมูลของการจำลองข้อมูลกรณีที่มีการแจกแจงแบบปกติหลายตัวแปร ทุกกรณีที่ทำการศึกษา โดยจำแนกตามระดับความสัมพันธ์ของตัวแปรอิสระ (ρ)

ρ		SVM Poly	SVM Gauss	SVM Laplace	BP_NN
$\rho=0$	AUC	0.8558	0.8946	0.9429	0.8696
	MCR	0.1937	0.1530	0.1220	0.1753
$\rho=0.3$	AUC	0.8316	0.8773	0.9347	0.8472
	MCR	0.2173	0.1684	0.1369	0.1949
$\rho=0.5$	AUC	0.8166	0.8709	0.9312	0.8344
	MCR	0.2363	0.1773	0.1412	0.2100
$\rho=0.9$	AUC	0.7948	0.8398	0.9183	0.8110
	MCR	0.2571	0.2052	0.1647	0.2301

ผลจากตารางจะได้ว่า ค่าเฉลี่ยของ AUC ในทุกการจำลองข้อมูลนั้น วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุด และเมื่อระดับความสัมพันธ์ของตัวแปรอิสระเพิ่มขึ้น ทุกวิธีการที่ทำการศึกษาจะมีค่าเฉลี่ยของ AUC ลดลงเล็กน้อยในทุกวิธีการ

ค่าเฉลี่ยของ MCR ในทุกการจำลองข้อมูลนั้น วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุด โดยที่วิธีการอื่น ๆ มีค่าเฉลี่ยของ MCR ใกล้เคียงกัน

บทที่ 5

สรุปผลการวิจัย และข้อเสนอแนะ

การศึกษางานวิจัยในครั้งนี้ มีวัตถุประสงค์เพื่อเปรียบเทียบความแม่นยำในการพยากรณ์จำแนกประเภทระหว่างวิธีโครงข่ายประสาทเทียมกับวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนล โดยทำการจำลองข้อมูลเพื่อศึกษาผลกระทบจากระดับค่าพารามิเตอร์ของการแจกแจงข้อมูล (d), ค่าระดับความสัมพันธ์ของตัวแปรอิสระ (ρ) และขนาดของกลุ่มตัวอย่าง (n_1, n_2) ทำการพิจารณาผลการศึกษาด้วย Receiver Operating Characteristic (ROC) ใช้เป็นเครื่องมือวัดประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูล โดยใช้พื้นที่ใต้โค้ง ROC และใช้อัตราความผิดพลาดในการจำแนก ประเภทข้อมูล (Misclassification Rate : MCR) เพื่อศึกษาว่าวิธีการใดมีความผิดพลาดในการ จำแนกประเภท ซึ่งในงานวิจัยนี้ทำการศึกษาผลของเหตุการณ์เกิดขึ้นสองเหตุการณ์ (dichotomous) สามารถสรุปผลการศึกษาในกรณีต่าง ๆ ได้ดังนี้

5.1 สรุปผลการศึกษา

ผลการศึกษาจากการจำลองข้อมูลตามขอบเขตที่กำหนด สามารถสรุปวิธีที่มีประสิทธิภาพการพยากรณ์จำแนกประเภทที่ดีที่สุดจากผลการศึกษาข้อมูลได้ดังนี้

ตารางที่ 5.1 แสดงค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC และค่าเฉลี่ยอัตราความผิดพลาดในการจำแนกประเภทข้อมูลของการจำลองข้อมูลกรณีที่มีตัวแปรอิสระ 1 ตัวแปร โดยจำแนกตามลักษณะของการแจกแจงของข้อมูล

	การแจกแจงแบบชี้กำลัง	**การแจกแจงแบบปัวส์ซง**	**การแจกแจงแบบปกติ**
AUC	SVM Laplace	BP – NN	SVM Laplace
MCR	SVM Laplace	SVM Laplace	SVM Laplace

ผลจากตารางจะได้ว่า ค่าเฉลี่ยของ AUC ในทุกการจำลองข้อมูลนั้น วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุด สำหรับกรณีที่ข้อมูลมีการแจกแจงแบบชี้กำลังและกรณีที่ข้อมูลที่มีการแจกแจงแบบ

ปกติ โดยที่วิธีโครงข่ายประสาทเทียมแบบย้อนกลับให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุด ในกรณีที่ข้อมูลมีการแจกแจงแบบปัวส์ซง และเมื่อพิจารณาค่าเฉลี่ยของ MCR จะได้ว่า วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุด ในทุกกรณีที่ทำการศึกษา

ตารางที่ 5.2 แสดงค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC และค่าเฉลี่ยอัตราความผิดพลาดในการจำแนกประเภทข้อมูลของการจำลองข้อมูลกรณีที่มีการแจกแจงแบบปกติหลายตัวแปร ทุกกรณีที่ทำการศึกษา โดยจำแนกตามระดับความสัมพันธ์ของตัวแปรอิสระ (ρ)

ρ	0	0.3	0.5	0.9
AUC	SVM Laplace	SVM Laplace	SVM Laplace	SVM Laplace
MCR	SVM Laplace	SVM Laplace	SVM Laplace	SVM Laplace

ผลจากตารางจะได้ว่า ในทุกการจำลองข้อมูลนั้น วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุด และเมื่อพิจารณาค่าเฉลี่ยของ MCR จะได้ว่า วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ค่าอัตราความผิดพลาดในการจำแนกประเภทข้อมูลได้ต่ำที่สุด ในทุกกรณีที่ทำการศึกษา

จากการศึกษา สามารถสรุปผลลัพธ์ทุกการแจกแจงของข้อมูลที่ทำการศึกษาจากผลกระทบต่าง ๆ ตามขอบเขตการศึกษา สามารถสรุปผลการศึกษาข้อมูลได้ดังนี้

5.1.1. ผลกระทบจากระดับค่าพารามิเตอร์ของการแจกแจงข้อมูล (d)

พิจารณาจากผลกระทบที่ได้รับจากค่าพารามิเตอร์ของการแจกแจงข้อมูล (d) เมื่อกำหนดจำนวนขนาดตัวอย่าง จะได้ว่า เมื่อระดับค่าพารามิเตอร์ของการแจกแจงข้อมูล (d) มีค่าเพิ่มมากขึ้น ประสิทธิภาพของการพยากรณ์จำแนกประเภทจะเพิ่มขึ้นในทุกวิธีการที่ทำการศึกษา

5.1.2. ผลกระทบจากจำนวนขนาดตัวอย่าง

5.1.2.1 เมื่อขนาดตัวอย่างในแต่ละกลุ่มเท่ากัน

พิจารณาจากผลกระทบที่ได้รับจากขนาดตัวอย่างที่เท่ากัน เมื่อกำหนดระดับค่าพารามิเตอร์ของการแจกแจงข้อมูลเป็นค่าคงที่ จะได้ว่า ทุกวิธีการที่ทำการศึกษา จะมีประสิทธิภาพของการพยากรณ์จำแนกประเภทจะใกล้เคียงเมื่อขนาดตัวอย่างในแต่ละกลุ่มมีจำนวนขนาดตัวอย่างเท่ากันและจะมีประสิทธิภาพของการพยากรณ์จำแนกประเภทจะใกล้เคียงกันมากยิ่งขึ้นเมื่อข้อมูลมีขนาดใหญ่

5.1.2.2 เมื่อขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน

พิจารณาจากผลกระทบที่ได้รับจากขนาดตัวอย่างที่ไม่เท่ากัน เมื่อกำหนดระดับค่าพารามิเตอร์ของการแจกแจงข้อมูลเป็นค่าคงที่ จะได้ว่า ประสิทธิภาพการพยากรณ์จำแนกประเภทของวิธีซัพพอร์ตเวกเตอร์แมชชีนจะลดลงเป็นอย่างมาก แต่วิธีซัพพอร์ตเวกเตอร์แมชชีนก็ยังคงมีประสิทธิภาพอยู่ ถ้าเลือกใช้ฟังก์ชันเคอร์เนลที่มีลักษณะการแจกแจงข้อมูลที่ใกล้เคียงกับข้อมูลที่ทำการศึกษา ส่วนวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ จะมีประสิทธิภาพของการพยากรณ์จำแนกประเภทเพิ่มมากขึ้น โดยเฉพาะจำนวนขนาดตัวอย่างในแต่ละกลุ่มมีขนาดแตกต่างกันมาก ๆ

5.1.3. ผลกระทบจากค่าระดับความสัมพันธ์ระหว่างตัวแปรอิสระ (ρ)

พิจารณาจากผลกระทบที่ได้รับจากค่าระดับความสัมพันธ์ระหว่างตัวแปรอิสระ เมื่อกำหนดระดับค่าพารามิเตอร์ของการแจกแจงข้อมูลและขนาดตัวอย่าง จะได้ว่า เมื่อระดับค่าความสัมพันธ์ระหว่างตัวแปรอิสระมีค่าเพิ่มมากขึ้น ประสิทธิภาพของการพยากรณ์จำแนกประเภทจะลดลงเล็กน้อยในทุกวิธีการที่ทำการศึกษา ซึ่งอาจกล่าวได้ว่า ระดับความสัมพันธ์ระหว่างตัวแปรอิสระไม่ส่งผลกระทบต่อประสิทธิภาพการพยากรณ์จำแนกประเภทกับวิธีที่ทำการศึกษา

วิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ

ข้อดีและข้อจำกัด (Advantages and Disadvantages)

ข้อดี

1. สามารถทำการวิเคราะห์ข้อมูลขนาดตัวอย่างใหญ่ได้ดี และมีความยืดหยุ่นในการสร้างรูปแบบและคุณลักษณะต่างๆ เพื่อรับรูปข้อมูลที่ไม่ชัดเจนหรือไม่สมบูรณ์

2. สามารถทำการวิเคราะห์ข้อมูลที่มีความแตกต่างของขนาดกลุ่มตัวอย่างได้ดี โดยที่มีจำนวนขนาดตัวอย่างกลุ่มที่หนึ่งมากกว่าจำนวนขนาดตัวอย่างกลุ่มที่สอง

3. มีความสามารถในการวิเคราะห์ข้อมูลที่ไม่เคยพบเห็นหรือไม่ทราบการแจกแจงตัวแปรอิสระ เนื่องจากมีความสามารถในการปรับให้เข้ากับการเปลี่ยนแปลงของสิ่งแวดล้อมได้ดี

4. ไม่มีข้อกำหนดของการแจกแจงตัวแปรอิสระ หรือไม่จำเป็นต้องทราบการแจกแจงของตัวแปรอิสระ และมีประสิทธิภาพดีแม้ว่าตัวแปรอิสระมีความสัมพันธ์กันเอง

ข้อเสีย

1. ตัวแบบของวิธีโครงข่ายประสาทเทียมไม่สามารถอธิบายความของตัวแบบที่แสดงผลลัพธ์ได้ หรือไม่สามารถอธิบายความสัมพันธ์ของตัวแปรอิสระต่อตัวแปรตามได้

2. ในกระบวนการวิเคราะห์ด้วยวิธีโครงข่ายประสาทเทียมจะใช้เวลานานในการฝึกอบรมเครือข่ายเพื่อให้ผลลัพธ์ของการพยากรณ์มีประสิทธิภาพความแม่นยำ โดยกระบวนการทำงานจะหยุด ก็ต่อเมื่อทำงานครบตามจำนวนรอบที่กำหนดหรือค่าเฉลี่ยของค่าความคลาดเคลื่อนมีค่าน้อยกว่าค่าที่กำหนดได้

3. ไม่สามารถรับประกันได้ว่าค่าเฉลี่ยของค่าความคลาดเคลื่อนจะมีค่าเท่ากับศูนย์ ซึ่งอาจเกิดปัญหาของการสร้างตัวแบบที่มีการประมาณค่าสูงหรือต่ำเกินไป

4. การกำหนดรูปแบบหรือโครงสร้างของเครือข่ายให้เหมาะสมกับชุดของข้อมูล เป็นเรื่องที่ผู้วิจัยจะต้องศึกษาให้เหมาะสมกับปัญหาในแต่ละชุดข้อมูล ในการกำหนดจำนวนชั้นของชั้นรับข้อมูล , ชั้นแฝง และชั้นแสดงผล ซึ่งจะให้ผลลัพธ์ในกระบวนการและใช้เวลาแตกต่างกัน รวมไปถึงการกำหนดน้ำหนักเริ่มต้นของชุดของมูลด้วย

วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วยฟังก์ชันเคอร์เนล

ข้อดีและข้อจำกัด (Advantages and Disadvantages)

ข้อดี

1. เส้นแบ่งแยกประเภทของข้อมูลที่ดีที่สุด (Optimal Separating Hyper plane) ทำให้ตัวแบบที่ได้มีประสิทธิภาพในการพยากรณ์แบ่งประเภทของข้อมูลได้มีความถูกต้องแม่นยำ

2. สามารถทำการวิเคราะห์ข้อมูลขนาดตัวอย่างเล็กมากได้ดี และมีความยืดหยุ่นในการสร้างรูปแบบและคุณลักษณะต่างๆ

3. ไม่มีข้อกำหนดของการแจกแจงตัวแปรอิสระ หรือไม่จำเป็นต้องทราบการแจกแจงของตัวแปรอิสระ และมีประสิทธิภาพดีแม้ว่าตัวแปรอิสระมีความสัมพันธ์กันเอง

4. สามารถทำให้ค่าเฉลี่ยของค่าความคลาดเคลื่อนจะมีค่าใกล้เคียงกับศูนย์ ซึ่งทำให้ตัวแบบที่ได้จากวิธีการซัพพอร์ตเวกเตอร์มีการประมาณที่เหมาะสม และมีระยะห่างระหว่างกลุ่มตัวอย่างที่ต่างกันมีขนาดกว้างที่สุด เพื่อไม่ให้ตัวแบบมีการประมาณค่าที่สูงหรือต่ำเกินไป

ข้อเสีย

1. ตัวแบบของวิธีซัพพอร์ตเวกเตอร์แมชชีนไม่สามารถอธิบายความของตัวแบบที่แสดงผลลัพธ์ได้ หรือไม่สามารถอธิบายความสัมพันธ์ของตัวแปรอิสระต่อตัวแปรตามได้

2. การกำหนดรูปแบบของฟังก์ชันเคอร์เนลให้เหมาะสมกับชุดของข้อมูล เป็นเรื่องที่ผู้วิจัยจะต้องศึกษาให้เหมาะสมกับปัญหาในแต่ละชุดข้อมูล ซึ่งจะให้ผลลัพธ์และประสิทธิภาพในการทำงานได้แตกต่างกันขึ้นอยู่กับการแจกแจงของข้อมูลในแต่ละชุด

3. ในกระบวนการวิเคราะห์ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีนจะใช้เวลานานในการหาวิธีซัพพอร์ตเวกเตอร์เพื่อใช้เป็นขอบเขตในการแบ่งประเภทของข้อมูล โดยกระบวนการทำงานจะหยุด ก็ต่อเมื่อเป็นไปตามเงื่อนไขที่กำหนด

5.2 ด้านการศึกษาวิจัย

เพื่อเป็นแนวทางให้ผู้ที่สนใจได้ศึกษาเพิ่มเติม ซึ่งในการศึกษาครั้งต่อไป สามารถนำแนวทางกรณีต่าง ๆ ไปพัฒนาต่อไปดังนี้นี้

1. ในงานวิจัยครั้งนี้ศึกษาตัวแปรอิสระมีลักษณะการแจกแจงข้อมูลเพียง 4 แบบเท่านั้น และตัวแปรตามเป็นข้อมูลเชิงกลุ่มที่อยู่ในระดับนามบัญญัติ (Nominal Scale) มีลักษณะของการเกิดเหตุการณ์เพียงสองเหตุการณ์เท่านั้น ดังนั้น ในงานวิจัยต่อไปอาจทำการศึกษาในกรณีที่ตัวแปรอิสระมาจากการแจกแจงแบบอื่นๆ และตัวแปรตามเป็นเป็นข้อมูลเชิงกลุ่มอยู่ในมาตรเรียงอันดับ (Ordinal Scale) ที่สามารถเกิดเหตุการณ์ได้หลากหลายมากยิ่งขึ้น

2. ในงานวิจัยครั้งนี้ได้ศึกษาผลกระทบจากค่าค่าเฉลี่ย (μ) ของข้อมูลที่มีการแจกแจงแบบกติเท่านั้น ไม่ได้มีการศึกษาผลกระทบของค่าความแปรปรวน (σ^2) และจากงานวิจัยนี้ศึกษาผลกระทบระดับค่าพารามิเตอร์ของการแจกแจงข้อมูล (d) ซึ่งทำการศึกษาเพียงลักษณะการกระจายของข้อมูลแตกต่างกันของข้อมูลเพียงสองกลุ่ม ดังนั้น ในงานวิจัยต่อไปอาจทำการศึกษาในกรณีที่มีการแจกแจงข้อมูลแบบผสมกันระหว่างการแจกแจงที่เหมือนกันและการแจกแจงที่ต่างกัน

3. ในงานวิจัยนี้ศึกษาตัวแปรอิสระเพียงสองตัวแปร และตัวแปรอิสระทั้งสองตัวแปรนั้นมีลักษณะการแจกแจงพารามิเตอร์ของข้อมูลเหมือนกัน ดังนั้น ในงานวิจัยต่อไปจึงน่าจะมีการสลับพารามิเตอร์ของการแจกแจงข้อมูลในตัวแปรตัวที่หนึ่งกับตัวแปรตัวที่สองที่แตกต่างกัน

4. ในงานวิจัยนี้ศึกษาฟังก์ชันเคอร์เนลเพียง 3 ฟังก์ชันเท่านั้น ซึ่งอาจมีฟังก์ชันเคอร์เนลใหม่ที่มีประสิทธิภาพในการสร้างระนาบเส้นแบ่งประเภทข้อมูลได้ดีกว่า ดังนั้น ในงานวิจัยต่อไปอาจทำการศึกษาฟังก์ชันเคอร์เนลเพิ่มเติม หรือมีการผสมผสานฟังก์ชันเคอร์เนลเพื่อให้มีประสิทธิภาพในการพยากรณ์จำแนกประเภทเพิ่มมากขึ้น

5. ในงานวิจัยนี้ไม่ได้มีการเก็บค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) ของการศึกษา เพื่อเป็นการวัดการกระจายของกลุ่มข้อมูล และค่าส่วนเบี่ยงเบนมาตรฐานนี้ยังสามารถบอกประสิทธิภาพของฟังก์ชันเคอร์เนลในแต่ละฟังก์ชัน ดังนั้น ในงานวิจัยครั้งต่อไป จึงความนำเสนอค่าส่วนเบี่ยงเบนมาตรฐานในงานวิจัยด้วย

5.3 ข้อเสนอแนะ

จากงานวิจัยนี้พบว่า วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ให้ประสิทธิภาพความแม่นยำในการพยากรณ์จำแนกประเภทของข้อมูลดีที่สุดเป็นส่วนใหญ่ และใช้เวลาในการคำนวณน้อยกว่าวิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับซึ่งมีขบวนการที่ซับซ้อนทำให้ใช้เวลานานในการประมวลผล

จากการศึกษางานวิจัยโดยส่วนใหญ่นั้น ใช้วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Gaussian Kernel ในการพยากรณ์จำแนกประเภท ซึ่งก็ให้ผลลัพธ์ของการพยากรณ์จำแนกประเภทที่ดี แต่เมื่อทำการศึกษาในครั้งนี้ ทำให้ทราบว่า วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel มีประสิทธิภาพการพยากรณ์จำแนกประเภทที่ดีกว่าวิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Gaussian Kernel เกือบทุกกรณีที่ทำการศึกษา ยกเว้นกรณีที่ข้อมูลของตัวแปรอิสระมีการแจกแจงแบบแบบปัวส์ซงเท่านั้น

เมื่อพิจารณางานวิจัยของ Yuan – chin Ivan Chang ที่นำเสนอปัญหาของการที่ข้อมูลในกลุ่มตัวอย่างมีขนาดไม่สมดุลกัน แล้วจะทำให้ประสิทธิภาพของวิธีซัพพอร์ตเวกเตอร์แมชชีนลดลง จนต้องเสนอวิธีการแก้ไขปัญหาดังกล่าว แต่เมื่อพิจารณาผลลัพธ์ของการศึกษา วิธีซัพพอร์ตเวกเตอร์แมชชีนด้วย Laplacian Kernel ก็ยังให้ประสิทธิภาพการพยากรณ์ที่ดีแม้ว่าขนาดของกลุ่มตัวอย่างที่สนใจกับกลุ่มตัวอย่างที่ไม่สนใจมีขนาดแตกต่างกันมาก ยกเว้นกรณีที่ข้อมูลของตัวแปรอิสระมีการแจกแจงแบบแบบปัวส์ซงเท่านั้น

รายการอ้างอิง

ภาษาไทย

กัลยา วานิชย์บัญชา. การวิเคราะห์ข้อมูลหลายตัวแปร . พิมพ์ครั้งที่ 3. กรุงเทพฯ : ธรรมสาร, 2551.

ฐิติ อ่วมสวัสดิ์. การเปรียบเทียบการพยากรณ์ในการวิเคราะห์การถดถอยลอจิสติคกับวิธีนิวรอลเน็ทเวิร์คแบบแพร่กระจายย้อนกลับ.วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาสถิติประยุกต์. สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ. 2545.

ภาษาอังกฤษ

Huang Z.,Chen H., Hsu C., Chen W., Wu S.. Credit rating analysis with support vector machines and neural networks: a market comparative study. Decision Support Systems. Vol. 37, pp. 543 – 558, 2004.

D.A. Salazar, J.I. Velez and J.C. Salazar. Comparison between SVM and Logistic Regression : Which One is Better to Discriminate. Revista Colombiana de Estadistica Numero especial en bioestadistica. Vol. 35 No. 2, pp. 223 – 237, 2012.

Yuan – chin Ivan Chang. Boosting SVM Classifiers with Logistic Regression. Institute of Statistical Science, Academia Sinica, Taipei, Taiwan.

บรรณานุกรม

ภาษาไทย

กัลยา วานิชย์บัญชา. การวิเคราะห์ข้อมูลหลายตัวแปร . พิมพ์ครั้งที่ 3. กรุงเทพฯ : ธรรมสาร, 2551.

ภาษาอังกฤษ

Rojas, R. Neural networks : A Systematic Introduction. Springer, Berlin, 1996.

Hastie T, Tibshirani R, and Friedman J. The Elements of Statistical Learning. Springer Verlag, New York, 2001.

Karatzoglou, A., Meyer, D. Support Vector Machines in R. Journal of Statistical Software. Vol. 15, No. 9, 2006.

Blanchard G, Bousquet O , and Massart P. Statistical performance of support vector machines. The Annals of Statistics. Vol. 36, pp. 489 – 531, 2008.

Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters. Vol. 27, pp. 861 – 874, 2006.

ภาคผนวก

คำสั่งที่ใช้ในการวิเคราะห์ข้อมูลจากการจำลองด้วยโปรแกรม R

ตัวอย่างกรณีที่มีตัวแปรอิสระ 1 ตัวแปรที่มีการแจกแจงแบบชี้กำลัง (Exponential Distribution) และตัวแปรตาม 2 กลุ่ม ซึ่งกำหนดจำนวนขนาดของตัวอย่างคือ $n_1 = 150$ และ $n_2 = 30$ โดยที่ค่าเฉลี่ยของแต่ละกลุ่มคือ $\beta_1 = 1$ และ $\beta_2 = 3$

```
####################################
###     Exponential  Distribution        ###
###          n1=150  +  n2=30            ###
####################################

library(class)
library(e1071)
library(kernlab)
library(grid)
library(MASS)
library(neuralnet)
library(gtools)
library(gdata)
library(caTools)
library(KernSmooth)
library(gplots)
library(ROCR)
library(plyr)
library(utils)
library(pROC)
```

```
##################################
###          Collect  the data set          ###
##################################

Best_parameters<-c()
Result_AUC_poly<-c()
Result_AUC_gauss<-c()
Result_AUC_laplac<-c()
Result_AUC_BP_NN<-c()

Result_MCR_poly<-c()
Result_MCR_gauss<-c()
Result_MCR_laplac<-c()
Result_MCR_BP_NN<-c()

##################################
###       Replication (i in 1:500)          ###
##################################

for (i in 1:500)
        {

#############################################
###      Define the parameters of the distribution         ###
#############################################

npos<-150                    ## The number of samples positive group
nneg<-30                     ## The number of samples negative group
n<-npos+nneg                 ## The number of samples
```

```
betapos<-1                    ## Beta of the positive  group
betaneg<-3                    ## Beta of the negative  group

###################################
###     Simulation of the data          ###
###     Exponential  Distribution       ###
###################################

xpos<-matrix(rexp(npos,rate =betapos),npos)       ## Create a positive group
xneg<-matrix(rexp(nneg,rate =betaneg),nneg)       ## Create a negative group
x<-rbind(xpos,xneg)                               ## Independent variables
y<-matrix(c(rep(1,npos),rep(-1,nneg)))            ## Dependent variable
Data<-data.frame(x,y)                             ## Data set

#####################################################################
###     Selection of the parameters of support vector machines is best        ###
#####################################################################

#####  Support vector machines with Polynomial kernel       #####
SVM_simple_P<- tune(svm, y~x, data = Data,ranges = list(degree = 2^(1:10),
                    cost = (1:50)),tunecontrol = tune.control(sampling = "fix"))
## Simulation data with support vector machines
best_para_P<-SVM_simple_P$best.parameters              ## The best parameters
best_degree<-best_para_P$degree                        ## The best degree
best_cost_P<-best_para_P$cost                          ## The best C

#####  Support vector machines with Gaussian kernel         #####
SVM_simple_G<- tune(svm, y~x, data = Data,ranges = list(gamma = 2^(-5:5),
                    cost = (1:50)),tunecontrol = tune.control(sampling = "fix"))
```

```
## Simulation data with support vector machines
best_para_G<-SVM_simple_G$best.parameters                ## The best parameters
best_gamma<-best_para_G$gamma                            ## The best gamma
best_cost_G<-best_para_G$cost                            ## The best C
#############################################################################
###                  Parameter estimation with support vector machine                  ###
#############################################################################
#################################################################
#####          Support vector machines with Polynomial kernel        #####
#################################################################

#####  Parameter estimation   #####
SVM_P<-ksvm(x,y,type="C-svc",kernel="polydot",kpar=list(degree=best_degree,
        scale=1,offset=1),C=best_cost_P,prob.model=TRUE)

#####  Predictive of classification       #####
        ypred_P<-predict(SVM_P,x)                        ## Prediction for y
        table_ypred_P<-table(y,ypred_P)                  ## Results of predictions
        MCR_P<-(1-(sum(diag(table_ypred_P))/sum(table_ypred_P)))
## Misclassification Rate

#####  Graphing and finding the area under the ROC curve            ######
                ypredscore_P<-predict(SVM_P,x,type="decision")
                table_ypredscore_P<-table(ypredscore_P>0,ypred_P)
pred_P<-prediction(ypredscore_P,y)
perf_P<-performance(pred_P,measure="tpr",x.measure="fpr")
## Graphing the ROC curve
AUC_P<-auc(y,ypredscore_P)                  ## Finding the area under the ROC curve
```

```
#####################################################################
#####          Support vector machines with Gaussian RBF Kernel          #####
#####################################################################

#####  Parameter estimation   #####
SVM_G<-ksvm(x,y,type="C-svc",kernel="rbfdot",kpar=
        list (sigma=best_gamma),C=best_cost_G,prob.model=TRUE)
#####  Predictive of classification       #####
        ypred_G<-predict(SVM_G,x)                          ## Prediction for y
        table_ypred_G<-table(y,ypred_G)                    ## Results of predictions
        MCR_G<-(1-(sum(diag(table_ypred_G))/sum(table_ypred_G)))
## Misclassification Rate

#####  Graphing and finding the area under the ROC curve          ######
                ypredscore_G<-predict(SVM_G,x,type="decision")
                table_ypredscore_G<-table(ypredscore_G>0,ypred_G)
pred_G<-prediction(ypredscore_G,y)
perf_G<-performance(pred_G,measure="tpr",x.measure="fpr")
## Graphing the ROC curve
AUC_G<-auc(y,ypredscore_G)                  ## Finding the area under the ROC curve

#################################################################
#####          Support vector machines with Laplacian Kernel          #####
#################################################################

#####  Parameter estimation   #####
SVM_L<-ksvm(x,y,type="C-svc",kernel="laplacedot",kpar=list(sigma=best_gamma),
        C=best_cost_G,prob.model=TRUE)
```

```
#####  Predictive of classification       #####
        ypred_L<-predict(SVM_L,x)                          ## Prediction for y
        table_ypred_L<-table(y,ypred_L)                    ## Results of predictions
        MCR_L<-(1-(sum(diag(table_ypred_L))/sum(table_ypred_L)))
## Misclassification Rate
#####  Graphing and finding the area under the ROC curve             ######
                ypredscore_L<-predict(SVM_L,x,type="decision")
                table_ypredscore_L<-table(ypredscore_L>0,ypred_L)
pred_L<-prediction(ypredscore_L,y)
perf_L<-performance(pred_L,measure="tpr",x.measure="fpr")
## Graphing the ROC curve
AUC_L<-auc(y,ypredscore_L)          ## Finding the area under the ROC curve

#########################################################################
###     Parameter estimation with the Backpropagation Artificial Neural Network       ###
#########################################################################
#####  Create a data set        #####
y_BP<-ifelse(y==1, 1, 0)
Data_BP_NN<-data.frame(x,y_BP)

#####  Parameter estimation   #####
BP_NN<-neuralnet(y_BP~x, data=Data_BP_NN, hidden = 1, threshold = 0.05,
        stepmax = 1e+05, rep = 1, startweights = NULL,learningrate.factor =
        list(minus=0.5,plus=1.2),lifesign.step = 1000, algorithm = "rprop+",err.fct = "sse",
        act.fct = "logistic")

#####  Predictive of classification       #####
Weights_BP<-BP_NN$weights                   ## Weight estimator
Result_bp_nn<-BP_NN$net.result              ## Result BP - NN
```

```
Result_BP<-Result_bp_nn[[1]]                    ## Prediction for y
#####  Predictive of classification       #####
y_BP_NN<-ifelse(Result_BP>=0.5, 1, 0)                ## Prediction for y
table_ypred_BP_NN<-table(y_BP,y_BP_NN)              ## Results of predictions
MCR_BP_NN<-(1-(sum(diag(table_ypred_BP_NN))/sum(table_ypred_BP_NN)))
## Misclassification Rate
#####  Graphing and finding the area under the ROC curve              ######
ROC_BP_NN<-roc(y_BP,Result_BP)              ## Graphing the ROC curve
AUC_BP_NN<-ROC_BP_NN$auc                    ## Finding the area under the ROC curve

###################################
###          Collect  the data set          ###
###################################

Best_parameters<-
c(Best_parameters,best_degree,best_cost_P,best_gamma,best_cost_G)
        Result_AUC_poly<-c(Result_AUC_poly,AUC_P)
        Result_AUC_gauss<-c(Result_AUC_gauss,AUC_G)
        Result_AUC_laplac<-c(Result_AUC_laplac,AUC_L)
        Result_AUC_BP_NN<-c(Result_AUC_BP_NN,AUC_BP_NN)

        Result_MCR_poly<-c(Result_MCR_poly,MCR_P)
        Result_MCR_gauss<-c(Result_MCR_gauss,MCR_G)
        Result_MCR_laplac<-c(Result_MCR_laplac,MCR_L)
        Result_MCR_BP_NN<-c(Result_MCR_BP_NN,MCR_BP_NN)

        }
```

```
###################################
###       Results of the process          ###
###################################

AUC_POLY<-matrix(Result_AUC_poly,nrow=500,ncol=1)
AUC_GAUSS<-matrix(Result_AUC_gauss,nrow=500,ncol=1)
AUC_LAPLAC<-matrix(Result_AUC_laplac,nrow=500,ncol=1)
AUC_BP_NN_a<-matrix(Result_AUC_BP_NN,nrow=500,ncol=1)

Parameters<-matrix(Best_parameters,nrow=500,ncol=4,byrow = TRUE,
                   dimnames = list(c(1:500),c("degree","C_degree","gamma","C_gamma")))
Result_AUC<-cbind(AUC_POLY,AUC_GAUSS,AUC_LAPLAC,AUC_BP_NN_a)
colnames(Result_AUC)<-c("AUC Poly","AUC Gauss","AUC Laplac","AUC BP_NN")
rownames(Result_AUC)<-c(1:500)

MCR_POLY<-matrix(Result_MCR_poly,nrow=500,ncol=1)
MCR_GAUSS<-matrix(Result_MCR_gauss,nrow=500,ncol=1)
MCR_LAPLAC<-matrix(Result_MCR_laplac,nrow=500,ncol=1)
MCR_BP_NN_a<-matrix(Result_MCR_BP_NN,nrow=500,ncol=1)

Result_MCR<-cbind(MCR_POLY,MCR_GAUSS,MCR_LAPLAC,MCR_BP_NN_a)
colnames(Result_MCR)<-c("MCR Poly","MCR Gauss","MCR Laplac","MCR BP_NN")
rownames(Result_MCR)<-c(1:500)

Parameters
Result_AUC
Result_MCR
```

ตัวอย่างกรณีที่มีตัวแปรอิสระ 2 ตัวแปรที่มีการแจกแจงแบบปกติหลายตัวแปร (The Multivariate Normal Distribution) และตัวแปรตาม 2 กลุ่ม ซึ่งกำหนดระดับความสัมพันธ์ของตัวแปรอิสระเป็น 0.3 $(\rho = 0.3)$ จำนวนขนาดของตัวอย่างคือ $n_1 = 150$ และ $n_2 = 30$ โดยที่ค่าเฉลี่ยของแต่ละกลุ่มคือ $\mu_1 = 0$ และ $\mu_2 = 0.5$

```
##########################################
###      Multivariate Normal Distribution        ###
###          n1=150  +  n2=30                    ###
##########################################

library(class)
library(e1071)
library(kernlab)
library(grid)
library(MASS)
library(neuralnet)
library(gtools)
library(gdata)
library(caTools)
library(KernSmooth)
library(gplots)
library(ROCR)
library(plyr)
library(utils)
library(pROC)
```

```
##################################
###       Collect  the data set              ###
##################################

Best_parameters<-c()
Result_AUC_poly<-c()
Result_AUC_gauss<-c()
Result_AUC_laplac<-c()
Result_AUC_BP_NN<-c()

Result_MCR_poly<-c()
Result_MCR_gauss<-c()
Result_MCR_laplac<-c()
Result_MCR_BP_NN<-c()

##################################
###       Replication (i in 1:500)          ###
##################################
for (i in 1:500)
        {
#################################################
###      Define the parameters of the distribution         ###
#################################################

npos<-150                    ## The number of samples positive group
nneg<-30                     ## The number of samples negative group
n<-npos+nneg                 ## The number of samples
sigmapos<-1                  ## Variances of the positive  group
sigmaneg<-1                  ## Variances of the negative  group
```

```
meanpos<-0                    ## Mean of the positive  group
meanneg<-0.5                  ## Mean of the negative group
###################################
###      Simulation of the data          ###
###       Normal  Distribution           ###
###################################

                z_pos1<-rnorm(npos,mean=0,sd=1)
## Create a positive group of the first independent variables
                z_pos2<-rnorm(nneg,mean=0,sd=1)
## Create a positive group of the second independent variable
                z_neg1<-rnorm(npos,mean=0,sd=1)
## Create a negative group of the first independent variables
                z_neg2<-rnorm(nneg,mean=0,sd=1)
## Create a negative group of the second independent variable

################################################
###      Correlation of the Independent Variables      ###
################################################

                Corr<-matrix(c(1,0.3,0.3,1),2,2)        ## The correlation matrix
                MC<-chol(Corr)                          ## The Choleski Decomposition

                Z_pos<-matrix(c(z_pos1,z_pos2),npos,2)
## The positive group of two independent variables
                Z_neg<-matrix(c(z_neg1,z_neg2),nneg,2)
## The negative group of two independent variables
        ZG_pos<-Z_pos%*%MC      ## Relationships between independent variables
        ZG_neg<-Z_neg%*%MC      ## Relationships between independent variables
```

```
################################################
###                  Data set                  ###
################################################

          Xg_pos<-(sqrt(sigmapos)*ZG_pos)+meanpos
## Create a positive group
          Xg_neg<-(sqrt(sigmaneg)*ZG_neg)+meanneg
## Create a negative group

     X1<-c(Xg_pos[,1],Xg_neg[,1])
## Matrix of the first independent variables
     X2<-c(Xg_pos[,2],Xg_neg[,2])
## Matrix of the second independent variables
     X<-as.matrix(cbind(X1,X2))              ## Independent variables
     y<-matrix(c(rep(1,npos),rep(-1,nneg)))  ## Dependent variable
     Data<-data.frame(X,y)                   ## Data set

#####################################################################
###  Selection of the parameters of support vector machines is best  ###
#####################################################################

#####  Support vector machines with Polynomial kernel   #####
SVM_simple_P<- tune(svm, y~X1+X2, data = Data,ranges = list(degree = 2^(1:10),
          cost = (1:50)),tunecontrol = tune.control(sampling = "fix"))
## Simulation data with support vector machines
best_para_P<-SVM_simple_P$best.parameters        ## The best parameters
best_degree<-best_para_P$degree                  ## The best degree
best_cost_P<-best_para_P$cost                    ## The best C
```

```
#####  Support vector machines with Gaussian kernel          #####
SVM_simple_G<- tune(svm, y~X1+X2, data = Data,ranges = list(gamma = 2^(-5:5),
                    cost = (1:50)),tunecontrol = tune.control(sampling = "fix"))
## Simulation data with support vector machines
best_para_G<-SVM_simple_G$best.parameters                  ## The best parameters
best_gamma<-best_para_G$gamma                              ## The best gamma
best_cost_G<-best_para_G$cost                              ## The best C

#########################################################################
###                  Parameter estimation with support vector machine                  ###
#########################################################################

#################################################################
#####            Support vector machines with Polynomial kernel        #####
#################################################################

#####  Parameter estimation   #####
SVM_P<-ksvm( y~X1+X2,type="C-svc",kernel="polydot",kpar=list(degree=best_degree,
          scale=1,offset=1),C=best_cost_P,prob.model=TRUE)

#####  Predictive of classification       #####
          ypred_P<-predict(SVM_P,X)                        ## Prediction for y
          table_ypred_P<-table(y,ypred_P)                  ## Results of predictions
          MCR_P<-(1-(sum(diag(table_ypred_P))/sum(table_ypred_P)))
## Misclassification Rate

#####  Graphing and finding the area under the ROC curve             ######
                    ypredscore_P<-predict(SVM_P,X,type="decision")
                    table_ypredscore_P<-table(ypredscore_P>0,ypred_P)
```

```
pred_P<-prediction(ypredscore_P,y)
perf_P<-performance(pred_P,measure="tpr",x.measure="fpr")
## Graphing the ROC curve
AUC_P<-auc(y,ypredscore_P)                    ## Finding the area under the ROC curve

#################################################################
#####          Support vector machines with Gaussian RBF Kernel          #####
#################################################################

#####  Parameter estimation   #####
SVM_G<-ksvm(y~X1+X2,type="C-svc",kernel="rbfdot",kpar=list (sigma=best_gamma),
          C=best_cost_G,prob.model=TRUE)

#####  Predictive of classification       #####
          ypred_G<-predict(SVM_G,X)                              ## Prediction for y
          table_ypred_G<-table(y,ypred_G)                        ## Results of predictions
          MCR_G<-(1-(sum(diag(table_ypred_G))/sum(table_ypred_G)))
## Misclassification Rate

#####  Graphing and finding the area under the ROC curve              ######
                    ypredscore_G<-predict(SVM_G,X,type="decision")
                    table_ypredscore_G<-table(ypredscore_G>0,ypred_G)
pred_G<-prediction(ypredscore_G,y)
perf_G<-performance(pred_G,measure="tpr",x.measure="fpr")
## Graphing the ROC curve
AUC_G<-auc(y,ypredscore_G)                    ## Finding the area under the ROC curve
```

```
#################################################################
#####          Support vector machines with Laplacian Kernel          #####
#################################################################

#####  Parameter estimation   #####
SVM_L<-ksvm(y~X1+X2,type="C-svc",kernel="laplacedot",kpar=
        list(sigma=best_gamma),C=best_cost_G,prob.model=TRUE)
#####  Predictive of classification       #####
        ypred_L<-predict(SVM_L,X)                          ## Prediction for y
        table_ypred_L<-table(y,ypred_L)                    ## Results of predictions
        MCR_L<-(1-(sum(diag(table_ypred_L))/sum(table_ypred_L)))
## Misclassification Rate

#####  Graphing and finding the area under the ROC curve            ######
                ypredscore_L<-predict(SVM_L,X,type="decision")
                table_ypredscore_L<-table(ypredscore_L>0,ypred_L)
pred_L<-prediction(ypredscore_L,y)
perf_L<-performance(pred_L,measure="tpr",x.measure="fpr")
## Graphing the ROC curve
AUC_L<-auc(y,ypredscore_L)                  ## Finding the area under the ROC curve

#########################################################################
###     Parameter estimation with the Backpropagation Artificial Neural Network      ###
#########################################################################

#####  Create a data set        #####
y_BP<-ifelse(y==1, 1, 0)
Data_BP_NN<-data.frame(X,y_BP)
```

```
#####  Parameter estimation   #####
BP_NN<-neuralnet(y_BP~X1+X2, data=Data_BP_NN, hidden = 2, threshold
        = 0.05,stepmax = 1e+05, rep = 1, startweights = NULL,learningrate.factor =
        list(minus=0.5,plus=1.2),lifesign.step = 1000, algorithm = "rprop+",err.fct = "sse",
        act.fct = "logistic")
#####  Predictive of classification       #####
Weights_BP<-BP_NN$weights                    ## Weight estimator
Result_bp_nn<-BP_NN$net.result               ## Result BP - NN
Result_BP<-Result_bp_nn[[1]]                 ## Prediction for y
#####  Predictive of classification       #####
y_BP_NN<-ifelse(Result_BP>=0.5, 1, 0)                    ## Prediction for y
table_ypred_BP_NN<-table(y_BP,y_BP_NN)                   ## Results of predictions
MCR_BP_NN<-(1-(sum(diag(table_ypred_BP_NN))/sum(table_ypred_BP_NN)))
## Misclassification Rate

#####          Graphing and finding the area under the ROC curve          ######
ROC_BP_NN<-roc(y_BP,Result_BP)                   ## Graphing the ROC curve
AUC_BP_NN<-ROC_BP_NN$auc                         ## Finding the area under the

###################################
###      Collect  the data set          ###
###################################

Best_parameters<-
c(Best_parameters,best_degree,best_cost_P,best_gamma,best_cost_G)
        Result_AUC_poly<-c(Result_AUC_poly,AUC_P)
        Result_AUC_gauss<-c(Result_AUC_gauss,AUC_G)
        Result_AUC_laplac<-c(Result_AUC_laplac,AUC_L)
        Result_AUC_BP_NN<-c(Result_AUC_BP_NN,AUC_BP_NN)
```

```
        Result_MCR_poly<-c(Result_MCR_poly,MCR_P)
        Result_MCR_gauss<-c(Result_MCR_gauss,MCR_G)
        Result_MCR_laplac<-c(Result_MCR_laplac,MCR_L)
        Result_MCR_BP_NN<-c(Result_MCR_BP_NN,MCR_BP_NN)

        }

##################################
###      Results of the process          ###
##################################

AUC_POLY<-matrix(Result_AUC_poly,nrow=500,ncol=1)
AUC_GAUSS<-matrix(Result_AUC_gauss,nrow=500,ncol=1)
AUC_LAPLAC<-matrix(Result_AUC_laplac,nrow=500,ncol=1)
AUC_BP_NN_a<-matrix(Result_AUC_BP_NN,nrow=500,ncol=1)

Parameters<-matrix(Best_parameters,nrow=500,ncol=4,byrow = TRUE, dimnames =
list(c(1:500),c("degree","C_degree","gamma","C_gamma")))

Result_AUC<-cbind(AUC_POLY,AUC_GAUSS,AUC_LAPLAC,AUC_BP_NN_a)
colnames(Result_AUC)<-c("AUC Poly","AUC Gauss","AUC Laplac","AUC BP_NN")
rownames(Result_AUC)<-c(1:500)

MCR_POLY<-matrix(Result_MCR_poly,nrow=500,ncol=1)
MCR_GAUSS<-matrix(Result_MCR_gauss,nrow=500,ncol=1)
MCR_LAPLAC<-matrix(Result_MCR_laplac,nrow=500,ncol=1)
MCR_BP_NN_a<-matrix(Result_MCR_BP_NN,nrow=500,ncol=1)
```

```
Result_MCR<-cbind(MCR_POLY,MCR_GAUSS,MCR_LAPLAC,MCR_BP_NN_a)
colnames(Result_MCR)<-c("MCR Poly","MCR Gauss","MCR Laplac","MCR BP_NN")
rownames(Result_MCR)<-c(1:500)
```

Parameters

Result_AUC

Result_MCR

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวนันทณัฐ พันธุ์สีดา เกิดวันจันทร์ที่ 9 มิถุนายน พ.ศ. 2529 สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิชาสถิติ ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา ในปีการศึกษา 2551 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตรมหาบัณฑิต (วท.ม.) สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2554