การแบ่งกลุ่มข้อมูลด้วยวิธีเอสวีเอ็ม

SUPPORT VECTOR MACHINE FOR CLASSIFICATION

นายภาณุพงศ์ เสนาธรรม รหัส 48380358

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวร

ปีการศึกษา 2552

ใบรับรองปริญญานิพนธ์

ชื่อหัวข้อโครงงาน การแบ่งกลุ่มข้อมูลด้วยวิธีเอสวีเอ็ม

ผู้ดำเนินโครงงาน นายภาณุพงศ์ เสนาธรรม รหัส 48380358

ที่ปรึกษาโครงงาน ดร. ศุภวรรณ พลพิทักษ์ชัย

สาขาวิชา วิศวกรรมไฟฟ้า

ภาควิชา วิศวกรรมไฟฟ้าและคอมพิวเตอร์

ปีการศึกษา 2552

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวร อนุมัติให้ปริญญานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า

.... ที่ปรึกษาโครงงาน
(ดร. ศุภวรรณ พลพิทักษ์ชัย)

.... กรรมการ
(ดร. มุฑิตา สงฆ์จันทร์)

.... กรรมการ
(ดร. นิพัทธ์ จันทรมินทร์)

ชื่อหัวข้อโครงงาน การแบ่งกลุ่มข้อมูลด้วยวิธีเอสวีเอ็ม

ผู้ดำเนินโครงงาน นายภาณุพงศ์ เสนาธรรม รหัส 48380358

ที่ปรึกษาโครงงาน ดร. ศุภวรรณ พลพิทักษ์ชัย

สาขาวิชา วิศวกรรมไฟฟ้า

ภาควิชา วิศวกรรมไฟฟ้าและคอมพิวเตอร์

ปีการศึกษา 2552

....

บทคัดย่อ

เอสวีเอ็มคือวิธีที่ใช้สำหรับหาความสัมพันธ์ระหว่างข้อมูลขาเข้าและขาออก ซึ่งในโครงงานนี้จะใช้เอสวีเอ็มเป็นเครื่องมือสำหรับการแบ่งกลุ่มข้อมูล โดยทั่วไปแล้วการแบ่งกลุ่มจะอาศัยหลักการหาไฮเปอร์เพลนซึ่งเป็นระนาบที่อยู่ตรงกลางระหว่างข้อมูลทั้งสองกลุ่ม วิธีเอสวีเอ็มจะเป็นวิธีการหาไฮเปอร์เพลนที่ทำงานโดยใช้ฟังก์ชันเคอร์เนล และสร้างโดยใช้หลักการหาขอบเขตที่กว้างที่สุดร่วมกับการแก้ปัญหาแบบมีข้อแม้ ข้อดีของเอสวีเอ็มคือลักษณะตัวมันที่ไม่เป็นเชิงเส้นจะสามารถใช้กับการแบ่งกลุ่มชนิดไม่เป็นเชิงเส้นได้

โครงงานนี้มีวัตถุประสงค์เพื่อศึกษาการแบ่งข้อมูลด้วยวิธีเอสวีเอ็ม โดยใช้ฟังก์ชันเรเดียลเบสิส กับฟังก์ชันโพลิโนเมียล เป็นฟังก์ชันเคอร์เนล การทดลองจะแสดงให้เห็นถึงเส้นไฮเปอร์เพลนที่ได้ทั้งกับข้อมูลที่แบ่งได้และแบ่งไม่ได้ด้วยเส้นตรง นอกจากนั้นจะยังแสดงให้เห็นถึงผลของค่าพารามิเตอร์ของฟังก์ชันเคอร์เนลว่ามีผลต่อไฮเปอร์เพลนอย่างไร

Project title	Support Vector Machine for Classification
Name	Mr. Panupong Senatham ID. 48380358
Project advisor	Ms. Supawan Phonphitakchai, Ph.D.
Major	Electrical Engineering
Department	Electrical and Computer Engineering
Academic year	2009

....

Abstract

SVM (Support vector machine) is the method to investigate a relation between given input and output data. In this project, SVM is used as a tool to classify the data. A framework of classification is a method to investigate for hyperplane which lies in the middle between 2 classes of data. SVM applies the idea of finding hyperplane with kernel function and also including the method of constraint optimization. The advantage of SVM is its nonlinear characteristic which can classify a nonlinear data.

This project aims to study the classification method with SVM using radial basis function and polynomial function as the kernel function. The experiments illustrate hyperplane in both linear and nonlinear data. The effects of parameters in kernel function on hyperplane are also shown.

กิตติกรรมประกาศ

โครงงานเรื่องการแบ่งกลุ่มข้อมูลด้วยวิธีเอสวีเอ็ม สำเร็จได้ด้วยดีด้วยความกรุณาจาก ดร. ศุภวรรณ พลพิทักษ์ชัย ซึ่งเป็นอาจารย์ที่ปรึกษาให้กับโครงงานนี้ ที่กรุณาให้คำปรึกษา คำแนะนำ และช่วยเหลือตรวจสอบแก้ไขข้อบกพร่องต่างๆ ตลอดจนให้ความรู้และข้อคิดเห็นที่เป็นประโยชน์ต่อโครงงานนี้ด้วยความเอาใจใส่เป็นอย่างดีจนกระทั่งโครงงานเสร็จสมบูรณ์

ขอขอบคุณ ดร. มุฑิตา สงฆ์จันทร์ และ ดร. นิพัทธ์ จันทรมินทร์ ซึ่งเป็นคณะกรรมการโครงงาน ที่ให้คำแนะนำตลอดระยะเวลาในการทำโครงงาน

ผู้จัดทำโครงงานขอกราบขอบพระคุณทุกท่านที่มีส่วนร่วมในการทำโครงงานนี้ จนทำให้โครงงานนี้สำเร็จลุล่วงไปได้ด้วยดี มา ณ โอกาสนี้

นายภาณุพงศ์ เสนาธรรม

สารบัญ

หน้า

สารบัญ (ต่อ)

สารบัญรูป

สารบัญรูป (ต่อ)

รูปที่ หน้า

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของโครงงาน

เครื่องการเรียนรู้ (Machine Learning) เป็นงานวิจัยเพื่อหาวิธีการทำให้ระบบคอมพิวเตอร์สามารถเรียนรู้ ปรับปรุงตัวเองได้ หรืออาจกล่าวได้ว่าการเรียนรู้คือ การศึกษาวิธีวิเคราะห์เพื่อจำแนก หรือแจกแจงข้อมูลจำนวนมาก การเรียนรู้เป็นการหาความสัมพันธ์ระหว่างข้อมูลขาเข้าและขาออก ซึ่งจะแสดงในรูปฟังก์ชันเบื้องหลัง (Underlying function) การหาความสัมพันธ์จะมีสองแบบคือ การจำแนก (Classification) คือการแบ่งกลุ่มระหว่างข้อมูลเช่นแบ่งระหว่างตัว x กับ o โดยมีเส้นระนาบเกิน (Hyper plane) ในการแบ่งข้อมูลออกจากกัน อย่างที่สองก็คือ การถดถอยเชิงเส้น (Regression) คือ การลากเส้นไปตามจุดต่างๆจะเป็นเชิงเส้น หรือไม่เป็นเชิงเส้น ก็ได้

ในปี ค.ศ. 1960 ได้มีผู้เสนอใช้เคอร์เนลโดยมีข้อแตกต่างกับวิธีการเรียนรู้ที่มีมาก่อนหน้านั้น ในแง่ของความสามารถในการหาค่าความผิดพลาดที่ต่ำที่สุดได้โดยที่การเรียนรู้แบบเก่าสามารถหาได้เพียงค่าต่ำสุดเฉพาะที่ (Local minima) ซึ่งการหาค่าความผิดพลาดต่ำสุดสามารถหาข้อพิสูจน์ได้แบบแน่นอน นอกจากนั้นยังสามารถใช้ได้กับการแบ่งกลุ่มชนิดไม่เป็นเชิงเส้นได้ถูกนำมาประยุกต์ใช้กับทางการแพทย์ เช่น แบ่งข้อมูลดีเอ็นเอ

ในงานด้านการแบ่งกลุ่มของข้อมูลวิธีเคอร์เนล (Kernel method) ถูกนำมาประยุกต์ใช้เพื่อคำนวณค่าผลคูณภายในบนปริภูมิแต่งเติม (Feature space) ที่มีมิติสูงขึ้นประสิทธิภาพของการจำแนกประเภทหรือการประมาณค่านั้น จะขึ้นอยู่กับเคอร์เนลที่เลือกใช้ ซึ่งสามารถปรับให้เหมาะสมกับปัญหาที่กำลังสนใจ และให้ผลการแบ่งกลุ่ม หรือการประมาณค่าที่ดีในปริภูมิแต่งเติม

เอสวีเอ็ม (SVM) ย่อมาจาก Support Vector Machine คือวิธีที่ใช้สำหรับการแบ่งกลุ่มประเภทแบบเส้นตรง (Linear classifier) ที่ทำงานโดยใช้ฟังก์ชันเคอร์เนล และสร้างโดยใช้หลักการหาขอบที่กว้างที่สุดและการแก้ปัญหาแบบมีข้อแม้ (Constrain optimization) ข้อดีของเอสวีเอ็ม คือสามารถหาค่าได้ในจุดที่ต่ำที่สุด (Global minima) ได้ดีกว่าวิธีโครงข่ายประสาท (Neural network) นอกนั้นยังสามารถใช้ได้กับการแบ่งกลุ่มชนิดไม่เป็นเชิงเส้น

ดังนั้นโครงงานนี้จึงนำเสนอวิธีการแบ่งกลุ่มที่เรียกว่า เอสวีเอ็ม ซึ่งเป็นการแบ่งกลุ่มข้อมูลต่างๆ โดยการสร้างอัลกอริทึมของเอสวีเอ็มจะทำบนโปรแกรมแมทแลป (MATLAB)

1.2 วัตถุประสงค์ของโครงงาน

1. เพื่อศึกษาวิธีการแบ่งกลุ่มข้อมูลด้วยวิธีเอสวีเอ็ม
2. เพื่อนำความรู้ที่ได้ไปประยุกต์ใช้กับการแบ่งกลุ่มของข้อมูลชนิดต่างๆ
3. เพื่อศึกษาโปรแกรมแมทแลป และนำความรู้ที่ได้มาประยุกต์ใช้กับการแบ่งกลุ่มวิธีเอสวีเอ็ม

1.3 ขอบเขตของโครงงาน

1. ศึกษาทฤษฎีของการแบ่งกลุ่มวิธีเอสวีเอ็ม
2. ศึกษาข้อมูลชนิดต่างๆและทำการแบ่งกลุ่มด้วยวิธีเอสวีเอ็ม
3. ใช้โปรแกรมแมทแลป ในการสร้างการแบ่งกลุ่มวิธีเอสวีเอ็ม และแสดงผลเปรียบเทียบ

1.4 ขั้นตอนการดำเนินงานและแผนการดำเนินงานตลอดโครงการวิจัย

รายละเอียด	ปี 2552							ปี 2553		
	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.
1. ศึกษาทฤษฎีการแบ่งกลุ่มวิธีเอสวีเอ็ม										
2. ศึกษาข้อมูลชนิดต่างๆมาทำการแบ่งกลุ่มด้วยวิธีเอสวีเอ็ม										
3. ใช้โปรแกรมแมทแลปในการสร้างการแบ่งกลุ่มวิธีเอสวีเอ็ม และแสดงผลเปรียบเทียบ										
4. จัดทำปริญญานิพนธ์ฉบับสมบูรณ์										

1.5 ประโยชน์ที่คาดว่าจะได้รับจากโครงงาน

1. สามารถแบ่งกลุ่มข้อมูลด้วยวิธีเอสวีเอ็มได้
2. สามารถนำความรู้ที่ได้ไปประยุกต์ใช้กับการแบ่งกลุ่มของข้อมูลชนิดต่างๆ
3. สามารถนำความรู้ที่ได้จากการเขียนโปรแกรมแมทแลป มาประยุกต์ใช้กับการแบ่งกลุ่มวิธีเอสวีเอ็ม
4. เป็นพื้นฐานในการศึกษาการแบ่งกลุ่มเอสวีเอ็ม ขั้นสูงต่อไป

1.6 งบประมาณ

1. ค่าถ่ายเอกสารและค่าเข้าเล่มรายงานฉบับสมบูรณ์	เป็นเงิน	1000	บาท
2. ค่าพิมพ์เอกสาร	เป็นเงิน	500	บาท
3. ค่าวัสดุคอมพิวเตอร์	เป็นเงิน	500	บาท
	รวมเป็นเงินทั้งสิ้น	2000	บาท
		(สองพันบาทถ้วน)	

หมายเหตุ ถัวเฉลี่ยทุกรายการ

บทที่ 2
การแบ่งกลุ่มข้อมูล

ปัญหาของการจัดแบ่งกลุ่มข้อมูลสามารถพิจารณาได้ว่าเป็นปัญหาชนิดสองกลุ่ม (Two class problem) เป้าหมายก็คือการแบ่งข้อมูลออกเป็น 2 กลุ่มโดยใช้ฟังก์ชันที่คำนวณได้จากตัวอย่างข้อมูลที่มีอยู่ พิจารณาจากรูปที่ 1 จะมีเส้นการจัดกลุ่มที่สามารถนำไปใช้ในการแบ่งกลุ่มข้อมูลได้หลายเส้น แต่มีเพียงหนึ่งเส้นที่ทำให้ระยะห่างระหว่างข้อมูล 2 ชนิดมีค่ามากที่สุด นั่นคือระยะห่างระหว่างเส้นแบ่งและข้อมูลที่ใกล้เส้นแบ่งที่สุดของแต่ละกลุ่มมีค่ามากที่สุด เส้นตรงดังกล่าวนี้เป็นเส้นที่เหมาะสมที่สุดในการแยกชนิดของข้อมูลเรียกว่า ไฮเปอร์เพลน (Hyperplane)

รูปที่ 2.1 แสดงการแบ่งข้อมูลด้วยไฮเปอร์เพลน

2.1 การแบ่งด้วยเส้นไฮเปอร์เพลนที่เหมาะสมที่สุด

กำหนดให้เวกเตอร์ของข้อมูลขาเข้าที่ต้องการแบ่งกลุ่ม ประกอบไปด้วยข้อมูล 2 ชนิดนั่นคือ

$$(y_l, x_l), \ldots, (y_l, x_l), x \in R^n, y \in \{-1, +1\} \tag{2.1}$$

โดยที่ y_l คือ คลาสของข้อมูลและ x คือ เวกแตอร์ของข้อมูลที่ต้องการแบ่งกลุ่มไฮเปอร์เพลนที่แบ่งข้อมูลออกเป็น 2 ชนิด สามารถแสดงได้ดังนี้

$$(w \bullet x) + b = 0 \tag{2.2}$$

ซึ่งเซตของเวคเตอร์เหล่านี้ถูกแบ่งอย่างเหมาะสมที่สุดโดยใช้เส้นแบ่งไฮเปอร์เพลน จากสมการที่ (2.2) ตัวแปร w, b จะมีเงื่อนไขบังคับดังนี้

$$\min_{x_i} |w \bullet x + b| = 1 \tag{2.3}$$

และไฮเปอร์เพลนที่เหมาะสมจะสอดคล้องกับสมการต่อไปนี้

$$y_i\ [(\mathbf{w} \bullet \mathbf{x}_i) + b] \geq 1,\ i = 1,\ldots,l \tag{2.4}$$

โดยที่ระยะห่าง d (w , b ; x) ของจุด x ถึงไฮเปอร์เพลน (w , b) คือ

$$d(\mathbf{w},b;\mathbf{x}) = \frac{|\mathbf{w} \bullet \mathbf{x} + b|}{\|\mathbf{w}\|} \tag{2.5}$$

ดังนั้นเส้นไฮเปอร์เพลนที่เหมาะสมที่สุดจะหาได้จากการหาค่าที่มากที่สุดของ ρ (w,b) โดยที่ต้องสอดคล้องกับสมการที่ (2.4) ถ้ามาร์จินถูกกำหนดโดย

$$\begin{aligned}
\rho(\mathbf{w},b) &= \min_{\{x_i : y_i = 1\}} d(\mathbf{w},b;\mathbf{x}_i) + \min_{\{x_j : y_j = -1\}} d(\mathbf{w},b;\mathbf{x}_j) \\
&= \min_{\{x_i : y_i = 1\}} \frac{|\mathbf{w} \bullet \mathbf{x}_i + b|}{\|\mathbf{w}\|} + \min_{\{x_j : y_j = -1\}} \frac{|\mathbf{w} \bullet \mathbf{x}_j + b|}{\|\mathbf{w}\|} \\
&= \frac{1}{\|\mathbf{w}\|} \left(\min_{\{x_i : y_i = 1\}} |\mathbf{w} \bullet \mathbf{x}_i + b| + \min_{\{x_j : y_j = -1\}} |\mathbf{w} \bullet \mathbf{x}_j + b| \right) \\
&= \frac{2}{\|\mathbf{w}\|}
\end{aligned} \tag{2.6}$$

ดังนั้นเส้นไฮเปอร์เพลนจึงเหมาะสมที่สุดในการแบ่งค่าข้อมูลจะเป็นค่าที่น้อยที่สุดของ

$$\Phi(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 \tag{2.7}$$

วิธีแก้ปัญหาสมการที่ (2.7) ภายใต้ข้อบังคับของสมการที่ (2.4) หาได้จากจุดอานม้าของลากรานจ์ฟังก์ชันนัล (Lagrange functional)

$$L(\mathbf{w},b,\alpha) = \frac{1}{2} \|W\|^2 - \sum_{i=1}^{l} \{[(\mathbf{x}_i \bullet \mathbf{W}) + b] y_i - 1\} \tag{2.8}$$

เมื่อ α_i คือ ตัวคูณลากรานจ์ ดังนั้นสมการที่ (2.8) จะถูกหาค่าที่ต่ำที่สุดเทียบกับ w,b และหาค่ามากที่สุดเทียบกับ $\alpha_i \geq 0$ จากวิธีการแก้ปัญหาที่ใช้โดยทั่วไป สมการที่ (2.8) ซึ่งเรียกว่าเป็นปัญหาไพรมอล (Primal problem) จะถูกเปลี่ยนไปเป็นปัญหาดูอัล (Dual problem) ซึ่งจะหาคำตอบง่ายกว่านั่นคือ

$$\max_{\alpha} W(\alpha) = \max_{\alpha} \{ \min_{\mathbf{w},b} L(\mathbf{w},b,\alpha) \} \tag{2.9}$$

จุดต่ำสุดเทียบกับ w , b ของ L คือ

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \quad \sum_{i=1}^{l} \alpha_i y_i = 0 \tag{2.10}$$

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \quad \mathbf{w} = \sum_{i=1}^{l} \alpha_i \mathbf{x}_i y_i$$

จากสมการที่ (2.8), (2.9) และ (2.10) ปัญหาดูอัลจะแสดงโดย

$$\max_{\alpha} W(\alpha) = \max_{\alpha} -\frac{1}{2} \sum_{i=1}^{l} \sum_{i=1}^{l} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \bullet \mathbf{x}_j) - \sum_{i=1}^{l} \alpha_i \tag{2.11}$$

และคำตอบที่ต้องการคือ

$$\overline{\alpha} = \arg\min \frac{1}{2} \sum_{i=1}^{l}\sum_{i=1}^{l} \alpha_i \alpha_j y_i y_j (x_i \bullet x_j) - \sum_{i=1}^{l} \alpha_i \tag{2.12}$$

ด้วยข้อบังคับ

$$\alpha_i \geq 0, \quad i = 1,\ldots,l \tag{2.13}$$

$$\sum_{i=1}^{l} \alpha_i y_i = 0$$

แก้ปัญหาสมการที่ (2.12) ด้วยข้อบังคับสมการที่ (2.13) จะได้ตัวคูณลากรานจ์ จากนั้นจะนำผลที่ได้มาสร้างเป็นไฮเปอร์เพลนที่ต้องการดังต่อไปนี้

$$\overline{w} = \sum_{i=1}^{l} \overline{\alpha}_i x_i y_i \tag{2.14}$$

$$\overline{b} = -\frac{1}{2} \overline{w} \bullet [x_r + x_s]$$

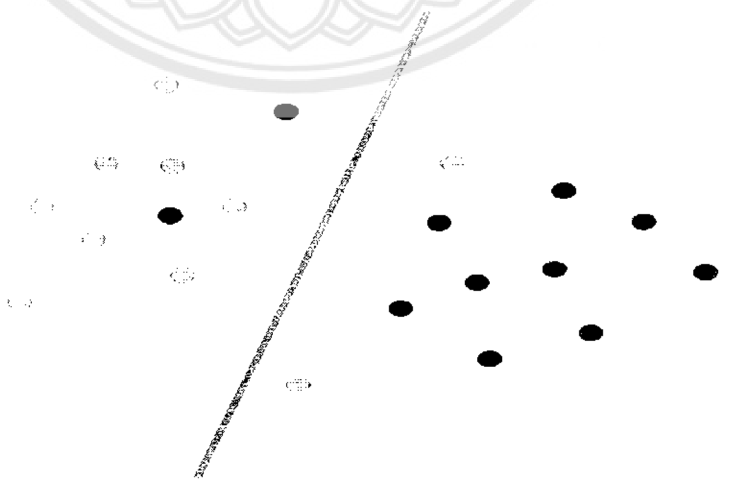
เมื่อ x_r และ x_s คือซัพพอร์ตเวกเตอร์ของแต่ละคลาสโดยที่

$$\overline{\alpha}_r, \overline{\alpha}_s > 0, \; y_r = 1, \; y_s = -1 \tag{2.15}$$

ดังนั้นจะได้ฟังก์ชันการจัดกลุ่ม

$$f(x) = \mathrm{sign}(\overline{w} \bullet x + \overline{b}) \tag{2.16}$$

2.2 การแบ่งด้วยเส้นไฮเปอร์เพลนที่เหมาะสมที่สุดในกรณีที่ข้อมูลไม่สามารถแบ่งแยกได้ด้วยเส้นตรง



รูปที่ 2.2 แสดงการแบ่งข้อมูลที่ไม่สามารถแบ่งได้แบบเส้นตรง

จากหัวข้อที่แล้วการแบ่งข้อมูลจะทำได้กรณีที่ข้อมูลสามารถถูกแบ่งกลุ่มได้ด้วยเส้นตรง อย่างไรก็ตามข้อมูลจริงมักจะไม่สามารถแบ่งแยกได้ด้วยเส้นตรงเนื่องจากสัญญาณ ดังนั้นจึงต้องมีการปรับปรุงวิธีการหาไฮเปอร์เพลน โดยการใช้ฟังก์ชันเพิ่มเติม และตัวแปร $\xi_i \geq 0$ ดังนั้น

$$F_\sigma(\xi) = \sum_{i=1}^{l} \xi_i^\sigma, \quad \sigma > 0, \tag{2.17}$$

เมื่อ ξ เป็นตัววัดความผิดพลาดของการจำแนก ข้อบังคับแสดงดังสมการที่ (7)จะถูกปรับปรุงเพื่อให้ใช้กับข้อมูลที่ไม่สามารถแบ่งได้ด้วยเส้นตรงดังนี้

$$y_i[(w \bullet x_i)+b] \geq 1-\xi_i, \; i = 1,\ldots,l \tag{2.18}$$

เมื่อ $\xi_i \geq 0$ เส้นไฮเปอร์เพลนจะถูกกำหนดโดยเวคเตอร์ w ที่ทำเป็นค่าต่ำที่สุดของฟังก์ชัน

$$\Phi(w,\xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i, \tag{2.19}$$

โดยที่ C เป็นค่าที่ถูกกำหนดให้เกี่ยวข้องกับสมการที่ (2.18)

วิธีการแก้ปัญหาค่าที่เหมาะสมที่สุดของสมการที่ (2.19) ภายใต้ข้อจำกัดของสมการที่ (2.18) จะเป็นจุดอานม้าของสมการ

$$L(w,b,\xi,\alpha,\beta) = \frac{1}{2}(w \bullet w) + C\sum_{i=1}^{l}\xi_i - \sum_{i=1}^{l}\alpha_i \; \{[(x_i \bullet w)+b]y_i - 1 + \xi_i\} - \sum_{i=1}^{l}\beta_i \; \xi_i, \tag{2.20}$$

เมื่อ α_i, β_i คือตัวคูณของลากรานจ์ และสมการที่ (31) จะถูกหาค่าต่ำสุดเทียบกับ w , b ,ξ และหาค่าสูงสุดเทียบกับ $\alpha_i, \beta_i \geq 0$ เช่นเดียวกับหัวข้อที่แล้ว สมการที่ (2.20) จะถูกทำให้อยู่ในรูปปัญหาดูอัล ซึ่งง่ายกว่าในการหาค่าคำตอบ นั่นคือ

$$\max_{\alpha,\beta} W(\alpha,\beta) = \max_{\alpha,\beta}\{\min_{w,b,\xi} L(w,b,\xi,\alpha,\beta)\} \tag{2.21}$$

หาค่าจุดต่ำสุดของสมการที่ (2.21) เทียบกับ w ,b, ξ จะได้

$$\begin{aligned} \frac{\partial L}{\partial b} &= 0 & \Rightarrow \quad & \sum_{i=1}^{l}\alpha_i y_i = 0 \\ \frac{\partial L}{\partial w} &= 0 & \Rightarrow \quad & w = \sum_{i=1}^{l}\alpha_i x_i y_i \\ \frac{\partial L}{\partial \xi_i} &= 0 & \Rightarrow \quad & \alpha_i + \beta_i = C \end{aligned} \tag{2.22}$$

ดังนั้นจากสมการที่ (2.20) , (2.21) และ (2.23) ปัญหาดูอัลจะแสดงโดย

$$\max_\alpha W(\alpha) = \max_\alpha -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j (x_i \bullet x_j) + \sum_{i=1}^{l}\alpha_i \tag{2.23}$$

และคำตอบคือ

$$\overline{\alpha} = \arg\min_\alpha \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j (x_i \bullet x_j) - \sum_{i=1}^{l}\alpha_i \tag{2.24}$$

ด้วยข้อบังคับ

$$0 \le \alpha_i \le C, \quad i = 1,\dots,l \tag{2.25}$$

$$\sum_{i=1}^{l} \alpha_i y_i = 0$$

การแก้ปัญหาจุดต่ำสุดวิธีนี้ เหมือนกับกรณีที่แล้วที่ข้อมูลสามารถแบ่งได้ด้วยเส้นตรง ยกเว้นการเปลี่ยนแปลงขอบเขตของตัวคูณลากรานจ์

2.3 การแบ่งแยกข้อมูลในมิติที่สูงขึ้น

ในกรณีนี้การแบ่งขอบเขตเส้นตรงด้วยวิธีเอสวีเอ็ม สามารถแมปเวกเตอร์ขาเข้า x ไปยังมิติที่สูงขึ้น เรียกว่าปริภูมิ z โดยใช้การแมปแบบไม่เป็นเส้นตรง ดังรูป

รูป2.3 แสดงการแมปข้อมูลไปยังมิติที่สูงกว่า

จากสมการที่ (2.24) ข้อมูลที่ใช้ในสมการจะถูกแมป ซึ่งส่วนใหญ่ฟังก์ชันที่นิยมใช้กันเช่น ฟังก์ชันโพลิโนเมียล ฟังก์ชันเส้นเฉียง และฟังก์ชันเคอร์เทรนซิกมอยด์ ปัญหาการแบ่งของสมการที่ (2.24) จะกลายเป็น

$$\overline{\alpha} = \arg\min_{\alpha} \frac{1}{2} \sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i \bullet x_j) - \sum_{i=1}^{l} \alpha_i \tag{2.26}$$

เมื่อ $K(x,y)$ คือ เคอร์เนลฟังก์ชัน ที่แสดงการแมปแบบไม่เป็นเส้นตรงไปยังปริภูมิเฟียเจอร์ แต่ไม่เปลี่ยนแปลงข้อจำกัด นั่นคือ

$$\alpha_i \ge 0, i = 1,\dots,l \tag{2.27}$$

$$\sum_{i=1}^{l} \alpha_i y_i = 0$$

แก้ปัญหาสมการที่ (2.26) ด้วยข้อจำกัดสมการที่ (2.27) จะได้ตัวคูณลากรานจ์และฟังก์ชันที่เหมาะสมในการแบ่งแยกข้อมูลในปริภูมิเฟียเจอร์ แสดงโดย

$$f(x) = \text{sign} \sum_{SVs} \overline{\alpha}_i y_i K(x_i, x) + \overline{b} \tag{2.28}$$

2.4 เคอร์เนลฟังก์ชัน

ในหัวข้อนี้จะแสดงถึงฟังก์ชันที่นิยมใช้ในการแมปปิ้งข้อมูลไปอยู่บนปริภูมิเฟียเจอร์

คุณสมบัติของฟังก์ชันเคอร์เนล คือ ผลคูณข้างในของฟังก์ชันเคอร์เนล สามารถแสดงได้เป็น

$$K(x,y) = k(x) \bullet k(y) \tag{2.29}$$

ถ้า K คือฟังก์ชันสมมาตรที่แน่นอน จะได้เงื่อนไขเมอร์เซอร์

$$K(x,y) = \sum_{m=1}^{\infty} \alpha_m \psi(x)\psi(y), \qquad \alpha_m \geq 0 \tag{2.30}$$

$$\iint K(x,y) g(x) g(y)\, dx dy > 0, \qquad \int g^2(x) dx < \infty$$

เมื่อเคอร์เนลคือผลคูณข้างในปริภูมิเฟียเจอร์

โพลิโนเมียลฟังก์ชัน

การแมปปิ้งโพลิโนเมียลนิยมใช้กับแบบไม่เป็นเส้นตรง

$$K(x,y) = ((x \bullet y)+1)^d \tag{2.31}$$

d หมายถึง ดีกรีของโพลิโนเมียล

เกาส์เซียนเรเดียลเบสิสฟังก์ชัน

ฟังก์ชันนี้ได้รับความสนใจอย่างแพร่หลาย

$$K(x,y) = \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right) \tag{2.32}$$

หมายถึงความกว้างของฟังก์ชันเคอร์เนล

เอกซ์โปเนนเชียลเรเดียลเบสิสฟังก์ชัน

$$K(x,y) = \exp\left(-\frac{|x-y|^2}{2\sigma^2}\right) \tag{2.33}$$

มัลติเลเยอร์เพอร์เซบตรอน (MLP)

$$K(\mathbf{x},\mathbf{y}) = \tanh(\text{scale}.(\mathbf{x} \bullet \mathbf{y})\text{-offset}) \tag{2.34}$$

ค่า scale และ offset เป็นค่าพารามิเตอร์ที่ต้องกำหนดค่าให้ เพื่อให้ได้ค่าที่แน่นอนของสเกลและออฟเซต ดังนั้นซัพพอร์ตเวกเตอร์จึงสอดคล้องกับชั้นแรก และผลคูณของลากรานจ์

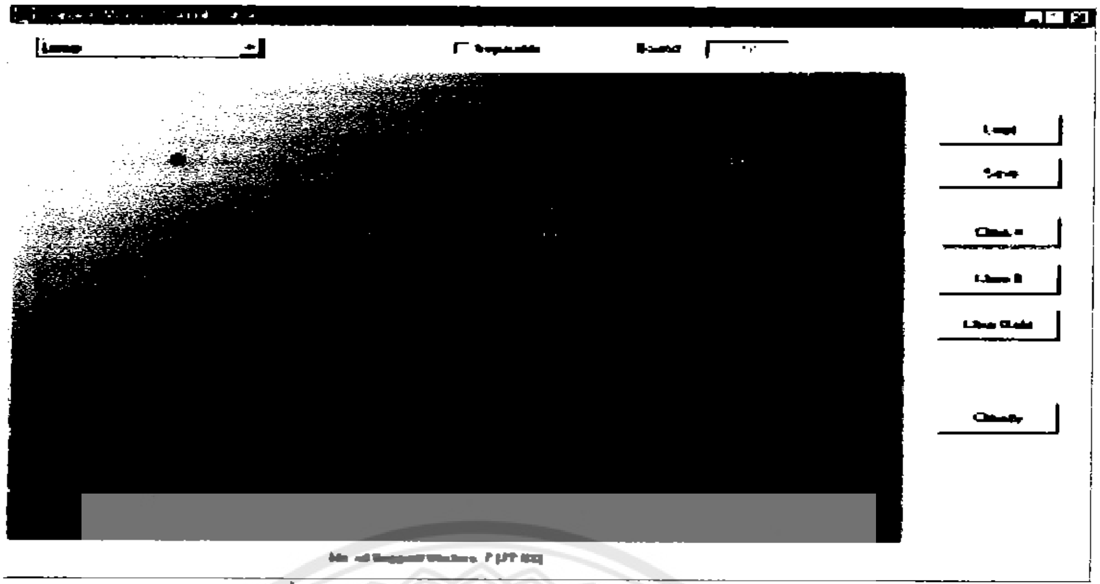
บทที่ 3
การแบ่งกลุ่มข้อมูลด้วยเอสวีเอ็ม

จากทฤษฎีในบทที่ 2 ได้กล่าวถึงวิธีการการแบ่งกลุ่มออกเป็น 2 กลุ่มด้วยวิธีเอสวีเอ็ม (SVM) โดยใช้สมการไฮเปอร์เพลน ในการแบ่งกลุ่มข้อมูล ในบทนี้จะกล่าวถึงแนวทางในการสร้างไฮเปอร์เพลนด้วยโปรแกรมแมทแลป

3.1 การแบ่งกลุ่มด้วยไฮเปอร์เพลน

การแบ่งกลุ่มด้วยวิธีนี้ เป็นวิธีที่ต้องให้ไฮเปอร์เพลนอยู่ระหว่างข้อมูล 2 กลุ่มและมีระยะห่างจากสองกลุ่มมากที่สุด จากรูปที่ 3.1 และ 3.2 แสดงตัวอย่างการแบ่งข้อมูลด้วยเส้นตรงที่แบ่งได้และไม่ได้

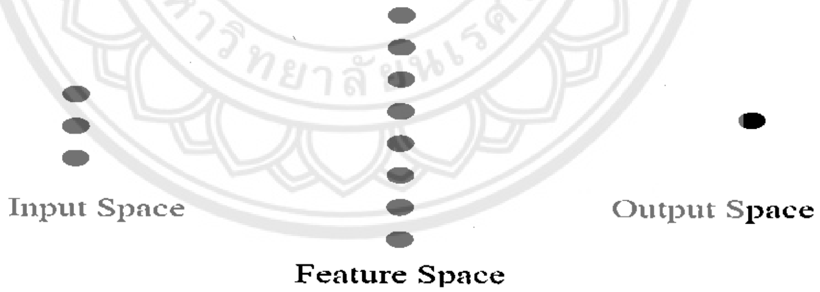
รูปที่ 3.1 ตัวอย่างการแบ่งข้อมูลด้วยเส้นตรง



รูปที่ 3.2 การแบ่งกลุ่มที่ไม่สามารถแบ่งได้ด้วยเส้นตรง

จากกรณีที่แบ่งด้วยเส้นตรงไม่ได้จะใช้วิธีเอสวีเอ็มในการแก้ปัญหา โดยทำการส่งข้อมูลไปอยู่ที่ปริภูมิเฟียเจอร์ ดังหัวข้อต่อไป

3.2 การแบ่งแยกข้อมูลในมิติที่สูงขึ้น (Generasation in High Dimention Feature Space)



รูปที่ 3.3 แสดงปริภูมิเฟียเจอร์

ไฮเปอร์เพลนจะหาได้จากสมการต่อไปนี้

$$\overline{\alpha} = \arg\min_{\alpha} \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i,\mathbf{x}_j) - \sum_{i=1}^{l}\alpha_i \qquad (3.1)$$

เมื่อ $K(\mathbf{x},\mathbf{y})$ คือเคอร์เนลฟังก์ชันแสดงการแมปแบบไม่เป็นเส้นตรง ไปยังปริภูมิเฟียเจอร์

$$\alpha_i \geq 0, i = 1,\ldots,l$$

$$\sum_{i=1}^{l}\alpha_i y_i = 0 \qquad (3.2)$$

การแก้สมการที่ (3.1) ด้วยข้อจำกัดของสมการที่ (3.2) โดยใช้ผลคูณลากรานจ์และใช้ไฮเปอร์เพลนในการแบ่งไปยังปริภูมิเฟียเจอร์

$$f(\mathbf{x}) = sign\left(\sum_{SVs} \overline{\alpha_i} y_i K(\mathbf{x}_i, \mathbf{x}) + \bar{b}\right) \tag{3.3}$$

เมื่อ

$$\overline{\mathbf{w}} \cdot \mathbf{x} = \sum_{SVs} \overline{\alpha_i} y_i K(\mathbf{x}_i, \mathbf{x})$$

$$\bar{b} = \frac{1}{2} \sum_{SVs} \overline{\alpha_i} y_i [K(\mathbf{x}_r, \mathbf{x}_i) + K(\mathbf{x}_s, \mathbf{x}_i)] \tag{3.4}$$

เมื่อแบ่งด้วยเส้นไฮเปอร์เพลนแล้ว ปริภูมิเฟียเจอร์จะส่งค่าที่ได้ไปยังส่วนเอาท์พุทสเปซ

รูปที่ 3.4 แสดงการแบ่งแบบไม่เป็นเส้นตรง

การแบ่งแยกข้อมูลด้วยเอสวีเอ็ม (SVM) สามารถสรุปเป็นขั้นตอนได้ดังต่อไปนี้

1. รับข้อมูลที่ต้องการแบ่งแยก
2. เลือกฟังก์ชันเคอร์เนล (k)
3. หาค่า α ที่จะใช้สร้างเป็นเส้นไฮเปอร์เพลนมาจากสมการที่ (3.1) โดยใช้วิธีที่เรียกว่า ควอดราติกโปรแกรมมิง (Quadratic Programming)
4. สร้างฟังก์ชันการแบ่งแยกจากสมการที่ (3.3)

เมื่อได้ฟังก์ชันการแบ่งแยกแล้ว จะสามารถแบ่งกลุ่มข้อมูลในอนาคตได้ว่าอยู่กลุ่มใด

บทที่ 4
ผลการแบ่งกลุ่มด้วยวิธีเอสวีเอ็ม

จากหลักการและวิธีแบ่งกลุ่มด้วยวิธีเอสวีเอ็มที่ได้กล่าวมาแล้วในบทที่ผ่านมา ในบทนี้จะเป็นการแสดงผลการแบ่งกลุ่มที่ได้ โดยใช้โปรแกรมแมทแลปในการแบ่งแยกข้อมูล

วิธีการที่ใช้ในการแบ่งคือ นำข้อมูลที่เรามีมาสร้างเป็นกลุ่ม โดยจะมีทั้งหมด 2 กลุ่ม ได้แก่ -1 และ 1 เมื่อได้กลุ่มในการแยกข้อมูลแล้ว เราก็จะนำข้อมูลที่เราสร้างเป็นกลุ่มมาใส่ในโปรแกรม แล้วดูว่าข้อมูลเหล่านั้นตกอยู่ในกลุ่มไหนของเส้นแบ่งแยกที่ได้ เราได้ทำการเขียนโปรแกรมให้ข้อมูลที่อยู่ในกลุ่มของ -1 เป็นสีแดง และเมื่อมีข้อมูลที่อยู่ในกลุ่มของ 1 เป็นสีน้ำเงิน

สมการของการแบ่งแยกข้อมูล คือ

$$f(x) = sign\left(\sum_{SVs} \overline{\alpha}_i y_i K(x_i, x) + \bar{b}\right) \tag{4.1}$$

4.1 กรณีแรกใช้ฟังก์ชันเคอร์เนลเป็นแบบเกาส์เซียน

สมการของเกาส์เซียน คือ

$$K(x,y) = \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right) \tag{4.2}$$

จากสมการจะเห็นว่ามีค่าพารามิเตอร์ σ ที่ต้องกำหนดค่าให้ก่อนที่จะเริ่มทำงาน ซึ่งการทดลองต่อไปนี้จะแสดงให้เห็นว่าค่า σ มีผลอย่างไรกับเส้นแบ่งกลุ่ม

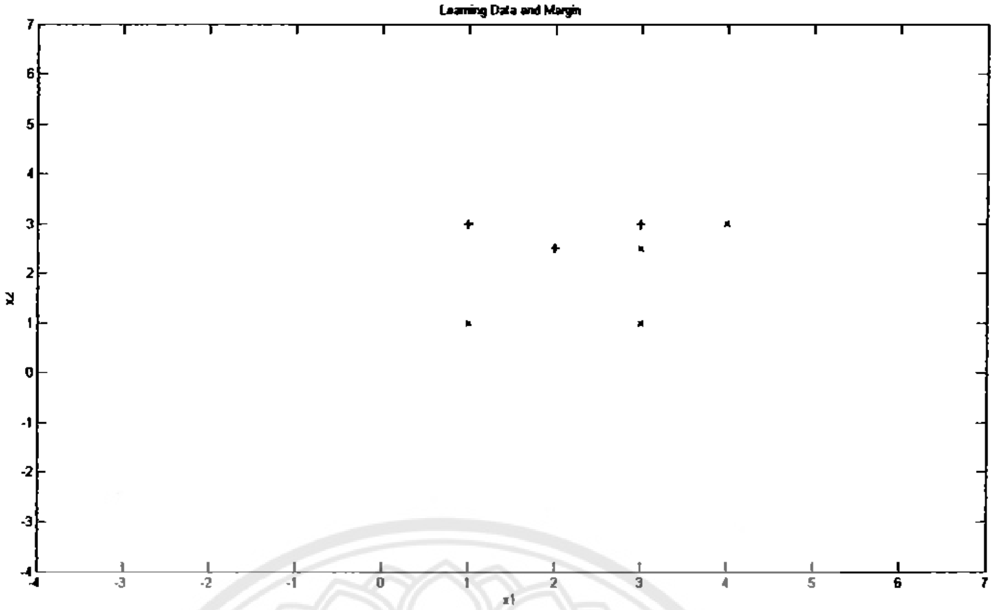
4.1.1 กรณีข้อมูลสามารถแบ่งได้ด้วยเส้นตรง

ข้อมูลที่ป้อนเข้าไปในโปรแกรม คือ

<table>
<tr><td>x1,x2</td><td>1</td><td>1</td><td>3</td><td>3</td><td>1</td><td>3</td><td>3</td><td>1</td><td>2</td><td>2.5</td><td>3</td><td>2.5</td><td>4</td><td>3</td></tr>
<tr><td>y</td><td colspan="2">-1</td><td colspan="2">1</td><td colspan="2">1</td><td colspan="2">-1</td><td colspan="2">1</td><td colspan="2">-1</td><td colspan="2">-1</td></tr>
</table>

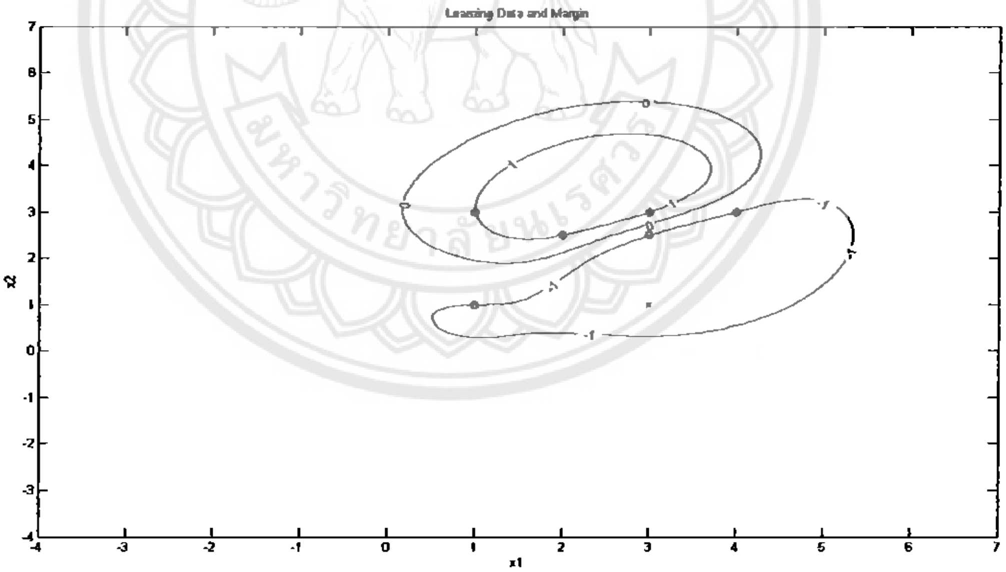
ค่า x คือ ข้อมูลขาเข้า และค่า y คือ กลุ่มของข้อมูล

ข้อมูลที่ต้องการแบ่งกลุ่มสามารถแสดงดังรูปที่ 4.1



รูปที่ 4.1 แสดงค่าข้อมูลที่นำมาใช้ในการแบ่งกลุ่ม

กรณีใช้ค่า $\sigma = 1$



รูปที่ 4.2 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า $\sigma = 1$

โดยเมื่อเราป้อนข้อมูล (x,y) เข้าไปในโปรแกรม โปรแกรมจะทำการสร้างกลุ่มของข้อมูลขึ้นมา นั่นคือมีเส้นแบ่งกลุ่มข้อมูลทั้งหมด 3 เส้น -1, 0, 1 โดย

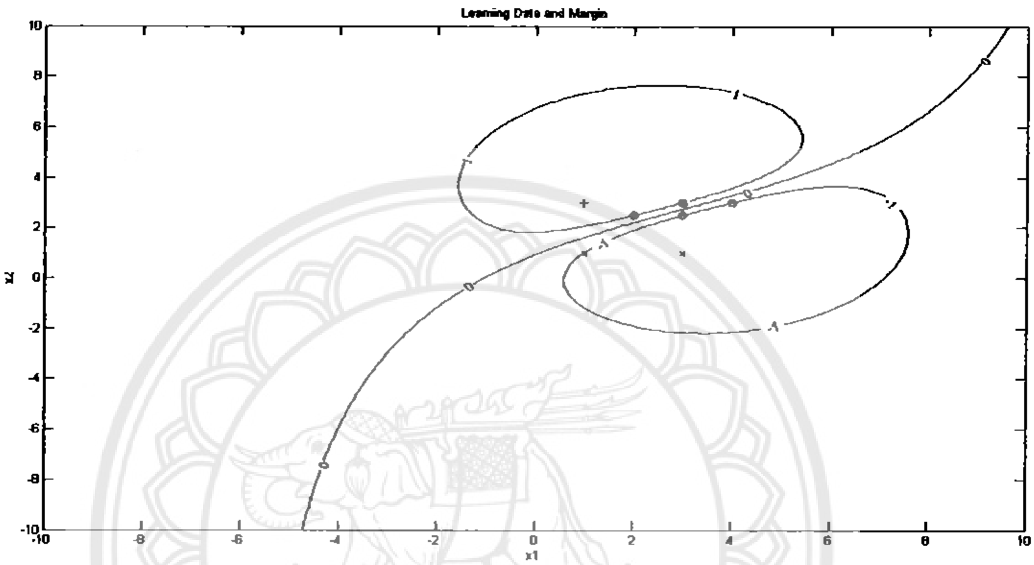
เส้น -1 คือ เส้นแบ่งเขตของกลุ่ม -1

เส้น 0 คือ เส้นที่แบ่งระหว่างกลุ่ม -1 กับกลุ่ม 1

เส้น 1 คือ เส้นแบ่งเขตของกลุ่ม 1

ข้อมูลที่มีค่ากลุ่ม -1 จะอยู่ในเส้น -1 และข้อมูลที่มีค่ากลุ่ม 1 จะอยู่ในเส้น 1 ซึ่งหมายถึงเส้นแบ่งกลุ่ม ถ้ามีข้อมูลที่ไม่รู้จักมาตกอยู่ภายในเส้นแบ่งกลุ่มใด ก็จะสามารถตีความได้ว่าข้อมูลนั้นอยู่ในกลุ่ม -1 หรือ 1

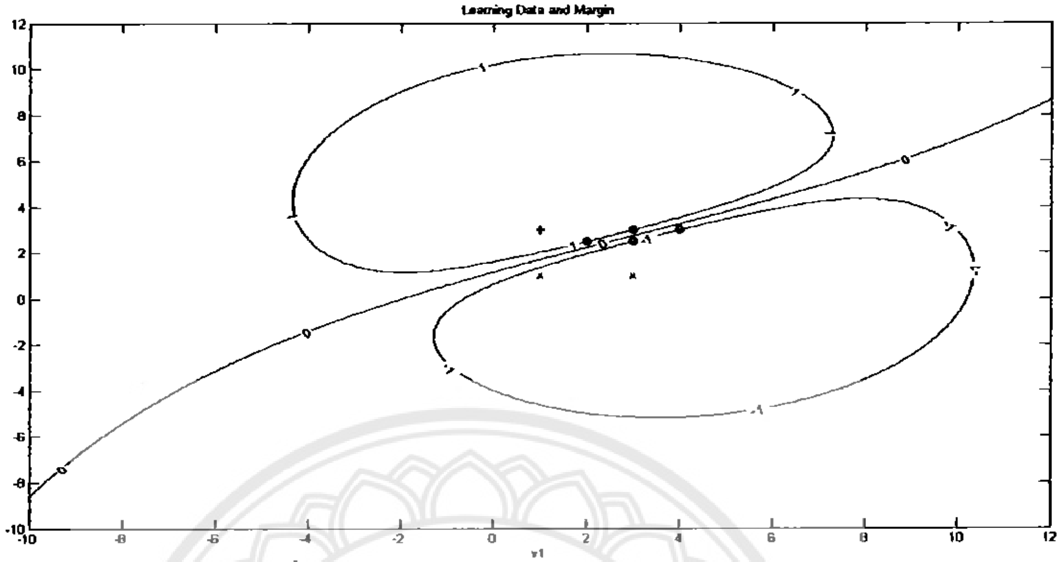
กรณีใช้ค่า $\sigma = 2$



รูปที่ 4.3 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า $\sigma = 2$

โดยเมื่อเราเพิ่มค่าของ σ เข้าไปในโปรแกรม ก็จะทำให้เส้นแบ่งกลุ่มเปลี่ยนแปลงไป แสดงดังรูปที่ 4.3 จะเห็นได้ว่าขอบเขตของเส้นแบ่งที่ใช้แบ่งกลุ่มจะมีขนาดใหญ่ขึ้น และเส้น 0 ที่ใช้ในการแบ่งกลุ่มของข้อมูลมีแนวเส้นที่ต่างจากกรณีที่ผ่านมา

กรณีใช้ค่า $\sigma = 3$



รูปที่ 4.4 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า $\sigma = 3$

เมื่อใช้ $\sigma = 3$ ความกว้างของฟังก์ชันเคอร์เนลมีมากขึ้น จึงทำให้เส้นแบ่งมีขนาดใหญ่ขึ้น แต่ก็ยังสามารถแบ่งกลุ่มข้อมูลได้อย่างถูกต้องเหมือนเดิม

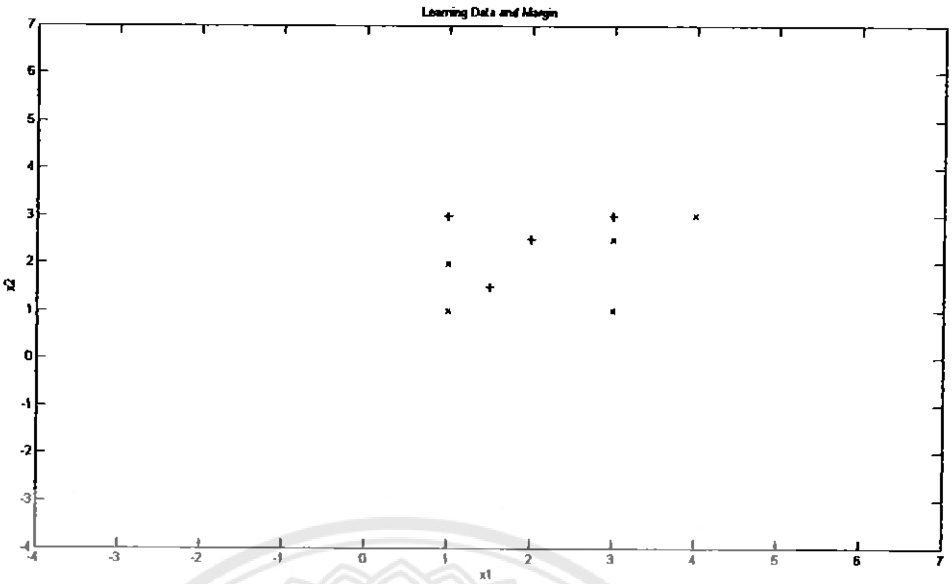
4.1.2 กรณีข้อมูลไม่สามารถแบ่งได้ด้วยเส้นตรง

ข้อมูลที่ป้อนเข้าไปในโปรแกรม คือ

x1,x2	1	1	3	3	1	3	3	1	2	2.5	3	2.5	4	3	1.5	1.5	1	2
y	-1		1		1		-1		1		-1		-1		1		-1	

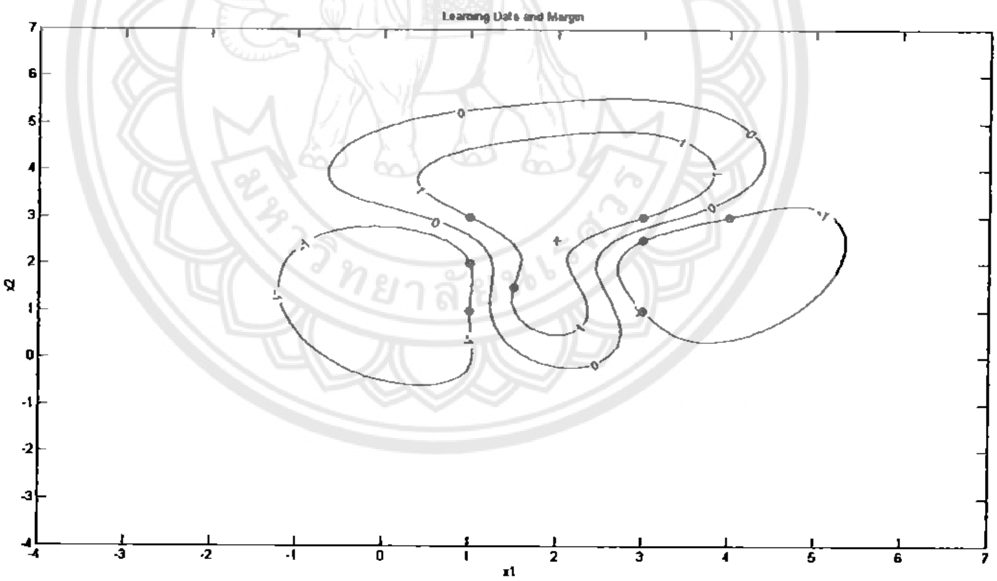
ค่า x คือ ข้อมูลขาเข้า และค่า y คือ กลุ่มของข้อมูล

รูปของข้อมูลแสดงได้ดังรูปที่ 4.5 จะเห็นได้ว่าข้อมูลมีลักษณะซับซ้อนขึ้นไม่สามารถแบ่งแยกได้ด้วยเส้นตรง



รูปที่ 4.5 แสดงค่าข้อมูลที่นำมาใช้ในการแบ่งกลุ่ม

กรณีใช้ค่า $\sigma = 1$



รูปที่ 4.6 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า $\sigma = 1$

โดยเมื่อเราป้อนข้อมูล (x,y) เข้าไปในโปรแกรม โปรแกรมจะทำการสร้างกลุ่มของข้อมูลขึ้นมา นั่นคือมีเส้นแบ่งกลุ่มข้อมูลทั้งหมด 3 เส้น -1, 0, 1 โดย

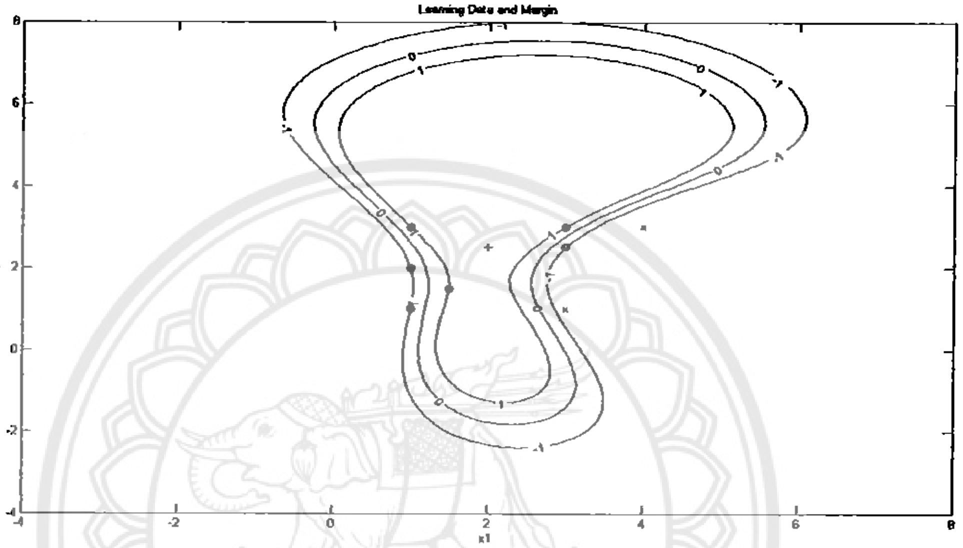
เส้น -1 คือ เส้นแบ่งเขตของกลุ่ม -1

เส้น 0 คือ เส้นที่แบ่งระหว่างกลุ่ม -1 กับกลุ่ม 1

เส้น 1 คือ เส้นแบ่งเขตของกลุ่ม 1

ข้อมูลที่มีค่ากลุ่ม -1 จะอยู่ในเส้น -1 และข้อมูลที่มีค่ากลุ่ม 1 จะอยู่ในเส้น 1 ซึ่งหมายถึงเส้นแบ่งกลุ่ม ถ้ามีข้อมูลที่ไม่รู้จักมาตกอยู่ภายในเส้นแบ่งกลุ่มใด ก็จะสามารถตีความได้ว่าข้อมูลนั้นอยู่ในกลุ่ม -1 หรือ 1

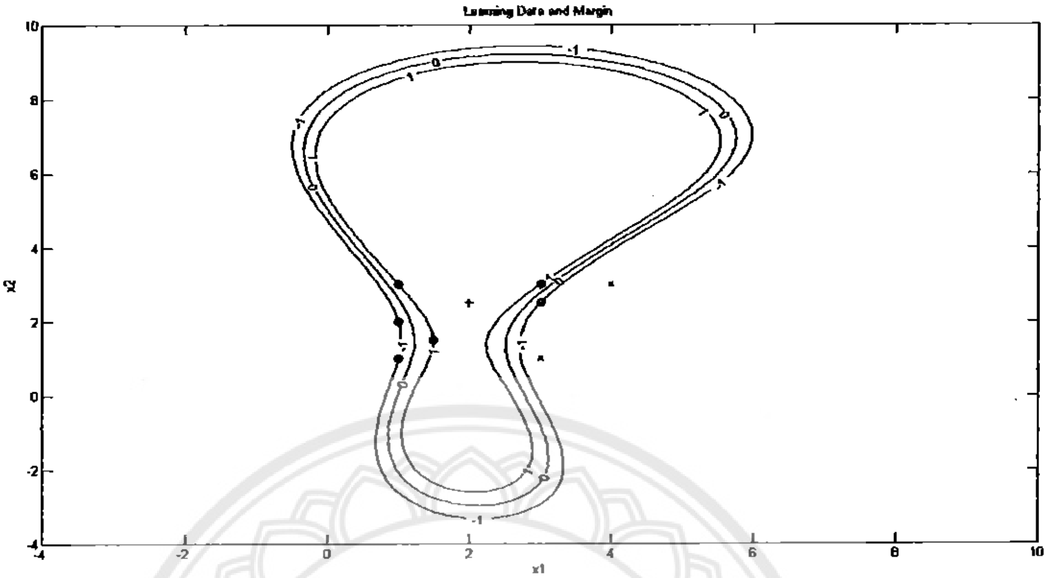
กรณีใช้ค่า $\sigma = 2$



รูปที่ 4.7 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า $\sigma = 2$

โดยเมื่อเราเพิ่มค่าของ σ เข้าไปในโปรแกรม ก็จะทำให้เส้นแบ่งกลุ่มเปลี่ยนแปลงไป ดัง **รูปที่ 4.7** จะเห็นได้ว่าขอบเขตของเส้นแบ่งที่ใช้แบ่งกลุ่มจะมีขนาดใหญ่ขึ้น และเส้น 0 ที่ใช้ในการแบ่งกลุ่มของข้อมูลมีแนวเส้นที่ต่างจากกรณีที่ผ่านมา

กรณีใช้ค่า $\sigma = 3$



รูปที่ 4.8 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า $\sigma = 3$

เมื่อใช้ $\sigma = 3$ ความกว้างของฟังก์ชันเคอร์เนลมีมากขึ้น จึงทำให้เส้นแบ่งมีขนาดใหญ่ขึ้น แต่ก็ยังสามารถแบ่งกลุ่มข้อมูลได้อย่างถูกต้องเหมือนเดิม

4.2 กรณีที่สองใช้ฟังก์ชันเคอร์เนลเป็นแบบโพลิโนเมียล

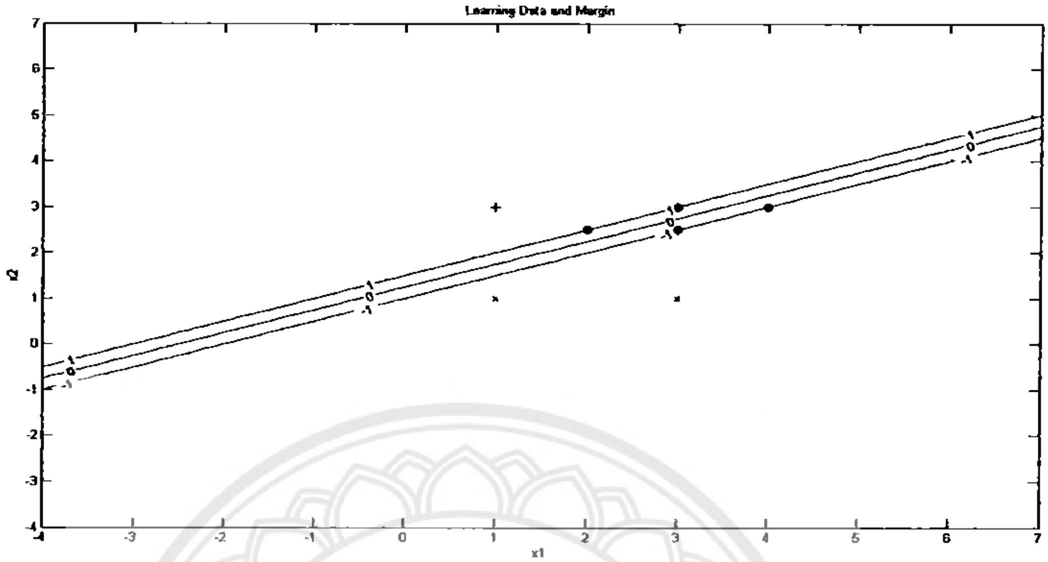
สมการ โพลิโนเมียล คือ

$$K(\mathbf{x},\mathbf{y}) = ((\mathbf{x} \cdot \mathbf{y})+1)^{d} \qquad (4.3)$$

4.2.1 กรณีข้อมูลสามารถแบ่งได้ด้วยเส้นตรง

ข้อมูลที่ป้อนเข้าไปในโปแกรม จะใช้เหมือนกับหัวข้อ 4.1.1

กรณีใช้ค่า d = 1



รูปที่ 4.9 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า d = 1

โดยเมื่อเราป้อนข้อมูล (x,y) เข้าไปในโปรแกรม โปรแกรมจะทำการสร้างกลุ่มของข้อมูลขึ้นมา นั่นคือมีเส้นแบ่งกลุ่มข้อมูลทั้งหมด 3 เส้น -1, 0, 1 โดย

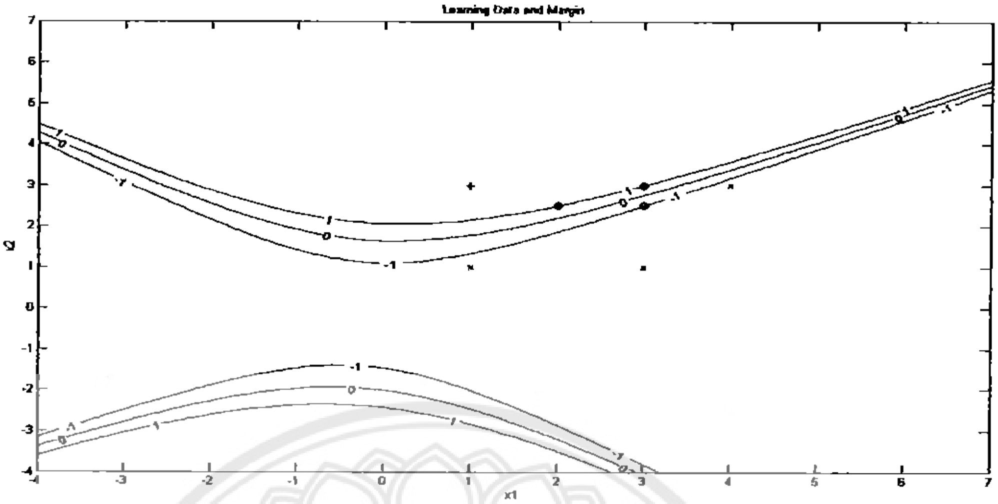
เส้น -1 คือ เส้นแบ่งเขตของกลุ่ม -1

เส้น 0 คือ เส้นที่แบ่งระหว่างกลุ่ม -1 กับกลุ่ม 1

เส้น 1 คือ เส้นแบ่งเขตของกลุ่ม 1

ข้อมูลที่มีค่ากลุ่ม -1 จะอยู่ในเส้น -1 และข้อมูลที่มีค่ากลุ่ม 1 จะอยู่ในเส้น 1 ซึ่งหมายถึงเส้นแบ่งกลุ่ม ถ้ามีข้อมูลที่ไม่รู้จักมาตกอยู่ภายในเส้นแบ่งกลุ่มใด ก็จะสามารถตีความได้ว่า ข้อมูลนั้นอยู่ในกลุ่ม -1 หรือ 1 จากรูปที่ 4.9 รูปนี้โปรแกรมสามารถแบ่งกลุ่มข้อมูลได้อย่างถูกต้องสมบูรณ์

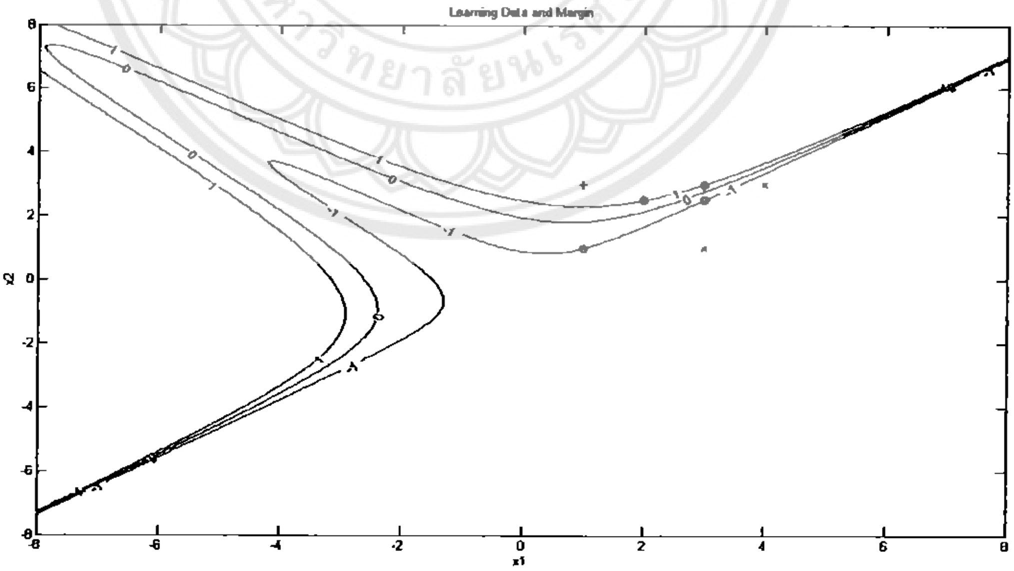
กรณีใช้ค่า d = 2



รูปที่ 4.10 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า d = 2

โดยเมื่อเราเพิ่มค่าของ d เข้าไปในโปรแกรม จะทำให้ฟังก์ชันเคอร์เนลมีความซับซ้อนขึ้น และทำให้เส้นแบ่งมีความซับซ้อนขึ้นด้วย

กรณีใช้ค่า d = 3



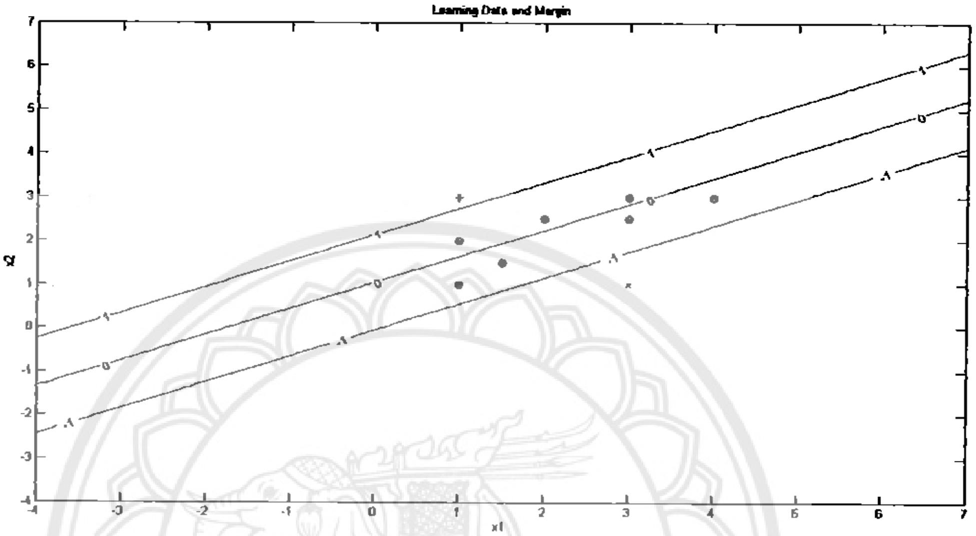
รูปที่ 4.11 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า d = 3

เมื่อใช้ d = 3 ความกว้างของฟังก์ชันเคอร์เนลมีมากขึ้น จึงทำให้เส้นแบ่งมีขนาดใหญ่ขึ้น แต่ก็ยังสามารถแบ่งกลุ่มข้อมูลได้อย่างถูกต้องเหมือนเดิม

4.2.2 กรณีข้อมูลไม่สามารถแบ่งได้ด้วยเส้นตรง

ข้อมูลที่ป้อนเข้าไปในโปรแกรม จะเหมือนกับข้อมูลในหัวข้อ 4.1.2

กรณีใช้ค่า d = 1



รูปที่ 4.12 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า d = 1

โดยเมื่อเราป้อนข้อมูล (x,y) เข้าไปในโปรแกรม โปรแกรมจะทำการ**สร้างกลุ่มของข้อมูลขึ้นมา** นั่นคือมีเส้นแบ่งกลุ่มข้อมูลทั้งหมด 3 เส้น -1, 0, 1 โดย

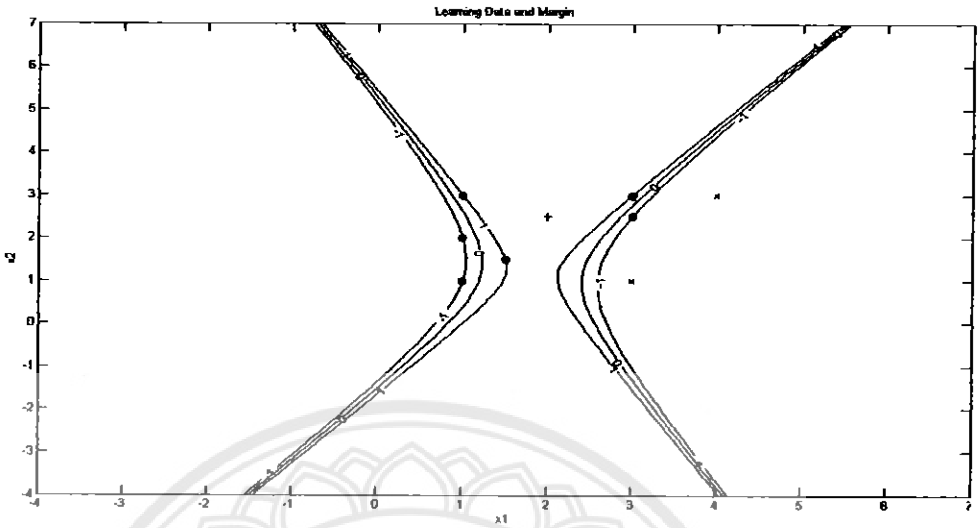
เส้น -1 คือ เส้นแบ่งเขตของกลุ่ม -1

เส้น 0 คือ เส้นที่แบ่งระหว่างกลุ่ม -1 กับกลุ่ม 1

เส้น 1 คือ เส้นแบ่งเขตของกลุ่ม 1

จากรูป 4.14 รูปนี้ไม่สามารถแบ่งด้วยเส้นตรงได้อย่างสมบูรณ์ จะเห็น**ได้ว่ามีข้อมูลของกลุ่ม -1 ตกไปอยู่ในกลุ่มของ 1 และมีข้อมูลของกลุ่ม 1 ตกไปอยู่ในกลุ่มของ -1**

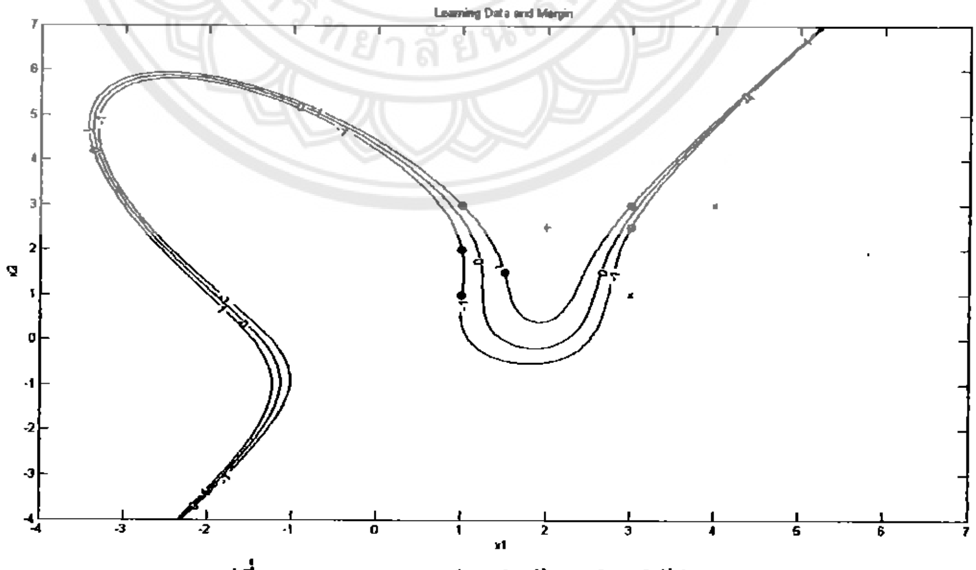
กรณีใช้ค่า d = 2



รูปที่ 4.13 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า d = 2

โดยเมื่อเราเพิ่มค่าของ d เข้าไปในโปรแกรม จะทำให้ฟังก์ชันเคอร์เนลมีความซับซ้อนขึ้น และทำให้เส้นแบ่งมีความซับซ้อนขึ้นด้วย จากเส้นแบ่งกรณีนี้สามารถแบ่งข้อมูลได้ถูกต้อง

กรณีใช้ค่า d = 3



รูปที่ 4.14 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า d = 3

เมื่อใช้ d = 3 ความกว้างของฟังก์ชันเคอร์เนลมีมากขึ้น จึงทำให้เส้นแบ่งมีขนาดใหญ่ขึ้น แต่ก็ยังสามารถแบ่งกลุ่มข้อมูลได้อย่างถูกต้องเหมือนเดิม

4.3 การทดสอบการแบ่งกลุ่มเมื่อเพิ่มจำนวนของข้อมูล

จากหัวข้อที่ 4.1 และ 4.2 เป็นการแบ่งกลุ่มข้อมูลที่ใช้จำนวนตัวแปรไม่มาก แต่ในหัวข้อนี้จะเป็นการทดสอบการแบ่งกลุ่ม โดยทำการเพิ่มข้อมูลให้กับโปรแกรม เพื่อศึกษาการทำงานของโปรแกรมว่าสามารถทำงานได้อย่างถูกต้องหรือไม่

4.3.1 ใช้เคอร์เนลฟังก์ชันแบบเกาส์เซียนกรณีสามารถแบ่งได้ด้วยเส้นตรง

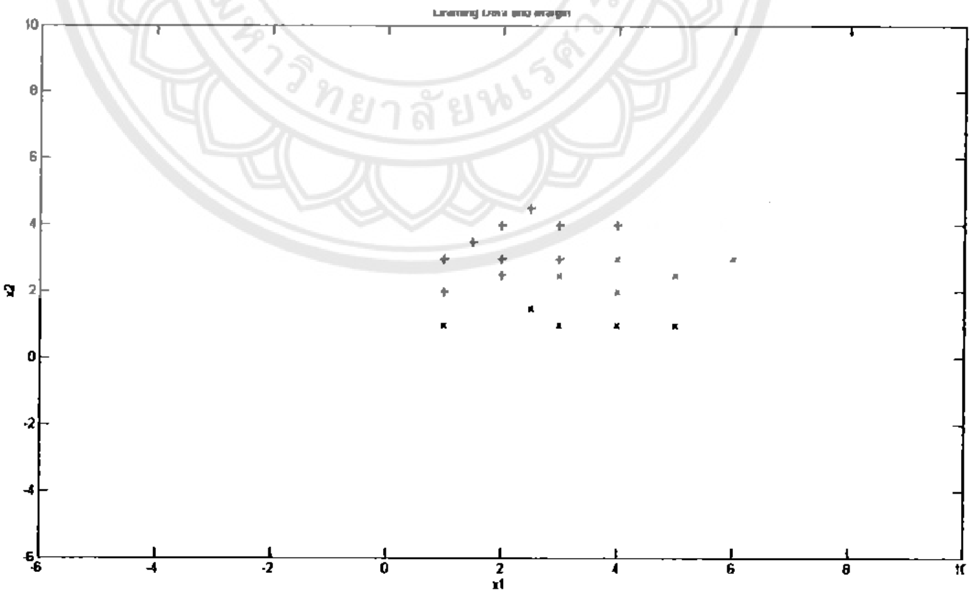
ข้อมูลที่ป้อนเข้าไปในโปรแกรมคือ

x1,x2	1	1	3	3	1	3	3	1	2	2.5	3	2.5	4	3	2	4	6	3	3	4
y	-1		1		1		-1		1		-1		-1		1		-1		1	

x1,x2	2.5	4.5	4	2	2.5	1.5	1	2	5	2.5	1.5	3.5	2	3	5	1	4	4	4	1
y	1		-1		-1		1		-1		1		1		-1		1		-1	

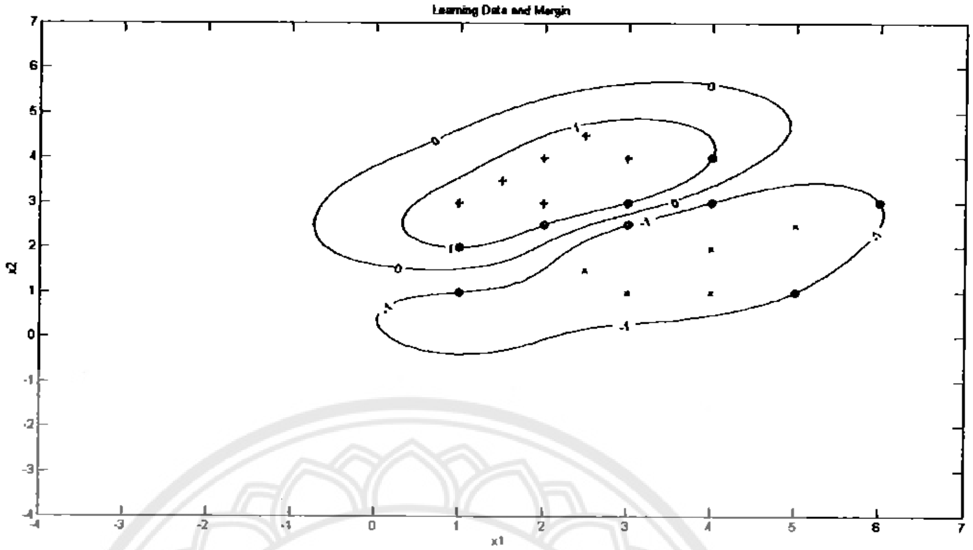
ค่า x คือข้อมูลขาเข้า และค่า y คือกลุ่มของข้อมูล

ข้อมูลที่ต้องการแบ่งกลุ่มสามารถแสดงดังรูปที่ 4.15



รูปที่ 4.15 แสดงค่าข้อมูลที่นำมาใช้ในการแบ่งกลุ่ม

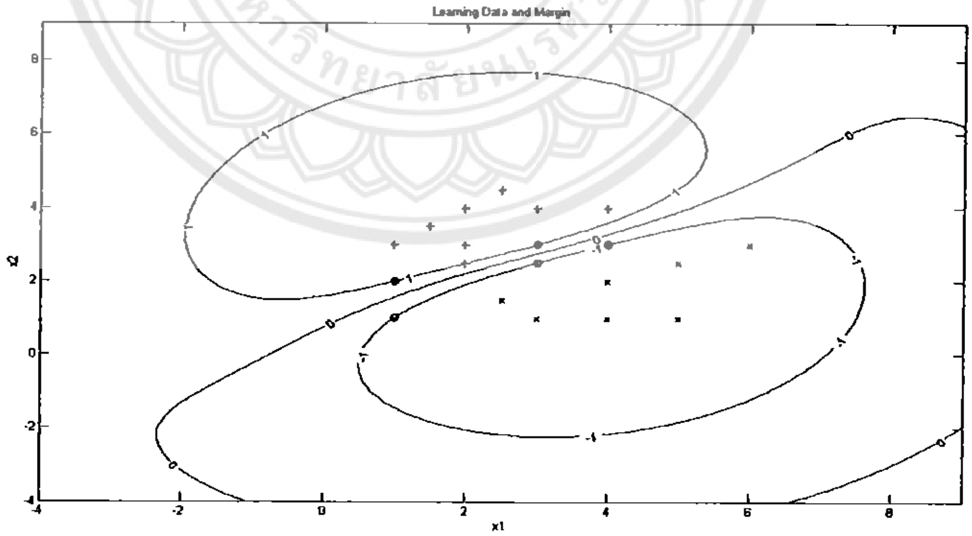
กรณีใช้ค่า $\sigma = 1$



รูปที่ 4.16 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า $\sigma = 1$

จากรูปที่ 4.16 ข้อมูลสามารถแบ่งกลุ่มได้อย่างสมบูรณ์

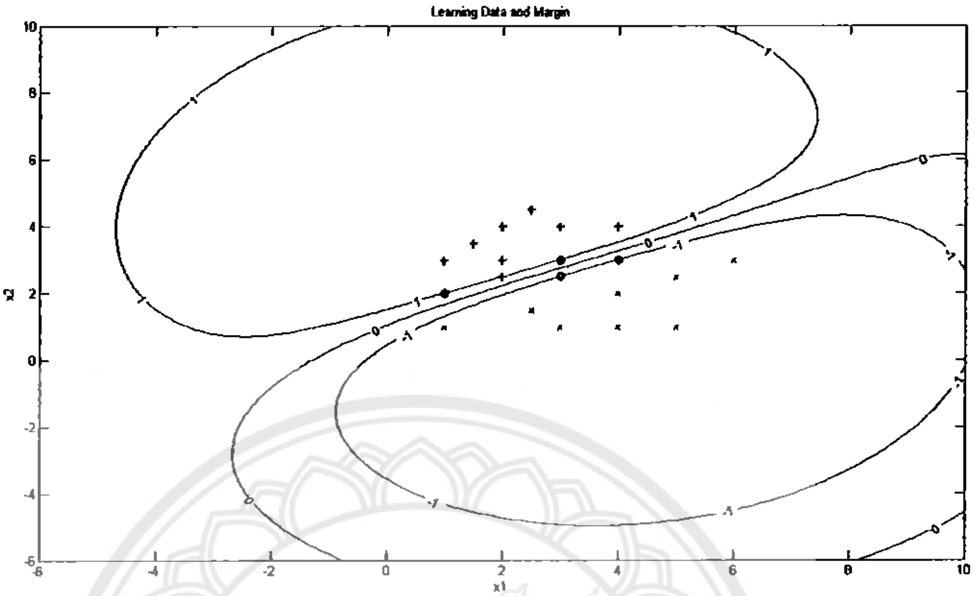
กรณีใช้ค่า $\sigma = 2$



รูปที่ 4.17 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า $\sigma = 2$

โดยเมื่อเราเพิ่มค่าของ σ เข้าไปในโปรแกรม ก็จะทำให้เส้นแบ่งกลุ่มเปลี่ยนแปลงไป แสดงดังรูปที่ 4.17 จะเห็นได้ว่าขอบเขตของเส้นแบ่งที่ใช้แบ่งกลุ่มจะมีขนาดใหญ่ขึ้น และเส้น 0 ที่ใช้ในการแบ่งกลุ่มของข้อมูลมีแนวเส้นที่ต่างจากกรณีที่ผ่านมา

กรณีใช้ค่า σ =3



รูปที่ 4.18 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า σ =3

เมื่อใช้ $\sigma = 3$ ความกว้างของฟังก์ชันเคอร์เนลมีมากขึ้น จึงทำให้เส้นแบ่งมีขนาดใหญ่ขึ้น แต่ก็ยังสามารถแบ่งกลุ่มข้อมูลได้อย่างถูกต้องเหมือนเดิม

4.3.2 ใช้เคอร์เนลฟังก์ชันแบบเกาส์เซียนกรณีไม่สามารถแบ่งได้ด้วยเส้นตรง

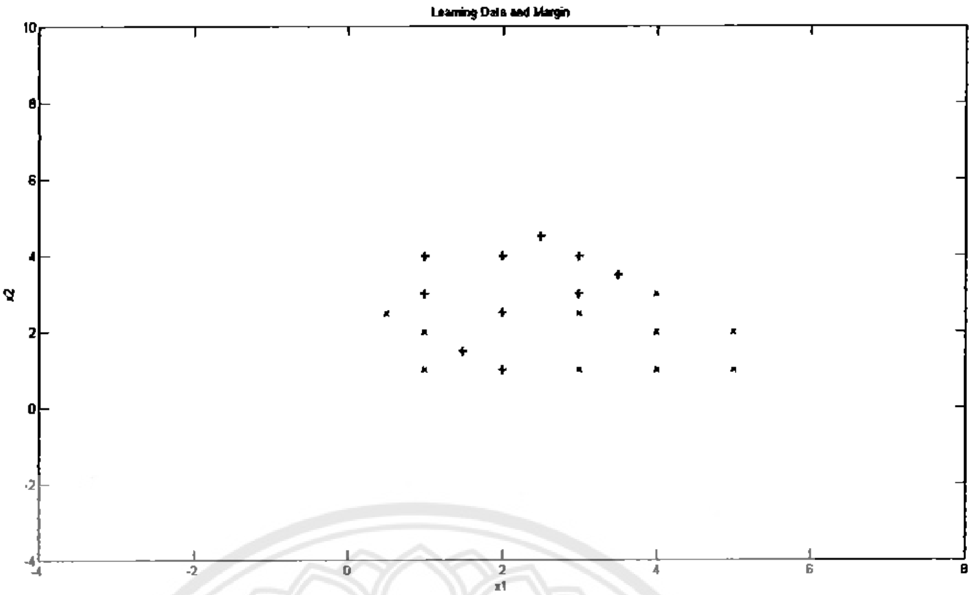
ข้อมูลที่ป้อนเข้าไปในโปรแกรม คือ

x1,x2	1	1	3	3	1	3	3	1	2	2.5	3	2.5	4	3	1.5	1.5	1	2	0.5	2.5
y	-1		1		1		-1		1		-1		-1		1		-1		-1	

x1,x2	2	1	2	4	4	2	5	1	1	4	3.5	3.5	4	1	3	4	5	2	2.5	4.5
y	1		1		-1		-1		1		1		-1		1		-1		1	

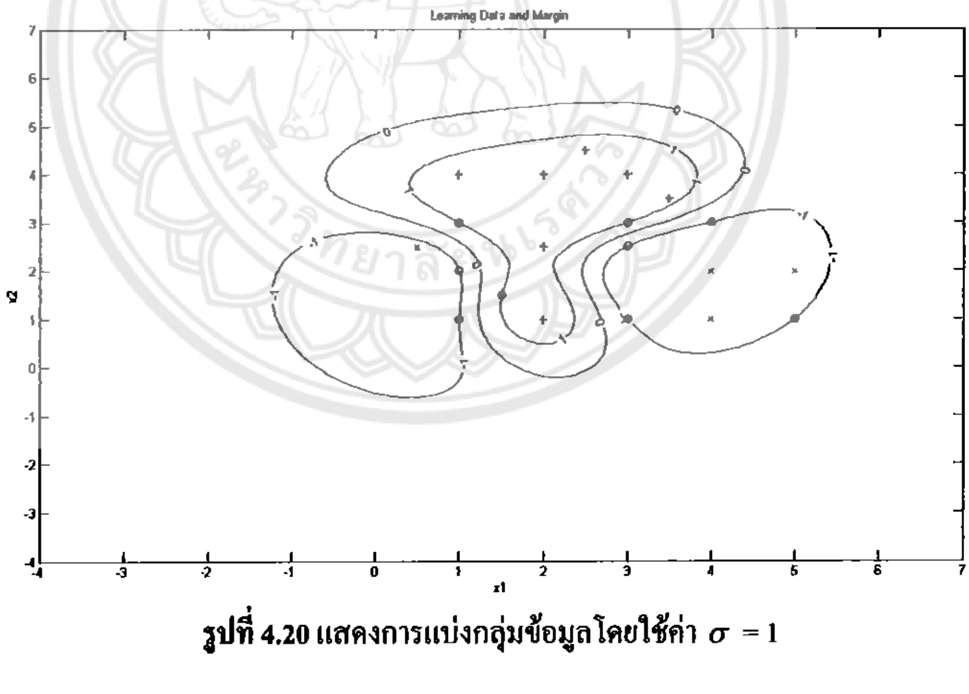
ค่า x คือข้อมูลขาเข้า และค่า y คือกลุ่มของข้อมูล

ข้อมูลที่ต้องการแบ่งกลุ่มสามารถแสดงดังรูปที่ 4.19



รูปที่ 4.19 แสดงค่าข้อมูลที่นำมาใช้ในการแบ่งกลุ่ม

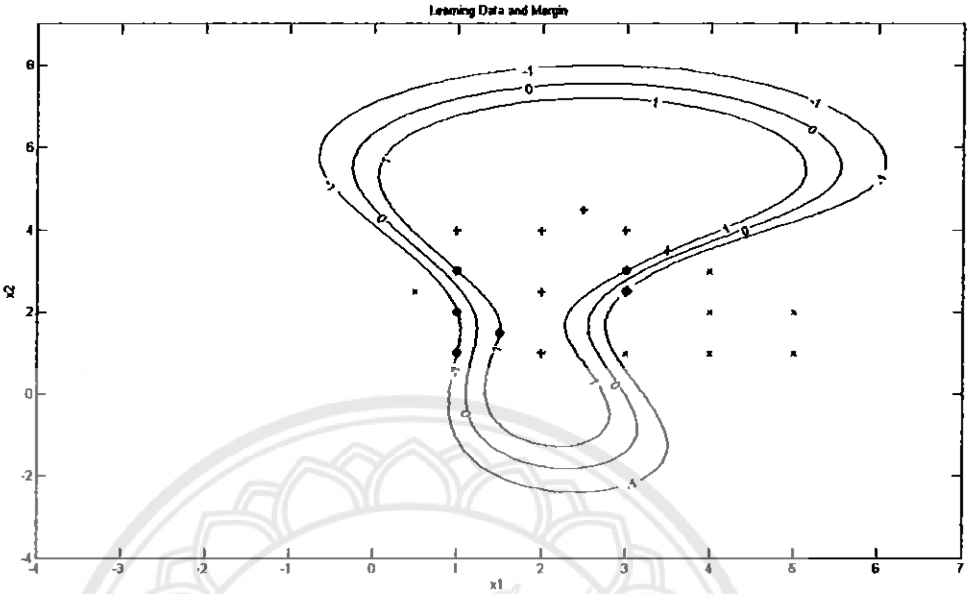
กรณีใช้ค่า $\sigma = 1$



รูปที่ 4.20 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า $\sigma = 1$

จากรูปที่ 4.20 ข้อมูลสามารถแบ่งกลุ่มได้อย่างสมบูรณ์

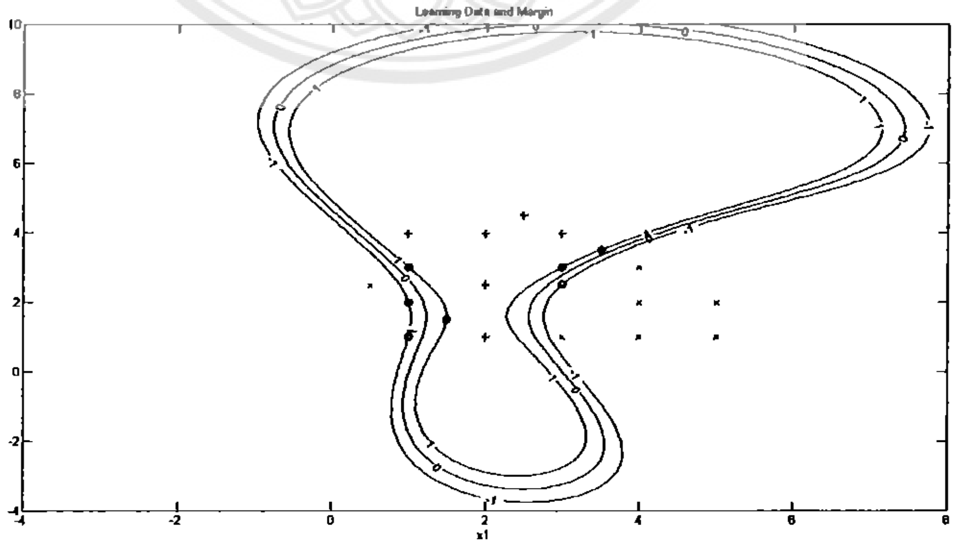
กรณีใช้ค่า σ = 2



รูปที่ 4.21 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า σ = 2

โดยเมื่อเราเพิ่มค่าของ σ เข้าไปในโปรแกรม ก็จะทำให้เส้นแบ่งกลุ่มเปลี่ยนแปลงไป ดังรูปที่ 4.21 จะเห็นได้ว่าขอบเขตของเส้นแบ่งที่ใช้แบ่งกลุ่มจะมีขนาดใหญ่ขึ้น และเส้น 0 ที่ใช้ในการแบ่งกลุ่มของข้อมูลมีแนวเส้นที่ต่างจากกรณีที่ผ่านมา

กรณีใช้ค่า σ = 3



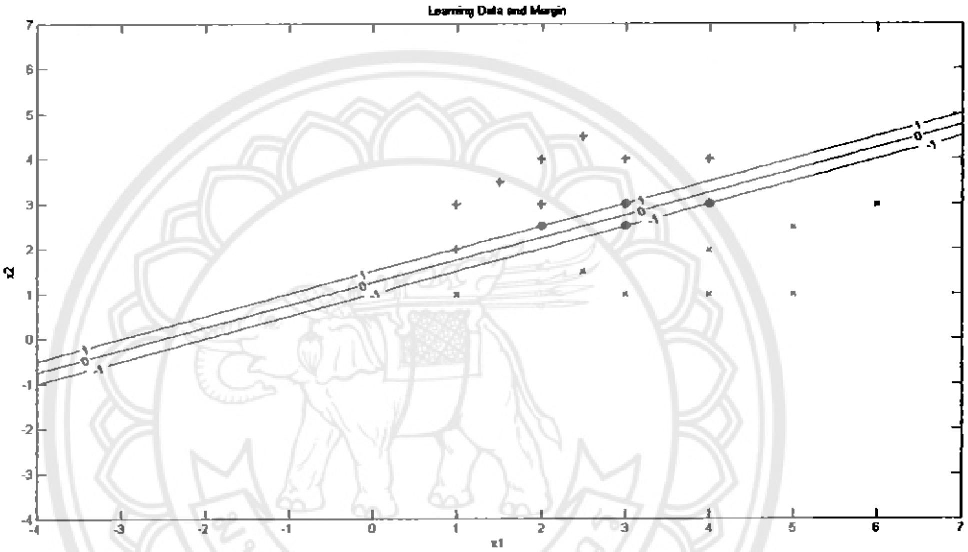
รูปที่ 4.22 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า σ = 3

เมื่อใช้ $\sigma = 3$ ความกว้างของฟังก์ชันเคอร์เนลมีมากขึ้นจึงทำให้เส้นแบ่งมีขนาดใหญ่ขึ้น แต่ก็ยังสามารถแบ่งกลุ่มข้อมูลได้อย่างถูกต้องเหมือนเดิม

4.3.3 ใช้เคอร์เนลฟังก์ชันโพลิโนเมียลกรณีสามารถแบ่งได้ด้วยเส้นตรง

ข้อมูลที่ป้อนเข้าไปในโปแกรม จะใช้เหมือนกับหัวข้อ 4.3.1

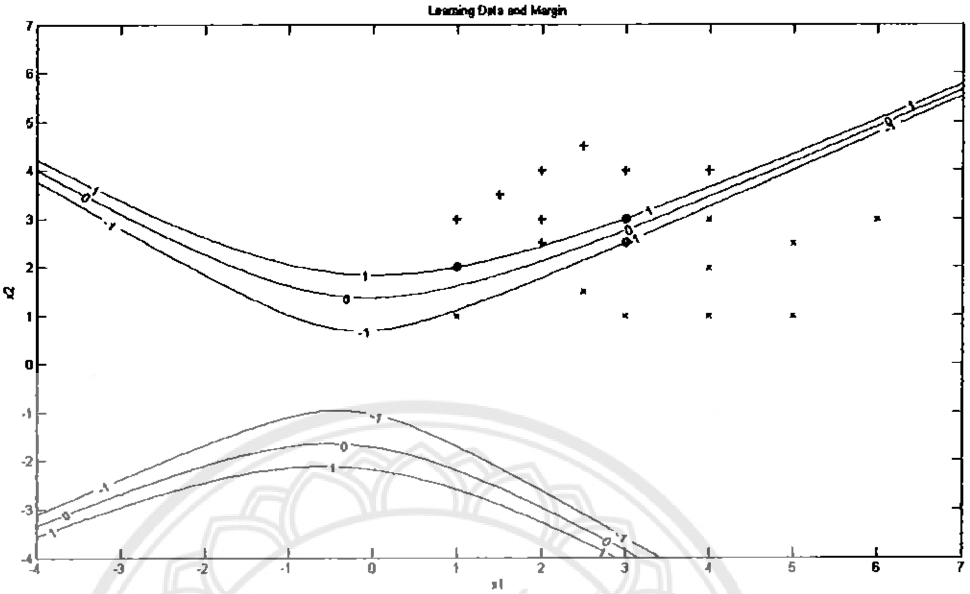
กรณีใช้ค่า d = 1



รูปที่ 4.23 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า d = 1

จากรูปที่ 4.23 ข้อมูลสามารถแบ่งกลุ่มได้อย่างสมบูรณ์

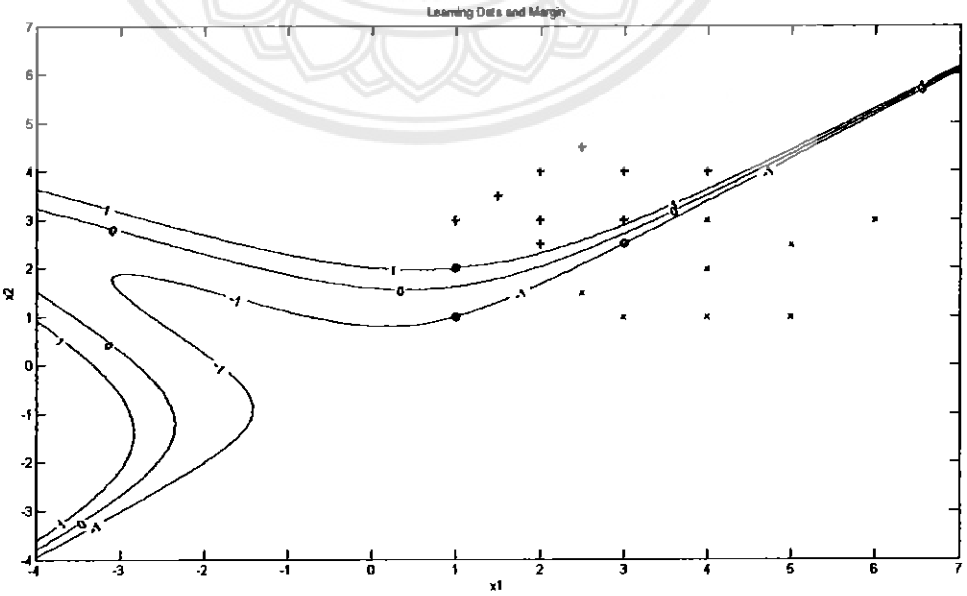
กรณีใช้ค่า d = 2



รูปที่ 4.24 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า d = 2

โดยเมื่อเราเพิ่มค่าของ d เข้าไปในโปรแกรม จะทำให้ฟังก์ชันเคอร์เนล**มีความซับซ้อนขึ้นและทำ**ให้เส้นแบ่งมีความซับซ้อนขึ้นด้วย

กรณีใช้ค่า d = 3



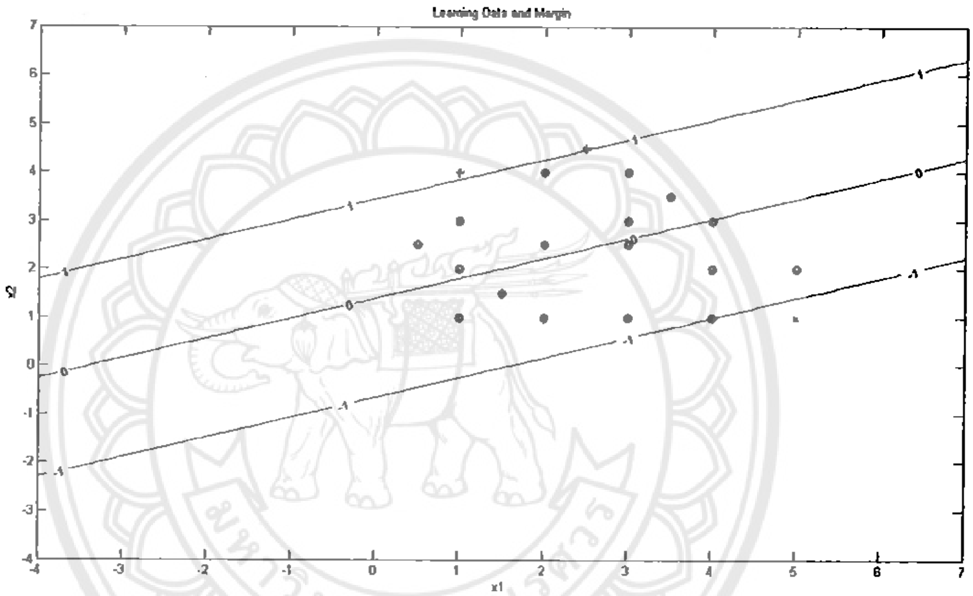
รูปที่ 4.25 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า d = 3

เมื่อใช้ d = 3 ความกว้างของฟังก์ชันเคอร์เนลมีมากขึ้น จึงทำให้เส้นแบ่งมีขนาดใหญ่ขึ้น แต่ก็ยังสามารถแบ่งกลุ่มข้อมูลได้อย่างถูกต้องเหมือนเดิม

4.3.4 ใช้เคอร์เนลฟังก์ชันโพลิโนเมียลกรณีไม่สามารถแบ่งได้ด้วยเส้นตรง

ข้อมูลที่ป้อนเข้าไปในโปแกรม จะใช้เหมือนกับหัวข้อ 4.3.2

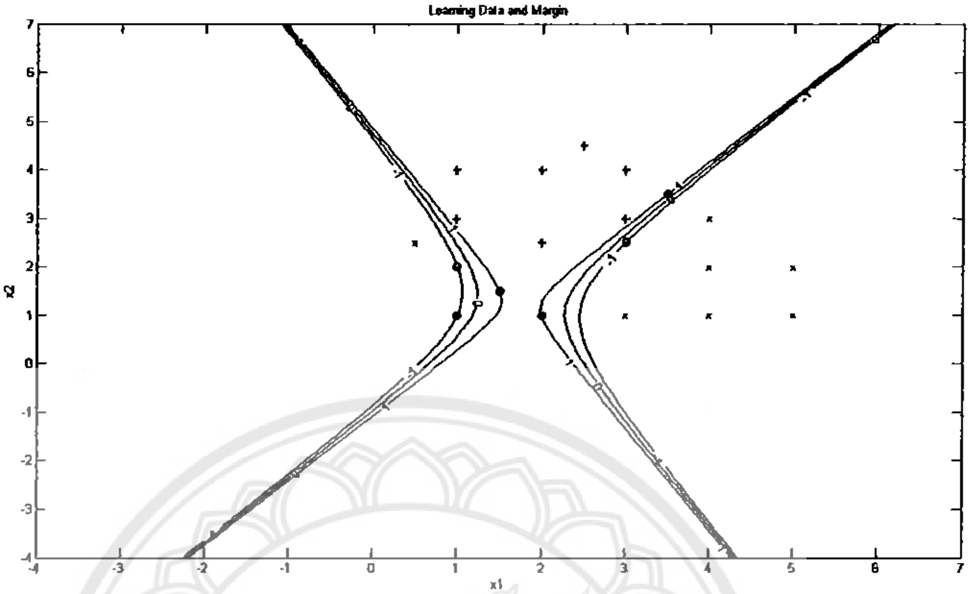
กรณีใช้ค่า d = 1



รูปที่ 4.26 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า d = 1

จากรูป 4.26 รูปนี้ไม่สามารถแบ่งด้วยเส้นตรงได้อย่างสมบูรณ์ จะเห็นได้ว่ามีข้อมูลของกลุ่ม -1 ตกไปอยู่ในกลุ่มของ 1 และมีข้อมูลของกลุ่ม 1 ตกไปอยู่ในกลุ่มของ -1

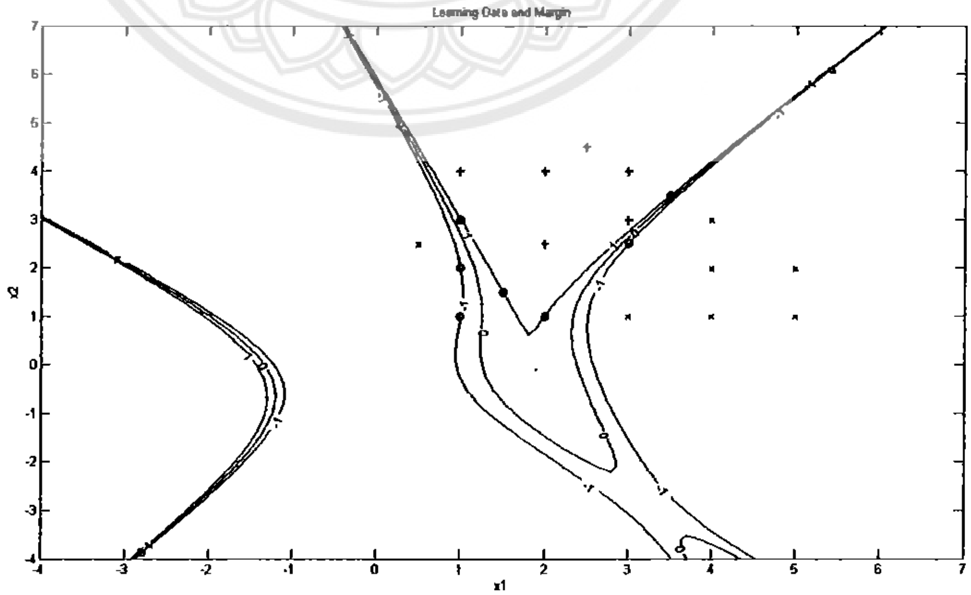
กรณีใช้ค่า d = 2



รูปที่ 4.27 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า d = 2

โดยเมื่อเราเพิ่มค่าของ d เข้าไปในโปรแกรม จะทำให้ฟังก์ชันเคอร์เนลมีความซับซ้อนขึ้น และทำให้เส้นแบ่งมีความซับซ้อนขึ้นและ สามารถแบ่งข้อมูลได้ถูกต้อง

กรณีใช้ค่า d = 3



รูปที่ 4.28 แสดงการแบ่งกลุ่มข้อมูลโดยใช้ค่า d = 3

เมื่อใช้ d = 3 ความกว้างของฟังก์ชันเคอร์เนลมีมากขึ้น จึงทำให้เส้นแบ่งมีขนาดใหญ่ขึ้น แต่ก็ยังสามารถแบ่งกลุ่มข้อมูลได้อย่างถูกต้องเหมือนเดิม

บทที่ 5
สรุปผลการแบ่งกลุ่มข้อมูลด้วยวิธีเอสวีเอ็ม

ในบทนี้เป็นการสรุปผลการแบ่งกลุ่มข้อมูลด้วยวิธีเอ็สวีเอ็ม เป็นการทดลองโดยเปลี่ยนเคอร์เนลฟังก์ชัน และพารามิเตอร์ที่ใช้ในเคอร์เนลฟังก์ชันทั้ง 2 ประเภทนั่นคือเคอร์เนลฟังก์ชันแบบเกาส์เซียน และแบบโพลิโนเมียลในการทดลองแบ่งกลุ่ม ในการทดลองจะมีการเปลี่ยนค่าของพารามิเตอร์ เพื่อดูผลที่เกิดขึ้นในการสร้างเส้นแบ่งแยก

5.1 การใช้เคอร์เนลฟังก์ชันแบบเกาส์เซียน

กรณีแรกที่สามารถแบ่งข้อมูลได้ด้วยเส้นตรง ผลที่ออกมาโปรแกรมสามารถแบ่งได้โดยไม่มีข้อผิดพลาด นอกจากนั้นเมื่อเราเพิ่มค่าของ σ ในสมการเกาส์เซียนให้เพิ่มมากขึ้น ขอบเขตการของเส้นแบ่งแยกก็จะมีมากขึ้นและโปรแกรมสามารถแบ่งแยกข้อมูลได้อย่างถูกต้องในทุกกรณี

กรณีที่ข้อมูลไม่สามารถแบ่งได้ด้วยเส้นตรง ผลที่ออกมาโปรแกรมสามารถแบ่งได้โดยไม่มีข้อผิดพลาดและเมื่อเราเพิ่มค่าของ σ ในฟังก์ชันเกาส์เซียนให้เพิ่มมากขึ้น ขอบเขตของเส้นการแบ่งแยกทำงานของสมการก็จะมีมากขึ้น

กรณีการแบ่งกลุ่มเมื่อเพิ่มจำนวนของข้อมูล เมื่อทำการเพิ่มจำนวนของข้อมูลแล้ว โปรแกรมสามรถแบ่งกลุ่มได้โดยไม่มีข้อผิดพลาดทั้งในกรณของแบ่งได้ด้วยเส้นตรงและ แบ่งไม่ได้ด้วยเส้นตรง

5.2 การใช้เคอร์เนลฟังก์ชันแบบโพลิโนเมียล

กรณีที่ข้อมูลสามารถแบ่งได้ด้วยเส้นตรง ผลที่ออกมาโปรแกรมสามารถแบ่งได้โดยไม่มีข้อผิดพลาด โดยเมื่อเราเพิ่มค่าของ d ในสมการโพลิโนเมียลให้มีค่าเพิ่มมากขึ้น ขอบเขตการทำงานของสมการก็จะมีมากขึ้น โปรแกรมของเราก็จะมีการทำงานที่ดีมากขึ้น เพราะขอบเขตที่ใช้ในการแบ่งกลุ่มก็จะซับซ้อน

กรณีที่ไม่สามารถแบ่งได้ด้วยเส้นตรง ผลที่ออกมาโปรแกรมไม่สามารถแบ่งข้อมูลได้อย่างถูกต้องในกรณีให้ค่า $d = 1$ เนื่องจากเส้นแบ่งแยกถูกสร้างในลักษณะเส้นตรง ซึ่งทำให้มีข้อมูลบางตัวแสดงผลผิดกลุ่ม ซึ่งสามารถแก้ปัญหาการแสดงผลของข้อมูลผิดกลุ่ม โดยทำการเพิ่มค่าของ d ในสมการโพลิโนเมียล ทำให้ขอบเขตการทำงานและความซับซ้อนของเส้นแบ่งแยกมีมากขึ้น ดังนั้นโปรแกรมจึงสามารถทำการแบ่งข้อมูลได้อย่างถูกต้อง

กรณีการแบ่งกลุ่มเมื่อเพิ่มจำนวนของข้อมูล เมื่อทำการเพิ่มจำนวนของข้อมูลแล้ว โปรแกรมสามารถแบ่งกลุ่มได้โดยไม่มีข้อผิดพลาดในกรณีแบ่งได้ด้วยเส้นตรง แต่ในกรณีไม่สามารถแบ่งได้

ด้วยเส้นตรง โปรแกรมไม่สามารถแบ่งข้อมูลได้อย่างถูกต้องในกรณีให้ค่า d = 1 แต่เมื่อทำการเพิ่มค่า d ข้อมูลสามารถแบ่งได้อย่างถูกต้อง

5.3 ข้อเสนอแนะ

5.3.1 ควรประยุกต์ใช้โปรแกรมกับข้อมูลจริง เช่น แบ่งแยกดีเอ็นเอของผู้ป่วยมะเร็ง

5.3.2 ควรจะมีการทดลองใช้โปรแกรมกับข้อมูลจำนวนมากๆ

เอกสารอ้างอิง

[1] ลัญฉกร วุฒิสิทธิกุลกิจ และคณะ. (2549).“**การใช้งานโปรแกรม Matlab เบื้องต้น**”, พิมพ์ครั้งที่ 3, สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2551.

[2] C. Cortes, and V. Vapnik. 1995. “**Support Vector Networks. Machine Learning**”, 20:273-297.

[3] Steve Gunn. “**Support Vector Machine for Classification and Regression**”, ISIS Technical Report, University of Southampton , UK , 1998.