

ซัพพอร์ตเวกเตอร์แมชชีนหลายประเภทโดยใช้เอนโทรปีและค่าประมาณความผิดพลาดโดยนัยทั่วไป



บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2559
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

MULTICLASS SUPPORT VECTOR MACHINES USING ENTROPY
AND GENERALIZATION ERROR ESTIMATION

Mr. Pittipol Kantavat



A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Computer Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2016

Copyright of Chulalongkorn University

พิตติพล คันธวัฒน์ : ซัพพอร์ทเวกเตอร์แมชชีนหลายประเภทโดยใช้เอนโทรปีและค่าประมาณความผิดพลาดโดยนัยทั่วไป (MULTICLASS SUPPORT VECTOR MACHINES USING ENTROPY AND GENERALIZATION ERROR ESTIMATION) อ.ที่ปรึกษา วิทยานิพนธ์หลัก: ศ. ดร. บุญเสริม กิจศิริกุล, 100 หน้า.

ตั้งแต่อดีตจนถึงปัจจุบันมีผู้ที่เสนอวิธีการจำแนกหลายประเภทด้วยซัพพอร์ทเวกเตอร์แมชชีนไว้หลายวิธี ถึงแม้ว่าจะมีการพัฒนาวิธีการจำแนกที่ให้ความแม่นยำสูงขึ้นมาแล้ว แต่วิธีการเหล่านั้นมักต้องใช้เวลาในการคำนวณมากและไม่เหมาะสมที่จะนำไปใช้กับปัญหาที่มีจำนวนประเภทข้อมูลมากๆ ดังนั้นจึงมีนักวิจัยหลายท่านพยายามพัฒนาหลากหลายเทคนิคที่มีเป้าหมายเพื่อลดเวลาที่ใช้ในการประมวลผลโดยยอมให้ความแม่นยำในการจำแนกลดลง ในงานวิจัยนี้ได้นำเสนอซัพพอร์ทเวกเตอร์แมชชีนรูปแบบใหม่ที่ใช้สถาปัตยกรรมแบบต้นไม้สำหรับใช้ในการจำแนกหลายประเภท ในแต่ละปมของต้นไม้เราได้เลือกตัวจำแนกสองประเภทที่เหมาะสมด้วยค่าเอนโทรปีและค่าประมาณความผิดพลาดโดยนัยทั่วไป จากนั้นจึงแบ่งประเภทข้อมูลออกเป็นกลุ่มบวกและกลุ่มลบด้วยตัวจำแนกดังกล่าว แล้วจึงทำการฝึกตัวจำแนกตัวใหม่เพื่อใช้ในขั้นตอนการจำแนก วิธีใหม่ที่เราได้นำเสนอสามารถทำงานในความซับซ้อนเชิงเวลาระหว่าง $O(\log_2 M)$ ถึง $O(N)$ โดยที่ N คือจำนวนประเภทข้อมูล เราได้ทำการเปรียบเทียบประสิทธิภาพการทำงานของวิธีที่เราได้นำเสนอกับเทคนิคดั้งเดิมด้วยข้อมูลจากพื้นที่เก็บข้อมูลการเรียนรู้ของเครื่องยูซีไอ โดยใช้วิธีการตรวจสอบไขว้สลับ ผลการทดลองแสดงให้เห็นว่าวิธีการที่เราได้นำเสนอมีประโยชน์มากกับปัญหาที่ต้องการความเร็วในการจำแนกหรือปัญหาที่มีจำนวนประเภทข้อมูลขนาดใหญ่ เนื่องจากวิธีที่เราได้นำเสนอสามารถทำงานเร็วกว่าวิธีการดั้งเดิมแต่ให้ความแม่นยำในการจำแนกเทียบเคียงกัน

ภาควิชา วิศวกรรมคอมพิวเตอร์

ลายมือชื่อนิสิต

สาขาวิชา วิศวกรรมคอมพิวเตอร์

ลายมือชื่อ อ.ที่ปรึกษาหลัก

ปีการศึกษา 2559

5371836721 : MAJOR COMPUTER ENGINEERING

KEYWORDS: SUPPORT VECTOR MACHINE (SVM) / MULTI-CLASS CLASSIFICATION / GENERALIZATION ERROR / ENTROPY / DECISION TREE

PITTIPOL KANTAVAT: MULTICLASS SUPPORT VECTOR MACHINES USING ENTROPY AND GENERALIZATION ERROR ESTIMATION. ADVISOR: PROF. DR. BOONSERM KIJSIRIKUL, 100 pp.

There have been several proposed methods of Support Vector Machines (SVMs) for multi-class classification problems. Though high-accuracy techniques have been developed, they usually consume high computational time and are not suitable for a problem with a large number of classes. Therefore, many researchers have attempted various techniques aiming to reduce the processing time, with possible sacrifice of classification accuracy. In this research, we propose new methods of SVMs using tree architecture for multi-class classification. In each node of the tree, we select an appropriate binary classifier using entropy and generalization error estimation, then group the examples into positive and negative classes based on the selected classifier and train a new classifier for using in the classification phase. The proposed methods can work in time complexity between $O(\log_2 N)$ to $O(N)$ where N is the number of classes. We compared the performance of our proposal methods to the traditional techniques on the UCI machine learning repository using 10-fold cross-validation. The experimental results show that our proposed methods are very useful for the problems that need fast classification time or problems with a large number of classes as the proposed methods run much faster than the traditional techniques but still provide comparable accuracy.

Department: Computer Engineering Student's Signature

Field of Study: Computer Engineering Advisor's Signature

Academic Year: 2016

กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จลุล่วงได้ด้วยความช่วยเหลือเป็นอย่างยิ่งของอาจารย์ที่ปรึกษา คือ ศาสตราจารย์ ดร. บุญเสริม กิจศิริกุล ซึ่งท่านได้ให้ความรู้ คำแนะนำ และข้อคิดเห็นอันเป็นประโยชน์อย่างยิ่งต่องานวิจัย จนทำให้งานวิจัยนี้สำเร็จลุล่วงด้วยดี

ขอขอบพระคุณอาจารย์กรรมการสอบวิทยานิพนธ์ ซึ่งประกอบด้วย ศาสตราจารย์ ดร. ประภาส จงสถิตย์วัฒนา รองศาสตราจารย์ ดร. ญาใจ ลิ้มปิยะกรรม รองศาสตราจารย์ ดร. ชลวิชนัทธี และอาจารย์ ดร. พีรพล เวทีกุล ที่ได้กรุณาสละเวลาอันมีค่า ทำการตรวจสอบ ให้คำแนะนำ และแก้ไขวิทยานิพนธ์ฉบับนี้

ขอขอบพระคุณศาสตราจารย์ ดร. มาชายุกิ นูมาโอะ และรองศาสตราจารย์ ดร. เคนอิชิ ฟุคุอิ แห่งมหาวิทยาลัยโอซาก้า ประเทศญี่ปุ่น ที่ให้คำแนะนำที่เป็นประโยชน์ และให้ใช้เครื่องคอมพิวเตอร์ประมวลผลงานวิจัยทั้งในขณะที่อยู่ที่ประเทศญี่ปุ่นและกลับมายังประเทศไทยแล้ว จนได้ผลการทดลองครบถ้วนและสำเร็จลุล่วง

ขอขอบคุณ ดร. ปทุมศิริ สงศิริ ที่ได้ให้ความช่วยเหลือเป็นอย่างยิ่งในงานวิจัยนี้ รวมถึงสมาชิกห้องปฏิบัติการ Machine Intelligence & Knowledge Management (MIND Lab) และห้องปฏิบัติการอื่นๆ ที่คอยให้กำลังใจและช่วยเหลือเป็นอย่างดี

ผู้วิจัยใคร่ขอกราบขอบพระคุณบิดา-มารดา ภรรยา รวมถึงทุกๆ คนในครอบครัว ที่เป็นกำลังใจและให้การสนับสนุนเรื่อยมาจนสำเร็จการศึกษา

งานวิจัยนี้ได้รับการสนับสนุนจากทุนปริญญาเอกกาญจนาภิเษก สำนักงานกองทุนสนับสนุนการวิจัย ซึ่งให้การสนับสนุนค่าใช้จ่ายในการวิจัยต่างๆ รวมถึงค่าใช้จ่ายในการเดินทางไปทำการวิจัยต่างประเทศด้วย

2.1.3.1 วิธีโอวีเอ (One-Versus-All (OVA))	29
2.1.3.2 วิธีโอวีโอ (One-Versus-One (OVO)).....	30
2.1.4 การใช้รหัสขอกแก้ไขข้อผิดพลาด (อีซีโอซี) (Error-Correcting Output Code (ECOC)).....	31
2.1.5 การใช้รหัสขอกแก้ไขข้อผิดพลาดแบบ 3 ค่า (Ternary Error-Correcting Output Code (Ternary ECOC)).....	33
2.2 งานวิจัยที่เกี่ยวข้อง.....	35
2.2.1 วิธีดีดีเอจี (Decision Directed Acyclic Graph SVM (DDAG)).....	35
2.2.2 วิธีเอดีเอจี (Adaptive Directed Acyclic Graph SVM (ADAG)).....	36
2.2.3 วิธีอาร์เอดีเอจี (Reordering Adaptive acyclic graph SVM (RADAG))	37
2.2.4 วิธีบีทีเอส (Binary Tree of SVM (BTS))	38
2.2.5 วิธีแตกครึ่งตามสารสนเทศ (Information-based Dichotimization (IBD)).....	40
2.2.6 วิธีต้นไม้ตัดสินใจเอสวีเอ็ม (Decision-tree-based SVM).....	43
2.2.7 วิธีเอสวีเอ็ม-บีดีที (SVM Binary Decision Tree (SVM-BDT))	44
2.2.8 การเปรียบเทียบวิธีการจำแนกหลายประเภทโดย Hsu และ Lin.....	45
บทที่ 3 ต้นไม้ตัดสินใจตามสารสนเทศและ ต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป	46
3.1 ต้นไม้ตัดสินใจตามสารสนเทศ (The Information Based Decision Tree (IB-DTree)).....	46
3.2 ต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป (The Information Based and Generalization Error Estimation Decision Tree (IBGE-DTree)).....	57
บทที่ 4 การทดลองและผล.....	62
4.1 ชุดข้อมูลที่ใช้ในการทดลอง.....	62

4.2 รายละเอียดการตั้งค่าต่างๆ ในการทดลอง.....	64
4.3 ผลการทดลอง.....	64
บทที่ 5 บทสรุปงานวิจัยและข้อเสนอแนะ	71
5.1 สรุปผลงานวิจัย	71
5.2 ข้อเสนอแนะ	72
ภาคผนวก ก. พารามิเตอร์ Gamma และ c ที่ใช้ในการทดลอง	73
ภาคผนวก ข. ผลการทดลองหาความสัมพันธ์ของร้อยละจำนวนตัวจำแนก ในวิธีการสร้างต้นไม้ ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป.....	88
รายการอ้างอิง	94
ประวัติผู้เขียนวิทยานิพนธ์	100

สารบัญรูปภาพ

หน้า

รูป 1 ตัวอย่างของพื้นผิวตัดสีนใจและระยะขอบ.....	21
รูป 2 ระนาบหลายมิติแบ่งแยกดีสุดและระยะขอบที่มากที่สุด	22
รูป 3 ตัวอย่างระนาบแบ่งหลายมิติที่ได้จากการฝึก	23
รูป 4 ชุดข้อมูลที่ข้อมูลที่ไม่สมบูรณ์	24
รูป 5 การปรับค่าข้อมูลที่ผิดพลาดด้วยตัวแปรหย่อน	24
รูป 6 การแมปข้อมูลจากปริภูมินำเข้าไปยังมิติปริภูมิอันดับสูง	26
รูป 7 ตัวอย่างระนาบหลายมิติสำหรับแบ่งแยก (ก) ปริภูมินำเข้า (ข) มิติปริภูมิอันดับสูง	26
รูป 8 ตัวอย่างเมทริกซ์รหัสค่าแบบ 3 ค่า.....	33
รูป 9 กระบวนการจำแนกหลายประเภทแบบดีดีเอจี้สำหรับปัญหา 4 ประเภท	36
รูป 10 กระบวนการจำแนกหลายประเภทแบบเอดีเอจี้สำหรับปัญหา 8 ประเภท	37
รูป 11 กระบวนการจำแนกหลายประเภทแบบอาร์เอดีเอจี้	38
รูป 12 ปัญหาหลายมิติแยกไม่สามารถจำแนกบางประเภทข้อมูลได้	39
รูป 13 ตัวอย่าง Binary Tree of SVM สำหรับปัญหา 6 ประเภทข้อมูล	39
รูป 14 ตัวอย่างข้อมูลที่กระจายสม่ำเสมอและต้นไม้แสดงการเข้ารหัสข้อมูล 4 ประเภทข้อมูล ..	40
รูป 15 ตัวอย่างข้อมูลที่กระจายไม่สม่ำเสมอและต้นไม้แสดงการเข้ารหัสข้อมูล 4 ประเภท ข้อมูล	41
รูป 16 ตัวอย่างการแบ่งบนปริภูมิคุณลักษณะ	43
รูป 17 ต้นไม้ตัดสีนใจของปัญหาในรูปที่ 16	43
รูป 18 ภาพของเอสวีเอ็ม-บีดีที.....	44
รูป 19 เอสวีเอ็ม-บีดีทีที่แบ่ง 7 ประเภทข้อมูล.....	44
รูป 20 การจับกลุ่มประเภทข้อมูลตามข้อมูลส่วนใหญ่	47
รูป 21 ภาพตัวอย่างต้นไม้ตัดสีนใจตามสารสนเทศ	48

รูป 22 ตัวอย่างปัญหาที่มีประเภทข้อมูล 4 ประเภท บนระนาบ 2 มิติ	49
รูป 23 เส้นระนาบหลายมิติของตัวจำแนก 1 vs 2.....	50
รูป 24 เส้นระนาบหลายมิติของตัวจำแนก 1 vs 3.....	51
รูป 25 เส้นระนาบหลายมิติของตัวจำแนก 1 vs 4.....	52
รูป 26 เส้นระนาบหลายมิติของตัวจำแนก 2 vs 3.....	53
รูป 27 เส้นระนาบหลายมิติของตัวจำแนก 2 vs 4.....	54
รูป 28 เส้นระนาบหลายมิติของตัวจำแนก 3 vs 4.....	55
รูป 29 ต้นไม้ตัดสินใจตามสารสนเทศสำหรับปัญหาตัวอย่างในรูปที่ 22.....	57
รูป 30 ต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป สำหรับปัญหา ตัวอย่างในรูปที่ 22	61



สารบัญตาราง

หน้า

ตาราง 1 ตัวอย่างของฟังก์ชันเคอร์เนล	28
ตาราง 2 ตัวอย่างรหัสที่มีการกระจาย 6 บิต	32
ตาราง 3 ตัวอย่างรหัสที่มีการกระจาย 15 บิต	32
ตาราง 4 จำนวนข้อมูลที่ถูกจำแนกในฝั่งบวกและลบโดยตัวจำแนก 1 vs 2	50
ตาราง 5 จำนวนข้อมูลที่ถูกจำแนกในฝั่งบวกและลบโดยตัวจำแนก 1 vs 3	51
ตาราง 6 จำนวนข้อมูลที่ถูกจำแนกในฝั่งบวกและลบโดยตัวจำแนก 1 vs 4	52
ตาราง 7 จำนวนข้อมูลที่ถูกจำแนกในฝั่งบวกและลบโดยตัวจำแนก 2 vs 3	53
ตาราง 8 จำนวนข้อมูลที่ถูกจำแนกในฝั่งบวกและลบโดยตัวจำแนก 2 vs 4	54
ตาราง 9 จำนวนข้อมูลที่ถูกจำแนกในฝั่งบวกและลบโดยตัวจำแนก 3 vs 4	55
ตาราง 10 รายละเอียดของชุดข้อมูลทดลอง	63
ตาราง 11 ค่าเฉลี่ยของผลความแม่นยำในการจำแนกพร้อมทั้งค่าส่วนเบี่ยงเบนมาตรฐาน	66
ตาราง 12 ค่าเฉลี่ยของผลความแม่นยำในการจำแนกพร้อมทั้งค่าส่วนเบี่ยงเบนมาตรฐาน	67
ตาราง 13 ผลการเปรียบเทียบความแม่นยำในการจำแนกด้วย Wilcoxon Signed Rank Test	68
ตาราง 14 จำนวนครั้งการตัดสินใจเฉลี่ยที่ใช้จำแนกประเภทข้อมูล	69

บทที่ 1

บทนำ

1.1. ที่มาและความสำคัญของปัญหา

ซัพพอร์ตเวกเตอร์แมชชีน - เอสวีเอ็ม (Support Vector Machine - SVM)¹⁻² เป็นการเรียนรู้ของเครื่อง (Machine Learning) แบบหนึ่ง ซึ่งใช้ในการจำแนกข้อมูล 2 ประเภท ทำงานโดยสร้างระนาบหลายมิติ (Hyperplane) ที่มีระยะขอบที่มากที่สุด (Maximizing Margin) ระหว่างข้อมูล 2 ประเภท แต่กับปัญหาที่มีประเภทข้อมูลมากกว่า 2 ประเภท เราจำเป็นต้องใช้เทคนิคบางประการเข้ามาประยุกต์ใช้ เพื่อให้สามารถใช้เอสวีเอ็มทำงานได้กับปัญหาที่มีประเภทมากกว่า 2 ประเภท ซึ่งเราเรียกเทคนิคดังกล่าวว่าการทำซัพพอร์ตเวกเตอร์แมชชีนแบบจำแนกหลายประเภท (Multi-class Support Vector Machine)

เทคนิคการจำแนกหลายประเภทนั้นสามารถแบ่งออกได้เป็น 2 วิธีการ คือ 1. มองเป็นการแก้ปัญหาการหาค่าเหมาะสมที่สุด (Optimization Problem) โดยพิจารณาปัญหาหลายประเภททั้งหมดในคราวเดียว 2. แแตกปัญหาดังต้นออกเป็นปัญหาย่อยหลายปัญหา แก้ปัญหาย่อยทีละส่วนแล้วจึงใช้คำตอบของปัญหาย่อยๆ ในการตอบปัญหาดังต้น วิธีการแรกได้มีนักวิจัยเสนอไว้หลายท่าน เช่น Vapnik¹, Bredensteiner และ Bennett³, Crammer และ Singer⁴ เป็นต้น อย่างไรก็ตาม Hsu และ Lin⁵ ได้ทำการวิเคราะห์และเปรียบเทียบเอาไว้ในงานวิจัยว่าวิธีการแรกไม่เหมาะกับการนำไปกับปัญหาจริง เนื่องจากวิธีการซับซ้อนเกินไปและใช้ทรัพยากรในการประมวลผลมากกว่าวิธีแรก ดังนั้นงานวิจัยภายหลังส่วนมากจึงมุ่งเน้นการพัฒนาในแนวทางที่สอง คือการรวมตัวจำแนกหลายตัวเข้าด้วยกันเพื่อใช้แก้ปัญหา

วิธีการรวมตัวจำแนกหลายประเภทเข้าด้วยกันมีหลายวิธี โดยพื้นฐานแล้วที่รู้จักกันทั่วไปและใช้กันอย่างแพร่หลาย คือ วิธีโอวีโอ (One-Versus-One (OVO))⁶⁻⁷ และ วิธีโอวีเอ (One-Versus-All (OVA))² โดยวิธีโอวีโอใช้ตัวจำแนก $N(N-1)/2$ ตัว ในการแก้ปัญหาที่มีจำนวนประเภทข้อมูล N ประเภท ซึ่งโดยปกติจะใช้กระบวนการประมวลผลข้อมูลขาออกที่เรียกว่า Max-Wins⁸ ซึ่งเลือกเอาประเภทข้อมูลที่ได้รับคะแนนการโหวตสูงสุดเป็นคำตอบ แต่หากเกิดกรณีที่มีประเภทข้อมูลมากกว่า

หนึ่งประเภทที่ได้รับคะแนนโหวตเท่ากัน จะสุ่มให้คำตอบเป็นประเภทที่ได้รับคะแนนมากที่สุดด้วยความน่าจะเป็นเท่าๆ กัน ส่วนวิธีโอวีเอจะใช้ตัวจำแนกทั้งสิ้น N ตัวสำหรับแก้ปัญหาที่มีจำนวนประเภทข้อมูล N ประเภท โดยในระยะฝึก (Training Phase) จะใช้ตัวจำแนกตัวที่ i เรียนรู้ประเภทข้อมูล i เป็นข้อมูลบวกและประเภทข้อมูลที่เหลือเป็นข้อมูลลบ และในการจำแนกข้อมูลทดสอบจะดูว่าตัวจำแนกใดที่มีคะแนนการจำแนกสูงสุดก็จะใช้ประเภทนั้นมาเป็นคำตอบ ถึงแม้ว่าทั้ง 2 วิธีนี้จะให้คำตอบที่น่าเชื่อถือและถูกใช้อย่างแพร่หลาย แต่อย่างไรก็ดี Hsu และ Lin⁵ ได้ทำการสรุปว่าในงานวิจัยว่าวิธีโอวีเอสามารถจำแนกได้ความแม่นยำสูงกว่าวิธีโอวีเอ อย่างไรก็ตาม Rifkin and Klautau⁹ ได้ศึกษาและทำการทดลองอย่างระมัดระวัง ได้ข้อสรุปว่าหากทำการปรับจูนพารามิเตอร์โอวีเออย่างดีก็สามารถให้ความแม่นยำในการจำแนกได้ไม่ต่างจากโอวีเอ

วิธีโอวีเอนั้นถึงแม้ว่าจะมีความแม่นยำในการจำแนกสูง แต่ต้องใช้เวลาในการประมวลผลค่อนข้างมาก ดังนั้นจึงมีผู้ที่เสนอวิธีการจำแนกซึ่งสร้างจากตัวจำแนกพื้นฐานของวิธีโอวีเอ แต่นำมาจัดเรียงเป็นโครงสร้างแบบมีลำดับชั้น เพื่อลดขั้นตอนการประมวลผลในการจำแนกประเภทข้อมูล เช่น Platt และคณะได้เสนอวิธีดีดีเอจี (Decision Directed Acyclic Graph (DDAG))¹⁰ ซึ่งเป็นการนำตัวจำแนกทั้ง $N(N-1)/2$ ตัว มาทำการจัดเรียงเป็นโครงสร้างสามเหลี่ยม ซึ่งมีจำนวนชั้นทั้งสิ้น $N-1$ ชั้น โดยหากเปรียบเทียบกับวิธีโอวีเอแล้ว ถึงแม้ว่าในระยะการฝึกจะต้องเตรียมตัวจำแนก $N(N-1)/2$ ตัว เหมือนกับวิธีโอวีเอ แต่ในระยะทดสอบใช้การจำแนกเพียง $N-1$ ครั้งเท่านั้น โดย Takahashi และ Abe¹¹ ได้เสนอวิธีการเพิ่มประสิทธิภาพของดีดีเอจีอีกด้วย ต่อมา Kijssirikul และคณะได้เสนอวิธีเอดีเอจี (Adaptive Directed Acyclic Graph (ADAG))¹² ซึ่งปรับปรุงประสิทธิภาพของวิธีดีดีเอจีด้วยการนำเอาตัวจำแนกจากวิธีโอวีเอมาจัดเรียงเป็นโครงสร้างสามเหลี่ยมเป็นแบบผกผัน โดยในระดับชั้นบนสุดจะสุ่มเลือกเอาคู่ประเภทมาจับคู่กันแบบ 1 ต่อ 1 จำนวนทั้งสิ้น $N/2$ ตัว โดยถึงแม้ว่าวิธีการนี้ต้องเตรียมตัวจำแนกไว้ $N(N-1)/2$ ตัว แต่ใช้จำนวนครั้งในระยะทดสอบ $N-1$ ครั้งเช่นเดียวกับดีดีเอจี ซึ่งในเวลาต่อมา Phetkeaw และคณะ ได้พัฒนาวิธีอาร์เอดีเอจี (Reordering Adaptive acyclic graph SVM (RADAG))¹³ ซึ่งปรับปรุงประสิทธิภาพของเอดีเอจีด้วยจับคู่ในแต่ละลำดับชั้นด้วยระเบียบวิธีการจับคู่สมบูรณ์แบบน้ำหนักน้อยที่สุด (Minimum Weight Perfect Matching)¹⁴ แทนการเลือกแบบสุ่ม แต่อย่างไรก็ดี ถึงแม้ว่าทั้งวิธีดีดีเอจี เอดีเอจี และ อาร์เอดีเอจี จะสามารถทำงานได้ในความซับซ้อนเชิงเวลาเป็น $O(N)$ แต่ก็ให้ความแม่นยำของการจำแนกต่ำกว่าวิธีโอวีเอ

ยังมีนักวิจัยอีกกลุ่มหนึ่ง ที่ได้เสนอวิธีการแก้ไขปัญหาคำแนะนำหลายประเภท ด้วยการเข้ารหัสข่าออกแก้ไขข้อผิดพลาด (Error correcting output code (ECOC)) เริ่มต้นจาก Dietterich และคณะ¹⁵⁻¹⁶ ได้ใช้แนวคิดจากพื้นฐานของทฤษฎีการเข้ารหัสมาประยุกต์ใช้กับการจำแนกหลายประเภท ในเวลาต่อมาจึงมีนักวิจัยจำนวนมากสนใจศึกษาต่อยอด เช่น James และ Hastie¹⁷ ได้ทำการศึกษาและประยุกต์ใช้รหัสข่าออกแก้ไขข้อผิดพลาดเข้ากับตัวจำแนกพื้นฐานแบบต่างๆ Ghani¹⁸ ประยุกต์วิธีการรหัสข่าออกแก้ไขข้อผิดพลาดเข้ากับปัญหาการจำแนกข้อความ (Text Classification) Allwein และคณะ¹⁹ ได้ทำการศึกษาและรวมกรอบงานของการจำแนกหลายประเภทแบบต่างๆ รวมถึงวิธีการรหัสข่าออกแก้ไขข้อผิดพลาดด้วย Crammer และ Singer⁴ สรุปว่าปัญหาการหาเมทริกซ์รหัส (code matrix) ที่ให้ค่าสูญเสียจากการทดลองน้อยที่สุด (minimized empirical loss) เป็นปัญหา NP-hard Masuli และ Valentini²⁰ สรุปว่าปัจจัยหลักที่ส่งผลกับประสิทธิภาพของการรหัสข่าออกแก้ไขข้อผิดพลาดคือความขึ้นต่อกัน (dependency) ระหว่างบิตของคำรหัส (codeword) Utschick and Weichselberger²¹ ได้เสนอวิธี Expectation Maximization เพื่อใช้ในการหาเมทริกซ์รหัส Kuncheva²² ได้เสนอวิธีการวัดความหลากหลาย (diversity) ในการสร้างเมทริกซ์รหัสข่าออกแก้ไขข้อผิดพลาดสำหรับกลุ่มก่อนตัวจำแนก (classifier ensembles) Escalera และคณะ²³ ได้เสนอ ECOC One ซึ่งเป็นขั้นตอนวิธีการออกแบบเมทริกซ์รหัสที่มีกระจายระยะทางแฮมมิง (Hamming distance) ระหว่างทุกๆ คู่คำรหัสเป็นอย่างดี โดยใช้ confusion matrix ในขั้นตอนฝึก

ต่อมา Escalera และคณะได้ทำการเสนอรหัสข่าออกแก้ไขข้อผิดพลาดแบบ 3 ค่า (Ternary ECOC)²⁴ ซึ่งเป็นการขยายกรอบงานของรหัสข่าออกแก้ไขข้อผิดพลาดแบบดั้งเดิม โดยการเพิ่มสัญลักษณ์ศูนย์เข้าไปในเมทริกซ์รหัส โดยรหัสข้อมูลที่ระบุเป็นศูนย์จะไม่ถูกนำไปพิจารณาจำแนกอยู่ทั้งในฝั่งบวกและลบ Pujol และคณะ²⁵ ได้แสดงวิธีการเขียนโอวีโอและโอวีเอในรูปแบบของการเข้ารหัสข่าออกแก้ไขข้อผิดพลาดแบบ 3 ค่า Escalera และคณะเสนอวิธีการออกแบบรหัสสุ่มประปราย (Sparse Random Code)²⁶ Escalera และคณะ²⁷ ได้ทำการศึกษาและวิเคราะห์ขั้นตอนการถอดรหัส (decoding step) ของการเข้ารหัสข่าออกแก้ไขข้อผิดพลาดแบบ 3 ค่า Bautista และคณะ²⁸ เสนอวิธีการออกแบบโดยใช้การประมวลผลเชิงวิวัฒนาการ (Evolutionary Computation) Bautista และคณะ²⁹ เสนอการออกแบบโดยใช้ขั้นตอนวิธีทางพันธุกรรมที่เข้ากันได้กับการรหัสข่าออกแก้ไขข้อผิดพลาด (ECOC compliant Genetic Algorithm) Bagheri และคณะ³⁰ เสนอวิธีการใช้ปริภูมิมิติย่อย (subspace) สำหรับการเข้ารหัสข่าออกแก้ไขข้อผิดพลาด Zhou และคณะ³¹ เสนอ

Data-Driven ECOC ซึ่งพิจารณาลักษณะการกระจายตัวของข้อมูลแล้วแบ่งเป็นกลุ่มย่อยก่อนสร้างเมตริกซ์รหัส Zhang และคณะ³² เสนอวิธีการแก้ไขปัญหาค่าจำแนกหลายประเภทด้วยรหัสขาออกแก้ไขข้อผิดพลาดที่อนุรักษ์โทโปโลยี (topology reserving)

ต่อมามีนักวิจัยอีกกลุ่มหนึ่งออกแบบและเสนอวิธีการจำแนกหลายประเภทบนพื้นฐานของวิธีโอวีโอ ที่สามารถตัดแยกประเภทข้อมูลที่ไม่เป็นคำตอบที่ต้องการออกได้ครั้งละหลายประเภทพร้อมๆ กัน ด้วยการนำโครงสร้างของต้นไม้ตัดสินใจเข้ามาประยุกต์ใช้ ซึ่งสามารถลดความซับซ้อนเชิงเวลาของการประมวลผล ให้น้อยลงจนอยู่ระหว่าง $O(\log_2 M)$ ถึง $O(N)$ ได้ เช่น Kijisrijul และคณะ³³ เสนอวิธีการแตกครึ่งแบบสมดุลซึ่งเป็นการสร้างต้นไม้ตัดสินใจโดยแบ่งประเภทข้อมูลให้สมดุลที่สุด Fei และ Lui ได้เสนอวิธีบีทีเอส (Binary Search Tree of SVM (BTS))³⁴ ซึ่งเลือกตัวจำแนกอย่างสุ่มหรือเลือกจากจุดศูนย์กลาง (centroid) ของประเภทข้อมูล Songsiri และคณะได้เสนอวิธีแตกครึ่งตามสารสนเทศ (Information Based Dichotomization (IBD))³⁵ ซึ่งเลือกตัวจำแนกด้วยเอนโทรปี Chen และคณะเสนอวิธีเอบีที (Adaptive Binary Tree (ABT))³⁶ ซึ่งเลือกตัวจำแนกโดยใช้จำนวนซัพพอร์ตเวกเตอร์เฉลี่ยต่ำสุด Bala และ Agrawal ได้เสนอ โอดีที-เอสวีเอ็ม (Optimal Decision Tree Based SVM (ODT-SVM))³⁷ ซึ่งใช้ตัววัดทางสถิติในการสร้างต้นไม้ตัดสินใจ เป็นต้น อย่างไรก็ตามวิธีเหล่านี้มีข้อเสียร่วมกัน นั่นคือตัวจำแนกที่เลือกมาอาจไม่สามารถจำแนกตัวอย่างจากบางประเภทข้อมูลได้อย่างสมบูรณ์ ดังนั้นบางเทคนิคจึงอนุญาตให้มีข้อมูลจากประเภทข้อมูลเหล่านั้นเข้ามามากกว่าหนึ่งปมในต้นไม้

นอกจากนี้ยังมีนักวิจัยบางกลุ่ม ที่ได้เสนอวิธีการสร้างต้นไม้ตัดสินใจสำหรับเอสวีเอ็มซึ่งจัดกลุ่มประเภทข้อมูลด้วยเทคนิควิธีการจัดกลุ่ม (clustering) เช่น Takahashi และ Abe ได้เสนอวิธีต้นไม้ตัดสินใจเอสวีเอ็ม (Decision-tree SVM)³⁸ ซึ่งใช้การคำนวณระยะทางแบบยูคลิด (Euclidean Distance) และมหาลาโนบิส (Mahalanobis Distance) เป็นตัวจำแนก Cheong และคณะ³⁹ ใช้แผนที่การจัดกลุ่มเอง (Self-Organizing Map - SOM) ในการแบ่งกลุ่มประเภทข้อมูลที่แต่ละปม Lei และ Govindaraju⁴⁰ สร้างต้นไม้ตัดสินใจที่ตัดประเภทข้อมูลออกครั้งละครึ่งหนึ่งของทั้งหมดโดยสร้างจากระยะทางเฉลี่ยระหว่างประเภทข้อมูล Liu และคณะ⁴¹ ใช้ K-means แบบก้าวหน้า (Progressive K-means) ซึ่งเป็นการทำ K-means บนปริภูมิอันดับสูงในการแบ่งประเภทข้อมูลสร้างต้นไม้ตัดสินใจ Madzarov และคณะเสนอวิธีเอสวีเอ็ม-บีดีที (SVM Binary Decision Tree (SVM-BDT))⁴² ซึ่งคำนวณ

ระยะทางบนปริภูมิเคอร์เนล (Kernel Space) เพื่อจัดกลุ่ม Dong และคณะ⁴³ สร้างต้นไม้ตัดสินใจโดยแบ่งกลุ่มประเภทข้อมูลจากความคล้ายคลึงกันของข้อมูล (Similarity Clustering) Lee และ Lee⁴⁴ ใช้ขั้นตอนวิธีเชิงพันธุกรรมในการแบ่งกลุ่มประเภทข้อมูลเพื่อสร้างต้นไม้ตัดสินใจ และมีนักวิจัยกลุ่มหนึ่งที่สร้างตัวจำแนกหลายประเภทบนพื้นฐานวิธีโอวีเอ เช่น Yang และคณะ⁴⁵ ใช้ฟังก์ชันแบ่ง (partition function) สามฟังก์ชันในการแยกประเภทข้อมูลออกจากกลุ่มครึ่งละหนึ่งประเภท Kumar และ Gopal⁴⁶ เสนอวิธีการตัดข้อมูลฝึกบางส่วนออกเพื่อลดเวลาในการฝึกโอวีเอ และยังคงให้ประสิทธิภาพการจำแนกเทียบเคียงกับโอวีเอดั้งเดิม García-Pedrajas และ Ortiz-Boyer⁴⁷ ได้เสนอวิธีการจำแนกหลายประเภทซึ่งผสมทั้งโอวีเอและโอวีโอเข้าด้วยกัน Soltani และคณะ⁴⁸ ได้พัฒนาวิธีโอวีโอโดยใช้ขั้นตอนวิธีเชิงพันธุกรรมแบบอาณานิคม (Ant Colony)

วิทยานิพนธ์นี้ได้ทำการศึกษาข้อดีและข้อเสีย ของวิธีจำแนกหลายประเภทด้วยเอสวีเอ็มแบบต่างๆ ที่มีนักวิจัยได้นำเสนอไว้ในปัจจุบัน และได้พัฒนาเทคนิคการจำแนกหลายประเภทแบบใหม่ 2 วิธี ซึ่งมีลักษณะเป็นโครงสร้างตัดสินใจแบบต้นไม้ ซึ่งสามารถจำแนกประเภทได้อย่างรวดเร็ว โดยทำงานในความซับซ้อนเชิงเวลาระหว่าง $O(\log_2 N)$ ถึง $O(N)$ สำหรับปัญหาที่มีจำนวนประเภทข้อมูลเป็น N ประเภท และให้ผลร้อยละความแม่นยำเทียบเคียงได้กับวิธีการดั้งเดิมที่นิยมใช้กัน โดยมีการทำการทดลองเพื่อเปรียบเทียบ ประสิทธิภาพของเทคนิคที่เสนอด้วยชุดข้อมูลจากพื้นที่เก็บข้อมูลการเรียนรู้ของเครื่องยูซีโอ (UCI Machine Learning Repository)⁴⁹ รวมถึงได้ทำการยืนยันผลการทดลอง ด้วยวิธีการทดสอบทางสถิติอีกด้วย ซึ่งเทคนิคใหม่ที่น่าเสนอในวิทยานิพนธ์นี้ จะเป็นประโยชน์อย่างมากกับปัญหาที่มีจำนวนประเภทข้อมูลขนาดใหญ่ หรือปัญหาที่ต้องการความรวดเร็วในการจำแนกสูงและความแม่นยำอยู่ในระดับเทียบเคียงกับวิธีดั้งเดิมได้

1.2. วัตถุประสงค์ของงานวิจัย

เพื่อทำการศึกษาข้อดีและข้อเสีย ของเทคนิคการจำแนกหลายประเภทด้วยเอสวีเอ็มแบบดั้งเดิมต่างๆ เพื่อออกแบบวิธีการใหม่ที่ทำงานได้อย่างรวดเร็ว และให้ผลการจำแนกเทียบเคียงได้กับวิธีการแบบดั้งเดิม

1.3. ขอบเขตของงานวิจัย

1. ทำการเปรียบเทียบประสิทธิภาพของเทคนิคที่นำเสนอใหม่กับเทคนิคดั้งเดิมต่างๆ ได้แก่ โอวีโอ โอวีเอ ดีดีเอจี เอดีเอจี บีทีเอส-จี และ ซี-บีทีเอส-จี ใช้วิธีการนับจำนวนครั้งเฉลี่ยที่ใช้ในการจำแนก และร้อยละโดยเฉลี่ยของความแม่นยำในการจำแนก
2. ทำการทดลองโดยใช้ชุดข้อมูลจากพื้นที่เก็บข้อมูลการเรียนรู้ของเครื่องยูซีไอ (UCI Machine Learning Repository) จำนวน 20 ชุดข้อมูล
3. ทำการจัดข้อมูลฝึกและข้อมูลทดสอบในแต่ละชุดข้อมูลด้วยวิธีตรวจสอบไขว้สิบพับ (10-fold cross validation)
4. ทำการยืนยันความแตกต่างอย่างมีนัยสำคัญทางสถิติด้วย Wilcoxon Signed Rank Test⁵⁰

1.4. ขั้นตอนและวิธีการดำเนินงานวิจัย

1. ศึกษาและเปรียบเทียบ เทคนิคแบบดั้งเดิมต่างๆ ได้แก่ โอวีโอ โอวีเอ ดีดีเอจี เอดีเอจี และ เทคนิคที่ใช้โครงสร้างแบบต้นไม้บางเทคนิค เช่น บีทีเอส และ ต้นไม้แตกกิ่งตามสารสนเทศ
2. ศึกษาวิธีการคำนวณและใช้งานค่าเอนโทรปี (Entropy) และวิธีการประมาณค่าความผิดพลาดโดยนัยทั่วไป (Generalization Error Estimation)
3. ทำการออกแบบ เทคนิคการจำแนกหลายประเภทที่มีโครงสร้างเป็นต้นไม้ตัดสินใจแบบใหม่ โดยใช้เอนโทรปีและค่าประมาณความผิดพลาดโดยนัยทั่วไป
4. ทำการทดลองเพื่อทดสอบประสิทธิภาพของเทคนิคที่ออกแบบขึ้น เพื่อเทียบกับวิธีการแบบดั้งเดิม
5. วิเคราะห์และเปรียบเทียบผลการทดลองเทียบกับวิธีการแบบดั้งเดิม
6. สรุปผลการวิจัยและจัดทำวิทยานิพนธ์

1.5. ประโยชน์ที่คาดว่าจะได้รับ

ได้เทคนิคการจำแนกหลายประเภทด้วยซัพพอร์ตเวกเตอร์แมชชีนแบบใหม่ที่สามารถจำแนกประเภทได้อย่างรวดเร็ว โดยทำงานในความซับซ้อนเชิงเวลาระหว่าง $O(\log_2 N)$ ถึง $O(N)$ สำหรับปัญหาที่มีจำนวนประเภทข้อมูลเป็น N ประเภท และให้ผลร้อยละความแม่นยำเทียบเคียงได้กับวิธีการ

ดั้งเดิม สามารถนำไปประยุกต์ใช้กับปัญหาที่มีจำนวนประเภทข้อมูลขนาดใหญ่ หรือปัญหาที่ต้องการความรวดเร็วในการจำแนกสูงและความแม่นยำอยู่ในระดับเทียบเคียงกับวิธีดั้งเดิม

1.6. ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์

วิทยานิพนธ์นี้แบ่งออกเป็น 5 บทดังนี้ บทที่ 1 บทนำ ซึ่งจะกล่าวถึงที่มาและความสำคัญของปัญหา รวมถึงวัตถุประสงค์และขั้นตอนการทำวิจัย บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการทำวิจัย บทที่ 3 กล่าวถึงรายละเอียดทั้งหมดของเทคนิคการจำแนกหลายประเภทที่ออกแบบขึ้นในงานวิจัยนี้ บทที่ 4 กล่าวถึงรายละเอียดการทดลองและผลการทดลอง บทที่ 5 เป็นข้อสรุปและข้อเสนอแนะจากงานวิจัย



บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีและงานวิจัยที่เกี่ยวข้อง โดยในส่วนของทฤษฎีที่เกี่ยวข้อง ประกอบด้วยทฤษฎีเนื้อหาพื้นฐานของซัพพอร์ตเวกเตอร์แมชชีน ประสิทธิภาพทั่วไปของซัพพอร์ตเวกเตอร์แมชชีน (Generalization Performance of SVM) ทฤษฎีพื้นฐานของซัพพอร์ตเวกเตอร์แมชชีนแบบจำแนกหลายประเภท (Multo-class SVM) การใช้รหัสขอกแก้ไขข้อผิดพลาด (Error-Correcting Output Code) แบบ 2 ค่า (Binary) การใช้รหัสขอกแก้ไขข้อผิดพลาดแบบ 3 ค่า (Ternary) และในเนื้อหาส่วนของงานวิจัยที่เกี่ยวข้อง ประกอบด้วยงานวิจัยต่างๆ ที่ได้ทำการศึกษา เพื่อให้ได้แนวคิดในการพัฒนาต่อยอดในงานวิจัยนี้

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine)

ซัพพอร์ตเวกเตอร์แมชชีน – เอสวีเอ็ม (Support Vector Machine - SVM)¹⁻² เป็นตัวจำแนกประเภทข้อมูลชนิดหนึ่งซึ่งมีวิธีการทำงาน คือ การพยายามจำแนกข้อมูล 2 ประเภทด้วยระนาบหลายมิติแบ่งแยกที่ดีที่สุด (optimal separating hyperplane) ซึ่งให้ระยะขอบระหว่างข้อมูลมากที่สุด (maximum margin)

2.1.1.1 เอสวีเอ็มสำหรับข้อมูลที่สามารถแบ่งแยกได้เชิงเส้น (SVM for Linearly Separable Data)

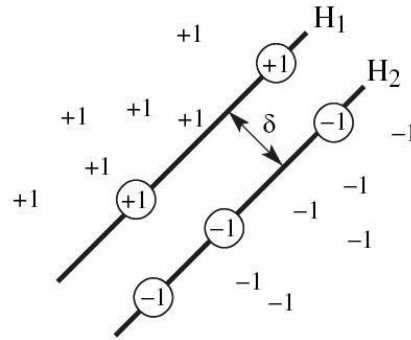
ลองพิจารณาปัญหาการแบ่งแยกสองประเภท กลุ่มเวกเตอร์ข้อมูลใช้ฝึกเป็น

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \quad (1)$$

โดยที่ $x_i \in R^N$ และ $y_i \in \{-1, +1\}$ สำหรับ $i=1, \dots, l$, เมื่อ x_i เป็นข้อมูลตัวอย่าง และ y_i เป็นฉลากประเภทข้อมูล เราจะได้พื้นผิวตัดสินใจเชิงเส้น (linear decision surface) ที่นิยามได้โดยสมการ

$$w \cdot x + b = 0 \quad (2)$$

ในการแบ่งข้อมูลออกเป็น 2 ฝั่ง อาจมีระนาบหลายมิติที่สามารถแบ่งข้อมูล 2 ฝั่งหลายระนาบ แต่ระนาบที่เป็นเป้าหมายของการเรียนรู้ที่เราต้องการ คือระนาบที่ให้ค่า $w \in R^N$ และปริมาณสเกลาร์ b ที่ให้มีระยะขอบระหว่างตัวอย่างบวกและตัวอย่างลบที่มากที่สุด ดังรูปที่ 1



รูป 1 ตัวอย่างของพื้นผิวตัดสลับใจและระยะขอบ [51]

จากรูปที่ 1 ระนาบหลายมิติสามารถแบ่งแยกได้ก็ต่อเมื่อ

$$w \cdot x_i + b > 0 \quad \text{ถ้า } y_i = +1 \quad (3)$$

$$w \cdot x_i + b < 0 \quad \text{ถ้า } y_i = -1 \quad (4)$$

ระนาบหลายมิติที่เราจะพิจารณานั้น ตัวแปร w กับ b ต้องเข้าเงื่อนไขที่ว่า $\min_i |w \cdot x_i + b| = 1$ และ เราจะได้ระนาบหลายมิติสำหรับจำแนกประเภทข้อมูลที่อยู่ในรูปแบบมาตรฐานซึ่งเป็นไปตามสมการ

$$w \cdot x_i + b \geq +1 \quad \text{ถ้า } y_i = +1 \quad (5)$$

$$w \cdot x_i + b \leq -1 \quad \text{ถ้า } y_i = -1 \quad (6)$$

ซึ่งสามารถรวมเป็นสมการเดียวได้ว่า

$$y_i (w \cdot x_i + b) \geq 1, \quad \forall i. \quad (7)$$

ระยะห่าง $d(w, b; x_i)$ ของจุด x_i จากระนาบหลายมิติ (w, b) คือ

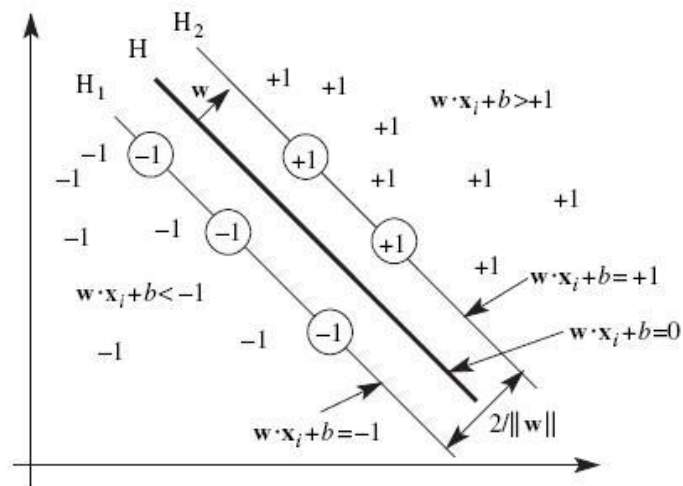
$$d(w, b; x_i) = \frac{|w \cdot x_i + b|}{\|w\|} \quad (8)$$

ระนาบหลายมิติที่ดีที่สุดจะให้ระยะขอบระหว่างข้อมูลทั้งสองประเภทมากที่สุด ρ ซึ่งเป็นไปตามสมการที่ (7) โดยระยะขอบหาได้จาก

$$\rho(w, b) = \min_{x_i, y_i = -1} d(w, b; x_i) + \min_{x_i, y_i = +1} d(w, b; x_i)$$

$$\begin{aligned}
&= \min_{x_i, y_i = -1} \frac{|w \cdot x_i + b|}{\|w\|} + \min_{x_i, y_i = +1} \frac{|w \cdot x_i + b|}{\|w\|} \\
&= \frac{1}{\|w\|} \left(\min_{x_i, y_i = -1} |w \cdot x_i + b| + \min_{x_i, y_i = +1} |w \cdot x_i + b| \right) \\
&= \frac{2}{\|w\|} \tag{9}
\end{aligned}$$

ระนาบหลายมิติแบ่งแยกที่ดีที่สุดและระยะขอบที่มากที่สุด สามารถแสดงให้เห็นได้ ดังรูปที่ 2



รูป 2 ระนาบหลายมิติแบ่งแยกที่ดีที่สุดและระยะขอบที่มากที่สุด [51]

ระนาบหลายมิติซึ่งแบ่งแยกได้ดีที่สุด จะให้ค่า $\frac{1}{2} \|w\|^2$ ให้น้อยที่สุดโดยมีเงื่อนไขคือ

$$y_i(w \cdot x_i + b) \geq 1 \text{ สำหรับ } i = 1, \dots, l$$

ปัญหานี้เป็นปัญหาการทำให้ดีที่สุดในสมการกำลังสอง (Quadratic optimization problem) ซึ่งเขียนออกมาในรูปของสมการลากรอง โดยมีตัวคูณลากรอง α_i , $i = 1, \dots, l$

$$L = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i ((y_i(w \cdot x_i + b)) - 1) \tag{10}$$

สิ่งที่เราต้องทำคือ หาค่าต่ำสุดเมื่อเทียบกับ w , b บนเงื่อนไข $\alpha_i \geq 0$ ซึ่งทำได้โดย

$$\frac{\partial L}{\partial b} = 0 \quad \rightarrow \quad \sum_{i=1}^l \alpha_i y_i = 0 \tag{11}$$

$$\frac{\partial L}{\partial w} = 0 \quad \rightarrow \quad w = \sum_{i=1}^l \alpha_i y_i x_i \tag{12}$$

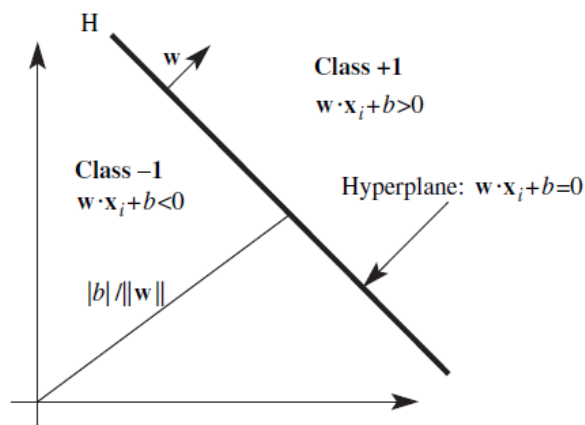
ปัญหาการหาค่าเหมาะที่สุดจะเป็นการทำ $L = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$ ให้มากที่สุด

ที่สอดคล้องเงื่อนไข $\sum_{i=1}^l \alpha_i y_i = 0$ โดยที่ $\alpha_i \geq 0, i = 1, \dots, l$

ตัวอย่างข้อมูลฝึกแต่ละตัวจะมีตัวคูณลากรองหนึ่งตัวอย่างที่มีค่า $\alpha_i \geq 0$ เรียกว่าซัพพอร์ต-เวกเตอร์ ซึ่งจากการแก้ปัญหาก็จะทำให้ได้ฟังก์ชันการตัดสินใจ คือ

$$f(x) = \text{sign} \left(\sum_{i=1}^l \alpha_i y_i (x_i \cdot x) + b \right) \quad (13)$$

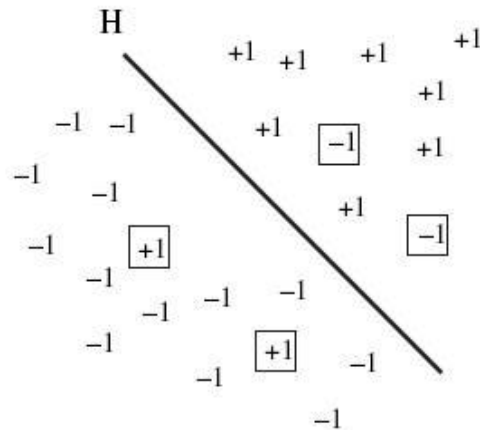
ตัวอย่างระนาบแบ่งหลายมิติที่ได้จากการฝึก เป็นดังรูปที่ 3



รูป 3 ตัวอย่างระนาบแบ่งหลายมิติที่ได้จากการฝึก [51]

2.1.1.2 เอสวีเอ็มแบบระยะขอบอ่อน (Soft Margin SVM)

เอสวีเอ็มที่ได้กล่าวไปข้างต้นนั้น สามารถทำงานได้กับข้อมูลที่แบ่งส่วนประเภทข้อมูลที่สมบูรณ์เท่านั้น แต่ในความเป็นจริง ข้อมูลฝึกย่อมมีความไม่สมบูรณ์ หรือมีข้อมูลรบกวน (noise) อยู่ด้วย ดังเช่นในรูปที่ 4



รูป 4 ชุดข้อมูลที่ข้อมูลที่ไม่สมบูรณ์ [51]

จึงมีการเพิ่มตัวแปร ξ ซึ่งเป็นตัวปรับข้อมูลที่ผิดพลาด เรียกว่า ตัวแปรหย่อน (Slack variable)

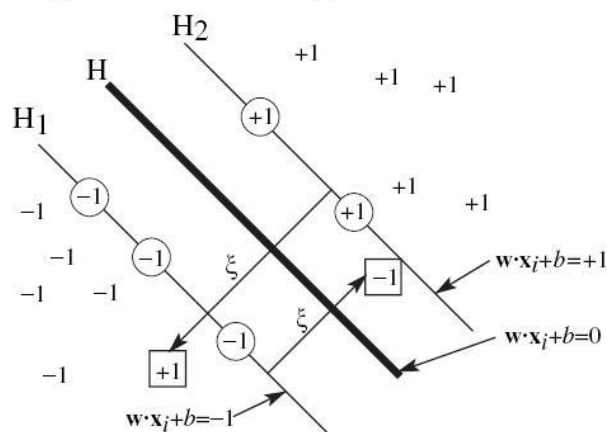
สำหรับประเภทข้อมูล + นิยาม ξ ว่า

$$\xi_i(w, b) = \begin{cases} 0 & \text{if } w \cdot x_i + b \geq +1 \\ 1 - (w \cdot x_i + b) & \text{if } w \cdot x_i + b \leq +1 \end{cases} \quad (14)$$

ส่วนสำหรับประเภทข้อมูล - นิยาม ξ ว่า

$$\xi_i(w, b) = \begin{cases} 0 & \text{if } w \cdot x_i + b \geq -1 \\ 1 + (w \cdot x_i + b) & \text{if } w \cdot x_i + b \leq -1 \end{cases} \quad (15)$$

การปรับค่าข้อมูลที่ผิดพลาดด้วยตัวแปรหย่อน แสดงได้ดังรูปที่ 5



รูป 5 การปรับค่าข้อมูลที่ผิดพลาดด้วยตัวแปรหย่อน [51]

เราสามารถรวมสมการที่ (14) และ (15) เข้าด้วยกันเป็น

$$\xi_i(w, b) = \begin{cases} 0 & \text{if } y_i(w \cdot x_i + b) \geq +1 \\ 1 - y_i(w \cdot x_i + b) & \text{if } y_i(w \cdot x_i + b) \leq -1 \end{cases} \quad (16)$$

เมื่อเรานำตัวแปรหย่อนไปรวมกับสมการที่ (3) และ (4) จะได้สมการเป้าหมายเป็น

$$\begin{aligned} w \cdot x_i + b &\geq +1 - \xi_i & \text{if } y_i = +1 \\ w \cdot x_i + b &\leq -1 + \xi_i & \text{if } y_i = -1 \\ \xi_i &> 0, \forall_i \end{aligned} \quad (17)$$

ปัญหาการหาค่าเหมาะสมที่สุดของจะกลายเป็นการทำให้ค่า $L = \frac{1}{2} \|w\|^2 - C \sum_{i=1}^l \xi_i$ มีค่าน้อยที่สุดโดยมีเงื่อนไข คือ $y_i(w \cdot x_i + b) \geq +1 - \xi_i$, $\xi_i \geq 0$ สำหรับ $i=1, \dots, l$

ปัญหานี้เป็นปัญหาการทำให้ดีที่สุดของสมการกำลังสอง (quadratic optimization problem) เขียนออกมาในรูปของสมการลากรอง โดยมีตัวคูณลากรองคือ α_i และ μ_i , $i=1, \dots, l$

$$L = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i ((y_i(w \cdot x_i + b)) - 1 + \xi_i) - \sum_{i=1}^l \mu_i \xi_i \quad (18)$$

สิ่งที่เราต้องทำคือ หาค่าต่ำสุดเมื่อเทียบกับ w, b, ξ_i บนเงื่อนไข $\alpha_i, \mu_i \geq 0$ ซึ่งทำได้โดย

$$\frac{\partial L}{\partial b} = 0 \quad \rightarrow \quad \sum_{i=1}^l \alpha_i y_i = 0 \quad (19)$$

$$\frac{\partial L}{\partial w} = 0 \quad \rightarrow \quad w = \sum_{i=1}^l \alpha_i y_i x_i \quad (20)$$

$$\frac{\partial L}{\partial \xi_i} = 0 \quad \rightarrow \quad C - \alpha_i - \mu_i = 0 \quad (21)$$

ปัญหาการหาค่าเหมาะสมที่สุดจะกลายเป็นการทำ $L = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$

ให้มากที่สุดบนเงื่อนไข $\sum_{i=1}^l \alpha_i y_i = 0$ โดยที่ $0 \leq \alpha_i \leq C$ และ $i=1, \dots, l$ และจะได้ฟังก์ชันการ

ตัดสินใจ คือ

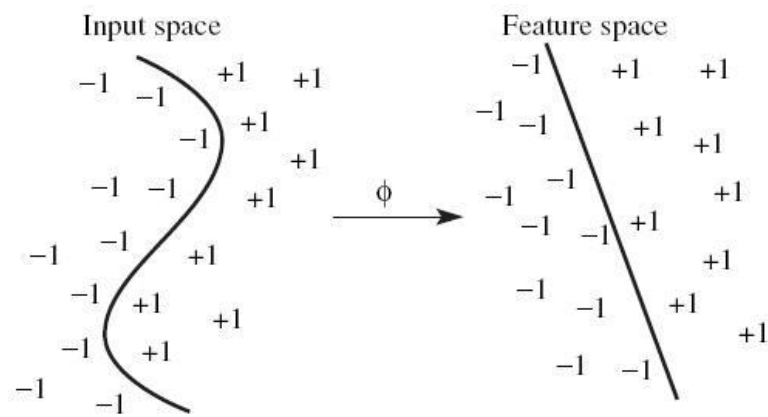
$$f(x) = \text{sign} \left(\sum_{i=1}^l \alpha_i y_i (x_i \cdot x) + b \right) \quad (22)$$

2.1.1.3 เอสวีเอ็มแบบไม่เชิงเส้นและฟังก์ชันเคอร์เนล (Nonlinear SVM and Kernel Function)

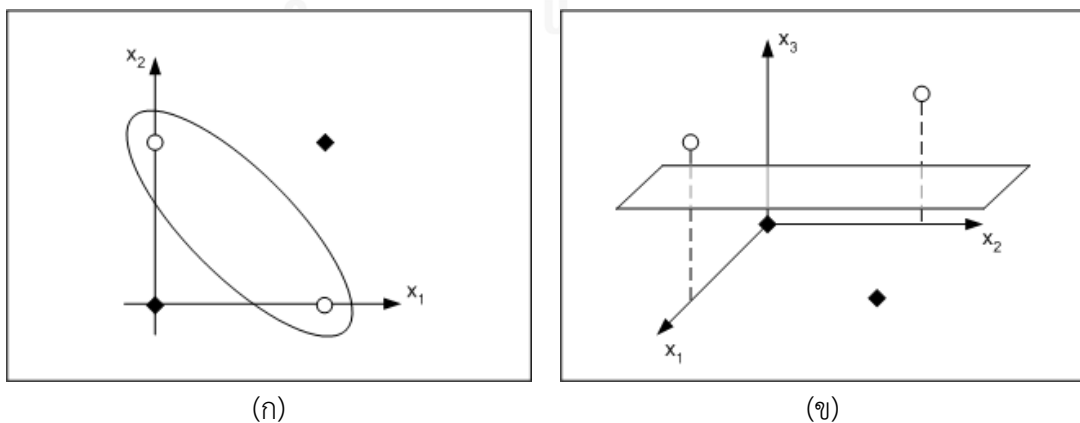
ในกรณีส่วนใหญ่ การหาระนาบหลายมิติเชิงเส้นที่เหมาะสมไม่สามารถทำได้ จึงต้องใช้เทคนิคเข้ามาช่วย นั่นคือการแมปข้อมูลตัวอย่างไปยังปริภูมิอันดับสูงโดยใช้ฟังก์ชันการแมป (Mapping Function) Φ แล้วจึงค่อยทำการฝึกและจำแนกประเภทข้อมูล (ดังรูปที่ 6 และ 7)

$$\Phi : R^N \rightarrow F$$

$$x \mapsto \Phi(x)$$



รูป 6 การแมปข้อมูลจากปริภูมินำเข้าไปยังมิติปริภูมิอันดับสูง [51]



รูป 7 ตัวอย่างระนาบหลายมิติสำหรับแบ่งแยก (ก) ปริภูมินำเข้า (ข) มิติปริภูมิอันดับสูง [51]

คุณสมบัติที่ดีอีกประการของเอสวีเอ็มคือ ไม่จำเป็นต้องรู้รูปแบบที่ชัดเจนของ Φ สิ่งที่เรา
 ต้องทำคือ นิยามผลคูณภายในในมิติปริภูมิอันดับสูงซึ่งเรียกว่า ฟังก์ชันเคอร์เนล (*kernel function*)
 เท่านั้น

$$K(x, y) = \Phi(x) \cdot \Phi(y) \quad (23)$$

ฟังก์ชันการตัดสินใจจะเป็นดังนี้

$$f(x) = \text{sign} \left(\sum_{i=1}^l \alpha_i y_i K(x_i, x) + b \right) \quad (24)$$

ตัวอย่างของฟังก์ชันเคอร์เนลแสดงดังตารางที่ 1



ตาราง 1 ตัวอย่างของฟังก์ชันเคอร์เนล

Kernel	Formula
Linear	$K(x, y) = x \cdot y$
Polynomial	$K(x, y) = (1 + x \cdot y)^d$
Sigmoid	$K(x, y) = \tanh(\alpha(x \cdot y) + \beta)$
Exponential RBF	$K(x, y) = \exp(-\gamma \ x - y\)$
Gaussian RBF	$K(x, y) = \exp(-\gamma \ x - y\ ^2)$
Spline	$K(x, y) = 1 + (x \cdot y) + \frac{1}{2}(x \cdot y) \min(x, y) - \frac{1}{6} \min(x, y)^3$
Multi-quadratic	$K(x, y) = -\sqrt{\ x - y\ ^2 + c^2}$
Wavelet	$K(x, y) = \prod_{i=1}^n \cos\left(1.75 \times \left(\frac{x_i - y_i}{a}\right)\right) \exp\left(-\frac{\ x_i - y_i\ ^2}{2a}\right)$
Moderate Decreasing	$K(x, y) = k \left[\exp\left(\frac{\gamma}{\ x - y\ ^2 + \sigma^2}\right) - 1 \right]$

ฟังก์ชันที่สามารถเป็นเคอร์เนลได้ จะต้องเป็นไปตามตั้งทฤษฎีของเมอร์เซอร์ (Mercer's theorem) ซึ่งกล่าวไว้ว่า สำหรับทุกๆ ฟังก์ชันที่สมมาตร $K(x, y)$ ในปริภูมินำเข้าสามารถแสดงผลคูณภายในของปริภูมิอันดับสูงนั้น ถ้า

$$\iint K(x, y) g(x) g(y) dx dy \geq 0 \quad (25)$$

สามารถหาได้สำหรับทุกๆ $g \neq 0$ สำหรับ $\int g^2(u) du < \infty$ แล้ว ฟังก์ชันเคอร์เนล K สามารถถูกขยายออกมาในรูปของ Φ_i

$$K(x, y) = \sum_{i=1}^{\infty} \lambda_i \Phi_i(x) \Phi_i(y) \quad (26)$$

ด้วย $\lambda_i \geq 0$ ในการนี้ ฟังก์ชันการแมปจากปริภูมินำเข้าไปยังปริภูมิอันดับสูงจะบรรยายได้ว่า

$$\Phi : x \rightarrow \left(\sqrt{\lambda_1} \Phi_1(x), \sqrt{\lambda_2} \Phi_2(x), \dots \right) \quad (27)$$

K สามารถเป็นผลคูณภายในได้ว่า

$$\Phi(x) \cdot \Phi(y) = \sum_{i=1}^{\infty} \lambda_i \Phi_i(x) \Phi_i(y) = K(x, y) \quad (28)$$

2.1.2 ประสิทธิภาพโดยนัยทั่วไปของซัพพอร์ตเวกเตอร์แมชชีน (Generalization Performance of Support Vector Machine)

ประสิทธิภาพโดยนัยทั่วไปในการเรียนรู้ของซัพพอร์ตเวกเตอร์แมชชีน^{49, 51} ด้วยความเชื่อมั่นภายใต้ ขอบเขตความผิดพลาด (error bound) จะขึ้นอยู่กับแนวคิดของหลักการความเสี่ยงเชิงโครงสร้างน้อยที่สุด (Structure Risk Minimization) ซึ่งสามารถประมาณได้ด้วยทั้งเทอมของความซับซ้อนของแบบจำลอง (มิติ VC ของฟังก์ชันประมาณค่า) และเทอมคุณภาพการพหุเหมาะสมกับข้อมูลฝึก (ความผิดพลาดเชิงประจักษ์) โดยหาก พิจารณาชุดข้อมูล X ซึ่งประกอบด้วยจำนวนข้อมูล m ตัว ในมิติ n จะได้ว่าค่าความเสี่ยง $R(\alpha)$ ด้วยความน่าจะเป็นน้อยสุด $1-\delta$ จะถูกจำกัดขอบเขตด้วยสมการ

$$R(\alpha) \leq \frac{l}{m} + \sqrt{\frac{c}{m} \left(\frac{R^2}{\Delta^2} \log^2 m + \log \frac{1}{\delta} \right)} \quad (29)$$

โดยที่ c เป็นค่าคงที่สำหรับความน่าจะเป็นการกระจาย l เป็นจำนวนตัวอย่างที่ถูกติดป้ายที่ขอบเขตน้อยกว่า Δ , R เป็นรัศมีที่น้อยที่สุดของทรงกลมที่ประกอบด้วยจุดข้อมูลทั้งหมด และ Δ เป็นระยะทางระหว่างระนาบหลายมิติ และจุดที่ใกล้ที่สุดของชุดข้อมูลฝึก (ขนาดของขอบเขต - margin size) ซึ่งสมการนี้แสดงโดยใช้ความรู้เรื่อง ความผิดพลาดเชิงประจักษ์ (empirical error) และมิติ VC (VC dimension)

2.1.3 ซัพพอร์ตเวกเตอร์แมชชีนแบบจำแนกหลายประเภท (Multiclass Support Vector Machine)

วิธีการพื้นฐานที่ใช้กันโดยทั่วไปในงานจำแนกหลายประเภท ประกอบด้วยวิธีโอวีเอ (One-Versus-All (OVA)) และโอวีโอ (One-Versus-One (OVO))

2.1.3.1 วิธีโอวีเอ (One-Versus-All (OVA))

วิธีการจำแนกหลายประเภทที่เก่าแก่ที่สุด คือวิธีโอวีเอ² โดยหลักการคือสร้างแบบจำลองเอสซีเอ็มจำนวน N ตัว โดยที่ N แทนจำนวนประเภท ให้เอสซีเอ็มตัวที่ i ถูกฝึกด้วยตัวอย่างที่มีประเภทข้อมูล i ด้วยป้ายบวก (positive label) และตัวอย่างฝึกอื่นๆ เป็นป้ายลบ (negative label) ดังนั้นจากการใช้ข้อมูลฝึกจำนวน l ตัว $(x_1, y_1), \dots, (x_l, y_l)$ โดยที่ $x_i \in R^n, i = 1, \dots, l$ และ $y_i \in \{1, \dots, k\}$ เป็นประเภทของ x_i แล้วเอสซีเอ็มตัวที่ i จะแก้ปัญหา :

$$\begin{aligned}
\min_{w^i, b^i, \xi^i} \quad & \frac{1}{2} (w^i)^T w^i + C \sum_{j=1}^l \xi_j^i (w^i)^T \\
& (w^i)^T \phi(x_j) + b^i \geq 1 - \xi_j^i \text{ if } y_i = i \\
& (w^i)^T \phi(x_j) + b^i \leq \xi_j^i - 1 \text{ if } y_i \neq i \\
& \xi_j^i \geq 0, j = 1, \dots, l
\end{aligned} \tag{30}$$

โดยที่ข้อมูลฝึก x_j จะถูกแมปไปสู่ปริภูมิอันดับสูงด้วยฟังก์ชัน ϕ และ C เป็นพารามิเตอร์ปรับ

การ minimize $\frac{1}{2} (w^i)^T w^i$ หมายความว่า เราต้องการ maximize $\frac{2}{\|w^i\|}$ ซึ่งก็คือพื้นที่ระหว่าง 2 กลุ่มข้อมูล ถ้าในกรณีข้อมูลที่ข้อมูลไม่สามารถแบ่งแยกได้เชิงเส้น จะมีเงื่อนไขปรับค่าของ $C \sum_{j=1}^l \xi_j^i$ ที่สามารถลดจำนวนความผิดพลาดจากการฝึก โดยแนวคิดพื้นฐานของเอสบีเอ็มคือการหาความสมดุลระหว่างเงื่อนไขปกติ $\frac{1}{2} (w^i)^T w^i$ และความผิดพลาดจากการฝึก

หลังจากแก้ปัญหา จะมีฟังก์ชันตัดสินใจจำนวน k ตัว คือ

$$(w^1)^T \phi(x) + b^1 \text{ ถึง } (w^N)^T \phi(x) + b^N \tag{31}$$

เรากล่าวได้ว่า x จะอยู่ในประเภทที่มีค่าจากฟังก์ชันตัดสินใจมากที่สุด

$$\text{Class of } X \equiv \arg \max_{i=1, \dots, N} (w^i)^T \phi(x) + b^i \tag{32}$$

โดยในทางปฏิบัติแล้ว เราแก้ปัญหาคู่ (dual problem) ของ (1) ที่ซึ่งจำนวนตัวแปรเท่ากับสมการที่ (31) แทน

2.1.3.2 วิธีโอวีโอ (One-Versus-One (OVO))

วิธีโอวีโอ (OVO)⁶⁻⁷ สร้างตัวจำแนก (classifier) จำนวน $N(N-1)/2$ ตัว โดยที่แต่ละตัวจะถูกฝึกด้วยข้อมูลจาก 2 ประเภท คือ ประเภทที่ i และประเภทที่ j แล้วแก้ไขปัญหา :

$$\begin{aligned}
\min_{w^{ij}, b^{ij}, \xi^{ij}} \quad & \frac{1}{2} (w^{ij})^T w^{ij} + C \sum_t \xi_t^{ij} (w^{ij})^T \\
& (w^{ij})^T \phi(x_t) + b^{ij} \geq 1 - \xi_t^{ij} \quad \text{if } y_t = i \\
& (w^{ij})^T \phi(x_t) + b^{ij} \leq \xi_t^{ij} - 1 \quad \text{if } y_t = j \\
& \xi_t^{ij} \geq 0
\end{aligned} \tag{33}$$

มีวิธีที่แตกต่างกันหลายวิธี ในการทดสอบข้อมูลหลังจากที่ตัวจำแนกจำนวน $N(N-1)/2$ ตัวถูกสร้างขึ้นแล้ว โดยวิธีที่มักใช้กันคือการโหวต (vote) โดยถ้าผลจาก $\text{sign}((w^i)^T \phi(x) + b)$ บอกว่า

x อยู่ในประเภทที่ i คะแนนการโหวตของประเภท i จะเพิ่มขึ้น ในทางตรงกันข้ามก็จะเพิ่มคะแนนโหวตให้กับประเภท j และหลังจากเปรียบเทียบทุกคู่แล้ว จะทำนายว่า x มีประเภทเป็นตัวที่ได้รับคะแนนโหวตสูงสุด โดยวิธีการนี้เรียกว่า “Max-Wins”⁸ หรือมากที่สุดชนะ โดยหากในกรณีที่มีมากกว่า 2 ประเภทได้รับคะแนนโหวตเท่ากันให้เลือกตอบแบบสุ่มด้วยความน่าจะเป็นเท่ากัน

โดยในทางปฏิบัติแล้ว เราแก้ปัญหาคู่ (dual problem) ของ (33) ที่ซึ่งจำนวนข้อมูลเหมือนกันกับจำนวนข้อมูลใน 2 ประเภท ดังนั้นถ้าโดยเฉลี่ยแล้วแต่ละประเภทมี l/N จุดข้อมูล เราต้องแก้สมการกำลังสองจำนวน $N(N-1)/2$ สมการ โดยที่แต่ละตัวมี $2l/N$ ตัวแปร

2.1.4 การใช้รหัสข่าออกแก้ไขข้อผิดพลาด (อีซีไอซี) (Error-Correcting Output Code (ECOC))

เราสามารถเข้ารหัสข่าออกแก้ไขข้อผิดพลาด (อีซีไอซี)¹⁵ เป็นตัวแทนของประเภทข้อมูลที่ต้องการจำแนก โดยแสดงอยู่ในรูปของรหัสข่าออกแบบกระจาย (distributed output code) โดยแต่ละประเภทจะถูกกำหนดให้แทนด้วยชุดของไบนารีสตริง (Binary String) ความยาว n ซึ่งเรียกสตริงเหล่านี้ว่า “คำรหัส” (codeword) โดยจะมีตัวจำแนกที่จะถูกเรียนรู้ตามแต่ละ bit ในคำรหัส โดยระหว่างการฝึกตัวอย่างประเภท i จะถูกระบุโดยคำรหัสสำหรับประเภท i

ค่าของ x จะถูกจำแนกโดยการประเมินแต่ละฟังก์ชันไบนารี n ที่ผลิตสตริง s ความยาว n บิตออกมา จากนั้นนำไปเปรียบเทียบกับแต่ละคำรหัส k และ x จะถูกจำแนกให้เป็นประเภทที่มีความใกล้เคียงกับคำรหัสของประเภทนั้นมากที่สุด (ด้วยวิธีการวัดระยะทางวิธีใดวิธีหนึ่ง)

ตัวอย่างเช่น ในตารางที่ 2 ซึ่งมีรหัสที่มีการกระจาย 6 บิต สำหรับประเภททั้งสิ้น 10 ประเภท จากการสังเกตจะเห็นว่าแต่ละประเภทมีรหัสไม่ซ้ำกัน โดยหากให้ทำนายประเภทของตัวอย่างใหม่ x ฟังก์ชัน $f_{vl}, f_{hl}, f_{dl}, f_{ol}$ และ f_{or} จะประเมินออกมาเป็นสตริง 6 บิต เช่น 110001 จากนั้นจะนำไปคำนวณระยะห่างกับคำรหัสทั้ง 10 จะได้ว่าคำรหัสที่ใกล้เคียงที่สุด (ตามวิธีการวัดระยะทางแบบ Hamming) คือ 110000 ซึ่งตรงกันกับประเภทที่ 4 ดังนั้นจึงทำนายมา x มีประเภทเป็น $f(x) = 4$

ตาราง 2 ตัวอย่างรหัสที่มีการกระจาย 6 บิต

Class	Code Word					
	vl	hl	dl	cc	ol	or
0	0	0	0	1	0	0
1	1	0	0	0	0	0
2	0	1	1	0	1	0
3	0	0	0	0	1	0
4	1	1	0	0	0	0
5	1	1	0	0	1	0
6	0	0	1	1	0	1
7	0	0	1	0	0	0
8	0	0	0	1	0	0
9	0	0	1	1	0	0

ในตารางที่ 3 ซึ่งมีรหัสแก้ไขความผิดพลาด 15 บิต สำหรับประเภททั้งสิ้น 10 ประเภท โดยหากให้ทำนายประเภทของตัวอย่างใหม่ x ฟังก์ชัน $f_0(x), \dots, f_{14}(x)$ จะประเมินออกมาเป็นสตริง 15 บิต แล้วจับคู่ไปยังค่ารหัสที่ใกล้เคียงที่สุด ซึ่งรหัสนี้สามารถแก้ไขความผิดพลาดได้ 3 จาก 15 บิต

ตาราง 3 ตัวอย่างรหัสที่มีการกระจาย 15 บิต

Class	Code Word														
	f(0)	f(1)	f(2)	f(3)	f(4)	f(5)	f(6)	f(7)	f(8)	f(9)	f(10)	f(11)	f(12)	f(13)	f(14)
0	1	1	0	0	0	0	1	0	1	0	0	1	1	0	1
1	0	0	1	1	1	1	0	1	0	1	1	0	0	1	0
2	1	0	0	1	0	0	0	1	1	1	1	0	1	0	1
3	0	0	1	1	0	1	1	1	0	0	0	0	1	0	1
4	1	1	1	0	1	0	1	1	1	0	0	0	0	0	1
5	0	1	0	0	1	1	0	1	1	1	0	0	0	0	1
6	1	0	1	1	1	0	0	0	0	1	0	1	0	0	1
7	0	0	0	1	1	1	1	0	1	0	1	1	0	0	1
8	1	1	0	1	0	1	1	0	0	1	0	0	0	1	1
9	0	1	1	1	0	0	0	0	1	0	1	0	0	1	1

จากตัวอย่างข้างต้น จะเห็นว่าวิธีการวัดคุณภาพของรหัสขาออกแก้ไขข้อผิดพลาด จะขึ้นอยู่กับระยะทางที่น้อยที่สุดระหว่างคำรหัสคู่ใดๆ หรือระยะทางแฮมมิง โดยหากค่านี้เป็น d แล้วจะได้ว่ารหัสสามารถแก้ไขความผิดพลาดได้อย่างน้อย $\lfloor (d-1)/2 \rfloor$ บิต โดยจากตารางที่ 3 จะเห็นว่ามีการใช้ระยะทาง Hamming น้อยที่สุดระหว่างรหัสคู่ใดๆ เป็น 7 ดังนั้นจึงสามารถรองรับความผิดพลาดได้ 3 บิต

2.1.5 การใช้รหัสขาออกแก้ไขข้อผิดพลาดแบบ 3 ค่า (Ternary Error-Correcting Output Code (Ternary ECOC))

การใช้รหัสขาออกแก้ไขข้อผิดพลาดแบบดั้งเดิมซึ่งมี 2 ค่า ได้ถูกต่อขยายออกเป็น 3 ค่า (Ternary)²⁴, 52

$M \in \{-1, 0, 1\}^{N \times n}$ โดยสัญลักษณ์ 0 หมายถึงประเภทนั้นไม่ได้รับการพิจารณาโดยตัวจำแนกประเภท (dichotomy) ซึ่งการทำเช่นนี้จะเพิ่มจำนวนที่เป็นไปได้ของตัวจำแนกประเภทที่สร้างขอบเขตการตัดสินใจที่ต่างกัน ซึ่งจะช่วยให้เพิ่มความแม่นยำในการจำแนกประเภทให้มากขึ้น

	h_1	h_2	h_3	h_4	h_5	h_6	h_7
c_1	Black	White	White	White	Black	White	White
c_2	Black	White	Black	Black	Black	Black	Black
c_3	White	Black	Black	Black	White	White	Black
c_4	Black	Black	White	Black	White	Black	White

รูป 8 ตัวอย่างเมทริกซ์รหัสค่าแบบ 3 ค่า

จากรูปที่ 8 อักษร c หมายถึงประเภท และอักษร h หมายถึงตัวจำแนกประเภท (dichotomy) ช่องสีขาว สีเทา สีดำ หมายถึงจำแนกประเภทให้เป็น 1, 0, -1 ตามลำดับ ซึ่งในการวัดระยะทางเพื่อถอดรหัส (decode) ว่าตัวอย่าง x อยู่ในประเภทใด มีผู้ที่เสนอเอาไว้หลายวิธี เช่น Hamming (HD) inverse Hamming (IHD) Attenuated Euclidean distance (AED) and Loss-based decoding (LB) เป็นต้น

Inverse Hamming Distance (IHD) ทำโดย ให้ $D(x) = [d(x, y^1), d(x, y^2), \dots, d(x, y^{N_c})]$ เป็นนิยามของเซตของระยะทางโดยประมาณจากรหัสที่ต้องการทดสอบไปยังรหัสของประเภทจำนวน N_c ประเภท และให้ Δ เป็นเมทริกซ์ที่ถูกสร้างขึ้นโดยระยะทางแฮมมิง ระหว่างคำรหัส M ซึ่งแต่ละตำแหน่งของ Δ นิยามโดย $\Delta(i, j) = d(y^i, y^j)$ ซึ่ง $d(y^i, y^j)$ นิยามระยะทางแฮมมิง ระหว่างคำรหัส i และ j ถ้าเซต D ถูกประเมินค่าโดยใช้ระยะทาง Hamming แล้ว Δ จะสามารถถูกแปลงเพื่อหาเวกเตอร์ $Q = [q_1, q_2, \dots, q_{N_c}]$ ซึ่ง $Q = \Delta^{-1}D^T$ โดยวิธีนี้มีพื้นฐานมาจากทฤษฎีทำให้ระยะทางแฮมมิงให้น้อยที่สุด (Hamming minimization Theory) ซึ่งมีคุณสมบัติเหมือนกันสำหรับกรณี 3 ค่า (ternary case)

Attenuated Euclidean distance (AED) วิธีนี้เป็นการประยุกต์วิธี Euclidean distance ให้สามารถรองรับค่าศูนย์ได้ โดยนิยามให้การถอดรหัสเป็น $d(x,y^i) = \sqrt{\sum_{j=1}^n |y_j^i| (x_j - y_j^i)^2}$ ซึ่งองค์ประกอบ y_j^i จะปฏิเสธความผิดพลาดที่ถูกรวบรวมจากสัญลักษณ์ศูนย์ที่คำรหัสของประเภท i ซึ่งด้วยเทคนิคนี้ เราสามารถพิจารณาความเกี่ยวข้องของข้อมูลโดยใช้ตำแหน่งของรหัสเท่านั้น และถ้าต้องการต่อขยายโดยใส่ความสำคัญของศูนย์เข้าไปด้วยเป็นกรณีความน่าจะเป็น สามารถใช้กลวิธี Loss-based decoding

Loss-based decoding (LB) วิธีนี้ ใช้ข้อมูลขาออกของตัวจำแนกประเภท 2 ค่ามาเป็นคะแนนของขอบเขต (margin) โดยมีข้อกำหนด 2 ข้อ คือ 1. คะแนนควรจะเป็นค่าบวกถ้าตัวอย่างที่ถูกแบ่งเป็นตัวอย่างบวกและเป็นค่าลบหากเป็นตัวอย่างลบ 2. ขนาดของคะแนนควรจะเป็นตัวชี้วัดความมั่นใจของการทำนาย

ให้ $f(l,j)$ เป็นคะแนนขอบเขตของตัวอย่าง l ของแมทริกซ์รหัส โดยแบ่งประเภทตรงกันกับคอลัมน์ j ของแมทริกซ์รหัส M สำหรับแต่ละแถว i ของ M และแต่ละตัวอย่าง l เราคำนวณระยะทางระหว่าง $f(l,j)$ และ $y^i = M(i,j) \forall j \in \{1, \dots, n\}$ $d^m(l,i) = \sum_{j=1}^n L(M(i,j) \cdot f(l,j))$ ที่ซึ่ง L เป็นฟังก์ชันสูญเสีย (loss function) ที่ขึ้นอยู่กับธรรมชาติของตัวจำแนก 2 ค่า (binary classifier) โดยฟังก์ชันที่ใช้กันโดยทั่วไปมากที่สุด 2 ฟังก์ชัน ได้แก่ $L(\mathbf{h}) = -\mathbf{h}$ และ $L(\mathbf{h}) = e^{-\mathbf{h}}$ ที่ซึ่ง $\mathbf{h} = M(i,j) \cdot f(l,j)$ เราตั้งป้ายกำกับแต่ละตัวอย่าง x ด้วยป้ายที่ให้ค่า d_L น้อยที่สุด โดยเทคนิคนี้จะเจือจาง (attenuate) ความผิดพลาดจากสัญลักษณ์ศูนย์ ขณะที่ยังคงรักษาน้ำหนักของตำแหน่งรหัสโดยไม่ขึ้นกับจำนวนศูนย์ของแต่ละรหัส โดยเทคนิคนี้เจือจางความผิดพลาดจากศูนย์ในการทำงานเดียวกับเหมือนกับ AED โดยต่อขยาย additive probabilistic model

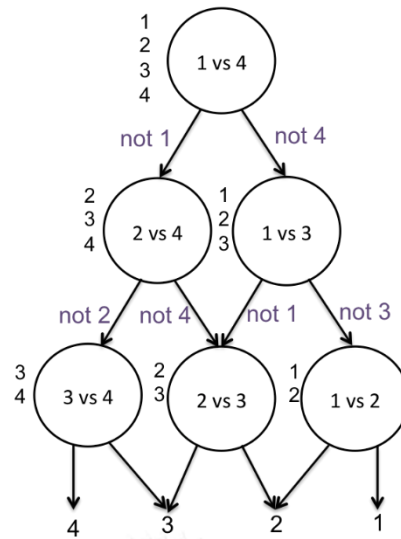
2.2 งานวิจัยที่เกี่ยวข้อง

จากเนื้อหาของซัพพอร์ตเวกเตอร์แมชชีนแบบจำแนกหลายประเภท ที่ได้กล่าวไปแล้วข้างต้น ได้มีผู้ที่นำแนวคิดไปพัฒนาต่อยอดหลายชิ้นงาน โดยในเนื้อหาบทนี้ประกอบด้วยวิธีการที่สร้างจากพื้นฐานของวิธีโอวีโอ ได้แก่ วิธีดีดีเอจี (Decision Directed Acyclic Graph SVM (DDAG)) วิธีเอดีเอจี (Adaptive Directed Acyclic Graph SVM (ADAG)) วิธีอาร์เอดีเอจี (Reordering Adaptive Directed Acyclic Graph SVM (RADAG)) วิธีบีทีเอส (Binary Tree SVM (BTS)) วิธีการแตกครึ่งตามสารสนเทศ (Information-based Dichotimization (IBD)) และวิธีการที่ประยุกต์ใช้การจัดกลุ่มประเภทข้อมูล (Clustering) ได้แก่ วิธีต้นไม้ตัดสินใจเอสวีเอ็ม (Decision-tree-based SVM) วิธีเอสวีเอ็ม-บีดีที (SVM Binary Decision Tree (SVM-BDT)) นอกจากนี้ยังมีงานวิจัยที่ได้ทำการวิเคราะห์เปรียบเทียบเทคนิคแบบดั้งเดิมไว้

2.2.1 วิธีดีดีเอจี (Decision Directed Acyclic Graph SVM (DDAG))

วิธีดีดีเอจี¹⁰ เป็นวิธีหนึ่งในการจำแนกหลายประเภทที่สร้างจากพื้นฐานของวิธีการโอวีโอ โดยนำเอาตัวจำแนกทั้ง $N(N-1)/2$ ตัว มาทำการจัดเรียงเป็นโครงสร้างสามเหลี่ยม จำนวน $N-1$ ชั้น ในกระบวนการทดสอบจะเริ่มจากชั้นบนสุดแล้วทำการทดสอบจากบนลงล่าง ซึ่งในแต่ละชั้นจะมีประเภทข้อมูลที่ไม่ใช่คำตอบถูกคัดออกหนึ่งประเภทตามเส้นทางการสำรวจ (Traversal Path) จนกระทั่งพบได้ประเภทข้อมูลคำตอบที่ชั้นล่างสุดของสามเหลี่ยม

โดยหากเปรียบเทียบกับวิธีการโอวีโอแล้ว ถึงแม้ว่าในระยะการฝึกจะต้องเตรียมตัวจำแนกของทุกๆ คู่ประเภท ซึ่งมีจำนวนทั้งสิ้น $N(N-1)/2$ ตัวเหมือนกับโอวีโอ แต่ต่างกันที่ในระยะทดสอบจะทำการทดสอบเพียง $N-1$ ครั้งเท่านั้น เพราะว่าในแต่ละชั้นของสามเหลี่ยม เราจะใช้ตัวจำแนกที่เตรียมไว้เพียง 1 ตัวเท่านั้น โดยตัวอย่างของวิธีดีดีเอจี สำหรับปัญหา 4 ประเภท แสดงไว้ในรูปที่ 9

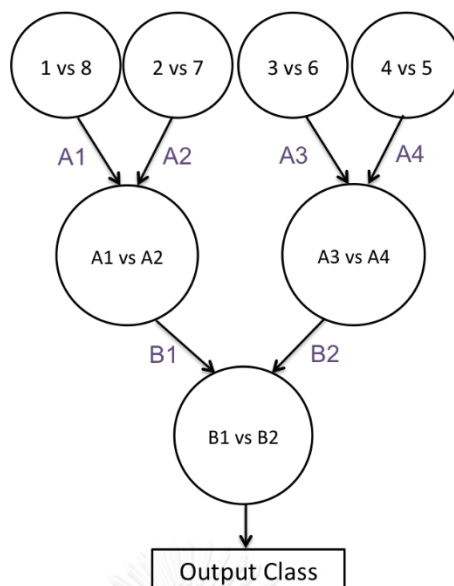


รูป 9 กระบวนการจำแนกหลายประเภทแบบดีดีเอจี้สำหรับปัญหา 4 ประเภท

อย่างไรก็ดี ในงานวิจัยเอดีเอจี้ (ซึ่งจะกล่าวเป็นลำดับถัดไป) ได้ทำการอธิบายให้เห็นถึงปัญหาของวิธีดีดีเอจี้ คือ อาจมีบางประเภทข้อมูลถูกพิจารณาจำแนกด้วยความถี่ที่สูงกว่าประเภทอื่นๆ เนื่องจากโครงสร้างในกราฟ ไม่ได้มีค่านิ่งถึงความถี่ของการถูกนำไปใช้งานของแต่ละประเภทข้อมูล ดังนั้นหากมีบางประเภทที่จำแนกไม่แม่นยำแต่ถูกจำแนกด้วยความถี่ที่สูง อาจมีความผิดพลาดสะสมในอัตราที่สูงตามไปด้วย และส่งผลต่อความแม่นยำของโครงสร้างโดยรวมได้

2.2.2 วิธีเอดีเอจี้ (Adaptive Directed Acyclic Graph SVM (ADAG))

วิธีเอดีเอจี้¹² เป็นวิธีการจำแนกหลายประเภทซึ่งปรับปรุงประสิทธิภาพของวิธีดีดีเอจี้ โดยการนำเอาตัวจำแนกแบบวิธีโอวีโอ มาทำการจัดเรียงเป็นโครงสร้างสามเหลี่ยมเป็นแบบผกผัน โดยในระดับชั้นบนสุดจะสุ่มเลือกเอาคู่ประเภทมาจับคู่กันแบบ 1 ต่อ 1 จำนวนทั้งสิ้น $N/2$ ตัว จากนั้นในระดับถัดไปจะนำเอาประเภทที่เป็นคำตอบจากชั้นบนสุดมาจับคู่กัน แล้วทดสอบจำแนกประเภทที่เหมาะสมเพื่อส่งต่อไปยังชั้นล่างต่อไป จนกระทั่งได้ประเภทที่เป็นคำตอบเพียงหนึ่งเดียวในระดับชั้นล่างสุด ทั้งนี้เพื่อเป็นกระจายความถี่ของการถูกนำไปใช้งานของแต่ละประเภทข้อมูลให้สม่ำเสมอ เพื่อให้ความผิดพลาดสะสมของแต่ละประเภทข้อมูลมีน้อยที่สุด ซึ่งวิธีการนี้ต้องเตรียมตัวจำแนกไว้ $N(N-1)/2$ ตัว แต่ใช้จำนวนครั้งในระยะทดสอบ $N-1$ ครั้งเช่นเดียวกับวิธีดีดีเอจี้ โดยตัวอย่างของวิธีการเอดีเอจี้สำหรับปัญหาจำแนกข้อมูล 8 ประเภท เป็นตามตัวอย่างในรูปที่ 10



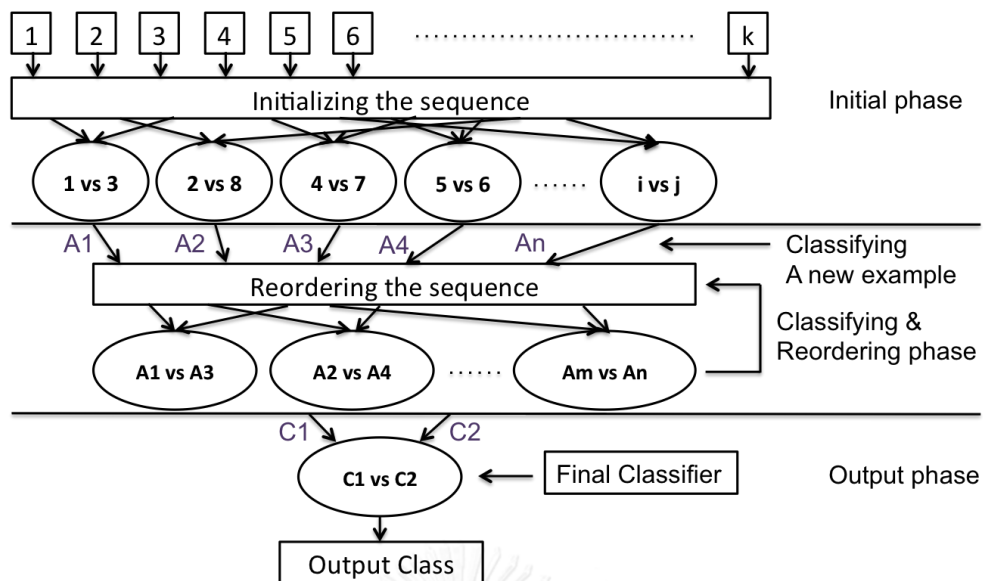
รูป 10 กระบวนการจำแนกหลายประเภทแบบเอตไอเอจี้สำหรับปัญหา 8 ประเภท

ถึงแม้ว่าวิธีเอตไอเอจี้จะปรับปรุงการจำแนกของดีเอตไอเอจี้ให้มีประสิทธิภาพมากขึ้นได้แล้ว แต่ก็ยังมีปัญหาเรื่องวิธีการเลือกและจัดลำดับของประเภทข้อมูลที่เหมาะสมในแต่ละชั้น จึงมีนักวิจัยหาวิธีการเลือกและจัดลำดับให้กับตัวจำแนกข้อมูลที่เหมาะสมในแต่ละชั้น เพื่อให้มีโอกาสจำแนกข้อมูลผิดพลาดต่ำที่สุดในหัวข้อถัดไป

2.2.3 วิธีอาร์เอตไอเอจี้ (Reordering Adaptive acyclic graph SVM (RADAG))

วิธีอาร์เอตไอเอจี้¹³ เป็นวิธีการจำแนกประเภทข้อมูลหลายประเภทที่พัฒนาปรับปรุงเอตไอเอจี้ ให้มีประสิทธิภาพมากขึ้น โดยจะมีการเลือกลำดับที่ทำให้เกิดความผิดพลาดน้อยสุดในการจำแนกโดยใช้ระเบียบวิธีการจับคู่สมบูรณ์แบบน้ำหนักน้อยที่สุด (Minimum Weight Perfect Matching)⁵³ ทั้งนี้เพื่อให้ค่าขอบเขตความผิดพลาดของการจับคู่มีค่าต่ำที่สุด โดยค่าขอบเขตความผิดพลาดแบบสองประเภทแต่ละตัว สามารถหาได้จากอาศัยข้อมูลจากประสิทธิภาพโดยนัยทั่วไปของซัพพอร์ตเวกเตอร์แมชชีน (Generalization Performance of Support Vector Machine)

รูปที่ 11 แสดงให้เห็นถึงการทำงานของอาร์เอตไอเอจี้ โดยในแต่ละชั้นจะทำการจัดคู่ที่เหมาะสม และเมื่อทำการจำแนกเสร็จเรียบร้อยแล้ว ในขั้นถัดไปก็จะทำการจัดลำดับใหม่ (Reordering) จนกระทั่งเลือกประเภทข้อมูลเดียวเป็นคำตอบในท้ายที่สุด



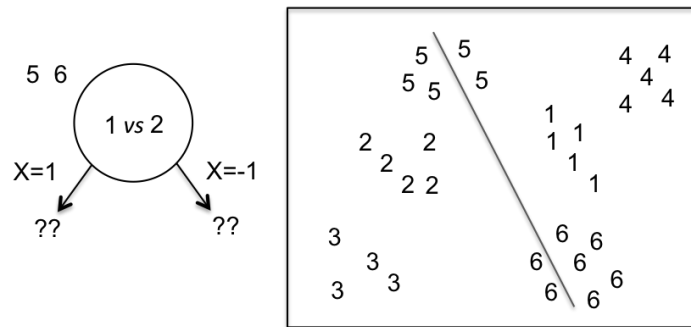
รูป 11 กระบวนการจำแนกหลายประเภทแบบอาร์เอทีเอจี

2.2.4 วิธีพีทีเอส (Binary Tree of SVM (BTS))

วิธีพีทีเอส³⁴ เป็นการจำแนกหลายประเภทข้อมูลที่ใช้โครงสร้างต้นไม้ตัดสินใจ ซึ่งสร้างขึ้นโดยใช้การประกอบกันของตัวจำแนกแบบสองประเภทข้อมูล (Binary Classifier) หรืออาจเรียกว่าเป็นสร้างบนพื้นฐานของวิธีโอวีโอ (OVO-based) โดย Fei และ Lui ได้เสนอวิธีการเลือกคู่ประเภทข้อมูลที่เหมาะสมในแต่ละปมของต้นไม้ไว้ 2 วิธี คือ

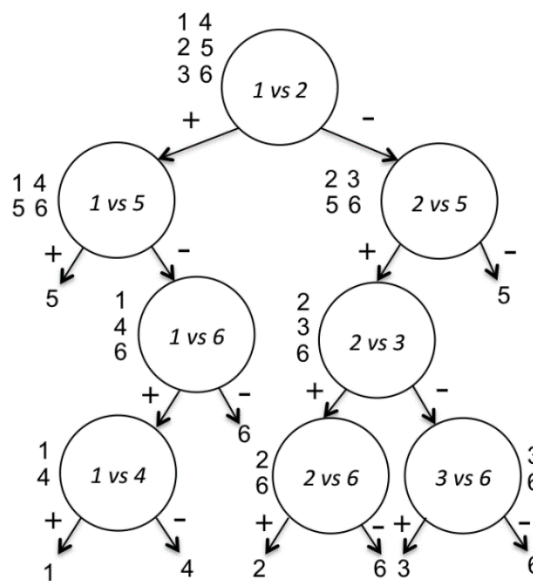
1. แบบบีทีเอส (BTS) ทำการเลือกคู่ประเภทแบบสุ่ม
2. แบบซี-บีทีเอส (c-BTS) ทำการเลือกคู่ประเภทโดยพิจารณาจุดศูนย์กลางของสองประเภทข้อมูลที่ระยะทาง ใกล้กับจุดศูนย์กลางของข้อมูลรวมทั้งหมด

ในการใช้ตัวจำแนกแบบสองประเภทข้อมูลในแต่ละปมของต้นไม้ นั้น อาจมีบางกรณีที่ประเภทข้อมูล k ไม่สามารถถูกแบ่งได้อย่างมีประสิทธิภาพโดยใช้ตัวจำแนกแบบสองประเภทข้อมูล i vs j เนื่องจากประเภทข้อมูล k อาจกระจายอยู่ที่ทั้ง 2 ฝั่งของระนาบหลายมิติแยก i vs j หรืออาจอยู่ใกล้กับระนาบหลายมิติมากเกินไป ดังนั้นในกรณีเช่นนี้ BTS จึงสามารถให้ประเภทข้อมูล k ถูกจำแนกไปทั้งสองฝั่งของปมลูกได้ ดังตัวอย่างในรูปที่ 12 สมมติว่าเราเลือกตัวจำแนกจากประเภทข้อมูล 1 vs 2 ซึ่งระนาบหลายมิติแยกที่ถูกรสร้างขึ้นสามารถแยกข้อมูลตัวอย่างประเภทข้อมูลที่ 1, 4 และ 2, 3 ออกจากกันได้ดี แต่ไม่สามารถแบ่งแยกประเภทข้อมูลที่ 5, 6 ได้ ในกรณีเช่นนี้ทั้งวิธีบีทีเอสและวิธีซี-บีทีเอส จึงให้ปมลูกทั้งด้านซ้ายและด้านขวามีข้อมูลจากประเภทข้อมูล 5, 6 เพื่อใช้พิจารณาในระดับถัดไปด้วย



รูป 12 ปัญหาหระนาบหลายมิติแยกไม่สามารถจำแนกบางประเภทข้อมูลได้

ดังรูปที่ 13 แสดงตัวอย่างการใช้วิธีบีทีเอส หรือ วิธีซี-บีทีเอส แก้ปัญหา 6 ประเภทข้อมูลที่อาจเป็นไปได้ โดยในระดับชั้นบนสุดไม่สามารถจำแนกประเภทข้อมูล 5, 6 ได้ จึงส่งผลให้ปมใบที่ได้คำตอบประเภทข้อมูล 5, 6 ปรากฏอยู่มากกว่า 1 ครั้งในต้นไม้ตัดสินใจ



รูป 13 ตัวอย่าง Binary Tree of SVM สำหรับปัญหา 6 ประเภทข้อมูล

ในการพิจารณาและตัดสินใจว่าจะนำข้อมูลที่ไม่สามารถจำแนกได้ มาทำซ้ำและกระจายไปทั้ง 2 ฝั่งของต้นไม้เล็กน้อยเพียงใด จะขึ้นกับแบบจำลองความน่าจะเป็นและการกำหนดค่า δ ของแบบจำลอง โดยหากกำหนดค่ามากก็จะส่งผลให้มีปมที่ประเภทข้อมูลซ้ำมาก แต่จะเพิ่มความแม่นยำของการจำแนกและเพิ่มระยะเวลาเฉลี่ยที่ใช้ในการจำแนกด้วย อย่างไรก็ตามก็ได้มีการพิสูจน์ว่าโดยเฉลี่ยแล้วทั้งวิธีบีทีเอสและวิธีซี-บีทีเอสใช้เวลาในการจำแนกเป็น $\log_{4/3}((N + 3)/4)$ โดยที่ N คือจำนวนประเภทข้อมูลของปัญหาที่กำลังพิจารณา

2.2.5 วิธีแยกครึ่งตามสารสนเทศ (Information-based Dichotimization (IBD))

วิธีแยกครึ่งตามสารสนเทศ³⁵ เป็นวิธีที่ใช้เอนโทรปี (Entropy) เป็นคะแนนในการเลือก ตัวจำแนกแบบสองประเภทข้อมูล (Binary Classifier) เพื่อนำมาประกอบกันเป็นโครงสร้างต้นไม้ตัดสินใจ โดยมีแนวคิดมาจากทฤษฎีสารสนเทศ (Information Theory) และการเข้ารหัส (Encoding)

เพื่ออธิบายหลักการทำงานของการเข้ารหัส สมมติว่าปัญหา 4 ประเภทข้อมูลที่มีความน่าจะเป็นที่จะเกิดขึ้นเท่ากัน จำนวนบิตที่น้อยที่สุดที่ใช้เป็นตัวแทนประเภทข้อมูล คือ 2 หลัก เช่น

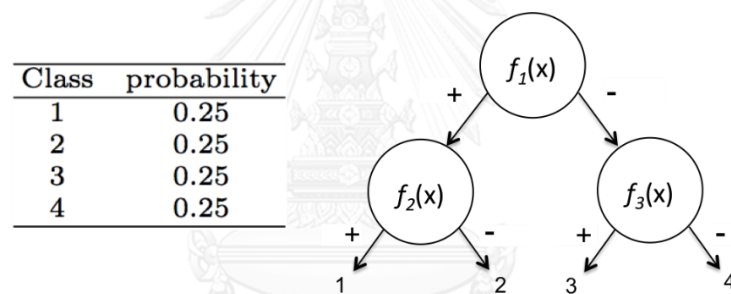
ประเภทข้อมูลที่ 1 สามารถแทนได้ด้วยรหัส 00,

ประเภทข้อมูลที่ 2 สามารถแทนได้ด้วยรหัส 01,

ประเภทข้อมูลที่ 3 สามารถแทนได้ด้วยรหัส 10 และ

ประเภทข้อมูลที่ 4 สามารถแทนได้ด้วยรหัส 11

ในการเข้ารหัสลำดับข้อมูล 12131412 สามารถแทนได้ด้วยตัวแทนแบบบิต 16 ตัว คือ 0001001000110001 ซึ่งสามารถใช้ต้นไม้ตัดสินใจเพื่อถอดรหัสประเภทข้อมูลจากบิตได้ ดังรูปที่ 14



รูป 14 ตัวอย่างข้อมูลที่กระจายสม่ำเสมอและต้นไม้แสดงการเข้ารหัสข้อมูล 4 ประเภทข้อมูล

อย่างไรก็ดี ปัญหาในโลกแห่งความจริง แต่ละประเภทข้อมูลสามารถเกิดขึ้นด้วยความน่าจะเป็นไม่เท่ากัน สมมติว่าประเภทข้อมูลที่ 1 ถึง 4 เกิดขึ้นด้วยความน่าจะเป็น $P(1) = 1/2$, $P(2) = 1/4$, $P(3) = 1/8$, $P(4) = 1/8$ โดยที่ $P(X)$ หมายถึงความน่าจะเป็นของประเภทข้อมูล X เราสามารถเข้ารหัสประเภทข้อมูล 1-4 ได้ดังนี้

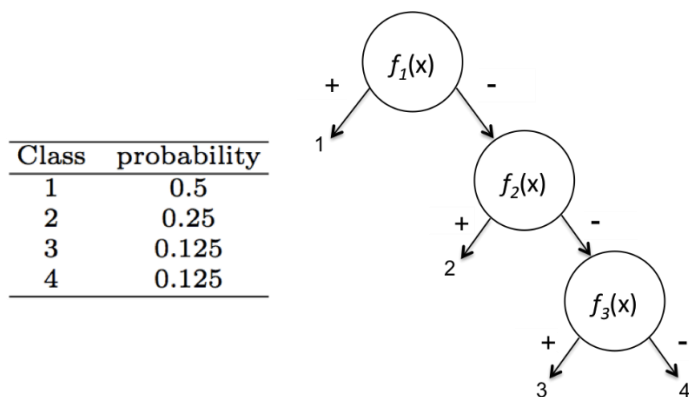
ประเภทข้อมูลที่ 1 สามารถแทนได้ด้วยรหัส 0,

ประเภทข้อมูลที่ 2 สามารถแทนได้ด้วยรหัส 10,

ประเภทข้อมูลที่ 3 สามารถแทนได้ด้วยรหัส 110 และ

ประเภทข้อมูลที่ 4 สามารถแทนได้ด้วยรหัส 111

ในการเข้ารหัสลำดับข้อมูล 12131412 เราจะสามารถแทนด้วยตัวแทนแบบบิตเพียง 14 ตัวเท่านั้น คือ 01001100111010 ซึ่งสามารถใช้ต้นไม้ตัดสินใจเพื่อถอดรหัสประเภทข้อมูลจากบิตได้ ดังรูปที่ 15



รูป 15 ตัวอย่างข้อมูลที่กระจายไม่สม่ำเสมอและต้นไม้แสดงการเข้ารหัสข้อมูล 4 ประเภทข้อมูล

ขั้นตอนวิธีการสร้างต้นไม้สำหรับการจำแนกหลายประเภทด้วยการแตกกิ่งตามสารสนเทศเป็นดังนี้

ขั้นตอนที่ 1 สร้างตัวจำแนกแบบสองประเภททุกแบบที่เป็นไปได้ ซึ่งหากมีจำนวนประเภทข้อมูล N ประเภท จะมีจำนวนจำแนกทั้งสิ้น $N(N-1)/2$ แบบ

ขั้นตอนที่ 2 เลือกตัวจำแนกแบบสองประเภทเพื่อนำมาสร้างปมของต้นไม้สำหรับการจำแนกหลายประเภท ซึ่งการพิจารณาตามสารสนเทศ จะใช้ค่าเอนโทรปี (entropy) แบ่งประเภทข้อมูลตามสมการที่ (34) คือ

$$\text{ค่าเอนโทรปีของตัวจำแนก (Entropy)} = \sum_{i=1}^n \frac{|t_i|}{T} I(t_i)$$

$$I(t_i) = \sum -p(m_{ij}) \log_2 p(m_{ij}) \quad (34)$$

โดยที่ n เป็นจำนวนกิ่งของต้นไม้ ในที่นี้มีค่าเป็น 2 ซึ่งเรียกว่ากิ่งบวกและกิ่งลบ

N เป็นจำนวนประเภทของข้อมูลทั้งหมด

$P(m_{ij})$ เป็นความน่าจะเป็นในการเกิดข้อมูลประเภท j ในกิ่ง t_i

ขั้นตอนที่ 3 ทำการฝึกสอนตามการแบ่งข้อมูลสอนเป็นฝั่งบวกและฝั่งลบตามที่ได้แบ่งไว้ในขั้นตอนที่ 2 ซึ่งจะได้ข้อมูลสอน 2 ส่วน คือ ข้อมูลสอนที่เป็นบวกและข้อมูลสอนที่เป็นลบ จากนั้นจึงพิจารณาข้อมูลสอนในแต่ละส่วนว่าในข้อมูลสอนนั้นอยู่ในกรณีใด โดยแบ่งเป็น 3 กรณี คือ

กรณีที่ 1 พบว่าเป็นประเภทเดียวแล้วก็ไม่ต้องดำเนินการต่อ

กรณีที่ 2 พบว่าข้อมูลที่พิจารณามีจำนวนเป็น 2 ประเภท ก็สามารถเลือกตัวจำแนกแบบสองประเภทดังกล่าวมาสร้างโนดได้ทันที

กรณีที่ 3 พบว่าข้อมูลที่พิจารณามีจำนวนประเภทมากกว่าสองประเภทจึงกลับไปทำซ้ำขั้นตอนที่ 1 และ 2

นอกจากนี้ในทางปฏิบัติ การจะหาระนาบหลายมิติที่สามารถแบ่งแยกข้อมูลประเภทเดียวกันไปฝั่งเดียวกัน (ฝั่งบวกหรือฝั่งลบ) ได้อย่างสมบูรณ์นั้นเป็นสิ่งที่สามารถทำได้ยาก เพื่อผ่อนคลาย (relax) สถานการณ์ดังกล่าววิธีแตกครึ่งตามสารสนเทศจึงมีการตัดเล็มต้นไม้ (Tree Pruning) โดยการลบตัวอย่างฝึกส่วนน้อยบางส่วน โดยกำหนดเป็นพารามิเตอร์ร้อยละการตัดเล็ม (pruning percentage) โดยหากข้อมูลที่อยู่ในฝั่งบวกหรือลบคิดเป็นร้อยละที่น้อยกว่าพารามิเตอร์ที่กำหนดไว้ ก็จะเข้าเงื่อนไขและทำการละข้อมูลส่วนน้อยทิ้งไป อย่างไรก็ตามหากข้อมูลประเภทเดียวกันกระจายทั้ง 2 ฝั่งของระนาบหลายมิติเป็นร้อยละมากกว่าค่าที่กำหนดไว้ จะไม่สามารถตัดข้อมูลทิ้งได้ และจะมีข้อมูลประเภทนั้นกระจายไปยังปมลูกทั้งด้านซ้ายและด้านขวา จึงส่งผลให้มีปมใบของประเภทข้อมูลนั้นปรากฏอยู่มากกว่า 1 ครั้งในต้นไม้ตัดสินใจ ซึ่งจะส่งผลให้เวลาเฉลี่ยที่ใช้ในการจำแนกสูงขึ้น ดังนั้นพารามิเตอร์ร้อยละการตัดเล็มจึงเป็นตัวกำหนดลักษณะของต้นไม้ที่สร้างขึ้น โดยหากยังกำหนดค่าสูงขึ้น หมายถึงว่าอนุญาตให้มีการตัดเล็มต้นไม้มากขึ้น

นอกจากนี้การตัดเล็มแล้ว วิธีแตกครึ่งตามสารสนเทศนั้นมีการใช้ค่าประมาณความผิดพลาดโดยนัยทั่วไป⁵¹ มาใช้ในการเลือกตัวจำแนกอีกด้วย โดยตัวจำแนกที่ถูกเลือกจะต้องมีค่าประสิทธิภาพโดยนัยทั่วไปอยู่ในช่วง R ที่กำหนด คือ $x_{min} \leq R \leq x_{mean} + sd$ เมื่อ x_{min} คือ ค่าต่ำสุดของขอบเขตความผิดพลาดของตัวจำแนกทั้งหมดที่พิจารณา และ $x_{mean} + sd$ คือ ผลรวมของค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของขอบเขตความผิดพลาดของตัวจำแนกทั้งหมดที่พิจารณา

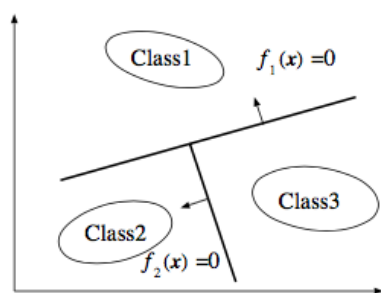
อย่างไรก็ตามในการกำหนดค่าพารามิเตอร์ของทั้งร้อยละการตัดเล็ม (P) และช่วงค่าของความผิดพลาด (R) นั้น จำเป็นต้องมีการคำนวณหาค่าที่เหมาะสมสำหรับพารามิเตอร์แต่ละตัว เนื่องจากถ้ากำหนดร้อยละการตัดเล็มมากเกินไปก็จะทำให้มีข้อมูลที่มีประโยชน์ถูกกำจัดทิ้งไปด้วย และทำให้การจำแนกผิดมีสูงขึ้น หรือถ้าเรากำหนดช่วงค่าที่เหมาะสมของความผิดพลาดน้อยเกินไปก็จะทำให้โอกาสได้ต้นไม้ที่สั้นน้อยลงด้วย

2.2.6 วิธีต้นไม้ตัดสินใจเอชวีเอ็ม (Decision-tree-based SVM)

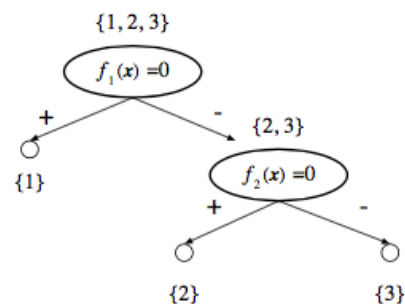
เป็นวิธีการจำแนกข้อมูลหลายประเภทที่มีโครงสร้างการเรียงตัวแบบต้นไม้ตัดสินใจที่ถูกเสนอขึ้นโดย Takahashi และ Abe³⁸ โดยในขั้นตอนฝึกจะพิจารณาระนาบหลายมิติ (Hyperplane) ที่เหมาะสมในการแบ่งประเภทข้อมูลออกทีละประเภท (หรือหลายประเภท) ด้วยการคำนวณจุดศูนย์กลางข้อมูล (Centroid) โดยเริ่มต้นจากระดับสูงสุด (Top node) และทำซ้ำลงไปจนกระทั่งเหลือประเภทเดียวในทุกๆ ปมของต้นไม้ โดยได้เสนอวิธีการพิจารณาและแบ่งแยกไว้ 4 แบบ ได้แก่

1. ทำการแบ่งแยกปมละหนึ่งประเภทข้อมูลด้วยระยะทางแบบยูคลิด (Euclidean Distance)
2. ทำการแบ่งแยกปมหลายประเภทข้อมูลด้วยระยะทางแบบยูคลิด (Euclidean Distance)
3. ทำการแบ่งแยกปมละหนึ่งประเภทข้อมูลด้วยระยะทางแบบมหาลาโนบิส (Mahalanobis Distance)
4. ทำการแบ่งแยกปมหลายประเภทข้อมูลด้วยระยะทางแบบมหาลาโนบิส (Mahalanobis Distance)

ในการนำไปใช้จำแนกประเภทข้อมูลทดสอบ จะเริ่มต้นทำการจำแนกที่ปมบนสุดของต้นไม้ แล้วทำการทดสอบลงมาตามทางเดิน (Path) ของต้นไม้จนกระทั่งถึงปมที่เป็นคำตอบที่เหมาะสมเพียงประเภทเดียว โดยตัวอย่างของ Decision-tree-based SVM สำหรับปัญหาจำแนก 3 ประเภทข้อมูลเป็นดังรูปที่ 16-17



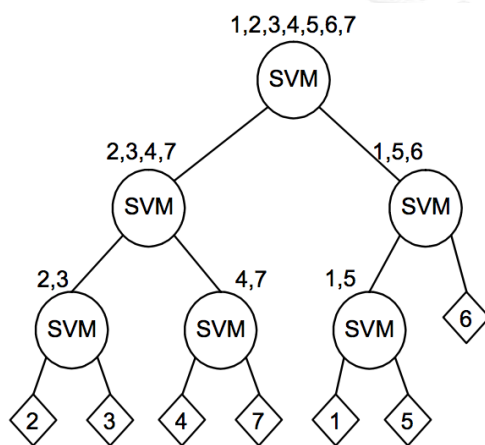
รูป 16 ตัวอย่างการแบ่งบนปริภูมิคุณลักษณะ



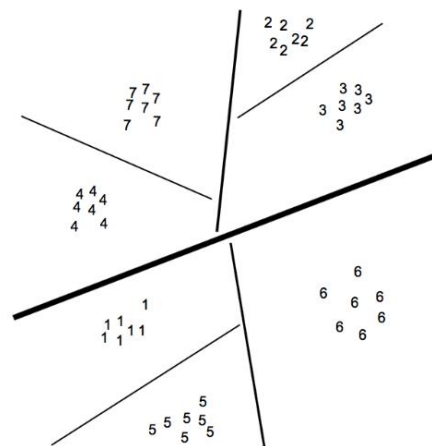
รูป 17 ต้นไม้ตัดสินใจของปัญหาในรูปที่ 16

2.2.7 วิธีเอสวีเอ็ม-บีดีที (SVM Binary Decision Tree (SVM-BDT))

วิธีเอสวีเอ็ม-บีดีที⁴² เป็นวิธีการสร้างต้นไม้ตัดสินใจแบบหนึ่ง ซึ่งแบ่งกลุ่มปมต้นไม้จาก ระยะทางที่วัดบนปริภูมิเคอร์เนล (Kernel Space) แทนที่จะวัดบนปริภูมิข้อมูลขาเข้า (Input Space) โดยในการแบ่งแต่ละครั้งจะแบ่งเป็น 2 กลุ่มที่ไม่ต่อกัน (Disjoint groups) คือ g_1 และ g_2 ซึ่งแต่ละกลุ่ม ได้จากการคำนวณจุดศูนย์กลางโน้มถ่วง (Gravity Centre) จำนวน N จุด เป็นตัวแทนของประเภท ข้อมูลทั้ง N ประเภท ในการเลือกประเภทข้อมูลเข้าสู่กลุ่ม g_1 และ g_2 จะเลือกคู่ประเภทที่อยู่ห่างกัน มากที่สุดเข้าสู่กลุ่ม g_1 และ g_2 ก่อน จากนั้นจึงจัดแบ่งประเภทที่มีระยะทางสั้นสุดเมื่อวัด จากประเภท ตั้งต้นใน g_1 และ g_2 สลับกันไปจนกระทั่งครบทุกประเภทข้อมูล ตัวอย่างของต้นไม้ที่ได้จากการสร้าง ด้วยวิธีเอสวีเอ็ม-บีดีที สำหรับแก้ปัญหาประเภทข้อมูล 7 ประเภท เป็นดังภาพที่ 18 และ 19



รูป 18 ภาพของเอสวีเอ็ม-บีดีที



รูป 19 เอสวีเอ็ม-บีดีทีที่แบ่ง 7 ประเภทข้อมูล

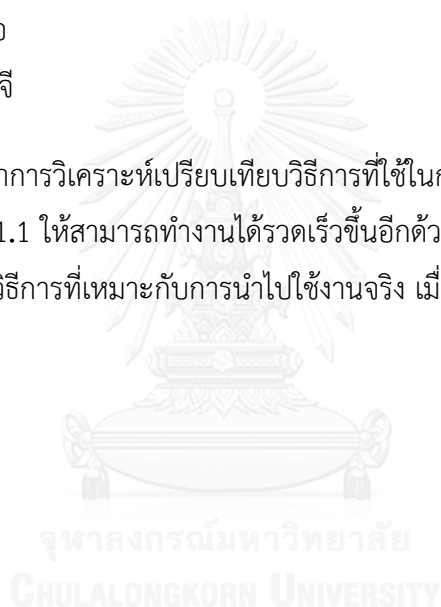
ในแง่ของความเร็วที่การนำไปใช้ทดสอบข้อมูล พบว่าวิธีเอสวีเอ็ม-บีดีทีสามารถทำงานได้เร็วกว่าวิธีโอวีโอ โอวีเอ ดีดีเอจี บีทีเอส เนื่องจากมีลักษณะเป็นต้นไม้ตัดสินใจที่มีปมเป็น $N-1$ ปม และใช้เวลาในการจำแนกเป็น $\lceil \log_2 N \rceil$ โดยที่ N เป็นจำนวนประเภทข้อมูล อย่างไรก็ตาม ในแง่ของประสิทธิภาพในการจำแนกนั้น ผลการทดลองแสดงให้เห็นว่าวิธีเอสวีเอ็ม-บีดีที มีประสิทธิภาพต่ำกว่าวิธีโอวีโอ โอวีเอ ดีดีเอจี บีทีเอส

2.2.8 การเปรียบเทียบวิธีการจำแนกหลายประเภทโดย Hsu และ Lin

Hsu และ Lin⁵ ได้ทำการวิจัยเพื่อเปรียบเทียบวิธีการจำแนกหลายประเภทไว้ โดยวิธีการที่ถูกนำมาวิเคราะห์และเปรียบเทียบ ได้แก่

1. วิธีแก้ปัญหาลำดับหลายประเภทขนาดใหญ่ครั้งเดียว 2 วิธี ได้แก่
 - 1.1 วิธีของ Vapnik ซึ่งนำมาประยุกต์ใช้โดย Weston และ Watkins
 - 1.2 วิธีของ Cramer และ Singer
2. วิธีการจากการผสมตัวจำแนกสองประเภทหลายตัวเข้าด้วยกัน 3 วิธี ได้แก่
 - 2.1 วิธีโอวีเอ
 - 2.2 วิธีโอวีโอ
 - 2.3 วิธีดีดีเอจี

ซึ่งทางผู้วิจัยได้ทำการวิเคราะห์เปรียบเทียบวิธีการที่ใช้ในการทดลอง และได้ทำการปรับปรุงประสิทธิภาพของวิธีการ 1.1 ให้สามารถทำงานได้รวดเร็วขึ้นอีกด้วย ซึ่งจากงานวิจัย ได้ข้อสรุปว่าวิธีโอวีโอและวิธีดีดีเอจี เป็นวิธีการที่เหมาะสมกับการนำไปใช้งานจริง เมื่อเปรียบเทียบกันในกลุ่มเทคนิคที่กล่าวถึงในงานวิจัย



บทที่ 3

ต้นไม้ตัดสินใจตามสารสนเทศและ

ต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป

ในบทนี้จะกล่าวถึงผลการออกแบบเทคนิคใหม่ และกล่าวถึงรายละเอียดของเทคนิคใหม่ที่เกิดขึ้นจากงานวิจัยนี้ โดยนำเสนอเทคนิคการจำแนกหลายประเภทแบบใหม่ 2 วิธี ได้แก่ วิธีต้นไม้ตัดสินใจตามสารสนเทศ (Information Based Decision Tree (IB-DTree)) และ วิธีต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป (The Information Based and Generalization Error Estimation Decision Tree (IBGE-DTree)) ซึ่งทั้ง 2 วิธีมีจุดมุ่งหมายให้สามารถจำแนกได้อย่างรวดเร็ว และอาจมีการยอมให้ความแม่นยำของการจำแนกลดลงบ้าง และคาดหวังให้สามารถทำงานได้ในความซับซ้อนเชิงเวลา $O(\log_2 N)$ ในกรณีที่ดีที่สุด และไม่เกิน $O(N)$ ในกรณีแย่งที่สุด โดยเทคนิคที่นำเสนอเหมาะสมในการใช้แก้ปัญหาที่มีจำนวนประเภทข้อมูลมากซึ่งไม่สามารถแก้ปัญหาได้อย่างมีประสิทธิภาพในทางปฏิบัติ ด้วยวิธีที่ใช้เวลาประมวลผล $O(N^2)$ ซึ่งมีรายละเอียดดังนี้

3.1 ต้นไม้ตัดสินใจตามสารสนเทศ (The Information Based Decision Tree (IB-DTree))

เป็นวิธีการจำแนกหลายประเภทที่มีพื้นฐานมาจากวิธีโอวีโอ โดยสร้างเป็นโครงสร้างต้นไม้ตัดสินใจ ซึ่งเลือกตัวจำแนกที่เหมาะสมในแต่ละปมของต้นไม้ด้วยวิธีการคำนวณค่าเอนโทรปี (Entropy) ซึ่งตัวจำแนกค่าเอนโทรปีต่ำสุดจะนำไปสู่การจำแนกที่รวดเร็ว เนื่องจากตามทฤษฎีสารสนเทศนั้น ค่าตอบที่มีความน่าจะเป็นสูงจะถูกแยกออกและสามารถค้นพบเป็นคำตอบได้ในไม่กี่ระดับชั้นจากปมที่เป็นราก โดยค่าเอนโทรปีนั้นสามารถคำนวณได้จากสมการ (1) โดยที่ $(p)^+$ และ $(p)^-$ คือ สัดส่วนจำนวนข้อมูลฝั่งบวก/ลบ ต่อจำนวนข้อมูลฝึกทั้งหมด โดยสัญลักษณ์ด้วยเครื่องหมาย (+) และ (-) ที่ด้านหลังวงเล็บหมายถึง ทำการคำนวณแต่ข้อมูลส่วนบวกหรือส่วนลบของระนาบหลายมิติเท่านั้น และในกรณีที่ไม่มีตัวอย่างบวก (หรือลบ) ที่ฝั่งนั้นของระนาบหลายมิติเลย ให้เทอม $(-p \log_2 p)$ มีค่าเป็น 0

$$Entropy \equiv p^+ \times [\sum_{i=1}^N -p_i^+ \log_2 p_i^+] + p^- \times [\sum_{i=1}^N -p_i^- \log_2 p_i^-] \quad (35)$$

กลไกที่สำคัญของต้นไม้ตัดสินใจตามสารสนเทศ คือ ขั้นตอนวิธี “การจับกลุ่มประเภทข้อมูลตามข้อมูลส่วนใหญ่” ซึ่งเป็นไปตามโค้ดเทียมแสดงขั้นตอนวิธีที่ 1 ซึ่งแทนที่เราใช้ตัวจำแนก h ที่แต่ละปมของต้นไม้โดยตรง ซึ่งจะพบกับปัญหาในการจำแนกประเภทข้อมูล k ด้วยตัวจำแนก $h = (i \text{ vs } j)$

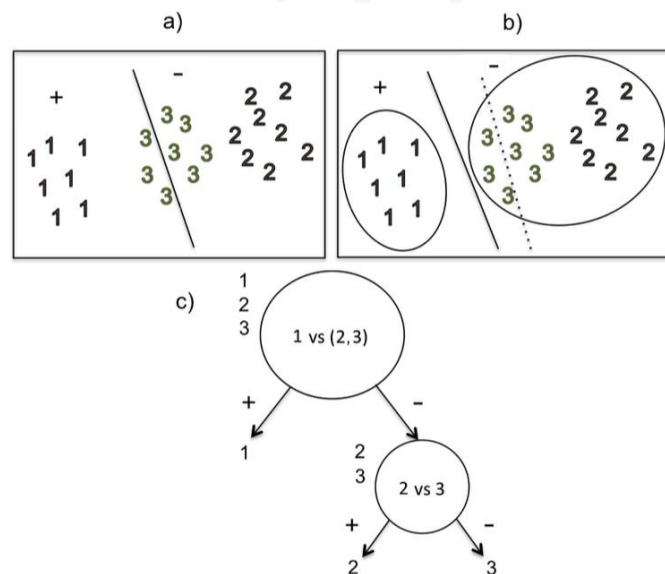
และอาจจำเป็นต้องทำการกระจายประเภทข้อมูล k ไปยังปมลูกทั้งสองฝั่ง แต่หากใช้ขั้นตอนวิธีนี้จะไม่เกิดกรณีนี้ขึ้น เนื่องจากประเภทข้อมูล k จะเป็นส่วนหนึ่งในกลุ่มข้อมูลบวกหรือลบ ในตัวจำแนก $h' = (P \text{ vs } N)$ เสมอ ดังนั้นจึงไม่จำเป็นต้องกระจายประเภทข้อมูล k ไปยังปมลูกซ้ายและขวา ซึ่งจะไม่ส่งผลให้ความลึกของต้นไม้เพิ่มขึ้นโดยไม่จำเป็น

Algorithm 1 Class-grouping-by-majority

- 1: **procedure** CLASS-GROUPING-BY-MAJORITY (SELECTED CLASSIFIER h , CANDIDATE CLASSES K)
 - 2: Initialize set of positive classes $P = \emptyset$ and set of negative classes $N = \emptyset$
 - 3: **for** each class $i \in K$ **do**:-
 - 4: Label all data of class i to (+) and (-) separated by initial classifier h
 - 5: $p \leftarrow \text{count}(+)$, $n \leftarrow \text{count}(-)$
 - 6: **if** ($p > n$) **then** $P \leftarrow P \cup \{i\}$
 - 7: **else** $N \leftarrow N \cup \{i\}$
 - 8: **end for**
 - 9: Train final classifier h' using all data if classes in P as positive examples, and in N as negative examples
 - 10: **return** h' , P , N
 - 11: **end procedure**
-

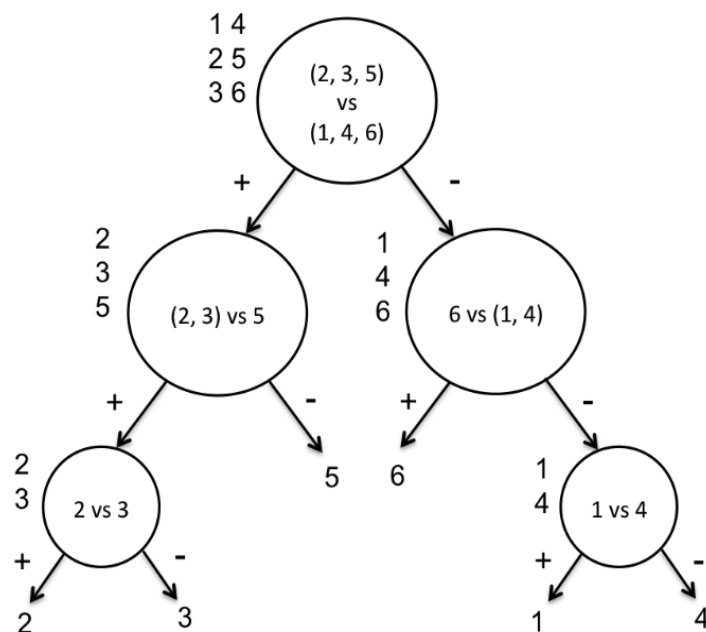
ขั้นตอนวิธีที่ 1 การจับกลุ่มประเภทข้อมูลตามข้อมูลส่วนใหญ่

ตัวอย่างของ “การจับกลุ่มประเภทข้อมูลตามข้อมูลส่วนใหญ่” สำหรับปัญหา 3 ประเภท ข้อมูลแสดงในภาพที่ 20 สมมติว่าเราเลือกตัวจำแนก 1 vs 2 เป็นตัวจำแนกของปมราก ในรูป 20(a) จะเห็นได้ว่าข้อมูลส่วนใหญ่ของประเภทที่ 3 ส่วนใหญ่อยู่ที่ฝั่งด้านลบของระนาบหลายมิติ ดังนั้นเราจึงจัดกลุ่มข้อมูลประเภทที่ 3 ทั้งหมดไปรวมกับกลุ่มลบ จากนั้นจึงฝึกตัวจำแนก 1 vs (2, 3) เพื่อใช้ในปมรากของต้นไม้ตัดดังรูป 20(b) และหลังจากกระบวนการทั้งหมดเสร็จสิ้นแล้วจะได้ต้นไม้ตัดสินใจดังรูป 20(c)



รูป 20 การจับกลุ่มประเภทข้อมูลตามข้อมูลส่วนใหญ่

เพื่อให้เห็นภาพการทำงานของต้นไม้ตัดสินใจตามสารสนเทศมากขึ้น หากเราประยุกต์ใช้ขั้นตอนวิธีนี้กับปัญหา 6 ประเภทข้อมูลในรูปที่ 13 ของเทคนิคบีทีเอส ซึ่งจะได้ต้นไม้ตัดสินใจตามสารสนเทศ ดังรูปที่ 21 ที่ปมรากเราเลือกตัวจำแนก 1 vs 2 เป็นตัวจำแนก h ซึ่งพบว่าข้อมูลส่วนใหญ่ของประเภท 3 และ 5 อยู่ในฝั่งบวกของระนาบ h ในขณะที่ประเภท 4 และ 6 อยู่ในฝั่งลบของระนาบ h ดังนั้นตัวจำแนก (2, 3, 5) vs (1, 4, 6) จึงถูกฝึกขึ้นเป็นตัวจำแนก h สำหรับปมราก กระบวนการเช่นนี้ดำเนินต่อไปเรื่อยๆ จนสิ้นสุด ซึ่งจะพบว่าไม่มีปมใบซึ่งเป็นประเภทข้อมูลเดียวกันซ้ำซ้อนในต้นไม้



รูป 21 ภาพตัวอย่างต้นไม้ตัดสินใจตามสารสนเทศ

ขั้นตอนวิธีการสร้างต้นไม้ตัดสินใจตามสารสนเทศถูกบรรยายไว้ ตามโค้ดเทียมแสดงขั้นตอนวิธีที่ 2 ซึ่งจะเห็นว่ากระบวนการสร้างต้นไม้ตัดสินใจเป็นกระบวนการทำวนซ้ำ (recursive process) โดยเริ่มจากปมรากที่บรรทัดที่ 1-7 โดยมีทุกประเภทข้อมูลถูกพิจารณาตั้งบรรทัดที่ 2 กระบวนการสร้างปมจะดำเนินไปตามกระบวนการในบรรทัดที่ 8-19 โดยขั้นตอนแรกจะทำการคำนวณค่าเอนโทรปีและเลือกตัวจำแนกที่ให้ค่าน้อยที่สุดเป็น h จากนั้นทุกประเภทจะถูกแยกออกไปรวมกับกลุ่มบวกหรือลบ ตามจำนวนส่วนใหญ่ของข้อมูลในประเภทข้อมูลนั้น และสุดท้ายกระบวนการนี้จะสั่งให้ปมลูกทั้งด้านซ้ายและขวาดำเนินการวนซ้ำต่อไป หรือหยุดดำเนินการหากเป็นปมใบเนื่องจากเหลือเพียงประเภทข้อมูลเดียวถูกพิจารณา

Algorithm 2 Information-Based Decision Tree SVM (IB-DTree)

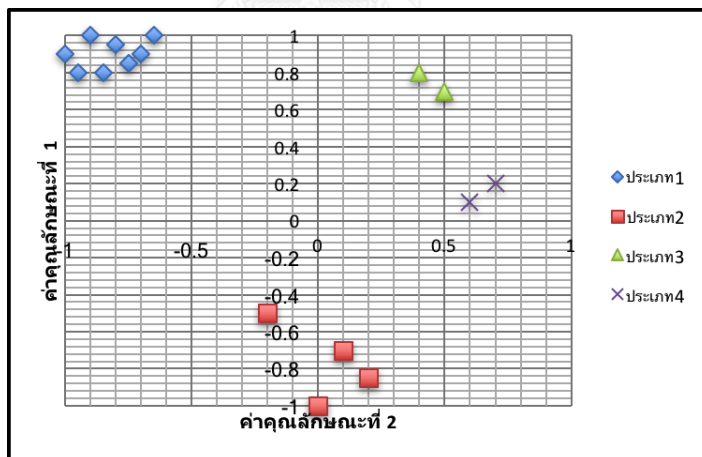
```

1: procedure IB-DTREE
2:   Initialize the tree  $T$  with root node  $Root$ 
3:   Initialize the set of candidate output classes  $S = \{1, 2, 3, \dots, N\}$ 
4:   Create the all binary classifiers ( $i$  vs  $j$ );  $i, j \in S$ 
5:   Construct Tree ( $Root, S$ )
6:   return  $T$ 
7: end procedure
8: procedure CONSTRUCT TREE (NODE  $D$ , CANDIDATE CLASSES  $K$ )
9:   for each binary classifier ( $i$  vs  $j$ );  $i, j \in K$ ;  $i < j$  do:-
10:    Calculate the entropy using training data of all classes in  $K$ 
11:   end for
12:   initial classifier  $h \leftarrow$  classifier ( $i$  vs  $j$ ) with the lowest entropy
13:   final classifier  $h'$ , positive classes  $P$ , negative classes  $N \leftarrow$  Class-grouping-by-majority( $h, K$ )
14:    $D.classifier \leftarrow h'$ 
15:   Initialize new node  $L$ ;  $D.left-child-node \leftarrow L$ 
16:   Initialize new node  $R$ ;  $D.right-child-node \leftarrow R$ 
17:   if  $|P| > 1$  then Construct Tree ( $L, P$ ) else  $L$  is the leaf node with answer class  $P$ 
18:   if  $|N| > 1$  then Construct Tree ( $R, N$ ) else  $R$  is the leaf node with answer class  $N$ 
19: end procedure

```

ขั้นตอนวิธีที่ 2 วิธีการสร้างต้นไม้ตัดสินใจตามสารสนเทศ

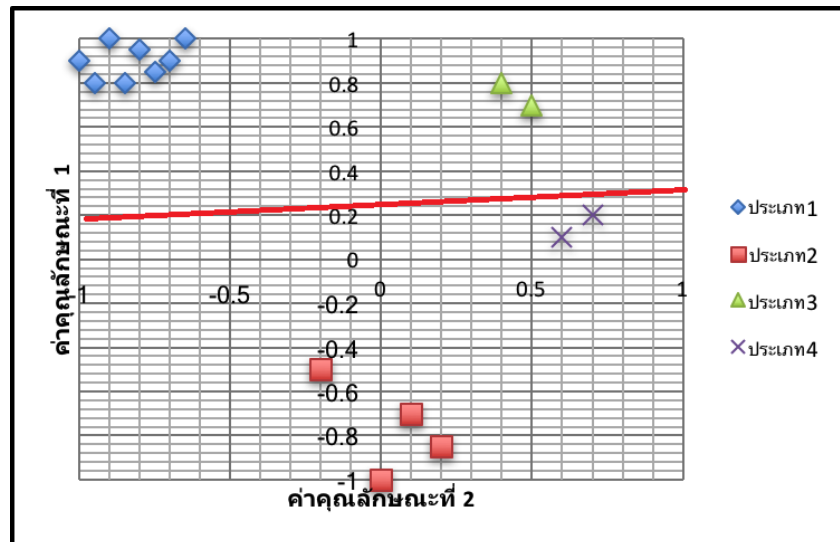
ตัวอย่างการฝึกเพื่อสร้างต้นไม้ตัดสินใจตามสารสนเทศสำหรับปัญหาที่มีประเภข้อมูล 4 ประเภท บนระนาบ 2 มิติ ซึ่งประกอบด้วยข้อมูลฝึกจำนวน 16 ตัว โดยมีจำนวนประเภทข้อมูลที่ 1, 2, 3, 4 เป็น 8, 4, 2, 2 ตัว ตามลำดับ ดังรูปที่ 22



รูป 22 ตัวอย่างปัญหาที่มีประเภข้อมูล 4 ประเภท บนระนาบ 2 มิติ

ในการฝึกและสร้างต้นไม้ตามขั้นตอนวิธีที่ 2 กระบวนการจะเริ่มสร้างที่ปมราก (Root node) ซึ่งมีตัวจำแนกที่ถูกพิจารณาทั้งสิ้น 6 ตัว คือ 1 vs 2, 1 vs 3, 1 vs 4, 2 vs 3, 2 vs 4, 3 vs 4 ซึ่งเมื่อนำไปประมวลผลด้วยโปรแกรม เพื่อคำนวณหาค่าเอนโทรปี

ค่าเอนโทรปีของตัวจำแนก 1 vs 2 คำนวณโดยการฝึกตัวจำแนก 1 vs 2 แล้วใช้ระนาบหลายมิติเป็นตัวแบ่งฝั่งข้อมูลบวกและลบของแต่ละตัวอย่างฝึกซึ่งจะได้ระนาบหลายมิติดังรูปที่ 23 ซึ่งจะได้จำนวนข้อมูลที่ถูกจำแนกในฝั่งบวกและลบของแต่ละประเภทตามตารางที่ 4



รูป 23 เส้นระนาบหลายมิติของตัวจำแนก 1 vs 2

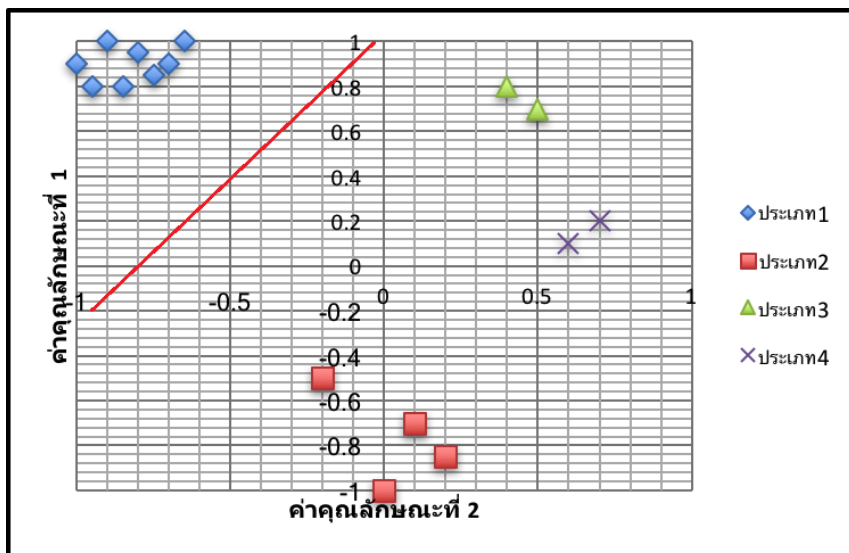
ตาราง 4 จำนวนข้อมูลที่ถูกจำแนกในฝั่งบวกและลบโดยตัวจำแนก 1 vs 2

ประเภทข้อมูล	จำนวนตัวอย่างบวก	จำนวนตัวอย่างลบ
1	8	0
2	0	4
3	2	0
4	0	2
รวม	10	6

จากตารางที่ 4 จะคำนวณค่า p^+ และ p^- ได้เป็น $10/16$ และ $6/16$ ตามลำดับ และคำนวณค่า $p_i^+ \log_2 p_i^+$ และ $p_i^- \log_2 p_i^-$ เพื่อใช้แทนค่าในสมการ (35) ซึ่งจะคำนวณได้ค่าเอนโทรปีของตัวจำแนก 1 vs 2 คือ

$$= \frac{10}{16} \left(-\frac{8}{10} \log_2 \frac{8}{10} - \frac{2}{10} \log_2 \frac{2}{10} \right) + \frac{6}{16} \left(-\frac{4}{6} \log_2 \frac{4}{6} - \frac{2}{6} \log_2 \frac{2}{6} \right) = 0.7956$$

ค่าเอนโทรปีของตัวจำแนก 1 vs 3 คำนวณโดยการฝึกตัวจำแนก 1 vs 3 แล้วใช้ระนาบหลายมิติเป็นตัวแบ่งฝั่งข้อมูลบวกและลบของแต่ละตัวอย่างฝึกซึ่งจะได้ระนาบหลายมิติดังรูปที่ 24 ซึ่งจะได้จำนวนข้อมูลที่ถูกจำแนกในฝั่งบวกและลบของแต่ละประเภทตามตารางที่ 5



รูป 24 เส้นระนาบหลายมิติของตัวจำแนก 1 vs 3

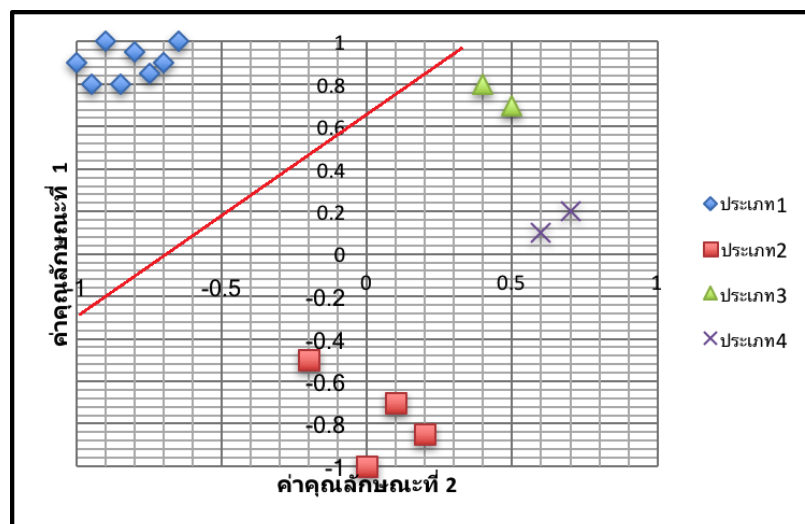
ตาราง 5 จำนวนข้อมูลที่ถูกจำแนกในฝั่งบวกและลบโดยตัวจำแนก 1 vs 3

ประเภทข้อมูล	จำนวนตัวอย่างบวก	จำนวนตัวอย่างลบ
1	8	0
2	0	4
3	0	2
4	0	2
รวม	8	8

จากตารางที่ 5 จะคำนวณค่า p^+ และ p^- ได้เป็น $8/16$ และ $8/16$ ตามลำดับ และคำนวณค่า $p_i^+ \log_2 p_i^+$ และ $p_i^- \log_2 p_i^-$ เพื่อใช้แทนค่าในสมการ (35) ซึ่งจะคำนวณได้ค่าเอนโทรปีของตัวจำแนก 1 vs 3 คือ

$$= \frac{8}{16} \left(-\frac{8}{8} \log_2 \frac{8}{8} \right) + \frac{8}{16} \left(-\frac{4}{8} \log_2 \frac{4}{8} - \frac{2}{8} \log_2 \frac{2}{8} - \frac{2}{8} \log_2 \frac{2}{8} \right) = 0.75$$

ค่าเอนโทรปีของตัวจำแนก 1 vs 4 คำนวณโดยการฝึกตัวจำแนก 1 vs 4 แล้วใช้ระนาบหลายมิติเป็นตัวแบ่งฝั่งข้อมูลบวกและลบของแต่ละตัวอย่างฝึกซึ่งจะได้ระนาบหลายมิติตั้งรูปที่ 25 ซึ่งจะได้จำนวนข้อมูลที่ถูกจำแนกในฝั่งบวกและลบของแต่ละประเภทตามตารางที่ 6



รูป 25 เส้นระนาบหลายมิติของตัวจำแนก 1 vs 4

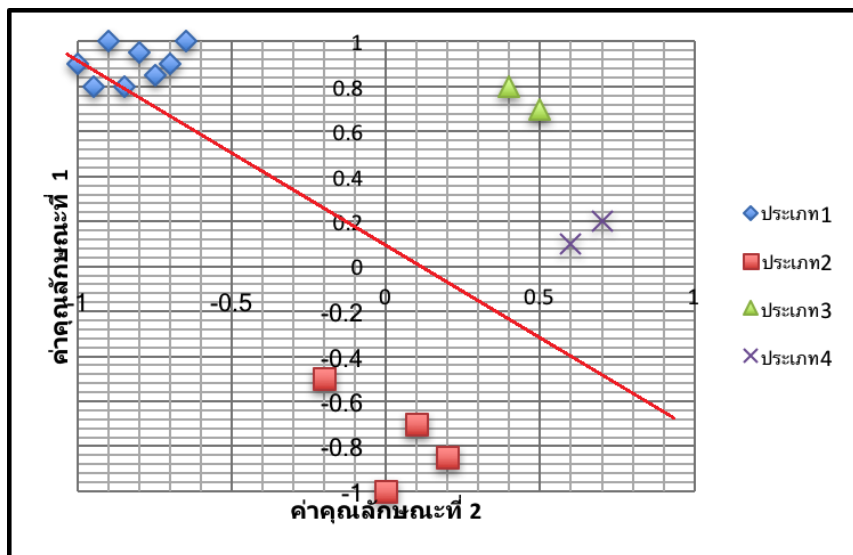
ตาราง 6 จำนวนข้อมูลที่ถูกจำแนกในฝั่งบวกและลบโดยตัวจำแนก 1 vs 4

ประเภทข้อมูล	จำนวนตัวอย่างบวก	จำนวนตัวอย่างลบ
1	8	0
2	0	4
3	0	2
4	0	2
รวม	8	8

จากตารางที่ 6 จะคำนวณค่า p^+ และ p^- ได้เป็น $8/16$ และ $8/16$ ตามลำดับ และคำนวณค่า $p_i^+ \log_2 p_i^+$ และ $p_i^- \log_2 p_i^-$ เพื่อใช้แทนค่าในสมการ (35) ซึ่งจะคำนวณได้ค่าเอนโทรปีของตัวจำแนก 1 vs 4 คือ

$$= \frac{8}{16} \left(-\frac{8}{8} \log_2 \frac{8}{8} \right) + \frac{8}{16} \left(-\frac{4}{8} \log_2 \frac{4}{8} - \frac{2}{8} \log_2 \frac{2}{8} - \frac{2}{8} \log_2 \frac{2}{8} \right) = 0.75$$

ค่าเอนโทรปีของตัวจำแนก 2 vs 3 คำนวณโดยการฝึกตัวจำแนก 2 vs 3 แล้วใช้ระนาบหลายมิติเป็นตัวแบ่งฝั่งข้อมูลบวกและลบของแต่ละตัวอย่างฝึกซึ่งจะได้ระนาบหลายมิติดังรูปที่ 26 ซึ่งจะได้จำนวนข้อมูลที่ถูกจำแนกในฝั่งบวกและลบของแต่ละประเภทตามตารางที่ 7



รูป 26 เส้นระนาบหลายมิติของตัวจำแนก 2 vs 3

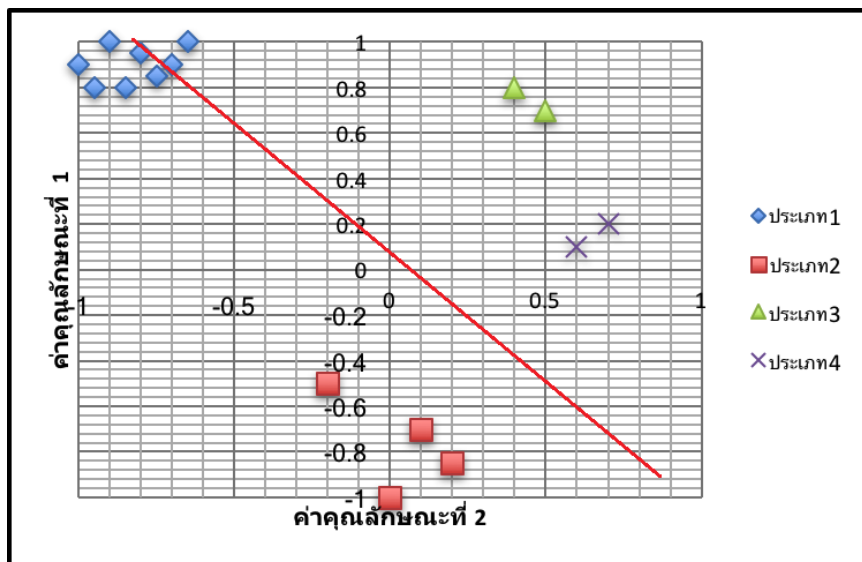
ตาราง 7 จำนวนข้อมูลที่ถูกจำแนกในฝั่งบวกและลบโดยตัวจำแนก 2 vs 3

ประเภทข้อมูล	จำนวนตัวอย่างบวก	จำนวนตัวอย่างลบ
1	2	6
2	4	0
3	0	2
4	0	2
รวม	6	10

จากตารางที่ 7 จะคำนวณค่า p^+ และ p^- ได้เป็น $6/16$ และ $10/16$ ตามลำดับ และคำนวณค่า $p_i^+ \log_2 p_i^+$ และ $p_i^- \log_2 p_i^-$ เพื่อใช้แทนค่าในสมการ (35) ซึ่งจะคำนวณได้ค่าเอนโทรปีของตัวจำแนก 2 vs 3 คือ

$$= \frac{6}{16} \left(-\frac{2}{6} \log_2 \frac{2}{6} - \frac{4}{6} \log_2 \frac{4}{6} \right) + \frac{10}{16} \left(-\frac{6}{10} \log_2 \frac{6}{10} - \frac{2}{10} \log_2 \frac{2}{10} - \frac{2}{10} \log_2 \frac{2}{10} \right) = 1.2012$$

ค่าเอนโทรปีของตัวจำแนก 2 vs 4 คำนวณโดยการฝึกตัวจำแนก 2 vs 4 แล้วใช้ระนาบหลายมิติเป็นตัวแบ่งฝั่งข้อมูลบวกและลบของแต่ละตัวอย่างฝึกซึ่งจะได้ระนาบหลายมิติดังรูปที่ 27 ซึ่งจะได้จำนวนข้อมูลที่ถูกจำแนกในฝั่งบวกและลบของแต่ละประเภทตามตารางที่ 8



รูป 27 เส้นระนาบหลายมิติของตัวจำแนก 2 vs 4

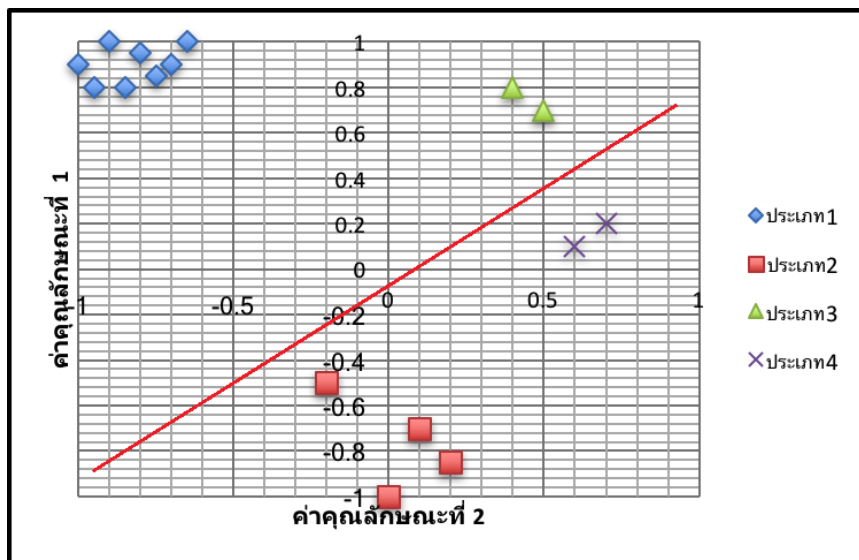
ตาราง 8 จำนวนข้อมูลที่ถูกจำแนกในฝั่งบวกและลบโดยตัวจำแนก 2 vs 4

ประเภทข้อมูล	จำนวนตัวอย่างบวก	จำนวนตัวอย่างลบ
1	6	2
2	4	0
3	0	2
4	0	2
รวม	10	6

จากตารางที่ 8 จะคำนวณค่า p^+ และ p^- ได้เป็น $10/16$ และ $6/16$ ตามลำดับ และคำนวณค่า $p_i^+ \log_2 p_i^+$ และ $p_i^- \log_2 p_i^-$ เพื่อใช้แทนค่าในสมการ (35) ซึ่งจะคำนวณได้ค่าเอนโทรปีของตัวจำแนก 2 vs 4 คือ

$$\begin{aligned}
 &= \frac{10}{16} \left(-\frac{6}{10} \log_2 \frac{6}{10} - \frac{4}{10} \log_2 \frac{4}{10} \right) \\
 &+ \frac{6}{16} \left(-\frac{2}{6} \log_2 \frac{2}{6} - \frac{2}{6} \log_2 \frac{2}{6} - \frac{2}{6} \log_2 \frac{2}{6} \right) = 1.2012
 \end{aligned}$$

ค่าเอนโทรปีของตัวจำแนก 3 vs 4 คำนวณโดยการฝึกตัวจำแนก 3 vs 4 แล้วใช้ระนาบหลายมิติเป็นตัวแบ่งฝั่งข้อมูลบวกและลบของแต่ละตัวอย่างฝึกซึ่งจะได้ระนาบหลายมิติดังรูปที่ 28 ซึ่งจะได้จำนวนข้อมูลที่ถูกจำแนกในฝั่งบวกและลบของแต่ละประเภทตามตารางที่ 9



รูป 28 เส้นระนาบหลายมิติของตัวจำแนก 3 vs 4

ตาราง 9 จำนวนข้อมูลที่ถูกจำแนกในฝั่งบวกและลบโดยตัวจำแนก 3 vs 4

ประเภทข้อมูล	จำนวนตัวอย่างบวก	จำนวนตัวอย่างลบ
1	8	0
2	0	4
3	2	0
4	0	2
รวม	10	6

จากตารางที่ 9 จะคำนวณค่า p^+ และ p^- ได้เป็น $10/16$ และ $6/16$ ตามลำดับ และคำนวณค่า $p_i^+ \log_2 p_i^+$ และ $p_i^- \log_2 p_i^-$ เพื่อใช้แทนค่าในสมการ (35) ซึ่งจะคำนวณได้ค่าเอนโทรปีของตัวจำแนก 3 vs 4 คือ

$$= \frac{10}{16} \left(-\frac{8}{10} \log_2 \frac{8}{10} - \frac{2}{10} \log_2 \frac{2}{10} \right) + \frac{6}{16} \left(-\frac{4}{6} \log_2 \frac{4}{6} - \frac{2}{6} \log_2 \frac{2}{6} \right) = 0.7956$$

จากการคำนวณพบว่า มีตัวจำแนกที่มีค่าเอนโทรปีต่ำที่สุด คือ 1 vs 3 และ 1 vs 4 ซึ่งระบบได้
 สุ่มเลือกตัวจำแนก 1 vs 3 เป็นตัวตั้งในการแบ่งกลุ่มข้อมูลบวกและลบ ซึ่งเมื่อได้ผลกลุ่มฝั่งบวกคือ
 ประเภทข้อมูลที่ 1 และฝั่งลบคือประเภทข้อมูลที่ 2, 3, 4 จึงได้ทำการฝึกตัวจำแนก 1 vs (2, 3, 4) เพื่อ
 ใช้สำหรับปมแรก

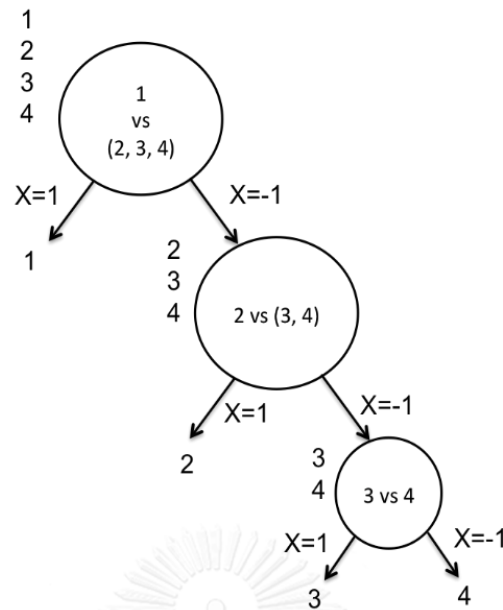
ที่ปมลูกฝั่งบวกของปมแรก มีเพียงประเภทข้อมูลเดียว คือประเภทที่ 1 จึงหยุดการสร้างกิ่ง
 ต้นไม้ ส่วนปมลูกฝั่งลบประกอบด้วยประเภทข้อมูล 3 ประเภท จึงทำการสร้างกิ่งต้นไม้ต่อไป ซึ่งมีตัว
 จำแนกที่ต้องพิจารณา 3 ตัว ได้แก่ 2 vs 3, 2 vs 4, 3 vs 4 ซึ่งมีข้อมูลเหลือให้พิจารณาจากประเภท
 ข้อมูลที่ 2, 3, 4 จำนวน 8 ตัว ซึ่งคำนวณได้ค่าเอนโทรปี ดังนี้

$$\begin{aligned} & \text{ค่าเอนโทรปีของตัวจำแนก 2 vs 3} \\ & = \frac{4}{8} \left(-\frac{4}{4} \log_2 \frac{4}{4} \right) + \frac{4}{8} \left(-\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4} \right) = 0.5 \end{aligned}$$

$$\begin{aligned} & \text{ค่าเอนโทรปีของตัวจำแนก 2 vs 4} \\ & = \frac{4}{8} \left(-\frac{4}{4} \log_2 \frac{4}{4} \right) + \frac{4}{8} \left(-\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4} \right) = 0.5 \end{aligned}$$

$$\begin{aligned} & \text{ค่าเอนโทรปีของตัวจำแนก 3 vs 4} \\ & = \frac{2}{8} \left(-\frac{2}{2} \log_2 \frac{2}{2} \right) + \frac{6}{8} \left(-\frac{4}{6} \log_2 \frac{4}{6} - \frac{2}{6} \log_2 \frac{2}{6} \right) = 0.6887 \end{aligned}$$

จากการคำนวณพบว่า มีตัวจำแนกที่มีค่าเอนโทรปีต่ำที่สุด คือ 2 vs 3 และ 2 vs 4 ซึ่งระบบได้
 สุ่มเลือกตัวจำแนก 2 vs 3 เป็นตัวตั้งในการแบ่งกลุ่มข้อมูลบวกและลบ ซึ่งเมื่อได้ผลกลุ่มฝั่งบวกคือ
 ประเภทข้อมูลที่ 2 และฝั่งลบคือประเภทข้อมูลที่ 3, 4 จึงได้ทำการฝึกตัวจำแนก 2 vs (3, 4) เพื่อใช้
 สำหรับปมนี้ ซึ่งจะเห็นว่าปมลูกฝั่งบวกมีประเภทข้อมูลเดียว คือประเภทที่ 2 ส่วนปมลูกด้านลบมี 2
 ประเภทข้อมูล คือ 3, 4 ซึ่งจึงใช้ตัวจำแนก 3 vs 4 สำหรับปมลูกฝั่งลบของปมนี้ เป็นอันสิ้นสุด
 กระบวนการสร้างต้นไม้ และได้โครงสร้างต้นไม้สำหรับปัญหาตัวอย่างนี้ ดังรูปที่ 28



รูป 29 ต้นไม้ตัดสินใจตามสารสนเทศสำหรับปัญหาตัวอย่างในรูปที่ 22

ต้นไม้ตัดสินใจตามสารสนเทศมีข้อดีหลายประการ ประการแรกคือ ประเภทข้อมูลที่เป็นคำตอบจะถูกค้นพบในไม่กี่ระดับชั้นจากปรกราก ประการที่สองคือ จะไม่มีปมใบที่มีประเภทข้อมูลซ้ำกัน ดังนั้นความสูงของต้นไม้จะถูกจำกัดให้ไม่มากเกินไป ประการสุดท้ายคือ ไม่มีการสูญหายของข้อมูล เนื่องจากไม่มีการตัดเล็มข้อมูล ซึ่งเราจะทำการทดสอบประสิทธิภาพการทำงานของต้นไม้ตัดสินใจตามสารสนเทศต่อไปในบทที่ 4

3.2 ต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป

(The Information Based and Generalization Error Estimation Decision Tree (IBGE-DTree))

เป็นเทคนิคการสร้างต้นไม้ตัดสินใจ ซึ่งปรับปรุงประสิทธิภาพของต้นไม้ตัดสินใจตามสารสนเทศด้วยการนำการประเมินค่าผิดพลาดโดยนัยทั่วไป (Generalization Error Estimation) มาร่วมกับค่าเอนโทรปีในการพิจารณาเลือกตัวจำแนกที่เหมาะสมในแต่ละปม

การสร้างต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป ถูกบรรยายไว้ตามโค้ดเทียมแสดงขั้นตอนวิธีที่ 3 โดยมีข้อแตกต่างจากต้นไม้ตัดสินใจตามสารสนเทศ คือ บรรทัดที่ 12-17 แทนที่จะเลือกตัวจำแนกที่มีค่าเอนโทรปีต่ำสุดในการประมวลผลเพื่อสร้างปม แต่ต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป จะทำการเรียงลำดับตัวจำแนกตามค่าเอนโทรปีจากน้อยไปมาก และหยิบเอาตัวจำแนกที่มีค่าต่ำจำนวนหนึ่ง ไปทำการ

แบ่งกลุ่มฝั่งบวกและลบแล้วทำการฝึกตัวจำแนกเหล่านั้น จากนั้นจึงเลือกเอาตัวจำแนกที่มีค่าการประมาณความผิดพลาดโดยนัยทั่วไป ต่ำสุดเป็นตัวจำแนกสำหรับบมนั้น โดยใช้กลุ่มประเภทข้อมูลบวกและลบจากตัวจำแนกที่ได้กับปมลูกทั้งด้านซ้ายและขวาด้วย

Algorithm 3 Information-Based and Generalization-Error Estimation Decision Tree SVM (IBGE-DTree)

```

1: procedure IBGE-DTREE
2:   Initialize the tree  $T$  with root node  $Root$ 
3:   Initialize the set of candidate output classes  $S = \{1, 2, 3, \dots, N\}$ 
4:   Create the all binary classifiers ( $i$  vs  $j$ );  $i, j \in S$ 
5:   Construct Tree ( $Root, S$ )
6:   return  $T$ 
7: end procedure
8: procedure CONSTRUCT TREE (NODE  $D$ , CANDIDATE CLASSES  $K$ )
9:   for each binary classifiers ( $i$  vs  $j$ );  $i, j \in K$ ;  $i < j$  do:-
10:    Calculate the entropy using training data of all classes in  $K$ 
11:   end for
12:   Sort the list of the initial classifiers ( $i$  vs  $j$ ) in ascending order by the entropy as  $h_1 \dots h_{all}$ 
13:   for each initial classifiers  $h_k$ ;  $k = \{1, 2, 3, \dots, n\}$ ,  $n$ =number of considering classifiers do
14:     final classifier  $h_k'$ , positive classes  $P_k$ , negative classes  $N_k \leftarrow$  Class-grouping-by-majority( $h_k, K$ )
15:     calculate generalization error estimation of final classifier  $h_k'$ 
16:   end for
17:    $D$ .classifier  $\leftarrow$  final classifiers with the lowest generalization error among  $h_1' \dots h_n'$ 
18:    $P' \leftarrow P$  used for training the final classifier with the lowest generalization error estimation
19:    $N' \leftarrow N$  used for training the final classifier with the lowest generalization error estimation
20:   Initialize new node  $L$ ;  $D$ .left-child-node  $\leftarrow L$ 
21:   Initialize new node  $R$ ;  $D$ .right-child-node  $\leftarrow R$ 
22:   if  $|P'| > 1$  then Construct Tree ( $L, P'$ ) else  $L$  is the leaf node with answer class  $P'$ 
23:   if  $|N'| > 1$  then Construct Tree ( $R, N'$ ) else  $R$  is the leaf node with answer class  $N'$ 
24: end procedure

```

ขั้นตอนวิธีที่ 3 วิธีการสร้างต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป

ค่าประมาณความผิดพลาดโดยนัยทั่วไป ของตัวจำแนกตามรายละเอียดในหัวข้อที่ 2.1.2 นั้นสามารถหาได้โดยตรงจากการประมาณค่าด้วยวิธีทดสอบไขว้ k พับ (k -fold cross-validation) หรืออีกวิธีการหนึ่งสามารถทำได้โดยการแทนค่าลงในสมการที่ 29 โดยตั้งสมมติฐานต่างๆ ที่เหมาะสม⁵⁴ โดยหากเราใช้วิธีนี้ เราจะสามารถเปรียบเทียบค่าความผิดพลาดโดยนัยทั่วไปโดยเปรียบเทียบกันภายในชุดข้อมูลเดียวกัน ภายใต้สภาพแวดล้อมเดียวกันได้ ซึ่งในบทที่ 4 เรากำหนดค่า $c = 0.1$ และ $\delta = 0.01$ สำหรับการทดลอง

ตัวอย่างของการฝึกเพื่อสร้างต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป สำหรับปัญหาที่มีประเภทข้อมูล 4 ประเภท บนระนาบ 2 มิติ ซึ่งประกอบด้วยข้อมูลฝึกจำนวน 16 ตัว ดังรูปที่ 22

ในการฝึกและสร้างต้นไม้ตามขั้นตอนวิธีที่ 3 กระบวนการจะเริ่มสร้างที่ปมราก (Root node) ซึ่งมีตัวจำแนกที่ถูกพิจารณาทั้งสิ้น 6 ตัว คือ 1 vs 2, 1 vs 3, 1 vs 4, 2 vs 3, 2 vs 4, 3 vs 4 ซึ่งเมื่อนำไปประมวลผลด้วยโปรแกรม เพื่อคำนวณหาค่าเอนโทรปี จะได้ค่าดังนี้

ค่าเอนโทรปีของตัวจำแนก 1 vs 2

$$= \frac{10}{16} \left(-\frac{8}{10} \log_2 \frac{8}{10} - \frac{2}{10} \log_2 \frac{2}{10} \right) + \frac{6}{16} \left(-\frac{4}{6} \log_2 \frac{4}{6} - \frac{2}{6} \log_2 \frac{2}{6} \right) = 0.7956$$

ค่าเอนโทรปีของตัวจำแนก 1 vs 3

$$= \frac{8}{16} \left(-\frac{8}{8} \log_2 \frac{8}{8} \right) + \frac{8}{16} \left(-\frac{4}{8} \log_2 \frac{4}{8} - \frac{2}{8} \log_2 \frac{2}{8} - \frac{2}{8} \log_2 \frac{2}{8} \right) = 0.75$$

ค่าเอนโทรปีของตัวจำแนก 1 vs 4

$$= \frac{8}{16} \left(-\frac{8}{8} \log_2 \frac{8}{8} \right) + \frac{8}{16} \left(-\frac{4}{8} \log_2 \frac{4}{8} - \frac{2}{8} \log_2 \frac{2}{8} - \frac{2}{8} \log_2 \frac{2}{8} \right) = 0.75$$

ค่าเอนโทรปีของตัวจำแนก 2 vs 3

$$= \frac{6}{16} \left(-\frac{2}{6} \log_2 \frac{2}{6} - \frac{4}{6} \log_2 \frac{4}{6} \right) + \frac{10}{16} \left(-\frac{6}{10} \log_2 \frac{6}{10} - \frac{2}{10} \log_2 \frac{2}{10} - \frac{2}{10} \log_2 \frac{2}{10} \right) = 1.2012$$

ค่าเอนโทรปีของตัวจำแนก 2 vs 4

$$= \frac{10}{16} \left(-\frac{6}{10} \log_2 \frac{6}{10} - \frac{4}{10} \log_2 \frac{4}{10} \right) + \frac{6}{16} \left(-\frac{2}{6} \log_2 \frac{2}{6} - \frac{2}{6} \log_2 \frac{2}{6} - \frac{2}{6} \log_2 \frac{2}{6} \right) = 1.2012$$

ค่าเอนโทรปีของตัวจำแนก 3 vs 4

$$= \frac{10}{16} \left(-\frac{8}{10} \log_2 \frac{8}{10} - \frac{2}{10} \log_2 \frac{2}{10} \right) + \frac{6}{16} \left(-\frac{4}{6} \log_2 \frac{4}{6} - \frac{2}{6} \log_2 \frac{2}{6} \right) = 0.7956$$

เมื่อนำค่าเอนโทรปีมาเรียงจากน้อยไปหามาก จะได้ตัวจำแนกตามลำดับ คือ 1 vs 3, 1 vs 4, 1 vs 2, 3 vs 4, 2 vs 3, 2 vs 4 ซึ่งสมมติเราตั้งคือ n ในบรรทัดที่ 13 ของขั้นตอนวิธี 3 ไว้ที่ร้อยละ 20 ของจำนวนตัวจำแนก ทั้งหมดที่พิจารณา ซึ่งในปมนี้ คือ 6 ตัว ดังนั้นในที่นี้จะได้อัตราส่วนค่า n ได้เป็น 1.2 ซึ่งเราจะปัดเศษขึ้นเป็น 2 และเมื่อนำตัวจำแนก 1 vs 3 กับ 1 vs 4 ไปทำการประมาณค่าความผิดพลาดโดยนัยทั่วไปแล้วนำมาเปรียบเทียบกันโดยการแทนค่าลงในสมการที่ (29) โดยค่า $m = 8$, $c = 0.1$, $\delta = 0.01$, l/m คือร้อยละความผิดพลาดจากการฝึกตัวจำแนก, $c/m = 0.1/8$, R^2/Δ^2 คือค่า VC dimension ซึ่งได้จากการฝึกสอนตัวจำแนก ซึ่งจากการคำนวณด้วยโปรแกรมพบว่าค่าประมาณของของตัวจำแนก 1 vs 3 ต่ำกว่า 1 vs 2 จึงใช้ตัวจำแนก 1 vs 3 เป็นตัวตั้งในการแบ่งกลุ่มข้อมูลบวกและลบ ได้ผลกลุ่มฝั่งบวกคือประเภทข้อมูลที่ 1 และฝั่งลบคือประเภทข้อมูลที่ 2, 3, 4 จึงได้ทำการฝึกตัวจำแนก 1 vs (2, 3, 4) เพื่อใช้สำหรับปมแรก

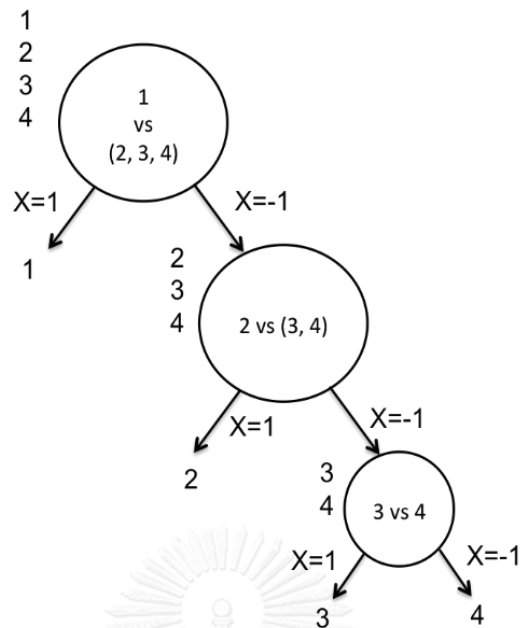
ที่ปมลูกฝั่งบวกของปมราก มีเพียงประเภทข้อมูลเดียว คือประเภทที่ 1 จึงหยุดการสร้างกิ่งต้นไม้ ส่วนปมลูกฝั่งลบประกอบด้วยประเภทข้อมูล 3 ประเภท จึงทำการสร้างกิ่งต้นไม้ต่อไป ซึ่งมีตัวจำแนกที่ต้องพิจารณา 3 ตัว ได้แก่ 2 vs 3, 2 vs 4, 3 vs 4 ซึ่งมีข้อมูลเหลือให้พิจารณาจากประเภทข้อมูลที่ 2, 3, 4 จำนวน 8 ตัว ซึ่งคำนวณได้ค่าเอนโทรปี ดังนี้

$$\begin{aligned} & \text{ค่าเอนโทรปีของตัวจำแนก 2 vs 3} \\ & = \frac{4}{8} \left(-\frac{4}{4} \log_2 \frac{4}{4} \right) + \frac{4}{8} \left(-\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4} \right) = 0.5 \end{aligned}$$

$$\begin{aligned} & \text{ค่าเอนโทรปีของตัวจำแนก 2 vs 4} \\ & = \frac{4}{8} \left(-\frac{4}{4} \log_2 \frac{4}{4} \right) + \frac{4}{8} \left(-\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4} \right) = 0.5 \end{aligned}$$

$$\begin{aligned} & \text{ค่าเอนโทรปีของตัวจำแนก 3 vs 4} \\ & = \frac{2}{8} \left(-\frac{2}{2} \log_2 \frac{2}{2} \right) + \frac{6}{8} \left(-\frac{4}{6} \log_2 \frac{4}{6} - \frac{2}{6} \log_2 \frac{2}{6} \right) = 0.6887 \end{aligned}$$

เมื่อนำค่าเอนโทรปีมาเรียงจากน้อยไปหามาก จะได้ตัวจำแนกตามลำดับ คือ 2 vs 3, 2 vs 4, 3 vs 4 ซึ่ง 20 ของจำนวนตัวจำแนกทั้งหมดที่พิจารณาในปมนี้ คือ 3 ตัว ดังนั้นในที่นี้จะได้อันดับค่า n ได้เป็น 0.6 ซึ่งเราจะปัดเศษขึ้นเป็น 1 ดังนั้นจึงเลือก 2 vs 3 เป็นตัวตั้งในการแบ่งกลุ่มข้อมูลบวกและลบ ซึ่งเมื่อได้ผลกลุ่มฝั่งบวกคือประเภทข้อมูลที่ 2 และฝั่งลบคือประเภทข้อมูลที่ 3, 4 จึงได้ทำการฝึกตัวจำแนก 2 vs (3, 4) เพื่อใช้สำหรับปมนี้ ซึ่งจะเห็นว่าปมลูกฝั่งบวกมีประเภทข้อมูลเดียว คือประเภทที่ 2 ส่วนปมลูกด้านลบมี 2 ประเภทข้อมูล คือ 3, 4 ซึ่งจึงใช้ตัวจำแนก 3 vs 4 สำหรับปมลูกฝั่งลบของปมนี้ เป็นอันสิ้นสุดกระบวนการสร้างต้นไม้ และได้โครงสร้างต้นไม้สำหรับปัญหาตัวอย่างนี้ ดังรูปที่ 29



รูป 30 ต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป
สำหรับปัญหาตัวอย่างในรูปที่ 22

ในแง่ของคุณสมบัติที่เป็นประโยชน์ เนื่องจากต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป เป็นเทคนิคที่ยกระดับความแม่นยำในการจำแนกของต้นไม้ตัดสินใจตามสารสนเทศ ดังนั้นข้อดีจึงคล้ายกันมาก แต่อย่างไรก็ดี การที่นำค่าประมาณความผิดพลาดโดยนัยทั่วไปมาผสมกับค่าเอนโทรปีเพื่อใช้เลือกตัวจำแนก จะสามารถเลือกตัวจำแนกในแต่ละปมของต้นไม้ที่มีประสิทธิภาพมากกว่า ซึ่งเราจะทดสอบประสิทธิภาพของต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป ในบทที่ 4

บทที่ 4

การทดลองและผล

ในบทนี้จะกล่าวถึงรายละเอียดของการทดลองและผลการทดลองที่ได้ โดยเป็นการทำการทดลองเพื่อทดสอบประสิทธิภาพการทำงานของต้นไม้ตัดสินใจตามสารสนเทศ และต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป เทียบกับเทคนิคอื่น ได้แก่ โอวีโอ โอวีเอ ดีดีเอ จี เอดีเอจี บีทีเอส-จี และ ซี-บีทีเอส-จี ด้วย

ในบทนี้จะแบ่งเนื้อหาออกเป็น 3 ส่วน ได้แก่ ชุดข้อมูลที่ใช้ในการทดลอง รายละเอียดการตั้งค่าต่างๆ ในการทดลอง และผลการทดลอง

4.1 ชุดข้อมูลที่ใช้ในการทดลอง

ในการทดลองนี้ใช้ชุดข้อมูลจากพื้นที่เก็บข้อมูลการเรียนรู้ของเครื่องยูซีไอ (UCI Machine Learning Repository)⁴⁹ ซึ่งมีรายละเอียดตามตารางที่ 5 โดยในตารางจะแสดงรายละเอียดของชุดข้อมูลทดลอง ได้แก่ ชื่อชุดข้อมูล (Dataset Name) จำนวนประเภทข้อมูล (#Classes) จำนวนคุณสมบัติ (#Attributes) และจำนวนตัวอย่าง (#Examples) โดยในทุกชุดข้อมูลหากมีทั้งชุดข้อมูลฝึกและชุดข้อมูลทดสอบจะนำมารวมกันเป็นชุดเดียว แล้วนำไปใช้ในการทดลองแบบตรวจสอบไขว้สิบพับ (10-fold cross validation) เพื่อประเมินความแม่นยำในการจำแนก นอกจากนี้เราได้ทำการปรับช่วงข้อมูลให้อยู่ระหว่าง $[-1, 1]$

ตาราง 10 รายละเอียดของชุดข้อมูลทดลอง

ชุดข้อมูล	จำนวน ประเภทข้อมูล	จำนวน คุณลักษณะ	จำนวน ตัวอย่าง
Page Block	5	10	5473
Segment	7	18	3210
Shuttle	7	9	58000
Arrhyth	9	255	438
Cardiotocography	10	21	2126
Mfeat-factor	10	216	2000
Mfeat-fourier	10	76	2000
Mfeat-karhunen	10	64	2000
Optdigit	10	62	5620
Pendigit	10	16	10992
Primary Tumor	13	15	315
Libras Movement	15	90	360
Abalone	16	8	4098
Chess (King-Rook vs King)	18	6	28056
Spectrometer	21	101	475
Isolet	26	34	7797
Letter	26	16	20052
Plant Margin	100	64	1600
Plant Shape	100	64	1600
Plant Texture	100	64	1599

4.2 รายละเอียดการตั้งค่าต่างๆ ในการทดลอง

ในส่วนของ SVM library ที่ใช้ นำมาจากชุดซอฟต์แวร์ SVMLight⁵⁵ รุ่น 6.02 โดยเราใช้ RBF Kernel ซึ่งกำหนดค่าแกมมา (γ) เป็น 0.001, 0.01, 0.1, 1, 10 และค่าพารามิเตอร์ปกติ C (regularization parameters C) เป็น 1, 10, 100, 1000 โดยทุกๆ เทคนิคที่มีพื้นฐานมาจากวิธีโอวีโอ เราจะเลือกค่าแกมมาและค่า C ที่ให้ค่าความผิดพลาดในการจำแนกเฉลี่ยต่ำสุดบนชุดข้อมูลฝึกพบที่ 1 ของวิธีโอวีโอไปประยุกต์ใช้กับทุกเทคนิคในทุกพบที่เหลือ ยกเว้นในวิธีโอวีโอเราจะใช้ค่าผิดพลาดในการจำแนกเฉลี่ยต่ำสุดบนเทคนิคโอวีโอแทน

เพื่อที่เราจะสามารถเปรียบเทียบประสิทธิภาพของวิธีต้นไม้ตัดสินใจตามสารสนเทศ และ วิธีต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป เทียบกับเทคนิคที่มีโครงสร้างเป็นต้นไม้อื่นๆ ได้ เราจึงได้ทำการออกแบบ บีทีเอส-จี (BTS-G) และ ซี-บีทีเอส-จี (c-BTS-G) ซึ่งเป็นรุ่นปรับปรุงของ บีทีเอส (BTS) และ บีทีเอส-จี (BTS-G) ดั้งเดิมโดยการนำเอา “การจับกลุ่มประเภทข้อมูลตามข้อมูลส่วนใหญ่” เข้าไปเพิ่มประสิทธิภาพของวิธีบีทีเอสและวิธีบีทีเอส-จี โดยสำหรับวิธีบีทีเอส-จี เราทำการสุ่มเลือกตัวจำแนกในแต่ละปมทั้งสิ้น 10 ครั้ง แบบเดียวกันกับบีทีเอสดั้งเดิม

สำหรับวิธีดีดีเอจีและวิธีเอดีเอจี เราทำการทดสอบด้วยการสุ่มเลือกลำดับการเลือก 50,000 ลำดับ แล้วจึงคำนวณหาค่าเฉลี่ย ในส่วนของต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป เราตั้งค่าจำนวนของตัวจำแนก n ในบรรทัดที่ 13 ของขั้นตอนวิธี 3 ที่ร้อยละ 20 ของจำนวนตัวจำแนกทั้งหมดที่เป็นไปได้ เช่น หากมีจำนวนประเภทข้อมูล 10 ประเภทที่ถูกพิจารณา จำนวนตัวจำแนกที่เป็นไปได้ คือ 45 ตัว ดังนั้นเราจะหยิบเอาตัวจำแนกจำนวน n ตัวที่จะพิจารณาด้วยค่าการประเมินความผิดพลาดโดยนัยทั่วไป 9 ตัว และหากคำนวณตัวเลข n ได้เลขจุดทศนิยม จะทำการปัดขึ้นเป็นเลขจำนวนเต็ม

4.3 ผลการทดลอง

ผลการทดลองเทคนิควิธีต้นไม้ตัดสินใจตามสารสนเทศ และ วิธีต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป เทียบกับวิธีการดั้งเดิมแสดงอยู่ในตารางที่ 11 ถึงตารางที่ 14 โดยตารางที่ 11 และตารางที่ 12 แสดงค่าเฉลี่ยความแม่นยำในการจำแนกของทุกชุดข้อมูล ตารางที่ 13 แสดงผลการเปรียบเทียบนัยสำคัญของความแตกต่างของความแม่นยำเฉลี่ยในการจำแนกด้วยวิธี Wilcoxon Signed Rank Test ส่วนตารางที่ 14 แสดงจำนวนครั้งในการตัดสินใจเฉลี่ยที่ใช้ในการตัดสินใจคำตอบของตัวอย่างทดสอบ

ในตารางที่ 11 และตารางที่ 12 แสดงค่าความแม่นยำในการจำแนกเฉลี่ยจากวิธีตรวจสอบไขว้ สิบพับ พร้อมกับส่วนเบี่ยงมาตรฐาน โดยตัวเลขหนาแสดงถึงว่าให้ความแม่นยำสูงสุดในบรรดาเทคนิค ที่ทำการเปรียบเทียบของชุดข้อมูลต่างๆ โดยลำดับของความแม่นยำ (rank) ของแต่ละเทคนิคจะระบุ อยู่ในวงเล็บ ผลการทดลองแสดงให้เห็นว่าเทคนิคที่ให้ความแม่นยำในการจำแนกสูงสุด คือวิธีโอวีโอ ตามด้วยวิธีเอดีเอจี โอวีเอ และ ดีดีเอจี ตามลำดับ ส่วนในกลุ่มที่ใช้โครงสร้างต้นไม้ด้วยกัน วิธีต้นไม้ ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป ให้ความแม่นยำในการจำแนก สูงสุด ตามด้วย วิธีต้นไม้ตัดสินใจตามสารสนเทศ บีทีเอส-จี และ ซี-บีทีเอส-จี ตามลำดับ



ตาราง 11 ค่าเฉลี่ยของผลความแม่นยำในการจำแนกหรือหาค่าส่วนเบี่ยงเบนมาตรฐาน

โดยลำดับของความแม่นยำ (rank) ของแต่ละเทคนิคระบุอยู่ในวงเล็บ

ชื่อชุดข้อมูล	OVA	OVO	DDAG	ADAG
Page Block	96.857 ± 0.478 (1)	96.735 ± 0.760 (3)	96.729 ± 0.764 (4)	96.740 ± 0.757 (2)
Segment	97.359 ± 1.180 (5)	97.431 ± 0.860 (3)	97.442 ± 0.848 (1)	97.436 ± 0.854 (2)
Shuttle	99.914 ± 0.053 (5)	99.920 ± 0.054 (1)	99.920 ± 0.054 (1)	99.920 ± 0.054 (1)
Arrhyth	72.603 ± 7.041 (2)	73.146 ± 6.222 (1)	67.375 ± 7.225 (8)	67.484 ± 7.318 (7)
Cardiotocography	83.208 ± 1.661 (5)	84.431 ± 1.539 (1)	84.241 ± 1.609 (3)	84.351 ± 1.607 (2)
Mfeat-factor	98.200 ± 1.033 (1)	98.033 ± 0.908 (3)	98.011 ± 0.941 (5)	98.019 ± 0.919 (4)
Mfeat-fourier	84.850 ± 1.528 (6)	85.717 ± 1.603 (1)	85.702 ± 1.589 (3)	85.708 ± 1.585 (2)
Mfeat-karhunen	98.000 ± 0.943 (1)	97.913 ± 0.750 (3)	97.894 ± 0.726 (5)	97.900 ± 0.722 (4)
Optdigit	99.324 ± 0.373 (2)	99.964 ± 0.113 (1)	99.288 ± 0.346 (3)	99.288 ± 0.346 (3)
Pendigit	99.554 ± 0.225 (4)	99.591 ± 0.203 (1)	99.569 ± 0.213 (3)	99.574 ± 0.211 (2)
Primary Tumor	46.667 ± 7.011 (3)	50.212 ± 7.376 (1)	39.278 ± 6.419 (8)	39.486 ± 6.483 (7)
Libras Movement	90.000 ± 2.986 (1)	89.074 ± 3.800 (2)	89.034 ± 3.729 (3)	89.017 ± 3.687 (4)
Abalone	16.959 ± 2.388 (8)	28.321 ± 1.516 (1)	24.093 ± 3.044 (7)	24.258 ± 3.154 (6)
Krkoft	85.750 ± 0.769 (1)	82.444 ± 0.628 (2)	81.952 ± 0.643 (4)	82.235 ± 0.634 (3)
Spectrometer	51.579 ± 6.256 (8)	68.421 ± 5.007 (1)	68.052 ± 4.706 (4)	68.392 ± 4.796 (2)
Isolet	94.947 ± 0.479 (1)	94.898 ± 0.648 (2)	94.872 ± 0.631 (4)	94.885 ± 0.643 (3)
Letter	97.467 ± 0.305 (4)	97.813 ± 0.382 (1)	97.746 ± 0.357 (3)	97.787 ± 0.360 (2)
Plant Margin	82.875 ± 2.655 (4)	84.401 ± 2.426 (1)	84.238 ± 2.516 (3)	84.341 ± 2.607 (2)
Plant Shape	70.938 ± 2.783 (3)	71.182 ± 3.295 (1)	70.922 ± 3.393 (4)	71.090 ± 3.313 (2)
Plant Texture	87.179 ± 2.808 (1)	86.387 ± 2.374 (2)	86.173 ± 2.519 (4)	86.259 ± 2.510 (3)
Average Rank	3.35	1.7	4.00	3.15

ตาราง 12 ค่าเฉลี่ยของผลความแม่นยำในการจำแนกพร้อมทั้งค่าส่วนเบี่ยงเบนมาตรฐาน

โดยลำดับของความแม่นยำ (rank) ของแต่ละเทคนิคระบุอยู่ในวงเล็บ (ต่อ)

ชื่อชุดข้อมูล	BTS-G	c-BTS-G	IB-DTree	IBGE-DTree
Page Block	96.622 ± 0.812 (5)	96.492 ± 0.884 (8)	96.565 ± 0.779 (7)	96.620 ± 0.852 (6)
Segment	97.273 ± 1.076 (7)	97.100 ± 0.957 (8)	97.316 ± 0.838 (6)	97.403 ± 1.100 (4)
Shuttle	99.914 ± 0.050 (5)	99.914 ± 0.053 (5)	99.916 ± 0.050 (4)	99.910 ± 0.053 (8)
Arrhyth	71.918 ± 5.688 (4)	71.918 ± 5.189 (4)	71.005 ± 5.836 (6)	72.146 ± 4.043 (3)
Cardiotocography	83.048 ± 2.147 (7)	82.926 ± 2.106 (8)	83.819 ± 1.710 (4)	83.161 ± 2.490 (6)
Mfeat-factor	97.810 ± 0.882 (8)	98.000 ± 0.888 (6)	98.000 ± 0.768 (6)	98.200 ± 0.816 (1)
Mfeat-fourier	84.235 ± 1.636 (8)	84.350 ± 1.700 (7)	85.200 ± 1.605 (4)	85.150 ± 1.717 (5)
Mfeat-karhunen	97.450 ± 0.832 (6)	97.050 ± 0.725 (8)	97.450 ± 0.879 (6)	97.950 ± 1.141 (2)
Optdigit	99.002 ± 0.266 (8)	99.039 ± 0.308 (7)	99.164 ± 0.288 (5)	99.093 ± 0.395 (6)
Pendigit	99.442 ± 0.184 (7)	99.427 ± 0.201 (8)	99.445 ± 0.198 (6)	99.454 ± 0.318 (5)
Primary Tumor	43.016 ± 3.824 (5)	40.635 ± 4.813 (6)	47.937 ± 4.567 (2)	44.762 ± 5.478 (4)
Libras Movement	87.861 ± 4.151 (8)	88.611 ± 3.715 (5)	88.056 ± 3.479 (6)	88.056 ± 3.057 (6)
Abalone	26.635 ± 1.236 (3)	26.013 ± 1.218 (4)	25.281 ± 0.904 (5)	26.745 ± 0.809 (2)
Kirkopt	77.137 ± 0.880 (8)	78.190 ± 0.783 (7)	79.006 ± 0.792 (6)	80.610 ± 1.039 (5)
Spectrometer	59.432 ± 5.563 (6)	52.421 ± 5.192 (7)	68.211 ± 3.397 (3)	67.789 ± 6.296 (5)
Isolet	92.850 ± 0.799 (7)	92.677 ± 0.702 (8)	93.639 ± 0.261 (6)	94.011 ± 0.640 (5)
Letter	96.174 ± 0.321 (7)	96.369 ± 0.423 (6)	96.135 ± 0.312 (8)	96.409 ± 0.344 (5)
Plant Margin	77.994 ± 1.946 (8)	78.188 ± 2.739 (7)	80.563 ± 3.638 (5)	79.313 ± 2.863 (6)
Plant Shape	63.219 ± 1.808 (7)	61.750 ± 3.594 (8)	67.000 ± 2.853 (5)	66.750 ± 2.408 (6)
Plant Texture	78.893 ± 2.547 (7)	78.174 ± 3.997 (8)	80.425 ± 3.602 (6)	80.863 ± 2.828 (5)
Average Rank	6.55	6.85	5.35	4.75

ในตารางที่ 13 แสดงถึงผลการเปรียบเทียบความแม่นยำในการจำแนกด้วย Wilcoxon Signed Rank Test⁵⁰ เพื่อทดสอบว่าระหว่างแต่ละคู่เทคนิคมีความแตกต่างกันอย่างมีนัยสำคัญหรือไม่ โดยตัวเลขหนาแสดงว่าแตกต่างกันอย่างมีนัยสำคัญด้วยระดับนัยสำคัญ (significant level) 0.05 ตัวเลขในวงเล็บระบุถึงผลการเปรียบเทียบจำนวนชุดทดสอบที่ ชนะ-แพ้-เสมอ ระหว่างเทคนิคที่เปรียบเทียบ เช่น ช่องซ้ายบนสุดของตารางซึ่งเทียบระหว่างโอวีเอ (ระบุทางด้านซ้ายของช่อง) และโอวีโอ (ระบุทางด้านบนของช่อง) ได้ผลเปรียบเทียบ (7-13-0) นั่นคือ โอวีเอให้ความแม่นยำชนะโอวีโอ 7 ชุดข้อมูล แพ้ 13 ชุดข้อมูล และไม่มีเสมอกัน

จากการทดสอบแสดงให้เห็นว่าวิธีโอวีโอชนะวิธีอื่นๆ อย่างมีนัยสำคัญ เมื่อเทียบกับในกลุ่มเทคนิคโครงสร้างต้นไม้ พบว่าวิธีต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป ดีที่สุดอย่างมีนัยสำคัญ นอกจากนี้วิธี ต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป ยังเทียบเคียงกับวิธีดั้งเดิม คือ โอวีเอ ดีดีเอจี เอดีเอจี และในส่วนของ พีทีเอส-จี ซี-พีทีเอส-จี ให้ความแม่นยำต่ำกว่าเทคนิคอื่นๆ อย่างมีนัยสำคัญ

ตาราง 13 ผลการเปรียบเทียบความแม่นยำในการจำแนกด้วย Wilcoxon Signed Rank Test โดยตัวเลขหนาแสดงว่าแตกต่างกันอย่างมีนัยสำคัญด้วยระดับนัยสำคัญ (significant level) 0.05 ตัวเลขในวงเล็บระบุผลการเปรียบเทียบ ชนะ-แพ้-เสมอ ระหว่างเทคนิคด้านซ้ายเทียบกับด้านบน

	OVO	DDAG	ADAG	BTS-G	c-BTS-G	IB-DTree	IBGE-DTree
OVA	0.1260 (7-13-0)	0.7642 (11-9-0)	1.0000 (10-10-0)	0.0160 (17-2-1)	0.0061 (17-2-1)	0.1443 (14-6-0)	0.0536 (15-4-1)
OVO		0.0002 (11-8-1)	0.0002 (17-2-1)	0.0001 (20-0-0)	0.0001 (20-0-0)	0.0001 (20-0-0)	0.0003 (18-2-0)
DDAG			0.0014 (2-16-2)	0.0188 (17-3-0)	0.0151 (17-3-0)	0.0574 (16-4-0)	0.1010 (15-5-0)
ADAG				0.0188 (17-3-0)	0.0124 (17-3-0)	0.0332 (17-3-0)	0.0536 (15-5-0)
BTS-G					0.2846 (12-7-1)	0.0264 (4-15-1)	0.0001 (2-18-0)
c-BTS-G						0.0466 (4-15-1)	0.0004 (0-2-18)
IB-DTree							0.4715 (8-11-1)

ในตารางที่ 14 แสดงจำนวนครั้งการตัดสินใจเฉลี่ยที่ใช้จำแนกประเภทข้อมูลขาออกของตัวอย่างทดสอบ โดยยิ่งตัวเลขเฉลี่ยมีค่าต่ำ แสดงว่ามีความรวดเร็วในการจำแนกสูง โดยผลการทดลองแสดงให้เห็นว่า ต้นไม้ตัดสินใจตามสารสนเทศและต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป มีความรวดเร็วในการจำแนกสูงสุดในบรรดาเทคนิคที่เปรียบเทียบ ส่วนวิธีโอวีโอ เป็นเทคนิคที่ทำงานช้าที่สุด

ตาราง 14 จำนวนครั้งการตัดสินใจเฉลี่ยที่ใช้จำแนกประเภทข้อมูล

โดยตัวเลขหนาแสดงว่าเป็นวิธีที่เร็วที่สุดในบรรดาเทคนิคเปรียบเทียบของชุดข้อมูลต่างๆ

ชื่อชุดข้อมูล	OVA	OVO	DDAG	ADAG	BTS-G	c-BTS-G	IB-DTree	IBGE-DTree
Page Block	5	10	4	4	3.628	3.801	3.790	3.831
Segment	7	21	6	6	3.630	3.882	2.858	3.009
Shuttle	7	21	6	6	4.703	5.370	4.998	5.019
Arrhyth	9	36	8	8	6.434	5.473	5.258	5.418
Cardiotocography	10	45	9	9	4.993	3.698	3.487	3.807
Mfeat-factor	10	45	9	9	4.224	3.643	3.473	3.754
Mfeat-fourier	10	45	9	9	4.512	3.796	3.522	3.786
Mfeat-karhunen	10	45	9	9	4.322	4.561	3.435	3.859
Optdigit	10	45	9	9	4.503	4.470	3.399	4.566
Pendigit	10	45	9	9	4.031	3.494	3.487	3.491
Primary Tumor	13	78	12	12	6.672	6.476	5.392	7.610
Libras Movement	15	105	14	14	5.493	5.114	4.325	4.411
Abalone	16	120	15	15	9.242	8.540	8.768	7.626
Krkopt	18	153	17	17	6.743	4.847	3.957	5.083
Spectrometer	21	210	20	20	6.728	6.080	4.411	4.613
Isolet	26	325	25	25	6.865	6.015	5.064	5.323
Letter	26	325	25	25	6.771	7.104	4.922	5.910
Plant Margin	100	4950	99	99	11.338	8.600	6.973	7.576
Plant Shape	100	4950	99	99	11.935	9.653	6.965	7.446
Plant Texture	100	4950	99	99	12.230	9.618	7.022	8.329

ผลการทดลองแสดงให้เห็นว่าวิธีต้นไม้ตัดสใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไปเป็นวิธีการที่มีประสิทธิภาพในการจำแนกสูงที่สุดในกลุ่มโครงสร้างต้นไม้ โดยใช้เวลาใกล้เคียงกับวิธีต้นไม้ตัดสใจตามสารสนเทศแต่ให้ผลความแม่นยำสูงกว่า นอกจากนี้วิธีต้นไม้ตัดสใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป ยังให้ผลการจำแนกไม่แตกต่างอย่างมีนัยสำคัญเมื่อเทียบกับวิธี โอวีเอ ดีดีเอจี เอดีเอจี แต่ใช้จำนวนครั้งในการจำแนกเฉลี่ยน้อยกว่ามาก และเมื่อเปรียบเทียบวิธีต้นไม้ตัดสใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป กับวิธี พีทีเอส-จี ซี-พีทีเอส-จี พบว่ามีประสิทธิภาพที่ดีกว่าทั้งในแง่ความแม่นยำและเวลาที่ใช้ในการประมวลผล สำหรับวิธีโอวีเอถึงแม้ว่าจะให้ความแม่นยำในการจำแนกสูงสุดแต่ก็ใช้เวลาในการประมวลผลมากเช่นกัน โดยเฉพาะอย่างยิ่งเมื่อประยุกต์ใช้กับปัญหาที่มีจำนวนประเภทข้อมูลหลายๆ เช่น Plant Margin, Plant Shape, Plant Texture เป็นต้น วิธีโอวีเอต้องใช้จำนวนครั้งในการตัดสใจ 4,950 ครั้ง ในขณะที่ต้นไม้ตัดสใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป ใช้จำนวนครั้งในการตัดสใจเฉลี่ยเพียง 7.45 ถึง 8.33 ครั้ง

ในส่วนของวิธีต้นไม้ตัดสใจตามสารสนเทศ เป็นวิธีที่มีประสิทธิภาพเชิงเวลาสูงที่สุดในกลุ่มทดลองเปรียบเทียบ แต่ให้ประสิทธิภาพเชิงความแม่นยำต่ำกว่าต้นไม้ตัดสใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป แต่ก็ยังคงสามารถเทียบเคียงได้กับวิธีโอวีเอ ดีดีเอจี และยิ่งดีกว่า พีทีเอส-จี ซี-พีทีเอส-จี อย่างมีนัยสำคัญอีกด้วย ดังนั้นถึงแม้ว่าโดยทั่วไปต้นไม้ตัดสใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป จะนำพิจารณานำไปใช้มากกว่าต้นไม้ตัดสใจตามสารสนเทศ ก็มีบางครั้งที่ควรเลือกใช้ต้นไม้ตัดสใจตามสารสนเทศ เช่น กรณีที่เวลาในการฝึกเป็นข้อจำกัดหนึ่งในการเลือกวิธีใช้แก้ปัญหา เนื่องจากต้นไม้ตัดสใจตามสารสนเทศใช้เวลาในการฝึกน้อยกว่าเพราะในการพิจารณาเลือกตัวจำแนกใช้ค่าเอนโทรปีเท่านั้น แต่ต้นไม้ตัดสใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป ต้องพิจารณาโดยใช้ทั้งค่าเอนโทรปีและค่าประมาณความผิดพลาดโดยนัยทั่วไป โดยเฉพาะในส่วนการหาค่าประมาณความผิดพลาดโดยนัยทั่วไปซึ่งใช้เวลานานกว่าการหาค่าเอนโทรปีเนื่องจากต้องทำการฝึกตัวจำแนกก่อนจึงจะหาค่าประมาณได้

บทที่ 5

บทสรุปงานวิจัยและข้อเสนอแนะ

5.1 สรุปผลงานวิจัย

เอสวีเอ็มแบบจำแนกหลายประเภท เป็นการเรียนรู้ของเครื่องสำหรับปัญหาที่มีประเภทข้อมูล มากกว่าสองประเภทขึ้นไป ซึ่งสามารถสร้างได้จากการผสมผสานเอสวีเอ็มแบบสองประเภทข้อมูลหลายตัว เข้าด้วยกันเพื่อสร้างเป็นเอสวีเอ็มแบบจำแนกหลายประเภทหนึ่งตัว ซึ่งมีผู้ที่เสนอเอาไว้หลายเทคนิค โดยวิธีโอวีโอเป็นวิธีดั้งเดิมที่ให้ความแม่นยำในการจำแนกสูงแต่ใช้เวลาในการจำแนกสูง คือ $O(N^2)$ โดยที่ N คือจำนวนประเภทข้อมูล ต่อมาผู้เสนอวิธีดีดีเอจีและวิธีเอดีเอจี ซึ่งสร้างอยู่บนพื้นฐานของวิธีโอวีโอ โดยให้ผลความแม่นยำในการจำแนกค่อนข้างดีและใช้เวลาในการจำแนกเป็น $O(N)$ และในเวลาถัดมาก็มีผู้เสนอวิธีบีทีเอสและวิธีแตกครึ่งตามสารสนเทศ ซึ่งสร้างบนพื้นฐานของวิธีโอวีโอ โดยเรียงเป็นโครงสร้างต้นไม้ตัดสินใจ ซึ่งใช้เวลาในการจำแนกระหว่าง $O(\log_2 N)$ ถึง $O(N)$ และให้ความแม่นยำในการจำแนกอยู่ในระดับที่ยอมรับได้

ในงานวิจัยนี้ เรานำเสนอวิธีต้นไม้ตัดสินใจตามสารสนเทศ และวิธีต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป ซึ่งนำเอาค่าเอนโทรปีและค่าประมาณความผิดพลาดโดยนัยทั่วไปมาผสมกัน เพื่อใช้ในการเลือกตัวจำแนกที่เหมาะสมแล้วสร้างเป็นโครงสร้างต้นไม้ โดยค่าเอนโทรปีเป็นการวัดความยุ่งเหยิง ของระบบซึ่งสามารถใช้ได้ดีกับต้นไม้ตัดสินใจ ค่าเอนโทรปีที่ต่ำสะท้อนให้เห็นถึงการเลือกตัวจำแนกที่เหมาะสม เนื่องจากจะส่งผลให้ต้นไม้ที่ได้จะจำแนกประเภทที่มีความถี่สูงออกก่อน ส่วนการประมาณค่าความผิดพลาดโดยนัยทั่วไปเป็นการวัดประสิทธิภาพของตัวจำแนก โดยตัวจำแนกที่มีค่าประมาณความผิดพลาดโดยนัยทั่วไปต่ำจะสามารถจำแนกได้ความแม่นยำสูง

ในต้นไม้ตัดสินใจตามสารสนเทศ และวิธีต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป ใช้กลวิธี “การจับกลุ่มประเภทข้อมูลตามข้อมูลส่วนใหญ่” ซึ่งเป็นกลวิธีการสร้างต้นไม้บนพื้นฐานของตัวจำแนกแบบโอวีโอ ซึ่งแบ่งกลุ่มประเภทข้อมูลบวก จะไม่ทำให้มีปมใบของประเภทข้อมูลซ้่ากระจายไปในต้นไม้ จึงทำให้ทั้งต้นไม้ตัดสินใจตามสารสนเทศ และต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป ทำการจำแนกประเภทของข้อมูลทดสอบได้ในเวลาไม่เกิน $O(N)$ โดยข้อดีของทั้งสองเทคนิคที่นำเสนอในงานวิจัยนี้มีหลายประการ เช่น 1. ประเภทข้อมูลที่ค่าตอบซึ่งมีความน่าจะเป็นสูงจะสามารถพบได้อย่างรวดเร็ว 2. การไม่มีปมใบประเภทข้อมูล

ซ้ำทำให้ขนาดของต้นไม้ไม่มากเกินไปจนความจำเป็น 3. ไม่มีการตัดเล็มข้อมูล จึงทำให้ไม่มีข้อมูลสูญหายระหว่างขั้นตอนการฝึกและสร้างต้นไม้จำแนก

เราได้ทำการทดลองเพื่อทดสอบประสิทธิภาพของวิธีที่นำเสนอ และเปรียบเทียบกับวิธีการดั้งเดิมโดยใช้ชุดข้อมูลจำนวน 20 ชุด จากพื้นที่เก็บข้อมูลการเรียนรู้ของเครื่องยูซีไอ (UCI Machine Learning Repository) ซึ่งสรุปได้ว่าทั้งวิธีต้นไม้ตัดสินใจตามสารสนเทศ และวิธีต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป เป็นวิธีที่มีประสิทธิภาพในการจำแนกเชิงเวลาสูง คือ ใช้จำนวนครั้งเฉลี่ยในการจำแนกต่ำและให้ความแม่นยำในการจำแนกเทียบเคียงได้กับวิธีการดั้งเดิม คือ โอวีเอ ดีดีเอจี ดีดีเอจี และยังให้ความแม่นยำสูงกว่าวิธีโครงสร้างต้นไม้ที่เปรียบเทียบกับในการทดลองด้วย โดยวิธีต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป ให้ความแม่นยำในการจำแนกสูงกว่าวิธีต้นไม้ตัดสินใจตามสารสนเทศเล็กน้อย แต่ต้นไม้ตัดสินใจตามสารสนเทศ ใช้เวลาในการฝึกน้อยกว่า ดังนั้นในกรณีทั่วไปจึงเหมาะสมในการเลือกใช้ต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป แต่ในกรณีที่เวลากการฝึกเป็นข้อจำกัด อาจเลือกใช้ต้นไม้ตัดสินใจตามสารสนเทศ

5.2 ข้อเสนอแนะ

1. ในการทดลองนี้ใช้พารามิเตอร์ที่หาได้จากพับที่ 1 ไปใช้กับพับที่ 2-10 ทั้งนี้เนื่องจากใช้สมมติฐานที่ว่า การกระจายข้อมูลทั้ง 10 พับเป็นไปอย่างสม่ำเสมอ ซึ่งอาจสามารถเพิ่มความแม่นยำในการทำการทดลองได้ โดยการทำการทดลองให้พับที่ 2-10 ใช้พารามิเตอร์แกมมาและ c ของแต่ละพับแยกกัน
2. ในการนำไปใช้งานจริง สามารถเพิ่มประสิทธิภาพในการจำแนกของต้นไม้ตัดสินใจตามสารสนเทศ และต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไปได้อีก โดยการปรับจูนพารามิเตอร์แกมมา และ c ให้แต่ละปม ใช้ค่าที่ปรับเหมาะที่สุดสำหรับปม
3. ในการทดลองนี้ วิธีต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป ได้ทำการตั้งค่าจำนวนตัวจำแนก k ตัวที่มีค่าเอนโทรปีเรียงจากต่ำไปสูง ซึ่งจะถูกนำมาฝึกและพิจารณาเลือกตัว ที่ให้ค่าความผิดพลาดโดยนัยทั่วไปต่ำสุดที่ร้อยละ 20 ของจำนวนตัวจำแนกที่เป็นไปได้ทั้งหมดในปม ซึ่งในการนำไปใช้จริง อาจปรับเพิ่มหรือลดจำนวนตัวจำแนกที่จะถูกนำไปใช้พิจารณาได้ตามความเหมาะสม

ภาคผนวก ก.

พารามิเตอร์ Gamma และ c ที่ใช้ในการทดลอง

1. สรุปพารามิเตอร์ Gamma และ C ที่ใช้ในการทดลอง

ชุดข้อมูล	จำนวนประเภท	โอวีโอ		โอวีเอ	
		C =	γ =	C =	γ =
Page Block	5	1000	0.1	1000	1
Segment	7	10	1	10	1
Shuttle	7	1000	10	1000	10
Arrhyth	9	10	0.01	10	0.01
Cardiotocography	10	100	0.1	10	1
Mfeat-factor	10	100	0.01	10	0.01
Mfeat-fourier	10	10	0.1	10	0.1
Mfeat-karhunen	10	10	0.1	10	0.1
Optdigit	10	10	0.1	10	0.1
Pendigit	10	100	0.1	10	1
Primary Tumor	13	10	0.01	1	0.1
Libras Movement	15	10	0.1	100	0.1
Abalone	16	100	0.1	100	0.1
Krkopt	18	100	1	10	10
Spectrometer	21	1000	0.001	100	0.01
Isolet	26	10	0.1	100	1
Letter	26	10	1	10	1
Plant Margin	100	100	0.1	10	0.1
Plant Shape	100	100	0.1	10	1
Plant Texture	100	10	0.1	100	0.1

2. รายละเอียดการหาพารามิเตอร์ γ และ c ที่ใช้ในการทดลองแบบโอวีโอ

Page Block

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	843	662	527	382
$\gamma = 0.01$	663	516	379	257
$\gamma = 0.1$	461	348	240	199
$\gamma = 1$	343	246	212	218
$\gamma = 10$	350	297	300	333

Segment

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	1011	426	229	161
$\gamma = 0.01$	418	227	145	95
$\gamma = 0.1$	204	125	85	88
$\gamma = 1$	126	74	74	83
$\gamma = 10$	272	256	261	261

Shuttle

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	8791	4586	2215	1748
$\gamma = 0.01$	4580	2201	1646	681
$\gamma = 0.1$	1836	771	283	146
$\gamma = 1$	371	172	120	96
$\gamma = 10$	150	105	82	56

Arrhyth

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	559	340	272	293
$\gamma = 0.01$	353	267	280	283
$\gamma = 0.1$	413	388	388	388
$\gamma = 1$	563	561	561	561
$\gamma = 10$	559	559	559	559

Cardiotocography

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	4286	1819	975	661
$\gamma = 0.01$	1850	962	621	505
$\gamma = 0.1$	911	536	461	511
$\gamma = 1$	641	565	584	586
$\gamma = 10$	3249	2909	2909	2909

Mfeat-factor

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	189	118	97	94
$\gamma = 0.01$	114	83	81	81
$\gamma = 0.1$	121	118	118	118
$\gamma = 1$	3944	3699	3699	3699
$\gamma = 10$	8012	7997	7997	7997

Mfeat-fourier

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	971	569	452	518
$\gamma = 0.01$	577	441	478	492
$\gamma = 0.1$	442	422	424	424
$\gamma = 1$	2882	2637	2637	2637
$\gamma = 10$	7971	7973	7973	7973

Mfeat-karhunen

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	405	201	156	195
$\gamma = 0.01$	196	148	160	160
$\gamma = 0.1$	121	100	100	100
$\gamma = 1$	886	831	831	831
$\gamma = 10$	7979	7970	7970	7970

Optdigit

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	341	163	137	154
$\gamma = 0.01$	125	78	91	91
$\gamma = 0.1$	67	64	64	64
$\gamma = 1$	5928	5541	5541	5541
$\gamma = 10$	22597	22595	22595	22595

Pendigit

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	2612	731	302	189
$\gamma = 0.01$	570	188	124	141
$\gamma = 0.1$	136	101	101	118
$\gamma = 1$	148	144	146	146
$\gamma = 10$	5143	4773	4773	4773

Primary Tumor

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	1018	651	515	597
$\gamma = 0.01$	684	511	603	618
$\gamma = 0.1$	518	581	585	586
$\gamma = 1$	766	766	766	765
$\gamma = 10$	782	785	785	781

Libras Movement

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	1375	357	204	169
$\gamma = 0.01$	328	146	129	126
$\gamma = 0.1$	124	113	113	113
$\gamma = 1$	522	486	486	486
$\gamma = 10$	1692	1658	1658	1658

Abalone

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	12809	9823	8946	8374
$\gamma = 0.01$	9839	8969	8340	8051
$\gamma = 0.1$	8900	8336	7903	7932
$\gamma = 1$	8432	7907	8240	8974
$\gamma = 10$	8231	8704	10004	10724

Krkopt

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	78995	62498	57133	45652
$\gamma = 0.01$	60674	46434	32490	27989
$\gamma = 0.1$	33253	26297	20570	17411
$\gamma = 1$	18161	11019	8858	9293
$\gamma = 10$	13889	10186	10185	10185

Spectrometer

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	2577	1014	346	285
$\gamma = 0.01$	1061	362	297	293
$\gamma = 0.1$	519	403	402	402
$\gamma = 1$	1855	1732	1732	1732
$\gamma = 10$	2847	2845	2845	2845

Isolet

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	11569	1262	749	716
$\gamma = 0.01$	1263	746	684	761
$\gamma = 0.1$	720	588	637	645
$\gamma = 1$	869	818	818	818
$\gamma = 10$	48206	45408	45408	45408

Letter

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	78088	22533	10908	7796
$\gamma = 0.01$	22493	10582	6868	4633
$\gamma = 0.1$	8964	3819	1960	1830
$\gamma = 1$	1524	862	933	935
$\gamma = 10$	2696	2615	2615	2615

Plant Margin

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	38085	6796	714	636
$\gamma = 0.01$	7264	714	629	629
$\gamma = 0.1$	765	599	590	590
$\gamma = 1$	1217	1176	1176	1176
$\gamma = 10$	61953	61504	61504	61504

Plant Shape

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	39440	21664	5079	2345
$\gamma = 0.01$	21906	5053	2288	1949
$\gamma = 0.1$	5025	2124	1781	1857
$\gamma = 1$	2392	1901	1961	1962
$\gamma = 10$	8546	8181	8182	8182

Plant Texture

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	39392	16242	1616	1126
$\gamma = 0.01$	17422	1620	1113	1124
$\gamma = 0.1$	1622	1041	1042	1042
$\gamma = 1$	2780	2699	2699	2699
$\gamma = 10$	55939	53712	53712	53712

3. รายละเอียดการหาพารามิเตอร์ γ และ c ที่ใช้ในการทดลองแบบโอวีเอ

Page Block

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	1008	936	717	618
$\gamma = 0.01$	935	722	605	468
$\gamma = 0.1$	717	554	407	372
$\gamma = 1$	521	379	356	348
$\gamma = 10$	448	371	379	422

Segment

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	1490	1015	578	369
$\gamma = 0.01$	1017	408	333	251
$\gamma = 0.1$	352	272	189	157
$\gamma = 1$	240	127	128	150
$\gamma = 10$	314	275	283	283

Shuttle

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	15676	12309	9977	16389
$\gamma = 0.01$	12308	9221	4861	1312
$\gamma = 0.1$	5199	1394	659	444
$\gamma = 1$	729	326	279	200
$\gamma = 10$	294	199	142	107

Arrhyth

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	348	254	222	278
$\gamma = 0.01$	265	220	283	289
$\gamma = 0.1$	262	274	274	274
$\gamma = 1$	348	346	346	346
$\gamma = 10$	348	348	348	348

Cardiotocography

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	1913	1870	1519	1272
$\gamma = 0.01$	1871	1468	1122	859
$\gamma = 0.1$	1335	847	696	757
$\gamma = 1$	777	660	738	742
$\gamma = 10$	1541	1367	1367	1367

Mfeat-factor

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	544	158	103	111
$\gamma = 0.01$	148	81	77	77
$\gamma = 0.1$	152	123	123	123
$\gamma = 1$	1792	1780	1780	1780
$\gamma = 10$	1792	1792	1792	1792

Mfeat-fourier

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	1680	859	678	742
$\gamma = 0.01$	865	643	671	689
$\gamma = 0.1$	614	594	600	600
$\gamma = 1$	1625	1601	1601	1601
$\gamma = 10$	1792	1792	1792	1792

Mfeat-karhunen

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	1800	412	216	232
$\gamma = 0.01$	417	197	189	200
$\gamma = 0.1$	145	101	101	101
$\gamma = 1$	1621	1526	1526	1526
$\gamma = 10$	1792	1790	1790	1790

Optdigit

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	1130	549	381	250
$\gamma = 0.01$	369	177	143	151
$\gamma = 0.1$	118	106	106	106
$\gamma = 1$	4873	4781	4781	4781
$\gamma = 10$	5058	5058	5058	5058

Pendigit

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	6313	2158	1389	647
$\gamma = 0.01$	1755	589	254	164
$\gamma = 0.1$	224	132	138	167
$\gamma = 1$	144	129	136	137
$\gamma = 10$	4862	4375	4375	4375

Primary Tumor

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	283	272	233	244
$\gamma = 0.01$	277	234	260	375
$\gamma = 0.1$	233	331	341	341
$\gamma = 1$	287	305	305	305
$\gamma = 10$	300	300	300	300

Libras Movement

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	324	290	212	156
$\gamma = 0.01$	290	181	115	119
$\gamma = 0.1$	137	81	80	81
$\gamma = 1$	231	197	197	197
$\gamma = 10$	276	276	276	276

Abalone

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	3688	3688	3688	3686
$\gamma = 0.01$	3688	3688	3686	3682
$\gamma = 0.1$	3688	3686	3682	3686
$\gamma = 1$	3685	3685	3722	4119
$\gamma = 10$	3683	3850	4561	5640

Krkopt

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	25250	25250	25250	25250
$\gamma = 0.01$	25250	25250	25250	22786
$\gamma = 0.1$	25250	22334	21315	52098
$\gamma = 1$	18980	14971	11929	13452
$\gamma = 10$	11870	8703	8704	8704

Spectrometer

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	427	409	370	367
$\gamma = 0.01$	409	368	350	458
$\gamma = 0.1$	366	365	397	397
$\gamma = 1$	409	392	392	392
$\gamma = 10$	427	427	427	427

Isolet

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	7017	4710	2115	1790
$\gamma = 0.01$	4743	2056	1630	1302
$\gamma = 0.1$	1877	1187	1118	1177
$\gamma = 1$	1292	1111	1111	1111
$\gamma = 10$	7017	7017	7017	7017

Letter

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	18046	15556	12876	11758
$\gamma = 0.01$	15512	12368	9357	5735
$\gamma = 0.1$	10026	5705	2813	1879
$\gamma = 1$	2792	1318	1424	1489
$\gamma = 10$	4201	3501	3501	3501

Plant Margin

	C=1	C=10	C=100	C=1000
$\gamma = 0.001$	302	273	115	46
$\gamma = 0.01$	276	109	36	38
$\gamma = 0.1$	101	21	21	21
$\gamma = 1$	192	173	173	173
$\gamma = 10$	302	302	302	302

Plant Shape

	C=1	C=10	C=100	C=1000
$\Psi = 0.001$	302	302	274	180
$\Psi = 0.01$	302	274	173	104
$\Psi = 0.1$	274	136	70	69
$\Psi = 1$	109	48	51	53
$\Psi = 10$	130	114	114	114

Plant Texture

	C=1	C=10	C=100	C=1000
$\Psi = 0.001$	301	300	207	97
$\Psi = 0.01$	301	204	81	76
$\Psi = 0.1$	197	56	51	51
$\Psi = 1$	117	88	88	88
$\Psi = 10$	301	301	301	301

ภาคผนวก ข.

ผลการทดลองหาความสัมพันธ์ของร้อยละจำนวนตัวจำแนก
ในวิธีการสร้างต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป

เป็นการทดลองเบื้องต้นเพื่อดูผลกระทบต่อร้อยละความแม่นยำการจำแนก และจำนวนครั้งเฉลี่ยที่ใช้ในการตัดสินใจ เมื่อทำปรับร้อยละของจำนวนตัวจำแนกที่นำไปใช้เป็นตัวเลือกพิจารณาหา ค่าประมาณความผิดพลาดเชิงนัย (ค่า n ในบรรทัดที่ 13 ของขั้นตอนวิธีที่ 3) ของวิธีการสร้างต้นไม้ตัดสินใจตามสารสนเทศและค่าประมาณความผิดพลาดโดยนัยทั่วไป หรือ ไอบีจีอี-ดีทรี (IBGE-DTree) โดยเลือกทำการทดลองกับบางชุดข้อมูล ได้แก่ Abalone Arrhyth, Cardiotocography, Isolet, Libras Movement, Mfeat_factor, Mfeat_fourier, Optdigit, Pendigit, Primary Tumor, Spectrometer

Abalone

ร้อยละ	ความแม่นยำ	จำนวนครั้งการตัดสินใจเฉลี่ย
0	25.2806	8.7679
10	26.4275	7.3260
20	26.7448	7.6257
30	26.2323	8.2418
40	26.8180	8.9109
50	26.5983	9.6254
60	26.3787	10.1362
70	25.7443	11.2638
80	25.9395	12.4048
90	25.9395	13.3992
100	25.8175	14.3819

Arrhyth

ร้อยละ	ความแม่นยำ	จำนวนครั้งการตัดสินใจเฉลี่ย
0	71.0046	5.2580
10	72.1461	5.2397
20	72.1461	5.4178
30	72.1461	6.2945
40	72.6027	6.3539
50	72.8311	6.5365
60	72.6027	6.7945
70	72.6027	7.4178
80	72.3744	7.4269
90	72.3744	7.4269
100	72.3744	7.4269

Cardiotocography

ร้อยละ	ความแม่นยำ	จำนวนครั้งการตัดสินใจเฉลี่ย
0	83.8194	3.4873
10	83.5842	3.7098
20	83.1609	3.8071
30	83.5372	4.4539
40	83.2549	4.8993
50	83.5842	4.8203
60	83.7723	5.1040
70	82.9727	5.7380
80	82.9727	6.6877
90	83.3020	6.8344
100	83.7253	7.0677

Isolet

ร้อยละ	ความแม่นยำ	จำนวนครั้งการตัดสินใจเฉลี่ย
0	93.6386	5.0640
10	94.0875	5.1007
20	94.0105	5.3230
30	94.2798	5.5193
40	94.0106	5.7033
50	94.0619	5.9085
60	94.0236	6.4406
70	94.1133	6.8121
80	94.1005	7.1254
90	94.1773	7.6261
100	94.1774	8.1852

Libras Movement

ร้อยละ	ความแม่นยำ	จำนวนครั้งการตัดสินใจเฉลี่ย
0	88.0556	4.3250
10	87.7778	4.4167
20	88.0556	4.4111
30	88.0556	4.4111
40	88.0556	4.5806
50	88.6111	4.6972
60	87.7778	4.7028
70	88.0556	4.7972
80	88.0556	4.9806
90	88.6111	5.4944
100	87.7778	6.0889

Mfeat_factor

ร้อยละ	ความแม่นยำ	จำนวนครั้งการตัดสินใจเฉลี่ย
0	98.0000	3.4725
10	98.0000	3.5010
20	98.2000	3.7535
30	98.1500	3.8230
40	98.2000	3.9260
50	98.2000	3.9875
60	98.1500	4.0085
70	98.1500	4.0085
80	98.1500	4.1160
90	98.1500	4.1160
100	98.1000	4.1365

Mfeat_fourier

ร้อยละ	ความแม่นยำ	จำนวนครั้งการตัดสินใจเฉลี่ย
0	85.2000	3.5220
10	85.5500	3.7465
20	85.1500	3.7855
30	85.2000	3.7780
40	85.4000	3.9985
50	85.2500	4.3375
60	85.3000	4.3685
70	85.2500	4.9220
80	84.9000	5.2265
90	84.9500	5.2270
100	84.9500	5.2270

Optdigit

ร้อยละ	ความแม่นยำ	จำนวนครั้งการตัดสินใจเฉลี่ย
0	99.1637	3.3988
10	99.1281	4.5656
20	99.0925	4.5656
30	99.0747	4.7690
40	99.0214	5.2998
50	99.0214	5.3836
60	99.0036	5.4361
70	99.0036	5.4359
80	99.0036	5.4359
90	99.0036	5.4368
100	98.9858	5.4343

Pendigit

ร้อยละ	ความแม่นยำ	จำนวนครั้งการตัดสินใจเฉลี่ย
0	99.4451	3.4866
10	99.4451	3.4972
20	99.4541	3.4914
30	99.4632	3.5370
40	99.4541	3.5585
50	99.4632	3.5725
60	99.4723	3.5731
70	99.4360	3.7102
80	99.4360	3.7434
90	99.4360	3.7418
100	99.4269	4.5656

Primary Tumor

ร้อยละ	ความแม่นยำ	จำนวนครั้งการตัดสินใจเฉลี่ย
0	47.9365	5.3905
10	46.9841	6.4794
20	44.7619	7.6095
30	45.0794	8.4508
40	44.7619	9.5778
50	46.3492	10.2921
60	45.7143	10.4286
70	45.3968	10.8794
80	45.0794	11.0286
90	44.7619	11.1587
100	45.0794	11.2952

Spectrometer

ร้อยละ	ความแม่นยำ	จำนวนครั้งการตัดสินใจเฉลี่ย
0	68.2105	4.4105
10	67.7895	4.3874
20	67.7895	4.6126
30	66.9474	4.7242
40	66.5263	4.8842
50	67.3684	5.0421
60	67.7895	5.2084
70	68.0000	5.4674
80	68.6316	6.0947
90	68.6316	6.7768
100	67.5789	7.6568

รายการอ้างอิง



1. Vapnik, V. N., Statistical Learning Theory. *Adaptive and learning systems for signal processing, communications, and control* 1998, 768.
2. Vapnik, V. N., An Overview of Statistical Learning Theory. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council* 1999, 10, 988-99.
3. Bredensteiner, E. J.; Bennett, K. P., Multicategory Classification by Support Vector Machines. *Computational Optimization* 1999, 12, 53-79.
4. Crammer, K.; Singer, Y., On the Learnability and Design of Output Codes for Multiclass Problems. *Machine Learning* 2002, 47, 201-233.
5. Hsu, C.; Lin, C., A Comparison of Methods for Multiclass Support Vector Machines. *Neural Networks, IEEE Transactions on* 2002, 13, 415-425.
6. Knerr, S.; Personnaz, L.; Dreyfus, G., Single-layer learning revisited: a stepwise procedure for building and training a neural network. *Neurocomputing* 1990, 41-50.
7. Hastie, T.; Tibshirani, R., Classification by Pairwise Coupling. *Annals of Statistics* 1998, 26, 451-471.
8. Friedman, J., Another Approach to Polychotomous classification. In *Technical Report*, Department of Statistic, Stanford University: 1996.
9. Rifkin, R.; Klautau, A.; Org, K. i., In Defense of One-Vs-All Classification. *Journal of Machine Learning Research* 2004, 5, 101-141.
10. Platt, J.; Cristianini, N.; Shawe-Taylor, J., Large Margin DAGs for Multiclass Classification. *Advances in Neural Information Processing Systems* 2000, 547-553.
11. Takahashi, F.; Abe, S., Optimizing Directed Acyclic Graph Support Vector Machines. *Proceedings of Artificial Neural Networks in Pattern Recognition (ANNPR 2003)* 2003, 166-70.
12. Kijirikul, B.; Ussivakulz, N.; Road, P., Multiclass Support Vector Machines Using Adaptive Directed Acyclic Graph. *International Joint Conference on Neural Networks* 2002, 2, 980-985.
13. Phetkaew, T.; Kijirikul, B.; Rivepiboon, W., Reordering Adaptive Directed Acyclic Graphs: An Improved Algorithm for Multiclass Support Vector Machines.

- Proceedings of the International Joint Conference on Neural Networks 2003*, 2, 1605-1610.
14. Cook, W.; Rohe, A., Computing Minimum-Weight Perfect Matching. *Technical Report 97863* 1997.
 15. Deitterich, T. G.; Bakiri, G., Solving Multiclass Learning problems via Error-Correcting Output Codes. *Journal of Artificial Intelligence Research* 1995, 263-283.
 16. Kong, E. B.; Dietterich, T. G. In *Error-Correcting Output Coding Corrects Bias and Variance*, Proceedings of the Twelfth International Conference on Machine Learning, 1995.
 17. James, G.; Hastie, T., The Error Coding Method and PICTs. *Journal of Computational and Graphical Statistics* 1998, 7 (3), 377-387.
 18. Ghani, R. In *Using Error-Correcting Codes For Text Classification*, Proceedings of the Seventeenth International Conference on Machine Learning, 2000.
 19. Allwein, E. L.; Schapire, R. E.; Singer, Y., Reducing multiclass to binary: a unifying approach for margin classifiers. *The Journal of Machine Learning Research* 2001, 1, 113-141.
 20. Masulli, F.; Valentini, G., An experimental analysis of the dependence among codeword bit errors in ECOC learning machines. *Neurocomputing* 2003.
 21. Utschick, W.; Weichselberger, W., Stochastic Organization of Output Codes in Multiclass Learning Problems. *Neural Computation* 2001, 13 (5).
 22. Kuncheva, L. I., Using diversity measures for generating error-correcting output codes in classifier ensembles. *Pattern Recognition Letters* 2005, 26 (1).
 23. Escalera, S.; Pujol, O. In *ECOC-ONE: A Novel Coding and Decoding Strategy*, 18th International Conference on Pattern Recognition, Hong Kong, China, Hong Kong, China, 2006.
 24. Escalera, S.; Pujol, O.; Radeva, P. In *Decoding of Ternary Error Correcting Output Codes*, Iberoamerican Congress on Pattern Recognition, Mexico, Mexico, 2006.
 25. Pujol, O.; Radeva, P.; Vitria, J., Discriminant ECOC: a heuristic method for application dependent design of error correcting output codes. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE* 2006.

26. Escalera, S.; Pujol, O.; Radeva, P., Separability of ternary codes for sparse designs of error-correcting output codes. *Pattern Recognition Letters* 2009.
27. Escalera, S.; Pujol, O.; Radeva, P., On the decoding process in ternary error-correcting output codes. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 2010.
28. Bautista, M. Á.; Escalera, S.; Baró, X.; Radeva, P.; Vitria, J.; Pujol, O., Minimal design of error-correcting output codes. *Pattern Recognition Letters* 2012, 33 (6).
29. Bautista, M. Á.; Escalera, S.; Baró, X.; Pujol, O., On the design of ECOC-compliant genetic algorithm. *Pattern Recognition* 2014, 42 (2).
30. Ali Bagheri, M.; Montazer, G. A.; Kabir, E., A subspace approach to error correcting output codes. *Pattern Recognition Letters* 2013, 34 (2).
31. Zhou, J.; Peng, H.; Suen, C. Y., Data-driven decomposition for multi-class classification. *Pattern Recognition* 2008, 41.
32. Zhang, J.; Zhao, X.; Du, L., Solving multi-class problems by data-driven topology-preserving output codes. *Neurocomputing* 2013, 121.
33. Kijirikul, B.; Boonsirisumpun, N.; Limpiyakorn, Y., Multiclass Support Vector Machines Using Balanced Dicotomization. In *The 8th Pacific Rim International Conference on Artificial Intelligence (PRICAI-2004)*, 2004.
34. Fei, B.; Liu, J., Binary Tree of SVM: A New Fast Multiclass Training and Classification Algorithm. *IEEE Transactions on Neural Networks* 2006, 17, 696-704.
35. Songsiri, P.; Kijirikul, B.; Phetkaew, T., Information-Based Dicotomizer: A Method for Multiclass Support Vector Machines. In *IJCNN*, 2008; pp 3284-3291.
36. Chen, J.; Wang, C.; Wang, R., Adaptive Binary Tree for Fast SVM Multiclass Classification. *Neurocomputing* 2009, 72, 3370-3375.
37. Bala, M.; Agrawal, R. K., Optimal Decision Tree Based Multi-class Support Vector Machine. *Informatica* 2011, 35, 197-209.
38. Takahashi, F.; Abe, S., Decision-tree-based Multiclass Support Vector Machines. In *Neural Information Processing IEEE, 2002. ICONIP'02. Proceedings of the 9th International Conference on 2002*, 3, 1418-1488.

39. Cheong, S.; Oh, S. H.; Lee, S.-Y., Support Vector Machines with Binary Tree Architecture for Multi-Class Classification. *Neural Information Processing 2004*, 2 (3).
40. Lei, H.; Govindaraju, V. In *Half-Against-Half multi-class support vector machines*, MCS'05 Proceedings of the 6th international conference on Multiple Classifier, Seaside, CA, Seaside, CA, 2005.
41. Liu, B.; Cao, L.; Yu, P. S.; Zhang, C. In *Multi-Space-Mapped SVMs for Multi-Class Classification*, IEEE International Conference on Data Mining, 2008.
42. Madzarov, G.; Gjorgjevikj, D.; Chorbev, I., A Multi-class SVM Classifier Utilizing Binary Decision Tree Support vector machines for pattern recognition. *Electrical Engineering 2009*, 33, 233-241.
43. Dong, C.; Zhou, B.; Hu, J. In *A hierarchical SVM based multiclass classification by using similarity clustering*, International Joint Conference on Neural Networks 2015.
44. Lee, Y.; Lee, J.; Quinlan, J. R.; Bala, M.; Agrawal, R. K., Binary tree optimization using genetic algorithm for multiclass support vector machine. *Expert Systems with Applications 2015*, 60, 81-106.
45. Yang, X.; Yu, Q.; He, L.; Guo, T., The One-against-all Partition Based Binary Tree Support Vector Machine Algorithms for Multi-class Classification. *Neurocomputing 2013*, 113, 1-7.
46. Kumar, M. A.; Gopal, M., Reduced one-against-all method for multiclass SVM classification. *Expert Systems with Applications 2011*, 38.
47. García-Pedrajas, N.; Ortiz-Boyer, D. In *Improving multiclass pattern recognition by the combination of two strategies*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006.
48. Soltani, A.; Ahadi, S. M.; Faraji, N.; Sharifian, S., Designing efficient discriminant functions for multi-category classification using evolutionary methods. *neurocomputing 2016*, 173 (P3).
49. Blake, C. L.; Merz, C. J., UCI Repository of Machine Learning Databases. *University of California 1998*, <http://archive.ics.uci.edu/ml/>.
50. Demšar, J., Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research 2006*, 7, 1-30.

51. Bartlett, P.; Shawe-Taylor, J., Generalization Performance of Support Vector Machines and Other Pattern Classifiers. *Advances in kernel methods* 1999.
52. S. Escalera; Pujol, O.; Radeva, P., Separability of ternary codes for sparse designs of error-correcting output codes. *Pattern Recognition Letters* 2009, 285–297.
53. Cook, W.; Rohe, A., Computing Minimum-Weight Perfect Matchings. *Infoms* 1999, 11 (2).
54. Songsiri, P.; Phetkaew, T.; Kijirikul, B., Enhancement of Multi-class Support Vector Machine Construction from Binary Learners Using Generalization Performance. *Neurocomputing* 2015, 151, 434-448.
55. Joachims, T., SVM light. 2008.



ประวัติผู้เขียนวิทยานิพนธ์

นายพิตติพล คັນธวัฒน์ จบการศึกษาทางด้านวิศวกรรมคอมพิวเตอร์จากคณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ระดับปริญญาบัณฑิตในปี พ.ศ. 2547 และระดับมหาบัณฑิตในปี พ.ศ. 2551 ปัจจุบันกำลังศึกษาระดับดุษฎีบัณฑิตทางด้านวิศวกรรมคอมพิวเตอร์จากคณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย มีความสนใจในงานวิจัยทางด้านปัญญาประดิษฐ์ การเรียนรู้ของเครื่อง การประยุกต์ใช้ปัญญาประดิษฐ์กับข้อมูลทางการเงิน และการลงทุน



