

การใช้พารามิเตอร์ที่สร้างจากภาคตัดกรวยในการจำแนกเลขที่บ้านจากภูเกิลสตรีตวิด้วยวิธีซัพพอร์ต  
เวกเตอร์แมชชีน



นายพงศ์ศรัณย์ เทียมตะวัน

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.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

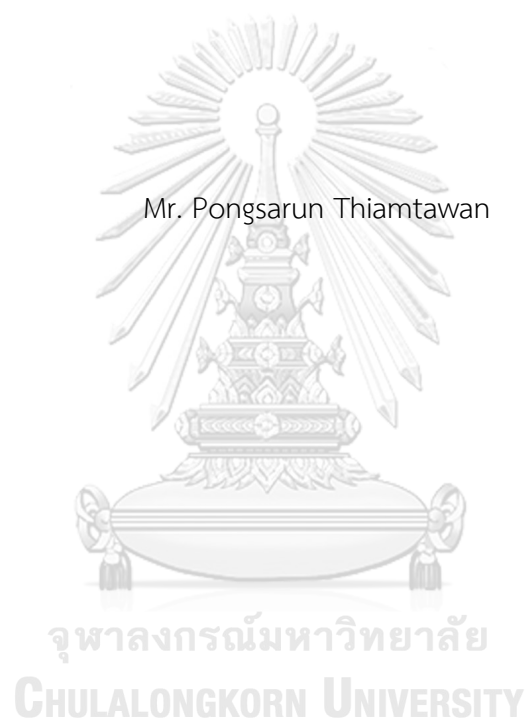
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2560

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

STREET VIEW HOUSE NUMBERS SUPPORT VECTOR MACHINE CLASSIFICATION USING  
PARAMETERS DERIVED FROM CONIC SECTIONS

Mr. Pongsarun Thiamtawan



A Thesis Submitted in Partial Fulfillment of the Requirements  
for the Degree of Master of Science Program in Statistics

Department of Statistics

Faculty of Commerce and Accountancy

Chulalongkorn University

Academic Year 2017

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์

การใช้พารามิเตอร์ที่สร้างจากภาคตัดกรวยในการจำแนก  
เลขที่บ้านจากกุเกิลสตรีตวิวด้วยวิธีซัพพอร์ตเวกเตอร์แม  
ชชีน

โดย

นายพงศ์ศรัณย์ เทียมตะวัน

สาขาวิชา

สถิติ

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

อาจารย์ ดร.อัครินทร์ ไพบุลย์พานิช

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้รับวิทยานิพนธ์  
ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

บัญชี

.....คณบดีคณะพาณิชยศาสตร์และการ  
(รองศาสตราจารย์ ดร.พสุ เดชะรินทร์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ  
(รองศาสตราจารย์ ดร.เสกสรร เกียรติสุไพบุลย์)  
.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก  
(อาจารย์ ดร.อัครินทร์ ไพบุลย์พานิช)

.....กรรมการ  
(อาจารย์ ดร.นันท กุลวานิช)

.....กรรมการภายนอกมหาวิทยาลัย  
(อาจารย์ ดร.อรุณี กำลั้ง)

พงศ์ศรีณย์ เทียมตะวัน : การใช้พารามิเตอร์ที่สร้างจากภาคตัดกรวยในการจำแนกเลขที่บ้านจากภูเก็ลสตรีตวิวด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน (STREET VIEW HOUSE NUMBERS SUPPORT VECTOR MACHINE CLASSIFICATION USING PARAMETERS DERIVED FROM CONIC SECTIONS) อ.ที่ปริกษาวิทยานิพนธ์หลัก: อ. ดร.อักรินทร์ ไพบูลย์พานิช, หน้า.

การรู้จำตัวเลขเป็นปัญหาสำคัญหนึ่งในสาขาคอมพิวเตอร์วิทัศน์และการรู้จำแบบ งานวิจัยนี้ทำการจำแนกตัวเลขจากฐานข้อมูลเลขที่บ้านจากภูเก็ล สตรีตวิว ซึ่งมีความสำคัญต่อการปรับปรุงความถูกต้องของแผนที่และระบบนำทางในปัจจุบัน (Netzer และคณะ, 2012) โดยใช้พารามิเตอร์ที่ได้จากลักษณะเชิงเรขาคณิตวิเคราะห์ของตัวเลขในการสร้างตัวแบบ และใช้ซัพพอร์ตเวกเตอร์แมชชีนในการจำแนกตัวเลข โดยเปรียบเทียบผลลัพธ์การจำแนกตัวเลขระหว่าง 4 เคอร์เนล ได้แก่ เคอร์เนลเชิงเส้น, เคอร์เนลพหุนาม, เคอร์เนลเรเดียล และเคอร์เนลซิกมอยด์ รวมถึงเปรียบเทียบความสำคัญของแต่ละลักษณะว่ามีผลต่อการจำแนกตัวเลขอย่างไร หลังจากเตรียมข้อมูลรูปภาพก่อนการสร้างตัวแบบ ในการดึงลักษณะจะใช้การแปลงฮัฟ (Hough transform) ในการค้นหารูปร่างภาคตัดกรวยที่พบในรูปขอบของตัวเลข 4 รูปแบบ ได้แก่ เส้นตรง, วงรี, พาราโบลาแนวตั้ง และพาราโบลาแนวนอน พบว่า เคอร์เนลเรเดียลให้ผลลัพธ์การจำแนกตัวเลขดีที่สุด คือ 72.17% และใช้เวลาในการสร้างตัวแบบ 22.36 นาที ลักษณะที่เกี่ยวข้องกับเส้นตรงมีความสำคัญต่อการจำแนกตัวเลขมากที่สุด รองลงมาคือ วงรี พาราโบลาแนวนอน และพาราโบลาแนวตั้ง ตามลำดับ

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

ภาควิชา สถิติ

ลายมือชื่อนิสิต .....

สาขาวิชา สถิติ

ลายมือชื่อ อ.ที่ปริกษาหลัก .....

ปีการศึกษา 2560

# # 5881547426 : MAJOR STATISTICS

KEYWORDS: STREET VIEW HOUSE NUMBERS (SVHN) / HOUGH TRANSFORM / CONIC SECTIONS / SUPPORT VECTOR MACHINE (SVM)

PONGSARUN THIAMTAWAN: STREET VIEW HOUSE NUMBERS SUPPORT VECTOR MACHINE CLASSIFICATION USING PARAMETERS DERIVED FROM CONIC SECTIONS. ADVISOR: AKARIN PHAIBULPANICH, Ph.D., pp.

Digits recognition is a major problem in computer visions and pattern recognition. In this study, we classify digits from the Street View House Numbers dataset (SVHN) which is important for improving map accuracy and navigation system (Netzer et al., 2012). We use parameters derived from digits' analytically geometric characteristics as features and classify them via support vector machine (SVM) comparing the accuracy rate of 4 kernels i.e. linear, polynomial, radial basis function (RBF), and sigmoid. Also, we measure each feature's importance for digit classification. After preprocessing image data, we apply Hough transform to detect 4 conic sections i.e. line, ellipse, vertical parabola, and horizontal parabola from the digit's edge images and extract corresponding parameters as features. We found that the best kernel is RBF yielding 72.17% accuracy rate and 22.36 minutes to train data. The most important features are line's features. The next ones are features of ellipse, horizontal parabola, and vertical parabola, respectively.

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

Department: Statistics

Student's Signature .....

Field of Study: Statistics

Advisor's Signature .....

Academic Year: 2017

## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ เสร็จสมบูรณ์ลงได้ด้วยดี ด้วยความช่วยเหลือและเอาใจใส่จาก อาจารย์ ดร. อัครินทร์ ไพบูลย์พานิช อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผู้วิจัยขอกราบขอบพระคุณ ท่านอาจารย์เป็นอย่างสูง ที่กรุณาสละเวลาให้คำปรึกษา อบรมสั่งสอน ให้คำแนะนำ ตลอดจนให้ความช่วยเหลือในการปรับปรุง แก้ไขวิทยานิพนธ์ และเป็นแบบอย่างที่ดีในการทำงาน จนกระทั่ง วิทยานิพนธ์สำเร็จลุล่วงไปได้ด้วยดี

ผู้วิจัยขอกราบขอบพระคุณท่านรองศาสตราจารย์ ดร. เสกสรร เกียรติสุไพบูลย์ ประธานกรรมการสอบวิทยานิพนธ์ อาจารย์ ดร. นันท กุลวานิช และอาจารย์ ดร. อรุณี กำลัง กรรมการสอบ วิทยานิพนธ์เป็นอย่างสูง ที่ท่านอาจารย์ทั้งสามท่านได้สละเวลาเพื่อตรวจสอบและให้คำแนะนำ เพื่อแก้ไขวิทยานิพนธ์ฉบับนี้ให้มีความสมบูรณ์มากยิ่งขึ้น อีกทั้งขอกราบขอบพระคุณ ผู้ช่วย ศาสตราจารย์ ดร. บุรชัย ภัทรโกศล ที่เอื้อเฟื้อระบบปฏิบัติการในการประมวลผลข้อมูล และผู้ช่วย ศาสตราจารย์ ดร. เกริก ภริมย์โสภาก ที่ให้คำปรึกษาในการจัดทำวิทยานิพนธ์ฉบับนี้

สุดท้ายนี้ ผู้วิจัยขอกราบขอบพระคุณครอบครัว ที่คอยให้กำลังใจ ห่วงใย ส่งเสริม และ สนับสนุนมาโดยตลอด ขอขอบคุณรุ่นพี่ เพื่อน ๆ และรุ่นน้องทุกคน ที่คอยช่วยเหลือ ให้คำแนะนำ และคอยให้กำลังใจแก่ผู้วิจัยเสมอมา

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

## สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฌ
สารบัญภาพ .....	ญ
บทที่ 1 บทนำ .....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย .....	3
1.3 สมมติฐานของการวิจัย.....	4
1.4 คำจำกัดความที่ใช้ในการวิจัย .....	4
1.5 ขอบเขตของการวิจัย.....	5
1.6 ประโยชน์ที่คาดว่าจะได้รับ .....	6
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง .....	7
2.1 การแบ่งกลุ่มข้อมูลแบบเคมีน (K-means clustering).....	7
2.2 การกำหนดสัญลักษณ์การเชื่อมต่อขององค์ประกอบ (Connected-components Labeling) .....	8
2.3 การแปลงฮัฟ (Hough Transform).....	11
2.3.1 เส้นตรง (Line).....	12
2.3.2 วงกลม (Circle).....	17
2.3.3 วงรี (Ellipse).....	20
2.3.4 พาราโบลา (Parabola).....	21
2.4 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines; SVM).....	23

บทที่ 3 วิธีการดำเนินงานวิจัย .....	30
3.1 แหล่งข้อมูล .....	30
3.2 ขั้นตอนในการดำเนินการวิจัย.....	32
3.2.1 กระบวนการเตรียมข้อมูลก่อนการสร้างตัวแบบ (Preprocessing data).....	32
3.2.2 กระบวนการคัดเลือกรูปภาพตัวเลข.....	34
3.2.3 กระบวนการดึงลักษณะ (Features extraction).....	37
3.2.4 กระบวนการสร้างตัวแบบจำแนกตัวเลขและทดสอบประสิทธิภาพของตัวแบบ .....	43
3.2.5 กระบวนการวัดความสำคัญของลักษณะ.....	43
3.3 ขั้นตอนการทำงานของโปรแกรม.....	45
บทที่ 4 ผลการวิจัย.....	47
4.1 ผลการปรับค่าพารามิเตอร์ของตัวแบบ SVM ทั้ง 4 เคอร์เนลด้วยวิธี 5-fold cross validation .....	47
4.2 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบ SVM ในการจำแนกตัวเลข.....	47
4.3 ผลการเปรียบเทียบความสำคัญของลักษณะที่มีผลต่อการจำแนกตัวเลข .....	49
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ .....	54
5.1 สรุปผลการวิจัย.....	54
5.2 ข้อเสนอแนะ .....	55
รายการอ้างอิง .....	57
ประวัติผู้เขียนวิทยานิพนธ์ .....	59



## สารบัญตาราง

ตารางที่ 3.1 ตัวอย่างค่าของตัวแปรที่ใช้ในการสร้างตัวแบบความถดถอยโลจิสติกทวิของตัวอย่าง รูปตัวเลขในภาพที่ 3.8 .....	36
ตารางที่ 3.2 จำนวนรูปภาพตัวเลขหลังจากการคัดเลือกแล้ว แยกตามแต่ละตัวเลข.....	37
ตารางที่ 3.3 พารามิเตอร์ที่ใช้อธิบายลักษณะของรูปตัวเลขแต่ละรูป .....	41
ตารางที่ 3.4 ตัวอย่างค่าพารามิเตอร์จากรูปภาคตัดกรวยของรูปตัวเลขในภาพที่ 3.12 .....	42
ตารางที่ 4.1 ค่าพารามิเตอร์ที่ดีที่สุดของตัวแบบ SVM ทั้ง 4 เคอร์เนล.....	47
ตารางที่ 4.2 อัตราการจำแนกถูกต้องและเวลาที่ใช้ในการสร้างตัวแบบของ SVM ทั้ง 4 เคอร์เนล ...	48
ตารางที่ 4.3 Confusion Matrix สำหรับเคอร์เนลเรเดียล.....	48
ตารางที่ 4.4 อัตราการจำแนกถูกต้องแยกตามแต่ละตัวเลข สำหรับเคอร์เนลเรเดียล .....	49
ตารางที่ 4.5 ลำดับความสำคัญของลักษณะ .....	51
ตารางที่ 4.6 ลำดับความสำคัญของรูปภาคตัดกรวยแต่ละรูป.....	52
ตารางที่ 4.7 ลำดับของลักษณะที่สำคัญที่สุด 3 อันดับแรกแยกตามแต่ละตัวเลข .....	52
ตารางที่ 4.8 ภาพตัวอย่างตัวเลขที่ได้รับผลกระทบจากการค้นหาขอบ .....	53

## สารบัญภาพ

ภาพที่ 1.1 ตัวอย่างรูปภาพตัวเลขจากฐานข้อมูล MNIST .....	2
ภาพที่ 1.2 ตัวอย่างรูปภาพตัวเลขที่บ้านจากกูเกิลสตรีตวิว.....	2
ภาพที่ 1.3 ภาพตัวอย่างรูปภาพตัดกรวยที่สามารถพบได้ในรูปตัวเลข.....	3
ภาพที่ 2.1 แผนภาพแสดงขั้นตอนการรู้จำตัวอักษรหรือตัวเลข.....	7
ภาพที่ 2.2 ตัวอย่างรูปภาพ binary.....	9
ภาพที่ 2.3 รูปภาพตามภาพที่ 2.2 หลังจากทำ connected-components labeling แล้ว.....	9
ภาพที่ 2.4 ตัวอย่างการกำหนดหมายเลข $\Delta$ หมายถึงพิกเซลที่ถูกกำหนดหมายเลขแล้ว 1 หมายถึงพิกเซลที่ยังไม่ถูกกำหนดหมายเลข และ $-$ หมายถึงพิกเซลพื้นหลังที่ทำเครื่องหมาย.....	10
ภาพที่ 2.5 ตัวอย่างพิกเซลที่เป็นขอบในขององค์ประกอบ $\Delta$ หมายถึงพิกเซลที่ถูกกำหนดหมายเลขแล้ว 1 หมายถึงพิกเซลที่ยังไม่ถูกกำหนดหมายเลข และ $-$ หมายถึงพิกเซลพื้นหลังที่ทำเครื่องหมาย.....	10
ภาพที่ 2.6 ตัวอย่างการกำหนดหมายเลขพิกเซลที่ไม่ใช่ขอบ $\Delta$ หมายถึงพิกเซลที่ถูกกำหนดหมายเลขแล้ว 1 หมายถึงพิกเซลที่ยังไม่ถูกกำหนดหมายเลข และ $-$ หมายถึงพิกเซลพื้นหลังที่ทำเครื่องหมาย.....	11
ภาพที่ 2.7 เส้นตรงบนระนาบคาร์ทีเซียนและปริภูมิพารามิเตอร์.....	13
ภาพที่ 2.8 พารามิเตอร์ปกติของเส้นตรง.....	13
ภาพที่ 2.9 เส้นตรงบนระนาบคาร์ทีเซียนและปริภูมิพารามิเตอร์ในรูปปกติ.....	14
ภาพที่ 2.10 ตัวอย่างพารามิเตอร์ปกติที่ค่าต่าง ๆ กันบนรูปภาพดิจิทัล.....	15
ภาพที่ 2.11 รูปเส้นตรงที่ $\rho$ เท่ากับขอบของเรนจ์.....	16
ภาพที่ 2.12 ภาพตัวอย่างการแปลงฮัฟสำหรับเส้นตรง.....	17
ภาพที่ 2.13 วงกลมบนระนาบคาร์ทีเซียน.....	18
ภาพที่ 2.14 การแปลงฮัฟสำหรับวงกลม ณ ค่ารัศมีคงที่.....	18
ภาพที่ 2.15 accumulator array 3 มิติ.....	19

ภาพที่ 2.16 ตัวอย่างวงรีบนระนาบคาร์ทีเซียน ที่แกนเอกขนานกับแกน x.....	20
ภาพที่ 2.17 วงรีในรูปพาราเมตริก .....	21
ภาพที่ 2.18 พาราโบลาที่แกนสมมาตรขนานกับแกน x.....	22
ภาพที่ 2.19 ตัวอย่างไฮเปอร์เพลนในปริภูมิ 2 มิติ แสดงด้วยเส้นตรง $1 + 2X_1 + 3X_2 = 0$ พื้นที่เหนือเส้นตรงคือเซตของจุดที่ทำให้ $1 + 2X_1 + 3X_2 > 0$ เป็นจริง ขณะที่พื้นที่ใต้เส้นตรงคือเซตของจุดที่ทำให้ $1 + 2X_1 + 3X_2 < 0$ เป็นจริง.....	23
ภาพที่ 2.20 ตัวอย่างของการใช้ไฮเปอร์เพลนในการแบ่งประเภทข้อมูล กรณีที่มีตัวแปรอิสระ 2 ตัว และมี 2 class แต่ละฝั่งของไฮเปอร์เพลนที่ใช้แบ่งข้อมูลจะแทนแต่ละ class .....	25
ภาพที่ 2.21 ตัวอย่างไฮเปอร์เพลนที่สามารถแบ่งประเภทข้อมูลได้.....	25
ภาพที่ 2.22 ตัวอย่างของ maximal margin classifier โดย optimal hyperplane แสดงด้วยเส้นประ ซัพพอร์ตเวกเตอร์แสดงด้วยข้อมูลที่อยู่ในกรอบสี่เหลี่ยมสีเทา ระยะห่างระหว่างไฮเปอร์เพลนที่ทับซัพพอร์ตเวกเตอร์ (เส้นทึบ) ของแต่ละ class คือมาร์จิน (แสดงด้วยลูกศร).....	26
ภาพที่ 2.23 ตัวอย่างของ support vector classifier.....	26
ภาพที่ 2.24 เปรียบเทียบตัวแบ่งประเภท maximal margin classifier กับ support vector classifier .....	27
ภาพที่ 3.1 ตัวอย่างรูปภาพ 2 รูปแบบ.....	30
ภาพที่ 3.2 การแปลงจากภาพสีเป็นภาพขาวดำ.....	32
ภาพที่ 3.3 รูปภาพตัวเลขหลังจากการทำ K-means clustering.....	33
ภาพที่ 3.4 รูปภาพตัวเลขหลังจากการกำหนดสัญลักษณ์การเชื่อมต่อขององค์ประกอบ.....	33
ภาพที่ 3.5 ลักษณะการกระจายของตัวแปรน้ำหนักรในแต่ละพิกเซล .....	33
ภาพที่ 3.6 รูปภาพตัวเลขหลังจากกำหนดน้ำหนักแล้ว .....	34
ภาพที่ 3.7 ภาพตัวอย่างรูปภาพบางส่วนหลังจากกำจัด noise และแปลงเป็นรูปภาพ binary .....	34
ภาพที่ 3.8 ตัวอย่างรูปภาพพร้อมกับเส้นแสดงพิกเซล .....	35
ภาพที่ 3.9 ตัวอย่างรูปภาพตัวเลขหลังค้นหาขอบ .....	37
ภาพที่ 3.10 ตัวอย่างการคำนวณหาความยาวเส้นตรงแนวเฉียง .....	38

ภาพที่ 3.11 พารามิเตอร์ของวงรีที่นิยามใหม่ ..... 39

ภาพที่ 3.12 ตัวอย่างรูปตัวเลขสำหรับสาธิตตัวอย่างค่าพารามิเตอร์จากภาคตัดกรวย..... 41

ภาพที่ 4.1 ภาพตัวอย่างการจำแนกชนิด..... 49



# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

การรู้จำตัวเลข (digits recognition) เป็นปัญหาสำคัญหนึ่งของคอมพิวเตอร์วิทัศน์ (computer vision) และการรู้จำแบบ (pattern recognition) ซึ่งมีการประยุกต์ใช้ที่หลากหลาย เช่น การรู้จำเลขทะเบียนรถยนต์ในภาพจากเครื่องบันทึกวิดีโอ การทดสอบประสิทธิภาพของ CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) การทดสอบผู้ใช้งานว่าเป็นมนุษย์จริงหรือไม่ การแปลงภาพถ่ายของเอกสารให้อยู่ในรูปแบบที่คอมพิวเตอร์สามารถค้นหาคำต่าง ๆ ได้ เป็นต้น ในการจำแนกตัวเลขว่าเป็นตัวเลขอะไรนั้น ลักษณะ (feature) ที่จะนำไปใช้สร้างตัวแบบจำแนกตัวเลขไม่จำเป็นต้องเป็นค่าของสี ซึ่งโดยปกติแล้วเป็นข้อมูลพื้นฐานของแต่ละพิกเซล (pixel) ในรูปภาพดิจิทัล ด้วยความหลากหลายของลักษณะในรูปภาพต่าง ๆ ส่งผลให้กระบวนการดึงลักษณะ (feature extraction) และตัวแบบที่ใช้จำแนกประเภท (classifier) ถือเป็นปัจจัยสำคัญที่มีผลต่อผลลัพธ์ของการรู้จำและจำแนกประเภท (Patel, Jagtap, & Kale, 2014)

ในช่วง 30 ปีที่ผ่านมา งานวิจัยเกี่ยวกับการจำแนกตัวเลขเพิ่มขึ้นอย่างต่อเนื่อง ทั้งในส่วนของลักษณะและตัวแบบที่ใช้ในการจำแนกตัวเลข เช่น Lecun, Bottou, Bengio, and Haffner (1998) เสนอการรู้จำตัวเลขที่เขียนด้วยลายมือจากฐานข้อมูล MNIST (Modified National Institute of Standards and Technology) โดยประยุกต์ใช้เทคนิคการเรียนรู้ที่ใช้เกรเดียนต์ในการหาค่าน้อยที่สุด (gradient-based learning) เพื่อประมาณค่าพารามิเตอร์ที่ทำให้ฟังก์ชันความสูญเสีย (loss function) มีค่าน้อยที่สุดด้วยวิธีการลดระดับเกรเดียนต์ (gradient descent algorithm) ซึ่ง LeCun และคณะทดลองใช้ตัวแบบจำแนกประเภทหลายแบบเปรียบเทียบกัน พบว่า โครงข่ายประสาทเทียมแบบสังวัตนาการ (convolutional neural network) ให้ผลลัพธ์ที่ดีที่สุด

ฐานข้อมูล MNIST ได้รับความนิยมจากนักวิจัยในการศึกษาพัฒนาตัวแบบจำแนกตัวเลขเป็นจำนวนมาก รวมถึง Giuliodori, Lillo, and Peña (2011) ใช้คุณลักษณะเฉพาะเชิงโครงสร้างของตัวเลขมาใช้เป็นลักษณะ เช่น ตำแหน่งและความยาวของเส้นตรง ตำแหน่งและรัศมีของวงกลม (2 ลักษณะดังกล่าวคำนวณโดยใช้การแปลงฮัฟ (Hough transform)) จำนวนรู (Holes) ตำแหน่งสุดขอบ (Extremes) เป็นต้น โดยเปรียบเทียบวิธีเชิงความน่าจะเป็นซึ่งอิงจากกฎของเบส์กับวิธี k-

nearest neighbor ในการจำแนกตัวเลข ซึ่ง Giuliodori และคณะ ใช้ข้อมูลทั้งจาก MNIST และ USPS (United State Postal Service ฐานข้อมูลตัวเลขที่เขียนด้วยลายมือบนซองจดหมายในสหรัฐอเมริกา) ตัวอย่างรูปภาพตัวเลขที่เขียนด้วยลายมือจากฐานข้อมูล MNIST แสดงดังภาพที่ 1.1

3 6 8 1 7 9 6 6 9 1 7 5 9 2 6 5 8 1 9 7  
 6 7 5 7 8 6 3 4 8 5 2 2 2 2 2 3 4 4 8 0  
 2 1 7 9 7 1 2 8 4 5 0 2 3 8 0 7 3 8 5 7  
 4 8 1 9 0 1 8 8 9 4 0 1 4 6 4 6 0 2 4 3  
 7 6 1 8 6 4 1 5 6 0 7 1 2 8 7 6 9 8 6 1

ภาพที่ 1.1 ตัวอย่างรูปภาพตัวเลขจากฐานข้อมูล MNIST

ที่มา: LeCun และคณะ (1998)

นอกจากฐานข้อมูล MNIST แล้ว ฐานข้อมูลรูปภาพตัวเลขที่กำลังได้รับความนิยมในหมู่นักวิจัยคือ ฐานข้อมูลเลขที่บ้านกูเกิล สตรีตวิว (The Street View House Number; SVHN) (Netzer et al., 2012) ซึ่งเป็นภาพถ่ายเลขที่บ้านจากสถานที่ต่าง ๆ ทั่วโลกด้วยเทคโนโลยีกูเกิล สตรีตวิว ของกูเกิล แสดงตัวอย่างดังภาพที่ 1.2



ภาพที่ 1.2 ตัวอย่างรูปภาพตัวเลขที่บ้านจากกูเกิลสตรีตวิว

ที่มา: ฐานข้อมูล SVHN

จากภาพที่ 1.1 และ 1.2 จะเห็นว่า แม้จะเป็นรูปภาพตัวเลขเหมือนกัน แต่ลักษณะเฉพาะแตกต่างกัน เช่น สีของตัวเลขจากฐานข้อมูล MNIST จะมีเพียงสีดำ และพื้นหลังเป็นสีขาวเท่านั้น ขณะที่ตัวเลขจากฐานข้อมูล SVHN อาจมีสีที่หลากหลายแตกต่างกันไป พื้นหลังก็จะมีสีแตกต่างกันไปตามแต่ละสถานที่ ขนาดและรูปร่างของตัวเลข MNIST จะค่อนข้างใกล้เคียงกัน ขณะที่ตัวเลขของ SVHN จะมีทั้งตัวบาง ตัวหนา นอกจากนี้ยังมีปัจจัยอื่นที่ทำให้การจำแนกตัวเลข SVHN ยากกว่าตัวเลขของ MNIST เช่น ความสว่างและความคมชัดที่แตกต่างกัน เป็นต้น การจำแนกตัวเลข SVHN ในปัจจุบันจึงมักจะเน้นการประยุกต์ใช้สถาปัตยกรรมการเรียนรู้เชิงลึก (deep learning architecture) เช่น โครงข่ายประสาทเทียมแบบสังวัตนาการ, โครงข่ายความเชื่อแบบลึก (deep belief networks) เป็นต้น ซึ่งให้ผลลัพธ์อัตราการจำแนกถูกต้องสูง อย่างไรก็ตาม การใช้สถาปัตยกรรมการเรียนรู้เชิงลึกมีข้อเสียคือ มีต้นทุนสูง ทั้งในแง่ของอุปกรณ์และเวลาในการประมวลผล นอกจากนี้ ไม่สามารถทราบได้ว่า ลักษณะใดของตัวเลขที่มีความสำคัญต่อการจำแนก

Netzer et al. (2012) ใช้ข้อมูลตัวเลขของ SVHN โดยที่ไม่ได้ใช้การเรียนรู้เชิงลึกในการจำแนกตัวเลข แต่จะเปรียบเทียบลักษณะ 4 แบบ โดยแบ่งเป็น hand crafted feature ได้แก่ Histograms of Oriented Gradients และ Weighted Direction Code Histogram กับ learned feature ได้แก่ Stacked Sparse Auto-Encoders และ K-means-based features โดยใช้ซัพพอร์ตเวกเตอร์แมชชีนในการจำแนกตัวเลข และไม่มีการวัดความสำคัญของลักษณะแต่ละตัวที่ใช้ในการจำแนกตัวเลข

ลักษณะของตัวเลขที่เห็นได้ชัดเจนรูปแบบหนึ่ง คือ ลักษณะเชิงเรขาคณิตวิเคราะห์ที่เป็นเอกลักษณ์เฉพาะตัว เช่น เลขศูนย์มักจะเป็นวงรี 1 วง, เลขหนึ่งมักจะเป็นเส้นตรง 1 เส้น, เลขสามมักจะมีประกอด้วยพาราโบลาแนวนอน 2 เส้น เป็นต้น ตัวเลขแต่ละตัวประกอบด้วยรูปร่างที่ตัดกรวยที่มีเอกลักษณ์เฉพาะ แสดงตัวอย่างดังภาพที่ 1.3



ภาพที่ 1.3 ภาพตัวอย่างรูปร่างที่ตัดกรวยที่สามารถพบได้ในรูปตัวเลข

งานวิจัยนี้ จะทดสอบประสิทธิภาพการจำแนกตัวเลขจากฐานข้อมูล SVHN โดยใช้รูปร่างที่ตัดกรวย ได้แก่ เส้นตรง วงกลม วงรี และพาราโบลา เป็นลักษณะ โดยประยุกต์ใช้การแปลงฮัฟ (Hough transform) ซึ่งเป็นเทคนิคในการค้นหารูปร่างต่าง ๆ ที่อยู่ในรูปภาพดิจิทัลวิธีหนึ่งในการดึงลักษณะ และใช้ซัพพอร์ตเวกเตอร์แมชชีน (SVM) ในการจำแนกตัวเลข อีกทั้งเปรียบเทียบลักษณะดังกล่าวแต่ละตัวในเชิงผลกระทบที่มีต่อการจำแนกตัวเลข

## 1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อพัฒนาลักษณะที่สร้างจากพารามิเตอร์ที่เกี่ยวข้องกับรูปร่างที่ตัดกรวยที่ปรากฏในรูปภาพตัวเลขจากฐานข้อมูล SVHN

1.2.2 เพื่อเปรียบเทียบผลลัพธ์การจำแนกตัวเลข SVHN ซึ่งใช้ลักษณะที่ได้จากการแปลงฮัฟด้วยซัพพอร์ตเวกเตอร์แมชชีน ระหว่าง 4 เคอร์เนล ได้แก่ เคอร์เนลเชิงเส้น, เคอร์เนลพหุนาม, เคอร์เนลเรเดียล และเคอร์เนลซิกมอยด์

1.2.3 เพื่อเปรียบเทียบความสำคัญของลักษณะที่ได้จากการแปลงฮัฟซึ่งใช้ในการจำแนกตัวเลขจากฐานข้อมูล SVHN

### 1.3 สมมติฐานของการวิจัย

ซีพพอร์ทเวกเตอร์แมชชีน เคอร์เนลเรเดียล ให้ผลลัพธ์การจำแนกตัวเลขสูงที่สุดเมื่อเทียบกับเคอร์เนลอื่น กล่าวคือ ลักษณะที่ได้จากการแปลงฮัฟมีความไม่เป็นเชิงเส้น

### 1.4 คำจำกัดความที่ใช้ในการวิจัย

1.4.1 รูปภาพแบบ binary (binary image) หมายถึง รูปภาพที่แต่ละพิกเซลแสดงด้วยค่า 0 หรือ 1 เท่านั้น

1.4.2 รูปร่างตัดกรวย ในที่นี้ หมายถึงรูปร่างตัดกรวยที่เกี่ยวข้องกับรูปตัวเลข ได้แก่ เส้นตรง วงกลม วงรี และพาราโบลา

1.4.3 การเตรียมข้อมูลก่อนสำหรับข้อมูลรูปภาพ (image preprocessing) หมายถึงการแปลงใด ๆ ให้กับข้อมูลพิกเซลก่อนที่จะทำการสร้างตัวแบบ (Prince, 2012)

1.4.4 ลักษณะ (feature) หมายถึง คุณลักษณะหรือคุณสมบัติของสิ่งที่เรากำลังสังเกตหรือศึกษาที่สามารถวัดได้

1.4.5 การค้นหาขอบ (edge detection) หมายถึง กระบวนการแปลงรูปภาพดิจิทัลให้กลายเป็นรูปภาพ binary โดยพิกเซลที่มีค่าเป็น 1 แสดงถึงขอบในรูปภาพนั้น (Prince, 2012)

1.4.6 Sobel operator หมายถึง วิธีการหนึ่งในการค้นหาขอบโดยการคำนวณค่าประมาณของการเปลี่ยนแปลงหรือเกรเดียนต์ของค่าสีของพิกเซลเทียบกับพิกเซลที่อยู่ติดกัน ทั้งในแนวขวางและแนวตั้ง โดย Sobel operator ประกอบด้วย 2 เคอร์เนล คือ  $M_x$  และ  $M_y$  สำหรับคำนวณเกรเดียนต์ในแนวขวางและแนวตั้ง ตามลำดับ โดย

$$M_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \quad \text{และ} \quad M_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

การคำนวณค่าประมาณของการเปลี่ยนแปลงของเกรเดียนต์ จะคำนวณจากผลประสาน (convolution) ของ Sobel operator กับรูปภาพดิจิทัล



## 1.5 ขอบเขตของการวิจัย

1.5.1 ข้อมูลรูปภาพตัวเลขนำมาจากฐานข้อมูลภาพเลขที่บ้านจากภูเก็ล สตรีตวิว (Street View House Numbers; SVHN) (Netzer et al., 2012) โดยจะใช้รูปภาพแบบ cropped digits เท่านั้น ซึ่งแต่ละภาพจะแสดงตัวเลขเพียงหนึ่งตัว ขนาด 32 x 32 พิกเซล

1.5.2 การกำหนดสัญลักษณ์การเชื่อมต่อขององค์ประกอบ (connected components labeling) ในขั้นตอนเตรียมข้อมูลก่อนจะใช้วิธี single pass (Chang, Chen, & Lu, 2004)

1.5.3 รูปตัวเลขที่จะใช้ในการสร้างตัวแบบและทดสอบนั้น จะทำการคัดเลือกหลังจากขั้นตอนการเตรียมข้อมูลรูปภาพ ด้วยการถดถอยโลจิสติกหิว โดยใช้เกณฑ์การเลือกรูปภาพ ความน่าจะเป็นมากกว่าหรือเท่ากับ 0.9

1.5.4 ก่อนทำการดึงลักษณะ จะค้นหาขอบของรูปตัวเลข ด้วย Sobel operator

1.5.5 ลักษณะของข้อมูลที่จะใช้เป็นตัวแปรอิสระในการจำแนกตัวเลขจะใช้พารามิเตอร์ที่สร้างจากรูปภาพตัดกรวย ได้แก่ เส้นตรง วงรี และพาราโบลา โดยใช้การแปลงฮัฟในการค้นหาค่าพารามิเตอร์ของลักษณะต่าง ๆ เหล่านี้จากรูปภาพตัวเลข (เนื่องจากรูปร่างกลมเป็นกรณีเฉพาะของรูปวงรี ดังนั้น งานวิจัยนี้จะค้นหาเฉพาะรูปวงรี)

1.5.6 ใช้ซัพพอร์ตเวกเตอร์แมชชีน (SVM) เป็นตัวแบบในการจำแนกตัวเลข โดยเปรียบเทียบ 4 เคอร์เนล ได้แก่

ก) เคอร์เนลเชิงเส้น

ข) เคอร์เนลพหุนาม เปรียบเทียบที่ดีกรี ( $d$ ) เท่ากับ 5, 10 และ 20

ค) เคอร์เนลเรเดียล เปรียบเทียบที่แกมมา ( $\gamma$ ) เท่ากับ 0.01, 0.1 และ 1

ง) เคอร์เนลซิกมอยด์ เปรียบเทียบที่แคปปา ( $\kappa$ ) เท่ากับ 0.01, 0.1 และ 1

สำหรับพารามิเตอร์ cost จะเปรียบเทียบที่  $C$  เท่ากับ 1, 10 และ 100 โดยใช้ 5-fold cross validation และเนื่องจากข้อมูลนี้มีทั้งหมด 10 class ในที่นี้จะใช้ SVM วิธี one-versus-one

## 1.6 ประโยชน์ที่คาดว่าจะได้รับ

1.6.1 เพื่อนำไปใช้ประโยชน์ในการจำแนกตัวเลขจากภาพถ่ายหรือตัวเลขที่เขียนด้วยลายมือ

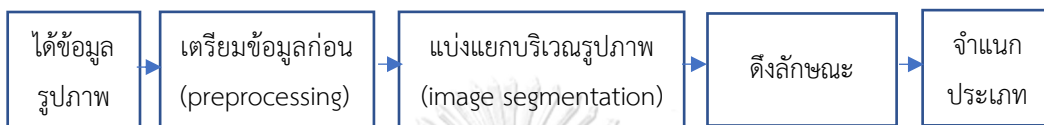
1.6.2 เพื่อเป็นแนวทางในการศึกษาวิจัยในการเพิ่มประสิทธิภาพของการรู้จำตัวเลขจากภาพถ่าย



## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ระบบการรู้จำรูปภาพตัวอักษรหรือตัวเลขโดยทั่วไปสามารถเขียนแผนภาพได้ดังนี้ (Patel et al., 2014)



ภาพที่ 2.1 แผนภาพแสดงขั้นตอนการรู้จำตัวอักษรหรือตัวเลข  
ที่มา: Patel et al. (2014)

งานวิจัยนี้จะทำการเตรียมข้อมูลก่อนด้วยการแปลงจากภาพสีเป็นภาพแบบ grayscale และแปลงเป็นรูปภาพ binary ด้วยการแบ่งกลุ่มข้อมูลแบบเคมีน (K-means clustering) และกำจัดปัจจัยรบกวนออกด้วยการประยุกต์ใช้การกำหนดสัญลักษณ์การเชื่อมต่อขององค์ประกอบ (connected-components labeling)

#### 2.1 การแบ่งกลุ่มข้อมูลแบบเคมีน (K-means clustering)

การแบ่งกลุ่มข้อมูลแบบเคมีนเป็นวิธีหนึ่งที่ใช้ในการแบ่งกลุ่มของข้อมูลออกเป็น  $K$  กลุ่ม โดยจะต้องกำหนดจำนวนกลุ่ม ( $K$ ) ที่ต้องการก่อนที่จะแบ่ง แนวคิดของวิธีการแบ่งแบบนี้คือ แต่ละกลุ่มจะต้องมีความแปรปรวนภายในกลุ่มน้อยที่สุด วิธีวัดความแปรปรวนภายในกลุ่มมีหลายวิธี วิธีหนึ่งที่เป็นที่นิยมจะวัดจากระยะห่างแบบยูคลิเดียนยกกำลังสอง (squared Euclidean distance) ซึ่งระยะห่างระหว่างเวกเตอร์นิยามดังนี้

$$d(x_i, x_{i'}) = \sum_{j=1}^p (x_{ij} - x_{i'j})^2 = \|x_i - x_{i'}\|^2 \quad (1)$$

เมื่อ  $p$  คือจำนวนตัวแปรอิสระ ดังนั้น หากกำหนดให้  $C_k$  คือเซตของเวกเตอร์ที่อยู่ในกลุ่มที่  $k$  สำหรับ  $\forall k = 1, \dots, K$  แล้ว จะนิยามความแปรปรวนภายในกลุ่มดังนี้

$$W(C_k) = \frac{1}{|C_k|} \sum_{i, i' \in C_k} \|x_i - x_{i'}\|^2 \quad (2)$$

เมื่อ  $|C_k|$  คือจำนวนของเวกเตอร์ในกลุ่มที่  $k$  หากเรากำหนดให้ค่าเฉลี่ยของกลุ่มที่  $k$  เท่ากับ  $\bar{x}_k = \frac{1}{|C_k|} \sum_{i \in C_k} x_i$  แล้ว ความแปรปรวนภายในกลุ่มจะสามารถเขียนได้ใหม่ว่า

$$W(C_k) = 2 \sum_{i \in C_k} \|x_i - \bar{x}_k\|^2 \quad (3)$$

การแบ่งกลุ่มข้อมูลแบบเคมีนสามารถเขียนในรูปของการแก้ปัญหาค่าต่ำสุดได้ดังนี้

$$\min_{C_1, \dots, C_K} \sum_{k=1}^K W(C_k) \quad (4)$$

นั่นคือ เป็นการกำหนดเลขกลุ่มให้กับเวกเตอร์ข้อมูลที่ทำให้ความแปรปรวนภายในกลุ่มมีค่าน้อยที่สุด  
ขั้นตอนวิธีโดยทั่วไปของการแบ่งกลุ่มข้อมูลแบบเคมีนเป็นดังนี้

1. กำหนดหมายเลขกลุ่ม (ตั้งแต่ 1 ถึง  $K$ ) แบบสุ่มให้แต่ละเวกเตอร์ในชุดข้อมูล
2. คำนวณจุดกลางกลุ่ม (centroid) ซึ่งหมายถึงเวกเตอร์ของค่าเฉลี่ยของแต่ละกลุ่ม ( $\bar{x}_k$ ) สำหรับทุกกลุ่ม
3. เปลี่ยนหมายเลขกลุ่มของข้อมูลให้เป็นหมายเลขของกลุ่มที่ข้อมูลนั้นอยู่ใกล้จุดกลางกลุ่มที่สุด โดยวัดระยะห่างด้วยวิธียูคลิดีเนียน (Euclidean distance)
4. ทำซ้ำขั้นตอนที่ 2 จนกว่าหมายเลขกลุ่มจะไม่เปลี่ยน

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

## 2.2 การกำหนดสัญลักษณ์การเชื่อมต่อขององค์ประกอบ (Connected-components Labeling)

การกำหนดสัญลักษณ์การเชื่อมต่อขององค์ประกอบ หรือการกำหนดหมายเลขให้ส่วนที่เชื่อมกัน (Connected-components Labeling) คือกระบวนการทางทฤษฎีกราฟ (graph theory) ที่ใช้กำหนดสัญลักษณ์ (label) ให้กับองค์ประกอบ (components) ที่อยู่ติดกันเมื่อเทียบกับค่าขององค์ประกอบนั้น ในสาขาคอมพิวเตอร์วิทัศน์ จะนำมาประยุกต์ใช้กับรูปภาพ binary โดยองค์ประกอบหมายถึงพิกเซล ซึ่งมีเพียง 2 ค่าคือ 0 กับ 1 หากกำหนดให้ 0 หมายถึงพื้นหลัง และ 1 หมายถึงจุดของภาพ การกำหนดสัญลักษณ์การเชื่อมต่อขององค์ประกอบของรูปภาพ binary จะทำให้พิกเซลที่มีค่าเป็น 1 ที่อยู่ติดกันมีสัญลักษณ์เดียวกัน โดยทั่วไป สัญลักษณ์ที่ใ้หมักเป็นตัวเลข เนื่องจากเข้าใจได้ง่าย

1	1	0	1	1	1	0	1
1	1	0	1	0	1	0	1
1	1	1	1	0	0	0	1
0	0	0	0	0	0	0	1
1	1	1	1	0	1	0	1
0	0	0	1	0	1	0	1
1	1	0	1	0	0	0	1
1	1	0	1	0	1	1	1

ก. เมทริกซ์รูปภาพ binary



ข. รูปภาพบนจอแสดงผล

ภาพที่ 2.2 ตัวอย่างรูปภาพ binary

ที่มา: Shapiro (2001)

1	1	0	1	1	1	0	2
1	1	0	1	0	1	0	2
1	1	1	1	0	0	0	2
0	0	0	0	0	0	0	2
3	3	3	3	0	4	0	2
0	0	0	3	0	4	0	2
5	5	0	3	0	0	0	2
5	5	0	3	0	2	2	2

ก. เมทริกซ์ของรูปภาพ



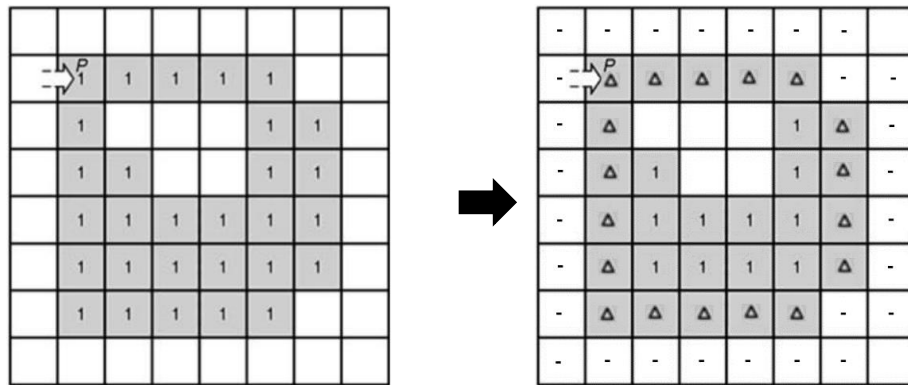
ข. รูปภาพบนจอแสดงผล

ภาพที่ 2.3 รูปภาพตามภาพที่ 2.2 หลังจากทำ connected-components labeling แล้ว

ที่มา: Shapiro (2001)

ขั้นตอนวิธีในการกำหนดสัญลักษณ์การเชื่อมต่อขององค์ประกอบมีหลายวิธี เช่น วิธี 2-pass, วิธี single pass เป็นต้น Chang et al. (2004) เสนอขั้นตอนวิธีกำหนดสัญลักษณ์แบบ single pass โดยใช้เทคนิคการติดตามขอบ (contour tracing) ซึ่งมีขั้นตอนดังนี้

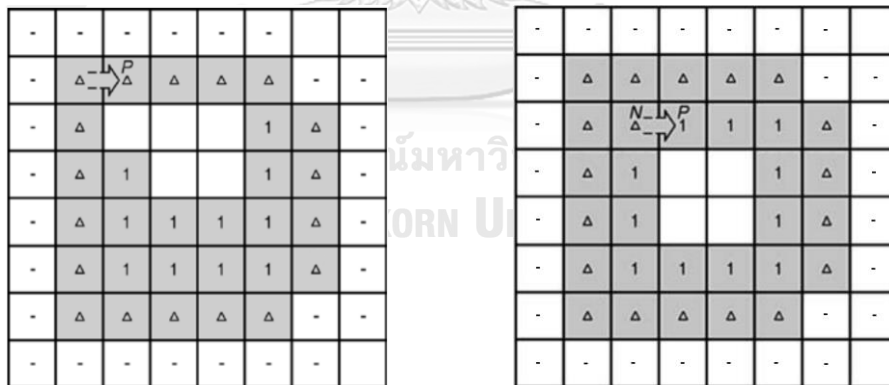
1. กำหนดให้ P เป็นพิกเซลที่กำลังพิจารณาอยู่ (พิกเซลปัจจุบัน) ถ้า P เป็นจุดของภาพ (นั่นคือ มีค่าเป็น 1) ที่ยังไม่ได้กำหนดหมายเลข และพิกเซลที่อยู่เหนือ P เป็นพื้นหลัง (มีค่าเป็น 0) แล้ว P จะเป็นขอบนอก (external contour) ขององค์ประกอบใหม่ของรูปภาพ เราจะกำหนดหมายเลขใหม่ให้กับ P รวมถึงขอบนอกทั้งหมดขององค์ประกอบด้วย โดยหมายเลขใหม่จะเพิ่มจากหมายเลขเดิมครั้งละ 1 และพิกเซลพื้นหลังที่อยู่รอบขอบนอกจะทำเครื่องหมาย - ไว้โดยกำหนดเป็นจำนวนเต็มลบ



ภาพที่ 2.4 ตัวอย่างการกำหนดหมายเลข  $\Delta$  หมายถึงพิกเซลที่ถูกกำหนดหมายเลขแล้ว 1 หมายถึงพิกเซลที่ยังไม่ถูกกำหนดหมายเลข และ - หมายถึงพิกเซลพื้นหลังที่ทำเครื่องหมาย

ที่มา: Chang et al. (2004)

2. ถ้าพิกเซลที่อยู่ใต้ P เป็นพื้นหลังที่ยังไม่ถูกทำเครื่องหมาย P จะเป็นขอบใน (internal contour) ขององค์ประกอบ ซึ่งเป็นไปได้ 2 แบบ แบบที่หนึ่งคือ P เป็นขอบนอกด้วย นั่นคือ ถูกกำหนดหมายเลขแล้ว ดังภาพที่ 2.5ก แบบที่สองคือ P ยังไม่ถูกกำหนดหมายเลข แต่พิกเซล N ที่อยู่ก่อนหน้าจะต้องถูกกำหนดหมายเลขแล้วแน่นอน ดังภาพที่ 2.5ข ให้กำหนดหมายเลขของ P ให้เป็นหมายเลขเดียวกับ N ทั้งนี้ ทั้งสองแบบจะต้องกำหนดหมายเลขให้กับขอบในขององค์ประกอบเป็นหมายเลขเดียวกับ P และ ทำเครื่องหมาย - ให้กับพิกเซลพื้นหลังที่อยู่รอบขอบในด้วย



ก. P เป็นพิกเซลที่เป็นทั้งขอบในและขอบนอกขององค์ประกอบ

ข. P เป็นพิกเซลที่เป็นเพียงขอบในขององค์ประกอบ

ภาพที่ 2.5 ตัวอย่างพิกเซลที่เป็นขอบในขององค์ประกอบ  $\Delta$  หมายถึงพิกเซลที่ถูกกำหนดหมายเลขแล้ว 1 หมายถึงพิกเซลที่ยังไม่ถูกกำหนดหมายเลข และ - หมายถึงพิกเซลพื้นหลังที่ทำเครื่องหมาย

ที่มา: Chang et al. (2004)

3. สำหรับ  $P$  ที่ไม่ใช่ขอบขององค์ประกอบ พิกเซล  $N$  ที่อยู่ก่อนหน้าจะต้องถูกกำหนดหมายเลขแล้วแน่นอน ให้กำหนดหมายเลขของ  $P$  ให้เป็นหมายเลขเดียวกับ  $N$

-	-	-	-	-	-	-	-	-
-	$\Delta$	$\Delta$	$\Delta$	$\Delta$	$\Delta$	$\Delta$	-	-
-	$\Delta$	$\Delta$	-	-	$\Delta$	$\Delta$	-	-
-	$\Delta$	$\Delta$	-	-	$\Delta$	$\Delta$	-	-
-	$\Delta$	$\Delta$	$\Delta$	$\Delta$	1	$\Delta$	-	-
-	$\Delta$	1	1	1	1	$\Delta$	-	-
-	$\Delta$	$\Delta$	$\Delta$	$\Delta$	$\Delta$	-	-	-
-	-	-	-	-	-	-	-	-

ภาพที่ 2.6 ตัวอย่างการกำหนดหมายเลขพิกเซลที่ไม่ใช่ขอบ  $\Delta$  หมายถึงพิกเซลที่ถูกกำหนดหมายเลขแล้ว 1 หมายถึงพิกเซลที่ยังไม่ถูกกำหนดหมายเลข และ - หมายถึงพิกเซลพื้นหลังที่ทำเครื่องหมายที่มา: Chang et al. (2004)

หลังจากเตรียมข้อมูลก่อนแล้ว หากเริ่มค้นหาลักษณะทันที อาจทำได้ยาก โดยเฉพาะกรณีวัตถุที่ต้องการค้นหาหรือจำแนกมีขนาดเล็กเมื่อเทียบกับขนาดของรูปภาพ ดังนั้น จึงอาจต้องแบ่งแยกบริเวณรูปภาพก่อน การแบ่งแยกบริเวณรูปภาพ (image segmentation) คือการแบ่งรูปภาพออกเป็นบริเวณเพื่อให้การค้นหาลักษณะของรูปภาพทำได้ง่ายขึ้น อย่างไรก็ตาม ฐานข้อมูลรูปภาพ SVHN มีรูปภาพ 2 รูปแบบ คือ แบบเต็มรูปกับแบบตัดส่วน ดังจะกล่าวโดยละเอียดในหัวข้อแหล่งข้อมูล งานวิจัยนี้ทำวิจัยเฉพาะรูปภาพแบบตัดส่วน ซึ่งจะละขั้นตอนการแบ่งแยกบริเวณภาพ

สำหรับการดึงลักษณะของรูปภาพในงานวิจัยนี้ จะใช้พารามิเตอร์ที่สร้างจากรูปภาคตัดกรวยของตัวเลขเป็นลักษณะ การค้นหารูปภาพตัดกรวยบนรูปภาพดิจิทัลจะประยุกต์ใช้การแปลงฮัฟ (Hough transform) ในการค้นหาและระบุพารามิเตอร์ของภาคตัดกรวยที่เหมาะสมที่สุดของตัวเลข

### 2.3 การแปลงฮัฟ (Hough Transform)

การแปลงฮัฟ คิดค้นโดย Hough (1962) และได้รับการพัฒนาต่อโดย Duda, Hart, and Newman (1972) เป็นเทคนิคที่ใช้ในการค้นหารูปร่างต่าง ๆ ในรูปภาพแบบ binary โดยใช้วิธีการนับคะแนน (vote) ในปริภูมิพารามิเตอร์ แล้วหาจุดสูงสุดสัมพัทธ์ (local maxima) ในปริภูมิดังกล่าวเพื่อนำมาพิจารณารูปร่างที่เป็นไปได้ในรูปภาพ

สำหรับรูปร่างที่จะกล่าวถึงในที่นี้ จะกล่าวเฉพาะรูปภาคตัดกรวยที่เกี่ยวข้องกับรูปตัวเลข ได้แก่ เส้นตรง วงกลม วงรี และพาราโบลา

### 2.3.1 เส้นตรง (Line)

รูปทั่วไปของสมการเส้นตรงในระบบพิกัดคาร์ทีเซียน (Cartesian coordinate system) อยู่ในรูปดังนี้

$$y = ax + b \quad (5)$$

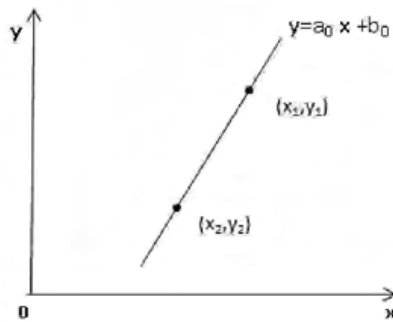
คู่อันดับ  $(x, y)$  ซึ่งสัมพันธ์กันตามสมการที่ 5 จะแสดงเป็นเส้นตรงที่มีความชันเท่ากับ  $a$  และจุดตัดแกน  $y$  ที่  $b$  บนระนาบคาร์ทีเซียน (Cartesian plane) ดังภาพที่ 2.7ก. จุด  $(x_1, y_1)$  และ  $(x_2, y_2)$  ต่างก็สัมพันธ์กันด้วยสมการ  $y = a_0x + b_0$  ดังนั้น จึงอยู่บนเส้นตรงเดียวกัน

จะเห็นว่า ที่จุด  $(x_i, y_i)$  ใด ๆ จะมีเส้นตรงผ่านได้เป็นจำนวนอนันต์ แต่ละเส้นจะมีค่าความชันค่าหนึ่งและจุดตัดแกน  $y$  ค่าหนึ่ง นั่นคือ เมื่อแสดงในปริภูมิพารามิเตอร์ (พารามิเตอร์ของเส้นตรงในที่นี้คือ ความชันและจุดตัดแกน  $y$ ) เส้นตรงหนึ่งเส้นบนระนาบคาร์ทีเซียนจะแทนด้วยจุดหนึ่งจุดในปริภูมิพารามิเตอร์ เพราะฉะนั้น เส้นตรงทุกเส้นที่สามารถผ่านจุด  $(x_i, y_i)$  บนระนาบคาร์ทีเซียนจะแสดงด้วยเส้นตรงหนึ่งเส้นในปริภูมิพารามิเตอร์ ซึ่งอยู่ในรูป

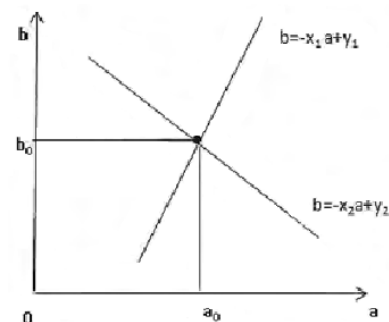
$$b = -x_i a + y_i \quad (6)$$

ดังแสดงตามภาพที่ 2.7ก. และ 2.7ข. เส้นตรง  $y = a_0x + b_0$  บนระนาบคาร์ทีเซียน แสดงด้วยจุด  $(a_0, b_0)$  บนปริภูมิพารามิเตอร์ ขณะเดียวกัน จุด  $(x_1, y_1)$  บนระนาบคาร์ทีเซียนมีเส้นตรงผ่านได้เป็นจำนวนอนันต์ นั่นคือ พารามิเตอร์  $a$  และ  $b$  มีค่าได้เป็นอนันต์ ดังนั้น เส้นตรงทุกเส้นที่สามารถผ่านจุด  $(x_1, y_1)$  บนระนาบคาร์ทีเซียนจึงแสดงด้วยเส้นตรง  $b = -x_1 a + y_1$  ในปริภูมิพารามิเตอร์ ทำนองเดียวกันกับจุด  $(x_2, y_2)$  ซึ่งเส้นตรงทุกเส้นที่สามารถผ่านจุด  $(x_2, y_2)$  เขียนแทนด้วยเส้นตรง  $b = -x_2 a + y_2$  ดังนั้น จุดที่เส้นตรง  $b = -x_1 a + y_1$  กับเส้นตรง  $b = -x_2 a + y_2$  ตัดกันในปริภูมิพารามิเตอร์ คือจุด  $(a_0, b_0)$  แสดงถึงพารามิเตอร์ของเส้นตรงที่จุด  $(x_1, y_1)$  กับจุด  $(x_2, y_2)$  อยู่บนเส้นเดียวกันบนระนาบคาร์ทีเซียน





ก. ระนาบคาร์ทีเซียน



ข. ปริภูมิพารามิเตอร์

ภาพที่ 2.7 เส้นตรงบนระนาบคาร์ทีเซียนและปริภูมิพารามิเตอร์

ที่มา: Giuliodori et al. (2011)

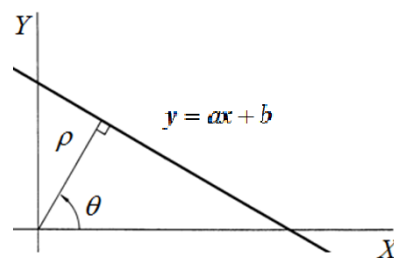
กล่าวโดยสรุปคือ เส้นตรงหนึ่งเส้นบนระนาบคาร์ทีเซียน สามารถเขียนแทนด้วยจุดหนึ่งจุดในปริภูมิพารามิเตอร์ ซึ่งส่งผลให้ เส้นตรงทุกเส้นที่สามารถผ่านจุดหนึ่งจุดบนระนาบคาร์ทีเซียน สามารถเขียนแทนด้วยเส้นตรงหนึ่งเส้นในปริภูมิพารามิเตอร์ และจุดตัดของเส้นตรงเหล่านั้นในปริภูมิพารามิเตอร์ แสดงถึงค่าพารามิเตอร์ของเส้นตรงบนระนาบคาร์ทีเซียน

อย่างไรก็ตาม หากความชันมีค่าเป็นอนันต์ จะไม่สามารถเขียนในรูปสมการที่ 5 ได้ Duda et al. (1972) จึงเสนอให้ใช้ normal parameterization ซึ่งสมการเส้นตรงจะอยู่ในรูป

$$x \cos \theta + y \sin \theta = \rho \quad (7)$$

เมื่อ  $\theta$  คือ มุมระหว่างแกน  $x$  กับเส้นที่สั้นที่สุดที่ลากจากจุดกำเนิดไปยังจุดบนเส้นตรงบนภาพ และ  $\rho$  คือ ระยะทางที่สั้นที่สุดระหว่างจุดกำเนิดกับจุดบนเส้นตรงในภาพ แสดงดังภาพที่ 2.8 ในรูปปกติ (normal form) นี้ พารามิเตอร์ของเส้นตรง ได้แก่  $\theta$  และ  $\rho$  เรียกว่า พารามิเตอร์ปกติ (normal parameters) เมื่อเทียบกับพารามิเตอร์ของรูปทั่วไปในสมการที่ 5 จะได้ว่า  $a = -\frac{1}{\tan \theta}$  และ

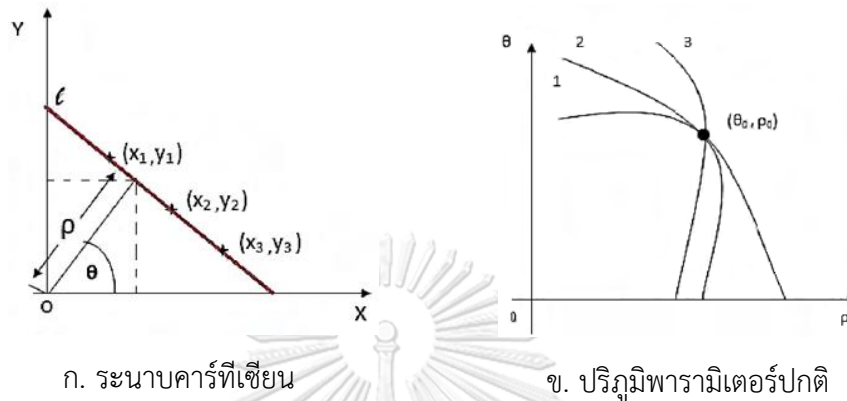
$$b = \frac{\rho}{\sin \theta}$$



ภาพที่ 2.8 พารามิเตอร์ปกติของเส้นตรง

ที่มา: Duda et al. (1972)

ในสมการรูปปกตินี้ เส้นตรงทุกเส้นที่สามารถผ่านจุดหนึ่งจุดบนระนาบคาร์ทีเซียนจะแทนด้วยเส้นโค้งหนึ่งเส้นในปริภูมิพารามิเตอร์ปกติ และหากกำหนดให้  $\theta$  อยู่ในช่วง  $[0, \pi)$  แล้ว เส้นตรงแต่ละเส้นบนระนาบคาร์ทีเซียน จะสามารถเขียนแทนด้วยจุดหนึ่งจุดที่ไม่ซ้ำกัน (unique point) ในปริภูมิพารามิเตอร์ปกติ

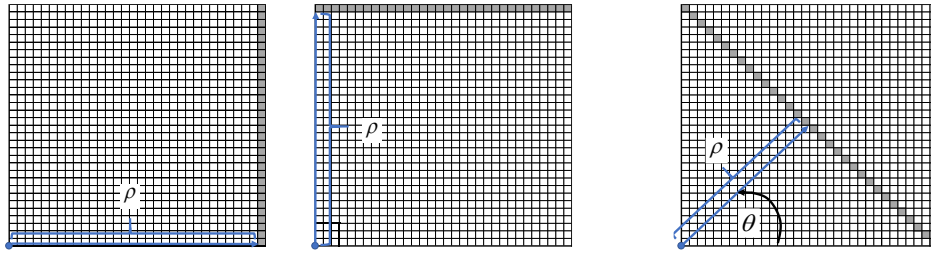


ภาพที่ 2.9 เส้นตรงบนระนาบคาร์ทีเซียนและปริภูมิพารามิเตอร์ในรูปแบบปกติ

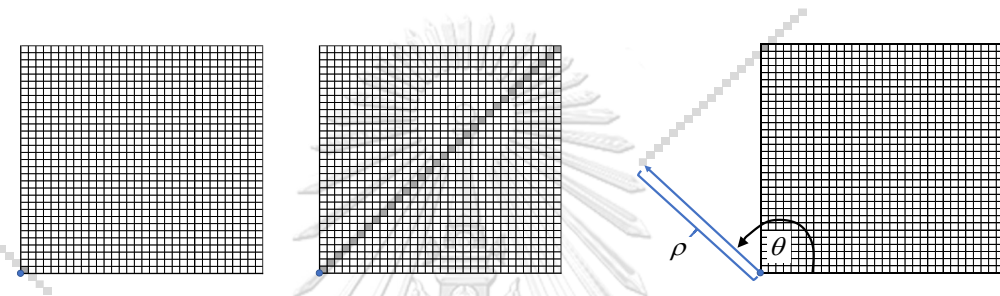
ที่มา: Giuliadori et al. (2011)

จากภาพที่ 2.9 เส้นตรงทั้งหมดที่สามารถผ่านจุด 3 จุด ได้แก่  $(x_1, y_1)$ ,  $(x_2, y_2)$  และ  $(x_3, y_3)$  บนระนาบคาร์ทีเซียนแทนด้วยเส้นโค้ง 3 เส้นในปริภูมิพารามิเตอร์ปกติ และจุดที่เส้นโค้ง 3 เส้นตัดกัน คือจุด  $(\theta_0, \rho_0)$  แสดงถึงพารามิเตอร์ของเส้นตรงบนระนาบคาร์ทีเซียนที่ทั้ง 3 จุดอยู่บนเส้นเดียวกัน ดังที่ได้กล่าวไปแล้วข้างต้น

สำหรับการค้นหาเส้นตรงในรูปภาพนั้น จะมองรูปภาพ binary เป็นระนาบคาร์ทีเซียน แต่ละพิกเซลจะแทนพิกัดของคู่อันดับบนระนาบ ในภาพที่ 2.10 แสดงตัวอย่างพารามิเตอร์ปกติที่มีค่าต่าง ๆ กันสำหรับรูปภาพที่มีขนาด  $N \times M$  พิกเซล โดย  $N$  คือจำนวนพิกเซลในแนวนอน และ  $M$  คือจำนวนพิกเซลในแนวตั้ง ในที่นี้กำหนดให้จุดกำเนิดบนระนาบคาร์ทีเซียนแทนด้วยจุดมุมซ้ายล่างของรูปภาพ พิกเซลที่เป็นสีเทาเข้มแทนจุดสีบนรูปภาพ ส่วนของเส้นตรงที่อยู่เกินขอบเขตของรูปภาพ แสดงด้วยสีเทาอ่อน



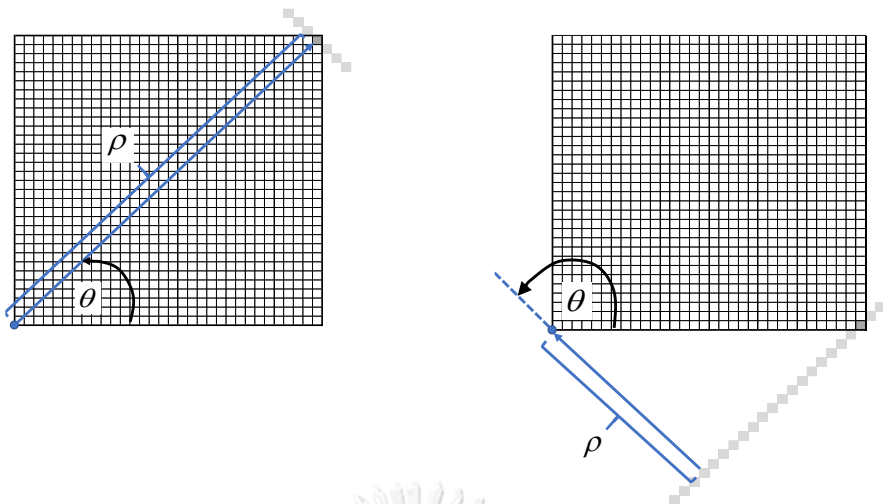
ก.  $\theta = 0$  และ  $\rho = N$    ข.  $\theta = \frac{\pi}{2}$  และ  $\rho = M$    ค.  $\theta = \frac{\pi}{4}$  และ  $\rho = \frac{\sqrt{N^2 + M^2}}{2}$



ง.  $\theta = \frac{\pi}{4}$  และ  $\rho = 0$    จ.  $\theta = \frac{3}{4}\pi$  และ  $\rho = 0$    ฉ.  $\theta = \frac{3}{4}\pi$  และ  $\rho = \frac{\sqrt{N^2 + M^2}}{2}$

ภาพที่ 2.10 ตัวอย่างพารามิเตอร์ปกติที่ค่าต่าง ๆ กันบนรูปภาพดิจิทัล

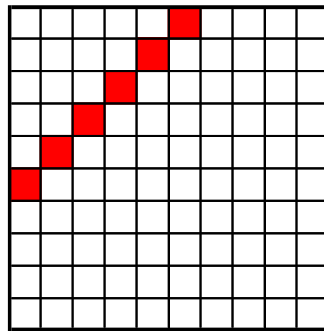
การแปลงฮัฟจะสร้าง แถวลำดับสะสม (Accumulator Array) หรือเรียกอีกอย่างหนึ่งว่า เมทริกซ์ฮัฟ (Hough matrix) โดยมีมิติของ accumulator array จะกำหนดโดยจำนวนพารามิเตอร์ ในที่นี้คือ  $\theta$  กับ  $\rho$  นั่นคือเป็นแถวลำดับ 2 มิติ และขนาดของแถวลำดับจะกำหนดโดยเรนจ์ของพารามิเตอร์ปกติ กล่าวคือ เรนจ์ของ  $\theta$  อยู่ในช่วง  $[0, \pi)$  และเรนจ์ของ  $\rho$  อยู่ในช่วง  $[-\frac{\sqrt{N^2 + M^2}}{2}, \sqrt{N^2 + M^2}]$  ภาพที่ 2.11 แสดงตัวอย่างรูปภาพที่เส้นตรงที่  $\rho$  เท่ากับขอบของเรนจ์



$$\text{ก. } \theta = \frac{\pi}{4} \text{ และ } \rho = \sqrt{N^2 + M^2} \quad \text{ข. } \theta = \frac{3}{4}\pi \text{ และ } \rho = -\frac{\sqrt{N^2 + M^2}}{2}$$

ภาพที่ 2.11 รูปเส้นตรงที่  $\rho$  เท่ากับขอบของเรนจ์

ขั้นตอนวิธีของการแปลงฮัฟสำหรับการค้นหาเส้นตรงในภาพ คือ แต่ละพิกเซลบนรูปภาพ จะเริ่มต้นจากการพิจารณาว่า พิกเซลนั้นเป็นจุดบนภาพหรือพื้นหลัง ถ้าเป็นจุดบนภาพ ก็จะพิจารณาเส้นตรงที่เป็นไปได้ทั้งหมดที่สามารถผ่านพิกเซลนั้นได้ ว่าแต่ละเส้นมีค่าพารามิเตอร์เท่าใด จากนั้น ก็จะไประบุในเซลล์ของ accumulator array ที่สอดคล้องกับค่าพารามิเตอร์ดังกล่าว ว่าพบ 1 พิกเซลที่เส้นตรงนี้สามารถผ่านได้ เช่น หากพิกเซล P เป็นจุดบนภาพ และเส้นตรงหนึ่งที่สามารถผ่านพิกเซล P ได้ มีพารามิเตอร์เท่ากับ  $(\theta_0, \rho_0)$  accumulator array ที่  $\theta = \theta_0$  และ  $\rho = \rho_0$  จะระบุค่า 1 ลงไป ทำเช่นนี้กับเส้นตรงทุกเส้นที่สามารถผ่านพิกเซล P ได้ เมื่อครบแล้วจึงพิจารณาพิกเซลถัดไป หากพิกเซลอื่น เช่น พิกเซล Q เป็นจุดบนภาพเหมือนกันแล้วพบว่า เส้นตรงที่มีพารามิเตอร์เท่ากับ  $(\theta_0, \rho_0)$  ก็สามารถผ่านพิกเซล Q ได้ ค่าใน accumulator array ที่  $\theta = \theta_0$  และ  $\rho = \rho_0$  จะเพิ่มขึ้นจาก 1 เป็น 2 หมายความว่า พบ 2 พิกเซลที่เส้นตรงนี้สามารถผ่านได้ นั่นคือเก็บค่าแบบสะสมไปเรื่อย ๆ อย่างไรก็ตาม หากพิกเซลนั้นเป็นพื้นหลังของภาพ นั่นคือ จะไม่มีเส้นตรงที่ผ่านพิกเซลดังกล่าว ก็จะพิจารณาพิกเซลถัดไปแทน หลังจากพิจารณาทุกพิกเซลบนรูปภาพแล้ว พารามิเตอร์ที่สอดคล้องกับเซลล์ที่มีค่ามากที่สุด ใน accumulator array จะหมายถึงพารามิเตอร์ของเส้นตรงในรูปภาพ



ก. รูปภาพแบบ binary

$\theta$										
179	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-
	-	-	1	1	1	-	-	-	-	-
135	-	-	-	1	1	1	-	-	-	-
	-	-	-	1	1	1	-	-	-	-
	-	-	-	-	1	1	1	-	-	-
90	-	-	-	-	1	1	1	-	-	-
	-	-	-	-	1	2	1	-	-	-
	-	-	-	-	-	4	-	-	-	-
45	-	-	-	-	-	6	-	-	-	-
	-	-	-	-	1	3	1	-	-	-
	-	-	-	-	1	2	1	-	-	-
0	-	-	-	-	-	-	-	-	-	-
										$\rho$

ข. Accumulator array

ภาพที่ 2.12 ภาพตัวอย่างการแปลงฮัฟสำหรับเส้นตรง

ภาพที่ 2.12 แสดงตัวอย่างของการใช้การแปลงฮัฟในการค้นหาเส้นตรงในรูปภาพแบบ binary (ในที่นี้กำหนดให้เซลล์ในแถวที่ 1 คอลัมน์ที่ 1 แทนจุดกำเนิดบนระนาบคาร์ทีเซียน) ในภาพที่ 2.12ก. มี 6 พิกเซลที่เรียงตัวเป็นเส้นตรง เมื่อแปลงเป็น accumulator array แล้ว จะได้ว่า เซลล์ที่มีค่ามากที่สุดมีค่าเท่ากับ 6 คือจุดที่  $\theta$  เท่ากับ 45 และ  $\rho$  เท่ากับ 4 นั่นคือ มี 6 พิกเซลที่เรียงตัวเป็นเส้นตรงที่มี  $\theta$  เท่ากับ 45 และ  $\rho$  เท่ากับ 4

### 2.3.2 วงกลม (Circle)

ในระบบพิกัดคาร์ทีเซียน วงกลมที่จุดศูนย์กลางอยู่ที่จุด  $(x_0, y_0)$  และมีรัศมีเท่ากับ  $r$  คือเซตของคู่อันดับ  $(x, y)$  ที่ทำให้

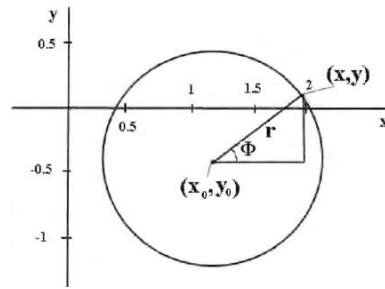
$$(x - x_0)^2 + (y - y_0)^2 = r^2 \quad (8)$$

หรือเขียนในรูปพาราเมตริกได้

$$x = x_0 + r \cos \phi \quad (9)$$

$$y = y_0 + r \sin \phi$$

เมื่อ  $\phi$  คือมุมที่วัดทวนเข็มนาฬิกาจากเส้น  $y = y_0$  ไปยังรัศมีวงกลมที่ลากไปยังจุด  $(x, y)$  ซึ่งอยู่ในช่วง  $[0, 2\pi)$  แสดงดังภาพที่ 2.13



ภาพที่ 2.13 วงกลมบนระนาบคาร์ทีเซียน

ที่มา: ดัดแปลงจาก Giulliodori et al. (2011)

จากสมการที่ 9 สามารถเขียนได้ใหม่เป็น

$$\begin{aligned} x_0 &= x - r \cos \phi \\ y_0 &= y - r \sin \phi \end{aligned} \quad (10)$$

นั่นคือ สมการของวงกลมสามารถมองได้ 2 แบบ คือ เป็นวงกลมที่มีจุดศูนย์กลางอยู่ที่จุด  $(x_0, y_0)$  จุดบนวงกลมอยู่ที่จุด  $(x, y)$  ในมุมมองของระบบพิกัดคาร์ทีเซียน (ดังภาพที่ 2.13) หรือ เป็นวงกลมที่มีจุดศูนย์กลางอยู่ที่จุด  $(x, y)$  จุดบนวงกลมอยู่ที่จุด  $(x_0, y_0)$  ในมุมมองของปริภูมิพารามิเตอร์  $x_0$  และ  $y_0$  กล่าวอีกนัยหนึ่งคือ จุดทุกจุดบนวงกลมใด ๆ บนระนาบคาร์ทีเซียนสามารถเป็นจุดศูนย์กลางของวงกลมในปริภูมิพารามิเตอร์ที่มีรัศมีเท่ากันได้ทั้งสิ้น

ในภาพที่ 2.14ก. แสดงตัวอย่างของจุด 3 จุดที่เป็นขอบของวงกลมบนระนาบคาร์ทีเซียน (หรือบนรูปภาพ binary image) ภาพที่ 2.14ข. แสดงวงกลมในปริภูมิพารามิเตอร์ที่มีจุดขอบเหล่านั้นเป็นจุดศูนย์กลาง จะเห็นว่า จุดที่วงกลม 3 วงนั้นตัดกันมากที่สุดคือตำแหน่งที่เป็นจุดศูนย์กลางของวงกลมบนระนาบคาร์ทีเซียน



ก. วงกลมบนระนาบคาร์ทีเซียน

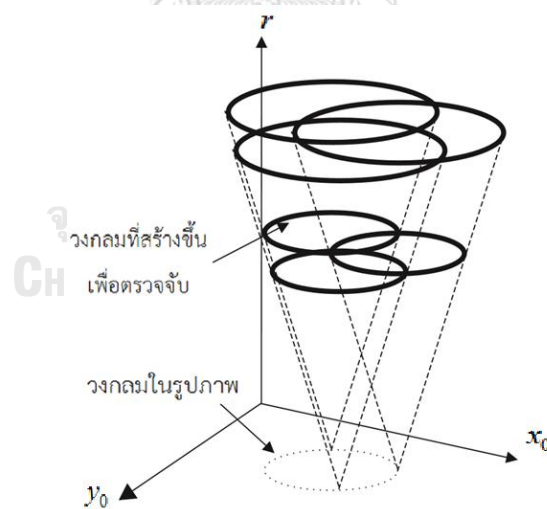
ข. ปริภูมิพารามิเตอร์

ภาพที่ 2.14 การแปลงฮัพสำหรับวงกลม ณ ค่ารัศมีคงที่

ที่มา: Aguado (2012)

การแปลงฮัฟจะอาศัยแนวความคิดดังกล่าวในการตรวจจับวงกลมในรูปภาพ หากเราทราบรัศมีของวงกลมในรูปภาพ (กล่าวคือ กำหนดให้รัศมีคงที่) พารามิเตอร์ที่ใช้ในการกำหนดรูปร่างวงกลมจะมีเพียงจุดศูนย์กลาง  $(x_0, y_0)$  ดังนั้น ปัญหาการตรวจจับวงกลมในรูปภาพ จะเทียบได้กับการค้นหาจุดศูนย์กลางของวงกลม มิติของ accumulator array สำหรับกรณีนี้จึงเท่ากับ 2 มิติ โดยเรนจ์ของพารามิเตอร์  $x_0$  และ  $y_0$  จะอยู่ในช่วง  $(0, M]$  และ  $(0, N]$  ตามลำดับ เมื่อ  $N \times M$  คือขนาดของรูปภาพ การแปลงฮัฟจะพิจารณาแต่ละพิกเซล ถ้าพิกเซลนั้นเป็นจุดบนภาพ ก็จะสร้างวงกลมโดยใช้จุดดังกล่าวเป็นจุดศูนย์กลาง ค่าที่เก็บใน accumulator array คือ จำนวนวงกลมที่ผ่านในแต่ละตำแหน่งของ  $x_0$  และ  $y_0$  เมื่อพิจารณาครบทุกพิกเซลแล้ว เซลล์ใน accumulator array ที่มีค่ามากที่สุดจะหมายถึงตำแหน่งของจุดศูนย์กลางของวงกลมในรูปภาพ

ในการตรวจจับวงกลมในรูปภาพโดยทั่วไปซึ่งไม่ทราบรัศมีของวงกลมนั้น พารามิเตอร์ที่กำหนดโดยวงกลมจะมี 3 ค่า ได้แก่  $x_0$ ,  $y_0$  และ  $r$  ดังที่ได้กล่าวไปแล้วว่า มิติของ accumulator array จะถูกกำหนดโดยจำนวนพารามิเตอร์ ดังนั้น accumulator array สำหรับวงกลมจะเป็นแถวลำดับ 3 มิติ ดังภาพที่ 2.15 ในแต่ละค่ารัศมีของวงกลม ขั้นตอนวิธีของการแปลงฮัฟจะเป็นเช่นเดียวกับที่กล่าวไปแล้วในกรณีรัศมีวงกลมคงที่ อย่างไรก็ตาม เรนจ์ของค่ารัศมีวงกลมที่ต้องการจะตรวจจับนั้น ขึ้นอยู่กับขนาดของวงกลมของวัตถุที่ต้องการจะตรวจจับด้วย



ภาพที่ 2.15 accumulator array 3 มิติ

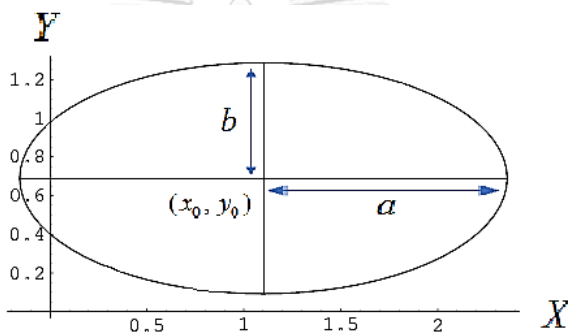
ที่มา: Aguado (2012)

### 2.3.3 วงรี (Ellipse)

สมการรูปแบบมาตรฐานในระบบพิกัดคาร์ทีเซียนของวงรีจะขึ้นอยู่กับว่า แกนเอกของวงรีขนานกับแกน  $x$  หรือแกน  $y$  กล่าวคือ สำหรับวงรีที่จุดศูนย์กลางวงรีอยู่ที่จุด  $(x_0, y_0)$  มีความยาวแกนเอกและแกนโทเท่ากับ  $2a$  และ  $2b$  ตามลำดับ เขียนสมการได้ดังนี้

แกนเอกขนานกับแกน $x$	แกนเอกขนานกับแกน $y$
$\frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} = 1$ (11.1)	$\frac{(x-x_0)^2}{b^2} + \frac{(y-y_0)^2}{a^2} = 1$ (11.2)

ตัวอย่างของวงรีที่แกนเอกขนานกับแกน  $x$  แสดงดังภาพที่ 2.16 ในที่นี้จะพิจารณาเฉพาะกรณีที่แกนเอกขนานกับแกน  $x$  หรือแกน  $y$  เท่านั้น

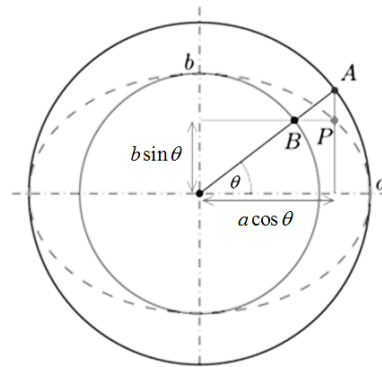


ภาพที่ 2.16 ตัวอย่างวงรีบนระนาบคาร์ทีเซียน ที่แกนเอกขนานกับแกน  $x$  สำหรับสมการวงรีในรูปพารามิตริกนั้น เขียนได้ในรูป

แกนเอกขนานกับแกน $x$	แกนเอกขนานกับแกน $y$
$x = x_0 + a \cos \theta$ $y = y_0 + b \sin \theta$ (12.1)	$x = x_0 + b \sin \theta$ $y = y_0 + a \cos \theta$ (12.2)

โดย  $\theta$  คือมุมที่วัดทวนเข็มนาฬิกาจากแกนที่ขนานกับแกน  $x$  ไปยังเส้นที่ลากจากจุดศูนย์กลางวงรีไปยังจุดบนวงกลมที่มีเส้นผ่านศูนย์กลางเท่ากับความยาวของแกนเอก และอยู่ในช่วง  $[0, 2\pi)$  แสดงดังภาพที่ 2.17 วงรีแสดงด้วยเส้นประ จุด  $P$  คือจุดบนวงรี ระยะทางจากจุดศูนย์กลางวงรีมายังจุดที่เป็นภาพฉาย (projection) ของจุด  $P$  บนแกน  $x$  และแกน  $y$  สามารถคำนวณโดยใช้วงกลม 2 วงที่มีรัศมีเท่ากับ  $a$  และ  $b$  มาช่วยได้





ภาพที่ 2.17 วงรีในรูปพาราเมตริก

ในทำนองเดียวกันกับวงกลม เราสามารถจัดรูปสมการที่ 12.1 และ 12.2 ให้จุดศูนย์กลางวงรีเป็นฟังก์ชันของจุดบนวงรีได้ นั่นคือ จุดทุกจุดบนวงรีบนระนาบคาร์ทีเซียนสามารถเป็นจุดศูนย์กลางวงรีที่มีความยาวแกนเอกและแกนโทเท่ากันในปริภูมิพารามิเตอร์ได้ การแปลงฮัฟสำหรับตรวจจับวงรีในรูปภาพก็จะคล้ายกับวงกลมเช่นกัน กล่าวคือ จุดที่วงรีตัดกันมากที่สุดในปริภูมิพารามิเตอร์จะหมายถึงจุดศูนย์กลางวงรีในรูปภาพ เนื่องจากพารามิเตอร์ที่ใช้กำหนดวงรี ได้แก่  $x_0$ ,  $y_0$ ,  $a$  และ  $b$  ดังนั้น accumulator array สำหรับวงรีจึงมี 4 มิติ เรนจ์ของ  $x_0$  และ  $y_0$  จะถูกกำหนดโดยขนาดของรูปภาพ และเรนจ์ของค่า  $a$  และ  $b$  ที่ต้องการจะตรวจจับนั้น ขึ้นอยู่กับขนาดของวัตถุที่ต้องการจะตรวจจับ เช่นเดียวกับวงกลม

#### 2.3.4 พาราโบลา (Parabola)

พาราโบลาที่มีจุดยอดอยู่ที่จุด  $(x_0, y_0)$  และระยะห่างระหว่างจุดยอดกับจุดโฟกัสเท่ากับ  $p$  สามารถเขียนสมการในระบบพิกัดคาร์ทีเซียนได้ดังนี้

แกนสมมาตรขนานกับแกน x	แกนสมมาตรขนานกับแกน y
$(y - y_0)^2 = 4p(x - x_0)$ (13.1)	$(x - x_0)^2 = 4p(y - y_0)$ (13.2)

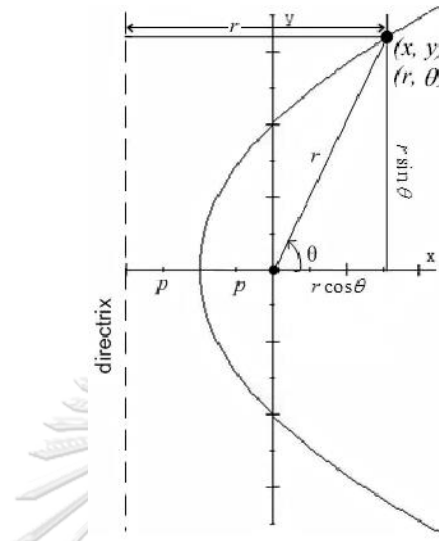
และหากเราทราบว่าจุดโฟกัสของพาราโบลาอยู่ที่จุด  $(x_f, y_f)$  จะสามารถเขียนในรูปพาราเมตริกได้ดังนี้

แกนสมมาตรขนานกับแกน x	แกนสมมาตรขนานกับแกน y
$x = x_f + r \cos \theta$ $y = y_f + r \sin \theta$ (14.1)	$x = x_f + r \sin \theta$ $y = y_f + r \cos \theta$ (14.2)

เมื่อ  $r = \frac{2p}{1 - \cos \theta}$  และ  $\theta$  คือมุมระหว่างแกนสมมาตรกับเส้นที่ลากจากจุดโฟกัสไปยังจุดบนพาราโบลาและอยู่ในช่วง  $[0, 2\pi)$  จากภาพที่ 2.18 เราทราบว่า ระยะห่างระหว่างจุดโฟกัสกับจุดบน

พาราโบลาคู่จะเท่ากับระยะห่างระหว่างจุดบนพาราโบลากับเส้นไดเรกทริกซ์เสมอ ดังนั้น จากรูปจะได้

$$\text{ว่า } r = 2p + r \cos \theta \text{ เมื่อจัดรูปใหม่จะได้ } r = \frac{2p}{1 - \cos \theta}$$



ภาพที่ 2.18 พาราโบลาคู่ที่แกนสมมาตรขนานกับแกน x

จากสมการที่ 14.1 และ 14.2 เราสามารถจัดรูปให้จุดโฟกัสเป็นฟังก์ชันของจุดบนพาราโบลาคู่ นั่นคือจุดทุกจุดบนพาราโบลาคู่บนระนาบคาร์ทีเซียนสามารถเป็นจุดโฟกัสของพาราโบลาคู่บนปริภูมิพารามิเตอร์ที่มีความยาวโฟกัส (focal length) เท่ากันได้ และจุดที่พาราโบลาคู่บนปริภูมิพารามิเตอร์ตัดกันมากที่สุดจะเป็นจุดโฟกัสบนระนาบคาร์ทีเซียน accumulator array สำหรับพาราโบลาคู่ 3 มิติ ได้แก่  $x_f$ ,  $y_f$  และ  $p$  เรนจ์ของ  $x_f$  และ  $y_f$  ถูกกำหนดโดยขนาดของรูปภาพ และเรนจ์ของค่า  $p$  ขึ้นอยู่กับขนาดของวัตถุที่ต้องการจะตรวจจับ

ในส่วนของงานวิจัยที่ใช้การแปลงฮัฟในการจำแนกตัวเลขนั้น Giuliodori et al. (2011) ใช้ Hough transform และ features อื่น ๆ ในการจำแนกข้อมูลตัวเลขที่เขียนด้วยลายมือ MNIST และ USPS โดยเปรียบเทียบวิธีเชิงความน่าจะเป็นซึ่งอิงจาก Bayes' rule กับวิธี k-nearest neighbor ในการจำแนกตัวเลข อัตราการทำนายผิด (test-error rate) ของวิธีเชิงความน่าจะเป็นเท่ากับ 4.3% และ 9.7% สำหรับข้อมูล MNIST และ USPS ตามลำดับ ขณะที่ test-error rate ของวิธี KNN เท่ากับ 3.65% และ 4.39% สำหรับข้อมูล MNIST และ USPS ตามลำดับ

นอกจากนี้ Garg and Ahuja (2013) เก็บตัวอย่างข้อมูลลายมือตัวเลขจำนวน 1200 ตัวอย่างจาก 24 คน และเปรียบเทียบลักษณะจากวิธี modified Hough transformation กับวิธี four view projection profiles รวมถึงเปรียบเทียบตัวแบบที่ใช้ในการจำแนกตัวเลขคือซัพพอร์ตเวกเตอร์แมชชีน (SVM) กับ Multi-Layer Perceptron (MLP) ผลลัพธ์อัตราการทำนายถูกต้องที่ได้คือ ลักษณะจากวิธี modified Hough transformation จะทายถูก 93.12% และ 72.5% เมื่อใช้ SVM และ

MLP ตามลำดับ ขณะที่ลักษณะจากวิธี four view projection profiles จะหายถูก 96.04% และ 98.73% เมื่อใช้ SVM และ MLP ตามลำดับ

เมื่อได้ลักษณะของรูปภาพตัวเลขแล้ว ก็จะทำกรจำแนกประเภท โดยงานวิจัยนี้จะใช้ซัพพอร์ตเวกเตอร์แมชชีนเป็นตัวแบบในการจำแนกตัวเลข

## 2.4 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines; SVM)

ซัพพอร์ตเวกเตอร์แมชชีน (SVM) คือตัวแบบหนึ่งที่ใช้ในการจำแนกประเภท โดยใช้แนวคิดเกี่ยวกับการใช้ไฮเปอร์เพลน (hyperplane) ในการแบ่งข้อมูล

ในปริภูมิ  $p$  มิติ ไฮเปอร์เพลนคือปริภูมิย่อย (subspace) ที่มี  $p-1$  มิติ เช่น บนระนาบซึ่งเป็นปริภูมิ 2 มิติ ไฮเปอร์เพลนจะเป็นปริภูมิย่อยที่มี 1 มิติ ซึ่งคือเส้นตรง เขียนได้ในรูป

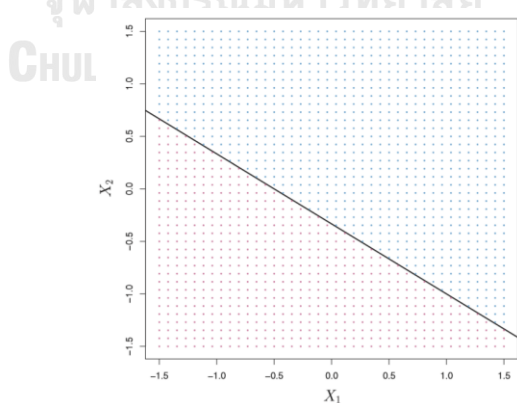
$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0 \quad (17)$$

โดยทั่วไปแล้ว ในปริภูมิ  $p$  มิติ ไฮเปอร์เพลนจะเขียนได้ในรูป

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p = 0 \quad (18)$$

หาก  $X = (X_1, X_2, \dots, X_p)^T$  ทำให้สมการที่ 18 เป็นจริง แสดงว่า  $X$  อยู่บนไฮเปอร์เพลน หาก  $X$  ไม่ได้อยู่บนไฮเปอร์เพลน  $X$  จะต้องอยู่ในฝั่งใดฝั่งหนึ่งของไฮเปอร์เพลน นั่นคือ

$$\begin{aligned} \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p &> 0 \\ \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p &< 0 \end{aligned} \quad \text{หรือ} \quad (19)$$



ภาพที่ 2.19 ตัวอย่างไฮเปอร์เพลนในปริภูมิ 2 มิติ แสดงด้วยเส้นตรง  $1 + 2X_1 + 3X_2 = 0$  พื้นที่เหนือเส้นตรงคือเซตของจุดที่ทำให้  $1 + 2X_1 + 3X_2 > 0$  เป็นจริง ขณะที่พื้นที่ใต้เส้นตรงคือเซตของจุดที่ทำให้  $1 + 2X_1 + 3X_2 < 0$  เป็นจริง

ที่มา: James, Witten, Hastie, and Tibshirani (2013)

สำหรับการจำแนกประเภทข้อมูลที่มีตัวแปรอิสระ  $p$  ตัว เทียบได้กับการใช้ไฮเปอร์เพลนในการแบ่งปริภูมิ  $p$  มิติ นั่นคือ หากมีข้อมูลจำนวน  $n$  คู่ คือ  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  โดย  $x_i$

เป็นเวกเตอร์ของตัวแปรอิสระ  $x_i = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{bmatrix}$  (กล่าวคือ  $x_i \in \mathbb{R}^p$ ,  $\forall i=1, \dots, n$ ) และ  $y_i$  เป็นตัวแปร

ตามที่ใช้ในการระบุ class (ในที่นี้จะยกตัวอย่างกรณีที่มีเพียง 2 class) ซึ่ง  $y_i \in \{-1, 1\}$  “ไฮเปอร์เพลนที่ใช้แบ่งข้อมูล (separating hyperplane)” จะนิยามโดย

$$\{x : f(x) = x^T \beta + \beta_0 = 0\} \quad (20)$$

เมื่อ  $\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$  และเป็นเวกเตอร์หนึ่งหน่วย (กล่าวคือ  $\|\beta\| = \sum_{j=1}^p \beta_j^2 = 1$ ) จากอสมการที่ 19 ทำให้

ทราบว่า การจำแนก class ของข้อมูลแต่ละเวกเตอร์ด้วยไฮเปอร์เพลนที่ใช้แบ่งข้อมูล จะดูจากเครื่องหมายของ  $f(x)$  นั่นคือ

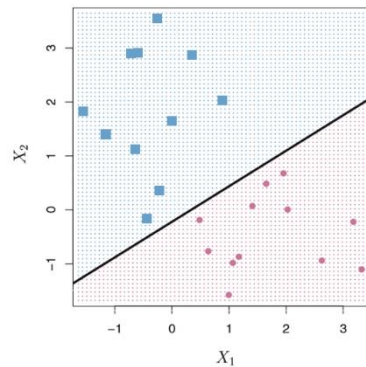
$$G(x) = \text{sign}[x^T \beta + \beta_0] \quad (21)$$

โดย  $G(x_i)$  จะเป็นบวก เมื่อ  $y_i = 1$  และเป็นลบ เมื่อ  $y_i = -1$  อีกนัยหนึ่งคือ

$$y_i(x_i^T \beta + \beta_0) > 0, \quad \forall i=1, \dots, n \quad (22)$$

ในขณะที่เครื่องหมายของ  $f(x)$  ระบุ class ของข้อมูล ขนาดของ  $f(x)$  จะบอกระยะทางจากจุดของข้อมูลมายังไฮเปอร์เพลนที่ใช้แบ่งข้อมูล หาก  $f(x_i)$  ใกล้เคียงกับ 0 นั่นคือข้อมูล  $x_i$  อยู่ใกล้กับไฮเปอร์เพลนที่ใช้แบ่งข้อมูล

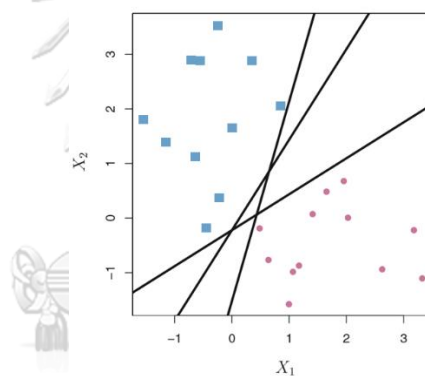
ตัวอย่างการใช้ไฮเปอร์เพลนที่ใช้แบ่งข้อมูลกับข้อมูลที่มีตัวแปรอิสระ 2 ตัว และมีจำนวน class เท่ากับ 2 แสดงดังภาพที่ 2.20



ภาพที่ 2.20 ตัวอย่างของการใช้ไฮเปอร์เพลนในการแบ่งประเภทข้อมูล กรณีที่มีตัวแปรอิสระ 2 ตัว และมี 2 class แต่ละฝั่งของไฮเปอร์เพลนที่ใช้แบ่งข้อมูลจะแทนแต่ละ class

ที่มา: ดัดแปลงจาก James et al. (2013)

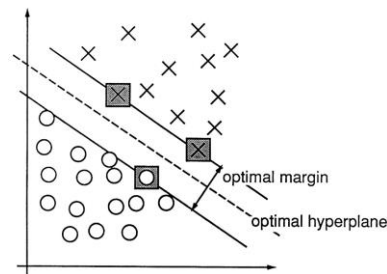
จะเห็นได้ว่า หากข้อมูลสามารถแบ่งได้ด้วยไฮเปอร์เพลนแล้ว ไฮเปอร์เพลนที่ใช้แบ่งข้อมูลจะมีจำนวนเป็นอนันต์ ดังภาพที่ 2.21



ภาพที่ 2.21 ตัวอย่างไฮเปอร์เพลนที่สามารถแบ่งประเภทข้อมูลได้

ที่มา: ดัดแปลงจาก James et al. (2013)

โดยทั่วไป จะเรียกเวกเตอร์ที่มีผลต่อไฮเปอร์เพลนที่ใช้แบ่งข้อมูลว่า ซัพพอร์ตเวกเตอร์ (support vector) ในกรณีที่ข้อมูลสามารถแบ่งได้ด้วยไฮเปอร์เพลน ซัพพอร์ตเวกเตอร์จะหมายถึงเวกเตอร์ของข้อมูลที่อยู่ใกล้กับไฮเปอร์เพลนที่ใช้แบ่งข้อมูลมากที่สุด ในแต่ละ class จะมีไฮเปอร์เพลนที่ลากผ่านซัพพอร์ตเวกเตอร์และขนานกับไฮเปอร์เพลนที่ใช้แบ่งข้อมูลเพียง 1 เส้น ระยะห่างระหว่างไฮเปอร์เพลนที่ลากผ่านซัพพอร์ตเวกเตอร์ของ class หนึ่งกับอีก class หนึ่งเรียกว่า มาร์จิน (margin) ไฮเปอร์เพลนที่ใช้แบ่งข้อมูลได้ดีที่สุด (optimal hyperplane) คือไฮเปอร์เพลนที่ใช้แบ่งข้อมูลที่มีมาร์จินมีค่ามากที่สุด (Cortes และ Vapnik, 1995) เรียกว่า maximal margin hyperplane (หรือ optimal separating hyperplane) และเรียกตัวจำแนกประเภทนี้ว่า maximal margin classifier



ภาพที่ 2.22 ตัวอย่างของ maximal margin classifier โดย optimal hyperplane แสดงด้วยเส้นประ ซัพพอร์ตเวกเตอร์แสดงด้วยข้อมูลที่อยู่ในกรอบสี่เหลี่ยมสีเทา ระยะห่างระหว่างไฮเปอร์เพลนที่ทับซัพพอร์ตเวกเตอร์ (เส้นทึบ) ของแต่ละ class คือมาร์จิ้น (แสดงด้วยลูกศร)

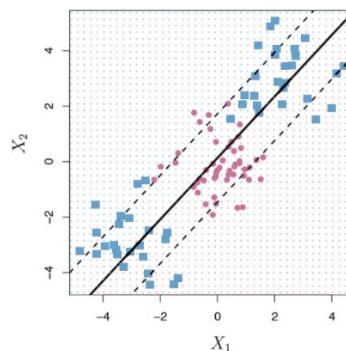
ที่มา: Cortes and Vapnik (1995)

หากกำหนดให้ความกว้างของมาร์จิ้นยาวเท่ากับ  $2M$  แล้ว maximal margin classifier คือ  $f(x) = x^T \beta + \beta_0$  โดยพารามิเตอร์  $\beta_0$  และ  $\beta$  หาได้จากการแก้ปัญหาค่าสูงสุดดังนี้

$$\begin{aligned} & \max_{\beta_0, \beta} M \\ \text{s.t.} & \sum_{j=1}^p \beta_j^2 = 1, \\ & y_i(x_i^T \beta + \beta_0) \geq M, \quad \forall i = 1, \dots, n \end{aligned} \quad (23)$$

ข้อจำกัด  $y_i(x_i^T \beta + \beta_0) \geq M$  จะเป็นการบังคับว่า ทุกเวกเตอร์จะต้องอยู่นอกบริเวณของมาร์จิ้น

กรณีที่มีข้อมูลไม่สามารถแบ่งได้ด้วยไฮเปอร์เพลนอย่างสมบูรณ์ การใช้ไฮเปอร์เพลนแบ่งประเภทโดยที่มีข้อมูลบางเวกเตอร์ไม่อยู่ใน class ที่ควรจะเป็น แสดงดังภาพที่ 2.23 เรียกตัวจำแนกประเภทนี้ว่า support vector classifier หรือ soft margin classifier



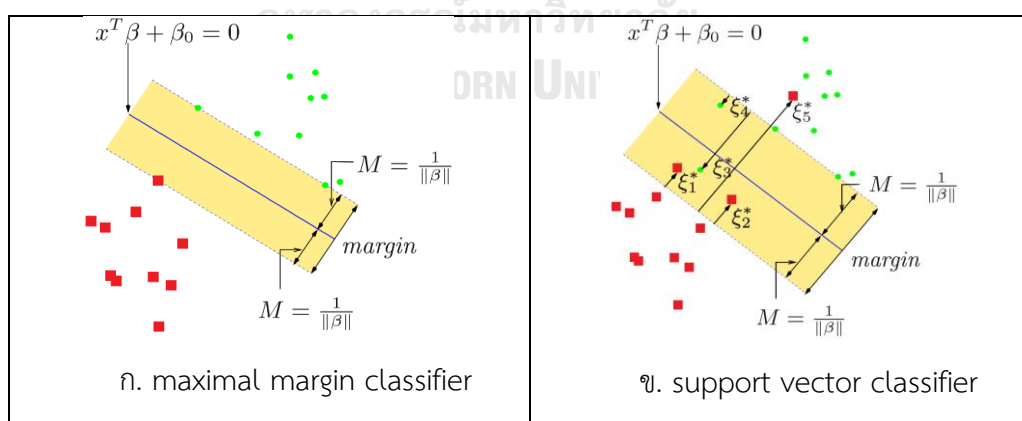
ภาพที่ 2.23 ตัวอย่างของ support vector classifier

ที่มา: ดัดแปลงจาก James et al. (2013)

เมื่อไม่สามารถแบ่งประเภทได้อย่างสมบูรณ์ จึงมีข้อมูลบางเวกเตอร์ที่ไม่ได้อยู่ในขอบริเวณมาร์จิ้นใน class ที่ถูกต้อง กล่าวคือ อาจอยู่ในบริเวณมาร์จิ้น แต่อยู่ใน class ที่ถูกต้องหรือไม่ถูกต้องก็ได้ หรืออาจอยู่นอกบริเวณมาร์จิ้นและอยู่ใน class ที่ไม่ถูกต้องก็ได้เช่นกัน เมื่อกำหนดให้  $x_i^*$  เป็นเวกเตอร์ที่มีอยู่นอกบริเวณมาร์จิ้นใน class ที่ถูกต้อง กำหนด  $\xi = (\xi_1, \xi_2, \dots, \xi_n)$  เมื่อ  $\xi_i$  คือระยะห่างจาก  $x_i^*$  ไปยังไฮเปอร์เพลนที่เป็นขอบของมาร์จิ้นใน class ของ  $x_i^*$  เมื่อเทียบกับความกว้างของมาร์จิ้น พารามิเตอร์ของ support vector classifier สามารถหาได้จากการแก้ปัญหาค่าสูงสุดดังนี้

$$\begin{aligned} & \max_{\beta_0, \beta, \xi} M \\ \text{s.t.} & \sum_{j=1}^p \beta_j^2 = 1, \\ & y_i(x_i^T \beta + \beta_0) \geq M(1 - \xi_i), \quad \forall i = 1, \dots, n, \\ & \sum_{i=1}^n \xi_i \leq C, \\ & \xi_i \geq 0, \quad \forall i = 1, \dots, n \end{aligned} \quad (24)$$

เมื่อ  $C$  หรือพารามิเตอร์ cost ใช้กำหนดระดับของการมีอยู่นอกบริเวณมาร์จิ้นใน class ที่ถูกต้องที่รับได้ นั่นคือ ยิ่ง  $C$  มีค่ามาก มาร์จิ้นก็ยิ่งกว้าง ในกรณีที่  $C > 0$  และเป็นจำนวนเต็ม จะหมายถึงจำนวนสูงสุดของเวกเตอร์ที่ support vector classifier อนุญาตให้อยู่ใน class ที่ไม่ถูกต้องได้ กรณีที่  $C = 0$  support vector classifier จะเทียบเท่ากับ maximal margin classifier



ภาพที่ 2.24 เปรียบเทียบตัวแบ่งประเภท maximal margin classifier กับ support vector classifier

ที่มา: ดัดแปลงจาก Hastie, Tibshirani, and Friedman (2001)

สำหรับ support vector classifier ซัพพอร์ตเวกเตอร์จะหมายถึงเวกเตอร์ทั้งหมดที่มีได้อยู่นอกบริเวณมาร์จินใน class ที่ถูกต้อง

สำหรับข้อมูลที่มีตัวแปรอิสระทั้งหมด  $p$  ตัว และมีข้อมูลจำนวน  $n$  คู่ เมื่อกำหนดให้  $\beta = \sum_{i=1}^n \alpha_i y_i x_i$  แล้ว support vector classifier ซึ่งเป็นตัวจำแนกประเภทเชิงเส้น สามารถเขียนได้ในรูป

$$f(x) = \beta_0 + \sum_{i=1}^n \alpha_i y_i \langle x, x_i \rangle \quad (25)$$

โดย  $x_i = (x_{i1}, \dots, x_{ip})^T$  คือเวกเตอร์ของข้อมูลที่ใช้สร้างตัวแบบ (train data) ตัวที่  $i$

$x = (x_1, \dots, x_p)^T$  คือเวกเตอร์ของข้อมูลที่ใช้ทดสอบประสิทธิภาพตัวแบบ (test data)

$\langle x_i, x_i \rangle$  คือผลคูณภายใน (inner product) ของข้อมูล  $x_i$  และ  $x_i$  กล่าวคือ

$$\langle x_i, x_i \rangle = \sum_{j=1}^p x_{ij} x_{i,j}$$

การประมาณค่าพารามิเตอร์  $\alpha_1, \dots, \alpha_n$  และ  $\beta_0$  จำเป็นต้องคำนวณผลคูณภายในของข้อมูลที่ใช้สร้างตัวแบบทุกคู่

ซัพพอร์ตเวกเตอร์แมชชีน (support vector machine; SVM) เป็นตัวจำแนกประเภทแบบไม่เชิงเส้น จะใช้ฟังก์ชันเคอร์เนล (kernel function) แทนผลคูณภายใน ดังนั้น สมการของ SVM จะอยู่ในรูป

$$f(x) = \beta_0 + \sum_{i=1}^n \alpha_i y_i K(x, x_i) \quad (26)$$

เมื่อ  $K(x_i, x_i)$  คือฟังก์ชันเคอร์เนล โดยฟังก์ชันเคอร์เนลที่เป็นที่นิยม ได้แก่

1. เคอร์เนลพหุนาม (polynomial kernel):  $K(x_i, x_i) = (1 + \sum_{j=1}^p x_{ij} x_{i,j})^d$  เมื่อ  $d$  คือจำนวนเต็มบวกใด ๆ แสดงถึงอันดับของพหุนาม หาก  $d$  เท่ากับ 1 จะกลายเป็นเคอร์เนลเชิงเส้น (linear kernel) ซึ่งหมายถึง support vector classifier ดังกล่าวข้างต้น



2. เคอร์เนลเรเดียล (radial basis function kernel): เคอร์เนลเรเดียลเรียกได้อีกอย่างหนึ่งว่า เคอร์

เนลเกาส์เซียน (Gaussian kernel) โดย  $K(x_i, x_{i'}) = \exp\left(-\frac{\sum_{j=1}^p (x_{ij} - x_{i'j})^2}{2\sigma^2}\right)$  โดยที่  $\sigma$  เป็น

พารามิเตอร์อิสระ โดยทั่วไปแล้ว จะกำหนดให้  $\gamma = \frac{1}{2\sigma^2}$  ซึ่งจะเขียนเคอร์เนลใหม่ได้เป็น

$$K(x_i, x_{i'}) = \exp\left(-\gamma \sum_{j=1}^p (x_{ij} - x_{i'j})^2\right) \text{ สำหรับ } \forall \gamma > 0$$

3. เคอร์เนลซิกมอยด์ (sigmoid kernel):  $K(x_i, x_{i'}) = \tanh\left(\kappa \sum_{j=1}^p x_{ij}x_{i'j} + r\right)$  สำหรับ  $\exists \kappa > 0$

โดย  $\kappa$  คือพารามิเตอร์สเกล (scale parameter) ของข้อมูล และ  $r$  คือพารามิเตอร์ที่ควบคุมระดับของการแปลง (Hsuan-tien & Chih-jen, 2013)

สำหรับการใช้ SVM จำแนกข้อมูลที่มี  $K$  class โดย  $K > 2$  แล้ว มีวิธีที่เป็นที่นิยม 2 วิธี ได้แก่

1. วิธี one-versus-one โดยจะสร้าง SVM ทั้งหมด  $\binom{K}{2}$  ตัวแบบ โดยแต่ละตัวแบบจะเทียบตัวแบบละ 2 class เช่น หากข้อมูลมี 3 class คือ ก, ข และ ค ก็จะสร้าง SVM ทั้งหมด  $\binom{3}{2} = 3$  ตัวแบบ คือ เทียบ ก กับ ข, เทียบ ก กับ ค และ เทียบ ข กับ ค จากนั้นจะนับจำนวนครั้งที่ข้อมูลแต่ละจุดถูกจำแนกให้แต่ละ class ผลลัพธ์การจำแนกสุดท้ายคือ class ที่ข้อมูลนั้นถูกจำแนกไปมากที่สุด

2. วิธี one-versus-all จะสร้าง SVM ทั้งหมด  $K$  ตัวแบบ โดยแต่ละตัวแบบจะเทียบระหว่าง 1 ใน  $K$  class นั้นกับ  $K-1$  class ที่เหลือ จากตัวอย่างข้างต้น จะได้ SVM ทั้งหมด 3 ตัวแบบ คือ เทียบ ก กับ ข และ ค, เทียบ ข กับ ก และ ค และ เทียบ ค กับ ก และ ข หากกำหนดให้  $\beta_{0k}, \beta_{1k}, \dots, \beta_{pk}$  แทนพารามิเตอร์ของ SVM เมื่อเทียบ class ที่  $k$  กับ class ที่เหลือ เวกเตอร์  $x^*$  จะถูกจำแนกไปยัง class ที่  $\beta_{0k} + \beta_{1k}x_1^* + \dots + \beta_{pk}x_p^*$  มีค่ามากที่สุด

Netzer et al. (2012) ใช้ SVM ในการจำแนกตัวเลข SVHN และเปรียบเทียบ features 4 แบบ ได้แก่ Histograms of Oriented Gradients, Weighted Direction Code Histogram, Stacked Sparse Auto-Encoders และ K-Means ได้อัตราการจำแนกถูกต้องเท่ากับ 85.0%, 63.3%, 89.7% และ 90.6% ตามลำดับ

### บทที่ 3 วิธีการดำเนินงานวิจัย

งานวิจัยนี้มีวัตถุประสงค์ในการรู้จำตัวเลขจากฐานข้อมูลภาพเลขที่บ้านจากกูเกิล สตรีตวิว (SVHN) โดยใช้การแปลงฮัฟในการค้นหารูปภาคตัดกรวยที่พบได้ในรูปตัวเลข ได้แก่ เส้นตรง วงรี และ พาราโบลา และนำค่าพารามิเตอร์ของรูปต่าง ๆ เหล่านี้ มาใช้เป็นตัวแปรอิสระในการจำแนกตัวเลข โดยตัวแบบที่ใช้ในการจำแนกตัวเลข คือ ซัพพอร์ตเวกเตอร์แมชชีน โดยเปรียบเทียบ 4 เคอร์เนล ได้แก่ เคอร์เนลเชิงเส้น, เคอร์เนลพหุนาม, เคอร์เนลเรเดียล และเคอร์เนลซิกมอยด์ นอกจากนี้ จะเปรียบเทียบความสำคัญของลักษณะที่ใช้ในการจำแนกตัวเลขว่ามีผลต่อการจำแนกตัวเลขแต่ละตัวอย่างไร โปรแกรมที่ใช้ในงานวิจัยนี้ คือ โปรแกรม R Version 3.3.3

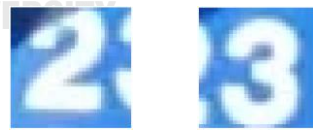
#### 3.1 แหล่งข้อมูล

ข้อมูลรูปภาพตัวเลขที่ใช้ในงานวิจัย นำมาจาก The Street View House Number Dataset (Netzer et al., 2012) ซึ่งประกอบด้วยตัวเลขจำนวน 73,257 ตัว สำหรับการสร้างตัวแบบ และ 26,032 ตัว สำหรับการทดสอบประสิทธิภาพของตัวแบบ

ชุดข้อมูลดังกล่าวประกอบด้วยรูปภาพ 2 รูปแบบ ได้แก่ แบบเต็มรูป (full numbers) และ แบบตัดส่วน (cropped digits) ขนาด 32 x 32 พิกเซล ในงานวิจัยนี้จะใช้เฉพาะแบบตัดส่วนเท่านั้น



ก. แบบเต็มรูป



ข. แบบตัดส่วน

ภาพที่ 3.1 ตัวอย่างรูปภาพ 2 รูปแบบ

ที่มา: ฐานข้อมูล SVHN

ดังที่ได้กล่าวไปแล้วว่า รูปภาพตัวเลขของ SVHN นั้นมีปัจจัยหลายประการที่ทำให้การจำแนกตัวเลขทำได้ยากกว่าข้อมูลตัวเลขของ MNIST รูปภาพดังต่อไปนี้ เป็นตัวอย่างรูปภาพที่มีประเด็นปัญหาเฉพาะ พร้อมกับสัดส่วนโดยประมาณต่อจำนวนรูปภาพทั้งหมด (ทั้งรูปภาพที่ใช้ในการสร้างตัวแบบและรูปภาพที่ใช้ทดสอบ รวม 99,289 รูป )

- 1) มีส่วนที่ไม่พึงประสงค์เป็นตัวรบกวนในภาพ (noise)

สัดส่วนโดยประมาณ: 50%



- 2) เงา

สัดส่วนโดยประมาณ: 5%



- 3) ตัวเลขจาง

สัดส่วนโดยประมาณ: 10%



หมายเหตุ ในรูปคือตัวเลข 2, 9, 3 และ 2 ตามลำดับ

- 4) ตัวเลขเอียง

สัดส่วนโดยประมาณ: 5%



- 5) ตัวเลขไม่ชัด

สัดส่วนโดยประมาณ: 5%



หมายเหตุ ในรูปคือตัวเลข 8, 3, 2 และ 6 ตามลำดับ

6) ตัวเลขบาง

สัดส่วนโดยประมาณ: 5%



7) ตัวเลขไม่อยู่ตรงกลาง

สัดส่วนโดยประมาณ: 0.1%



หมายเหตุ ในรูปคือตัวเลข 2, 7, 3 และ 1 ตามลำดับ

8) ตัวเลขผิด (หมายถึง ค่า  $y$  ที่ระบุว่าเป็นตัวเลขอะไรในฐานข้อมูลไม่ตรงกับรูปภาพ)

สัดส่วนโดยประมาณ: 0.3%

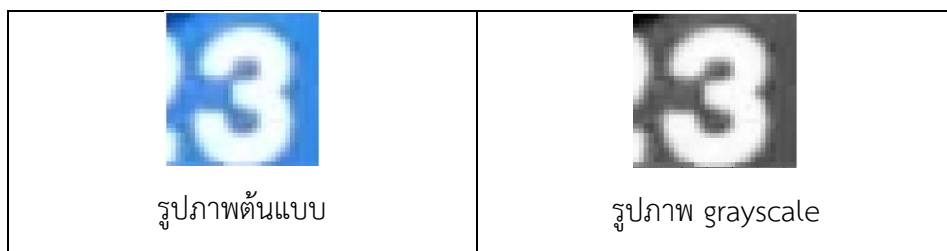


หมายเหตุ ค่า  $y$  ที่ระบุไว้ในฐานข้อมูลคือตัวเลข 4, 1, 9 และ 4 ตามลำดับ

### 3.2 ขั้นตอนในการดำเนินการวิจัย

#### 3.2.1 กระบวนการเตรียมข้อมูลก่อนการสร้างตัวแบบ (Preprocessing data)

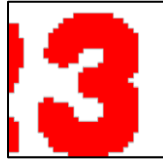
3.2.1.1 แปลงข้อมูลจากรูปสีซึ่งมี 3 channels (Red, Green และ Blue) ให้เป็น 1 channel (Grayscale)



ภาพที่ 3.2 การแปลงจากภาพสีเป็นภาพขาวดำ

3.2.1.2 rescale ค่าสีของพิกเซลจาก [0, 255] เป็น [0, 1]

3.2.1.3 ทำการแบ่งกลุ่มพิกเซลด้วย K-means clustering โดยใช้ค่าของสีที่ rescale แล้ว เป็นเกณฑ์ในการแบ่ง โดยกำหนดให้จำนวนกลุ่มเท่ากับ 2 กลุ่ม



ภาพที่ 3.3 รูปภาพตัวเลขหลังจากการทำ K-means clustering

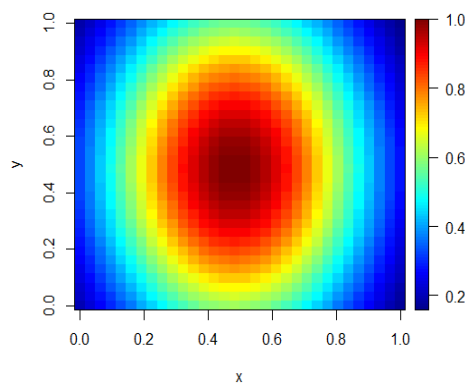
3.2.1.4 แปลงหมายเลข cluster จาก 1 และ 2 ให้กลายเป็น 0 (พื้นหลัง) กับ 1 (จุดภาพ) ตามลำดับ จากนั้น ทำให้เลข 1 ที่อยู่ติดกันเป็นกลุ่มเดียวกัน โดยใช้เทคนิคการกำหนดสัญลักษณ์การเชื่อมต่อขององค์ประกอบ (connected components labeling)



ภาพที่ 3.4 รูปภาพตัวเลขหลังจากการกำหนดสัญลักษณ์การเชื่อมต่อขององค์ประกอบ

3.2.1.5 เนื่องจากรูปภาพตัวเลขที่ทำการตัดส่วนแล้ว ตัวเลขจะอยู่ตรงกลางของภาพ ดังนั้น ในการกำจัด noise จะใช้วิธีให้น้ำหนักที่ขึ้นอยู่กับบริเวณภายในภาพ โดยกำหนดให้น้ำหนักมีการแจกแจงแบบปกติ 2 ตัวแปร (bivariate normal distribution) เนื่องจากรูปภาพมีขนาด 32 x 32 พิกเซล ดังนั้น ฟังก์ชันความหนาแน่นของความน่าจะเป็นของการแจกแจงแบบปกติ 2 ตัวแปรที่พิกเซลแถวที่  $i$  คอลัมน์ที่  $j$  หรือ  $f(i, j)$  จะกำหนดให้เวกเตอร์ของค่าเฉลี่ยเท่ากับ  $\begin{bmatrix} 16 \\ 16 \end{bmatrix}$  และเมทริกซ์ความแปรปรวนร่วมเท่ากับ  $\begin{bmatrix} 100 & 0 \\ 0 & 225 \end{bmatrix}$  และกำหนดตัวแปรน้ำหนัก  $w(i, j)$  โดยคำนวณจาก

$$\frac{f(i, j)}{\max f(i, j)}$$
 ลักษณะการกระจายของตัวแปรน้ำหนักในแต่ละพิกเซลแสดงดังภาพที่ 3.5



ภาพที่ 3.5 ลักษณะการกระจายของตัวแปรน้ำหนักในแต่ละพิกเซล

3.2.1.6 คำนวณน้ำหนักเฉลี่ยแยกตามแต่ละกลุ่มที่ได้จากการกำหนดสัญลักษณ์การเชื่อมต่อขององค์ประกอบ

3.2.1.7 ทำการ binarize อีกครั้งโดยกำหนดให้กลุ่มที่น้ำหนักเฉลี่ยมากที่สุดเป็นเลข 1 (รูปตัวเลข) ส่วนกลุ่มที่เหลือเป็นเลข 0 (พื้นหลัง)

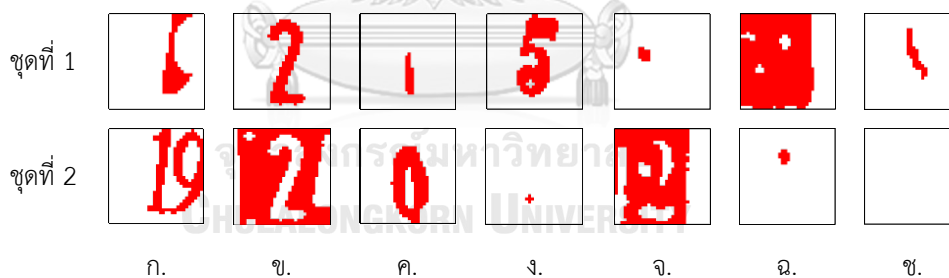


ภาพที่ 3.6 รูปภาพตัวเลขหลังจากกำหนดน้ำหนักแล้ว

3.2.1.8 เนื่องจากเราไม่สามารถกำหนดหมายเลข cluster ในขั้นตอนการทำ K-means clustering ได้ ดังนั้น ให้เริ่มทำขั้นตอนที่ 3.2.1.4 ซ้ำอีกครั้ง แต่ให้ cluster หมายเลข 2 เป็น 0 (พื้นหลัง) และ cluster หมายเลข 1 เป็น 1 (จุดภาพ)

3.2.2 กระบวนการคัดเลือกรูปภาพตัวเลข

หลังจากกำจัด noise และแปลงเป็นรูปภาพ binary แล้ว การทำ K-means clustering 2 ครั้งกับข้อมูลชุดเดียวกัน ทำให้ได้รูปภาพหลังเตรียมข้อมูล 2 ชุด แสดงตัวอย่างบางส่วนดังภาพที่ 3.7



ภาพที่ 3.7 ภาพตัวอย่างรูปภาพบางส่วนหลังจากกำจัด noise และแปลงเป็นรูปภาพ binary

จากตัวอย่างรูปภาพข้างต้น มีข้อสังเกต 2 ประการ คือ

1. เนื่องจากการเลือกกลุ่มของ K-means clustering เป็นแบบสุ่ม ดังนั้น รูปตัวเลขที่เหมาะสมแก่การนำไปตั้งลักษณะต่อจึงมีได้อยู่ในภาพชุดใดชุดหนึ่ง (จากรูปภาพตัวอย่าง ภาพที่เหมาะสมในชุดที่ 1 ได้แก่ ข. และ ง. ในขณะที่ภาพที่เหมาะสมในชุดที่ 2 ได้แก่ ค.)

2. มีรูปตัวเลขบางรูปที่แม้ว่าจะเตรียมข้อมูลก่อนแล้ว ก็ยังไม่เหมาะสมต่อการดึงลักษณะ เนื่องจากยังมี noise หลงเหลืออยู่หรือส่วนที่เป็นตัวเลขถูกกำจัดออกไปด้วย (จากรูปภาพตัวอย่าง ได้แก่ ก., ฉ. และ ช.) นั่นคือ ในรูปภาพทั้ง 2 ชุด ไม่มีรูปที่เหมาะสมแก่การนำไปดึงลักษณะเลย

จากข้อสังเกตทั้ง 2 ประการนี้ จึงต้องทำการคัดเลือกรูปภาพที่จะนำไปใช้จำแนก ในที่นี้ จะใช้การถดถอยโลจิสติกทวิ (binary logistic regression) ในการสร้างตัวแบบสำหรับการคัดเลือกรูปภาพ โดยมีขั้นตอนดังนี้

3.2.2.1 คัดเลือกรูปภาพตัวเลข ทั้งแบบที่เหมาะสมและไม่เหมาะสมต่อการดึงลักษณะ ทั้งจากชุดที่ 1 และชุดที่ 2 เพื่อนำมาใช้สร้างตัวแบบถดถอยโลจิสติกทวิ ในที่นี้ คัดเลือกมาทั้งหมด 4,852 รูป โดยแบ่งเป็นรูปที่เหมาะสมและไม่เหมาะสมอย่างละเท่ากัน คือ อย่างละ 2,426 รูป

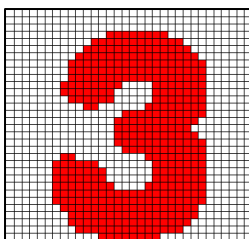
3.2.2.2 รายละเอียดของตัวแปรที่ใช้ในการสร้างตัวแบบมีดังนี้

ก. ตัวแปรอิสระ ได้แก่

1. จำนวนของพิกเซลที่มีค่าเท่ากับ 1 ในรูปภาพ (กล่าวคือ จำนวนของพิกเซลที่เป็นรูปตัวเลข) แทนด้วย  $S$
2. จำนวนของพิกเซลที่มีค่าเท่ากับ 1 ที่อยู่ติดกับขอบของรูปภาพ แทนด้วย  $R$
3. ผลรวมของน้ำหนัก  $w(i, j)$  ที่คำนวณได้ในขั้นตอนที่ 3.2.1.3.3 ของพิกเซลที่มีค่าเท่ากับ 1 แทนด้วย  $SW$
4. Interaction term ระหว่างตัวแปรแต่ละคู่ ได้แก่  $S \cdot R$ ,  $R \cdot SW$  และ  $S \cdot SW$
5. Interaction term ระหว่างตัวแปรทั้งสามตัว หรือ  $S \cdot R \cdot SW$

ข. ตัวแปรตาม ( $Y$ ) แสดงถึงความเหมาะสมของรูปภาพสำหรับการนำไปใช้ดึงลักษณะ โดยกำหนดให้ 0 คือไม่เหมาะสม และ 1 คือเหมาะสม

ภาพที่ 3.8 และตารางที่ 3.1 แสดงตัวอย่างรูปภาพและค่าของตัวแปรแต่ละตัวที่ใช้ในการสร้างตัวแบบความถดถอยโลจิสติกทวิ



ภาพที่ 3.8 ตัวอย่างรูปภาพพร้อมกับเส้นแสดงพิกเซล

ตารางที่ 3.1 ตัวอย่างค่าของตัวแปรที่ใช้ในการสร้างตัวแบบความถดถอยโลจิสติกทวิของตัวอย่างรูป  
ตัวเลขในภาพที่ 3.8

$S$	$R$	$SW$	$S \cdot R$	$R \cdot SW$	$S \cdot SW$	$S \cdot R \cdot SW$	$Y$
402	8	295.04	3,216	2,360.32	118,606.08	948,848.64	1

3.2.2.3 ตัวประมาณพารามิเตอร์ของตัวแบบความถดถอยโลจิสติกทวิได้ผลลัพธ์ดังนี้

$$\ln\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = -6.81 + 0.21S - 0.19R - 0.11SW - 0.02S \cdot R + 0.02R \cdot SW \\ - 0.0003S \cdot SW + 0.00001S \cdot R \cdot SW$$

สำหรับเกณฑ์ในการคัดเลือกรูปภาพ จะพิจารณาจากความน่าจะเป็นที่รูปภาพนั้นเหมาะสม  
ต่อการนำไปใช้ หรือ  $P(Y=1)$  โดยต้องเข้าเกณฑ์ทั้ง 2 ข้อดังนี้

1. เปรียบเทียบค่าความน่าจะเป็นของรูปภาพทั้ง 2 ชุด และเลือกชุดที่ให้ค่าความน่าจะเป็น  
สูงกว่า
2. ค่าความน่าจะเป็นที่ได้จากข้อ 1 ต้องมีค่ามากกว่าหรือเท่ากับ 0.9

เมื่อทำการคัดเลือกรูปภาพตัวเลขแล้ว จะได้รูปภาพสำหรับการสร้างตัวแบบจำนวน 41,180  
รูป (คิดเป็น 56.21% ของจำนวนรูปที่ใช้สร้างตัวแบบทั้งหมด) และรูปภาพสำหรับทดสอบ  
ประสิทธิภาพของตัวแบบจำนวน 16,071 รูป (คิดเป็น 61.74% ของจำนวนรูปที่ใช้ทดสอบ  
ประสิทธิภาพของตัวแบบทั้งหมด) จำนวนรูปภาพตัวเลขหลังจากการคัดเลือกแล้ว แยกตามแต่ละ  
ตัวเลข แสดงดังตารางที่ 3.2



ตารางที่ 3.2 จำนวนรูปภาพตัวเลขหลังจากการคัดเลือกแล้ว แยกตามแต่ละตัวเลข

ตัวเลข	จำนวนรูปภาพสำหรับสร้างตัวแบบ	สัดส่วน	จำนวนรูปภาพสำหรับทดสอบประสิทธิภาพของตัวแบบ	สัดส่วน
1	7,083	17.2%	2,669	16.6%
2	5,592	13.6%	2,577	16.0%
3	4,698	11.4%	1,843	11.5%
4	4,503	10.9%	1,740	10.8%
5	3,914	9.5%	1,527	9.5%
6	3,445	8.4%	1,250	7.8%
7	3,029	7.4%	1,271	7.9%
8	3,092	8.5%	1,088	6.8%
9	2,804	6.8%	1,059	6.6%
0	3,020	7.3%	1,047	6.5%
<b>รวม</b>	<b>41,180</b>		<b>16,071</b>	

### 3.2.3 กระบวนการดึงลักษณะ (Features extraction)

3.2.3.1 ทำการค้นหาขอบ (edge detection) ของตัวเลขในรูปภาพ โดยใช้ Sobel operator ตัวอย่างรูปภาพตัวเลขหลังค้นหาขอบแสดงดังภาพที่ 3.9



ภาพที่ 3.9 ตัวอย่างรูปภาพตัวเลขหลังค้นหาขอบ

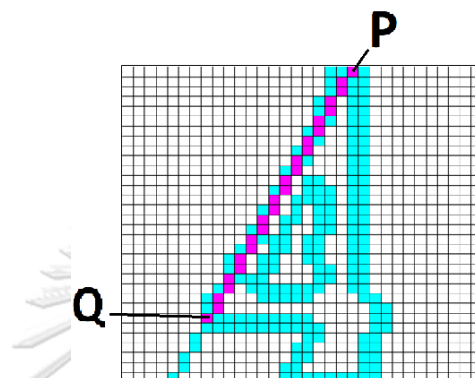
3.2.3.2 กำหนดเกณฑ์การดึงลักษณะด้วยการแปลงฮัฟในการค้นหารูปภาพตัดกรวย กำหนดให้เซลล์ในแถวที่ 1 คอลัมน์ที่ 1 แทนจุดกำเนิดบนระนาบคาร์ทีเซียน การค้นหาแบ่งออกเป็น 3 รูปร่างหลัก ได้แก่ เส้นตรง วงรี และพาราโบลา

#### ก. เส้นตรง

รูปตัวเลขที่มีเส้นตรงมาเกี่ยวข้อง ได้แก่ 1, 2, 4, 5 และ 7 จะเห็นว่า จำนวนเส้นตรงที่มากที่สุดที่เป็นไปได้คือ 3 เส้น (เลข 4) ดังนั้น จะกำหนดให้ค้นหาเส้นตรง 3 เส้น ได้แก่

- 1) เส้นตรงแนวนอน ( $\theta$  เท่ากับ 90 องศา)
- 2) เส้นตรงแนวตั้ง ( $\theta$  เท่ากับ 0 องศา)
- 3) เส้นตรงแนวเฉียง ( $\theta$  มีค่าตั้งแต่ 11 ถึง 80 องศา)

พารามิเตอร์ของเส้นตรงแต่ละเส้นที่ต้องการ ได้แก่  $\theta$  (เฉพาะเส้นตรงแนวเฉียง),  $\rho$  และความยาวของเส้นตรง การจะคำนวณความยาวของเส้นตรงได้ จำเป็นต้องทราบพารามิเตอร์ปกติคือ  $\theta$  และ  $\rho$  ก่อน เมื่อได้ค่าพารามิเตอร์ปกติแล้ว นำมาสร้างเส้นตรงในรูปภาพตัวเลข ความยาวของเส้นตรงนั้นจะเท่ากับระยะทางระหว่างจุดเริ่มต้นกับจุดสิ้นสุดของส่วนที่ทับซ้อนกับรูปตัวเลข แสดงตัวอย่างดังภาพที่ 3.10



ภาพที่ 3.10 ตัวอย่างการคำนวณหาความยาวเส้นตรงแนวเฉียง

จากภาพที่ 3.10 ตัวเลขสาธิตประกอบคือตัวเลข 4 โดยเส้นตรงแนวเฉียงที่พบในรูปมีค่า  $\theta$  เท่ากับ 26 และ  $\rho$  เท่ากับ 19 เมื่อวาดเส้นตรงทับลงไปบนรูปตัวเลขแล้ว จะได้ว่า จุด P และจุด Q เป็นจุดปลายของส่วนที่ทับซ้อนกันระหว่างรูปตัวเลขและรูปเส้นตรงที่สร้างขึ้น ความยาวของเส้นตรงที่จะนำไปใช้เป็นลักษณะจะคำนวณจากระยะทางระหว่างจุด P และจุด Q

#### ข. วงรี จุฬาลงกรณ์มหาวิทยาลัย

รูปตัวเลขที่มีวงรีเข้ามาเกี่ยวข้อง ได้แก่ 0, 6, 8 และ 9 จะเห็นว่าตำแหน่งของวงรีของแต่ละตัวเลขไม่ซ้ำกัน กล่าวคือ เลข 6 วงรีอยู่ทางครึ่งล่างของภาพ, เลข 9 วงรีอยู่ทางครึ่งบนของภาพ, เลข 8 วงรีอยู่ทั้งครึ่งบนและครึ่งล่าง และเลข 0 เป็นวงรีขนาดเต็มรูปภาพ ดังนั้น จะกำหนด scale ในการค้นหาวงรี 3 วง คือ

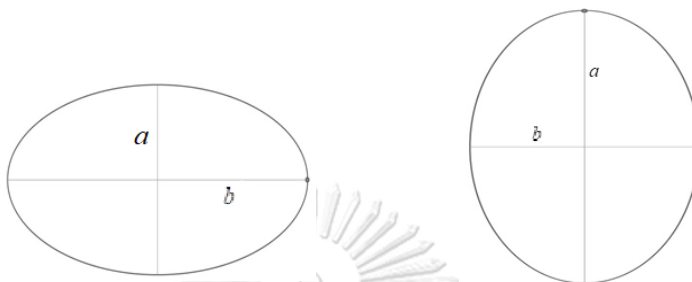
- 1) วงรีขนาดเต็มรูปภาพ
- 2) วงรีบนรูปภาพครึ่งบน
- 3) วงรีบนรูปภาพครึ่งล่าง

การค้นหาวงรีในที่นี้ จะใช้สมการพาราเมตริกของวงรี คือ

$$x = x_0 + b \sin \theta$$

$$y = y_0 + a \cos \theta$$

ซึ่งจะนิยาม  $a$  และ  $b$  ใหม่ โดย  $a$  หมายถึง ความยาวของเส้นที่ลากจากจุดศูนย์กลางวงรีไปยังจุดขอบของวงรีซึ่งขนานกับแกน  $y$  และ  $b$  หมายถึง ความยาวของเส้นที่ลากจากจุดศูนย์กลางวงรีไปยังจุดขอบของวงรีซึ่งขนานกับแกน  $x$



ภาพที่ 3.11 พารามิเตอร์ของวงรีที่นิยามใหม่

การนิยาม  $a$  และ  $b$  ใหม่นี้ ช่วยให้การค้นหาวงรีไม่ต้องแบ่งเป็นวงรีที่แกนเอกขนานกับแกน  $x$  และวงรีที่แกนเอกขนานกับแกน  $y$

สำหรับเรนจ์ของ  $a$  และ  $b$  ที่ต้องการจะค้นหา จะแตกต่างกันตาม scale ที่แบ่งไว้ ซึ่งแต่ละรูปมีเกณฑ์ดังนี้

#### 1) วงรีขนาดเต็มรูปภาพ

เรนจ์ของ  $a$  และ  $b$  กำหนดให้ค้นหาตั้งแต่ 1 ถึง 16 อย่างไรก็ตาม เพื่อลดเวลาในการประมวลผล ดังนั้น จะใช้ลักษณะของเลข 0 เป็นสมมติฐานในการกำหนดเงื่อนไขเพิ่มเติม คือ  $1 \leq \frac{a}{b} \leq 3$  และ  $\frac{1}{4} \leq \frac{\pi ab}{1024} \leq \frac{3}{4}$  เมื่อ  $\pi ab$  คือพื้นที่ของวงรี และ 1024 คือจำนวนพิกเซลต่อ 1 ภาพ (32 x 32 พิกเซล)

#### 2) วงรีบนรูปภาพครึ่งบน

กำหนดให้ค้นหา  $a$  และ  $b$  ตั้งแต่ 1 ถึง 8 และใช้ลักษณะครึ่งบนของเลข 8 และ 9 เป็นสมมติฐานในการกำหนดเงื่อนไขเพิ่มเติม คือ  $1 \leq \frac{a}{b} \leq 3$  และ  $\frac{1}{6} \leq \frac{\pi ab}{512} \leq \frac{1}{2}$  เมื่อ 512 คือจำนวนพิกเซลต่อครึ่งรูปภาพ (16 x 32 พิกเซล)

#### 3) วงรีบนรูปภาพครึ่งล่าง

กำหนดให้ค้นหา  $a$  และ  $b$  ตั้งแต่ 1 ถึง 8 และใช้ลักษณะครึ่งล่างของเลข 6 และ 8 เป็นสมมติฐานในการกำหนดเงื่อนไขเพิ่มเติม คือ  $1 \leq \frac{a}{b} \leq 3$  และ  $\frac{1}{6} \leq \frac{\pi ab}{512} \leq \frac{1}{2}$

พารามิเตอร์ของวงรีแต่ละวงที่ต้องการ ได้แก่ จุดศูนย์กลางวงรี  $(x_0, y_0)$ ,  $a$  และ  $b$  (วงกลมคือกรณีเฉพาะของวงรีที่  $a$  เท่ากับ  $b$  ดังนั้น ในงานวิจัยนี้จะค้นหาเฉพาะรูปร่างวงรี)

### ค. พาราโบลา

รูปตัวเลขที่มีพาราโบลาเข้ามาเกี่ยวข้อง ได้แก่ 2, 3, 5, 6 และ 9 จะเห็นว่าตำแหน่งของรูปพาราโบลาบนตัวเลขจะมีทั้งแบบที่อยู่ครึ่งบนของภาพ (เลข 2 และเลข 6), อยู่ครึ่งล่างของภาพ (เลข 5 และเลข 9) หรืออยู่ทั้งครึ่งบนและครึ่งล่าง (เลข 3) ดังนั้น จะกำหนด scale ในการค้นหาพาราโบลา 2 ส่วน คือ

1) พาราโบลารูปภาพครึ่งบน

2) พาราโบลารูปภาพครึ่งล่าง

ในการค้นหารูปพาราโบลา จะค้นหาทั้งพาราโบลาแนวตั้ง (แกนสมมาตรขนานแกน  $y$ ) และพาราโบลาแนวนอน (แกนสมมาตรขนานแกน  $x$ ) โดยเรนจ์ของ  $p$  (ความยาวโฟกัส) จะแตกต่างกันตามรูปพาราโบลา ดังนี้

1) พาราโบลาแนวตั้ง

อาศัยลักษณะครึ่งบนของเลข 2 กับเลข 6 และครึ่งล่างของเลข 9 เป็นสมมติฐานในการกำหนดเรนจ์ของ  $p$  โดยให้หา  $p$  ตั้งแต่ -3 ถึง 3 เนื่องจากเรนจ์ไม่กว้างมากนัก ดังนั้น จะแบ่งช่วง  $[-3, 3]$  ออกเป็น 20 ช่วงเท่า ๆ กัน เพื่อให้การแปลงฮัฟฟ์แมนละเอียดขึ้น และกำหนดเงื่อนไขเพิ่มเติมคือ  $p \geq 1$  หรือ  $p \leq -1$  โดยใช้เงื่อนไขเดียวกันทั้งรูปภาพครึ่งบนและรูปภาพครึ่งล่าง

2) พาราโบลาแนวนอน

เนื่องจากรูปพาราโบลาแนวนอนที่ปรากฏในเลข 3 และเลข 5 มีความยาวโฟกัสไม่มากนัก ดังนั้น จะกำหนดเรนจ์ของ  $p$  โดยให้หา  $p$  ตั้งแต่ -2 ถึง 2 และแบ่งช่วง  $[-2, 2]$  ออกเป็น 10 ช่วงเท่า ๆ กัน (เนื่องจากเรนจ์แคบกว่าพาราโบลาแนวตั้ง) และกำหนดเงื่อนไขเพิ่มเติมคือ  $p \geq 1$  หรือ  $p \leq -1$  โดยใช้เงื่อนไขเดียวกันทั้งรูปภาพครึ่งบนและรูปภาพครึ่งล่างพารามิเตอร์ของพาราโบลาที่ต้องการ ได้แก่ จุดโฟกัส  $(x_f, y_f)$  และ  $p$

นอกจากพารามิเตอร์พื้นฐานดังกล่าวแล้ว จะใช้อัตราส่วนของการทับซ้อน (overlapping ratio; OR) ของรูปตัวเลขกับรูปภาคตัดกรวยเป็นลักษณะด้วย โดยคำนวณดังนี้

$$OR = \frac{\text{จำนวนพิกเซลที่อยู่ทั้งในรูปตัวเลขและรูปภาคตัดกรวย}}{\text{จำนวนพิกเซลที่อยู่ในรูปภาคตัดกรวย}}$$

เมื่อทำการค้นหารูปร่างตามเกณฑ์ดังกล่าวแล้ว จะได้พารามิเตอร์ทั้งหมด 41 ค่า ดังตารางที่

3.3

ตารางที่ 3.3 พารามิเตอร์ที่ใช้อธิบายลักษณะของรูปตัวเลขแต่ละรูป

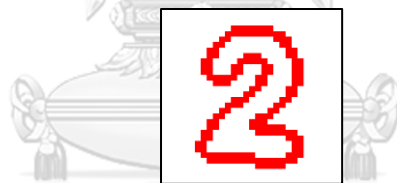
รูปภาคตัดกรวย	พารามิเตอร์ที่ใช้อธิบายลักษณะ ของรูปตัวเลขแต่ละรูป	จำนวนรูปภาคตัด กรวยที่ค้นหา	จำนวน พารามิเตอร์รวม
เส้นตรง	$\theta$ , $\rho$ , ความยาว และ OR	3	10*
พาราโบลาแนวนอน	$x_f$ , $y_f$ , $p$ และ OR	2	8
พาราโบลาแนวตั้ง	$x_f$ , $y_f$ , $p$ และ OR	2	8
วงรี	$x_0$ , $y_0$ , $a$ , $b$ และ OR	3	15

\*พารามิเตอร์  $\theta$  ของเส้นตรง จะหาเฉพาะของเส้นตรงแนวเฉียงเท่านั้น เนื่องจากเส้นตรงแนวตั้งและแนวนอนมีค่า  $\theta$  เท่ากับ 0 และ 90 องศา ตามลำดับ เท่ากันทุกภาพ ดังนั้น จำนวนพารามิเตอร์สำหรับเส้นตรงจึงเหลือ 10 ค่า

3.2.3.3 ทำการดึงลักษณะจากรูปภาพตัวเลข ทั้งข้อมูลที่ใช้ในการสร้างตัวแบบและข้อมูลที่ใช้ทดสอบประสิทธิภาพของตัวแบบ

ตัวอย่างค่าพารามิเตอร์จากรูปภาคตัดกรวยของรูปตัวเลขดังภาพที่ 3.12 แสดงดังตารางที่

3.4



ภาพที่ 3.12 ตัวอย่างรูปตัวเลขสำหรับสกัดตัวอย่างค่าพารามิเตอร์จากภาคตัดกรวย

ตารางที่ 3.4 ตัวอย่างค่าพารามิเตอร์จากรูปภาคตัดกรวยของรูปตัวเลขในภาพที่ 3.12

รูปภาคตัดกรวย	พารามิเตอร์	รูปแบบ / ตำแหน่ง	ค่าของพารามิเตอร์	
เส้นตรง	$\rho$	เส้นตรงแนวเฉียง	23	
		เส้นตรงแนวตั้ง	26	
		เส้นตรงแนวนอน	32	
	$\theta$	เส้นตรงแนวเฉียง	45	
		ความยาว	เส้นตรงแนวเฉียง	28.28
			เส้นตรงแนวตั้ง	22
	เส้นตรงแนวนอน		13	
	OR	เส้นตรงแนวเฉียง	0.31	
		เส้นตรงแนวตั้ง	0.69	
เส้นตรงแนวนอน		0.44		
พาราโบลาแนวนอน	$p$	ครึ่งบนของรูป	1.11	
		ครึ่งล่างของรูป	-1.11	
	$y_f$	ครึ่งบนของรูป	10	
		ครึ่งล่างของรูป	23	
	$x_f$	ครึ่งบนของรูป	9	
		ครึ่งล่างของรูป	18	
OR	ครึ่งบนของรูป	0.48		
	ครึ่งล่างของรูป	0.40		
พาราโบลาแนวตั้ง	$p$	ครึ่งบนของรูป	1.11	
		ครึ่งล่างของรูป	1.11	
	$y_f$	ครึ่งบนของรูป	13	
		ครึ่งล่างของรูป	25	
	$x_f$	ครึ่งบนของรูป	11	
		ครึ่งล่างของรูป	16	
OR	ครึ่งบนของรูป	0.43		
	ครึ่งล่างของรูป	0.38		
วงรี	$a$	เต็มรูป	14	
		ครึ่งบนของรูป	6	
		ครึ่งล่างของรูป	7	
	$b$	เต็มรูป	6	
		ครึ่งบนของรูป	5	
		ครึ่งล่างของรูป	6	
	$y_0$	เต็มรูป	18	
		ครึ่งบนของรูป	10	
		ครึ่งล่างของรูป	25	
	$x_0$	เต็มรูป	21	
		ครึ่งบนของรูป	14	
		ครึ่งล่างของรูป	13	
OR	เต็มรูป	0.61		
	ครึ่งบนของรูป	0.68		
	ครึ่งล่างของรูป	0.62		

### 3.2.4 กระบวนการสร้างตัวแบบจำแนกตัวเลขและทดสอบประสิทธิภาพของตัวแบบ

งานวิจัยนี้ใช้ SVM เป็นตัวแบบในการจำแนกตัวเลข โดยเปรียบเทียบ 4 เคอร์เนล ได้แก่ เคอร์เนลเชิงเส้น, เคอร์เนลพหุนาม, เคอร์เนลเรเดียล และเคอร์เนลซิกมอยด์ มีขั้นตอนในการสร้างตัวแบบดังนี้

3.2.4.1 ปรับค่าพารามิเตอร์ (tuning parameters) สำหรับ SVM แต่ละเคอร์เนล ด้วยวิธี 5-fold cross validation โดยใช้ชุดข้อมูลรูปภาพตัวเลขที่ใช้ในการสร้างตัวแบบ ค่าพารามิเตอร์แต่ละเคอร์เนลที่ต้องปรับมีดังนี้

ก. เคอร์เนลพหุนาม ได้แก่ ดีกรี ( $d$ ) ทำการ tuning ที่  $d$  เท่ากับ 5, 10 และ 20

ข. เคอร์เนลเรเดียล ได้แก่ แกมมา ( $\gamma$ ) ทำการ tuning ที่  $\gamma$  เท่ากับ 0.01, 0.1 และ 1

ค. เคอร์เนลซิกมอยด์ ได้แก่ แคปปา ( $\kappa$ ) ทำการ tuning ที่  $\kappa$  เท่ากับ 0.01, 0.1 และ 1

สำหรับพารามิเตอร์ cost จะทำการ tuning ที่  $C$  เท่ากับ 1, 10 และ 100 ทุกเคอร์เนล

3.2.4.2 สร้างตัวแบบ SVM ในการจำแนกตัวเลข โดยใช้ชุดข้อมูลรูปภาพตัวเลขที่ใช้ในการสร้างตัวแบบจำนวน 41,180 รูป และกำหนดค่าพารามิเตอร์ที่ได้จากการ tuning สำหรับแต่ละเคอร์เนล

3.2.4.3 ใช้ข้อมูลทดสอบจำนวน 16,071 รูปในการวัดประสิทธิภาพของตัวแบบ SVM แต่ละเคอร์เนล โดยมีตัวชี้วัดดังนี้

ก. อัตราการจำแนกถูกต้อง คำนวณจาก 
$$\frac{\text{จำนวนข้อมูลทดสอบที่จำแนกถูกต้อง}}{\text{จำนวนข้อมูลทดสอบทั้งหมด}}$$

ข. ระยะเวลาที่ใช้ในการสร้างตัวแบบ

3.2.4.4 เปรียบเทียบผลลัพธ์ของแต่ละเคอร์เนล นำตัวแบบที่ให้ผลลัพธ์ดีที่สุดมาวัดความสำคัญของลักษณะที่มีผลต่อการจำแนกตัวเลข

### 3.2.5 กระบวนการวัดความสำคัญของลักษณะ

ตัวชี้วัดความสำคัญของลักษณะแต่ละตัวจะวัดจากการลดลงของอัตราการจำแนกถูกต้องของตัวแบบที่มีลักษณะครบทุกตัว (กล่าวคือ ตัวแบบที่ให้อัตราการจำแนกถูกต้องสูงที่สุดจากขั้นตอนที่ 3.2.4.4) เมื่อเทียบกับอัตราการจำแนกถูกต้องของตัวแบบที่ไม่มีลักษณะนั้น ยิ่งลดลงมาก แสดงว่าลักษณะนั้นมีความสำคัญต่อการจำแนกตัวเลข ขั้นตอนการวัดความสำคัญของลักษณะมีดังนี้

3.2.5.1 สร้างตัวแบบขึ้นใหม่ (โดยที่พารามิเตอร์ของ SVM คงเดิม) โดยรูปภาพที่จะใช้สร้างตัวแบบตัดลักษณะออกไป 1 ตัว

3.2.5.2 ทดสอบประสิทธิภาพของตัวแบบใหม่โดยรูปภาพที่ใช้ทดสอบก็ตัดลักษณะตัวเดียวกันออก

3.2.5.3 วัดการลดลงของอัตราการจำแนกถูกต้องของตัวแบบที่มีลักษณะครบทุกตัวเมื่อเทียบกับอัตราการจำแนกถูกต้องของตัวแบบที่ไม่มีลักษณะนั้น

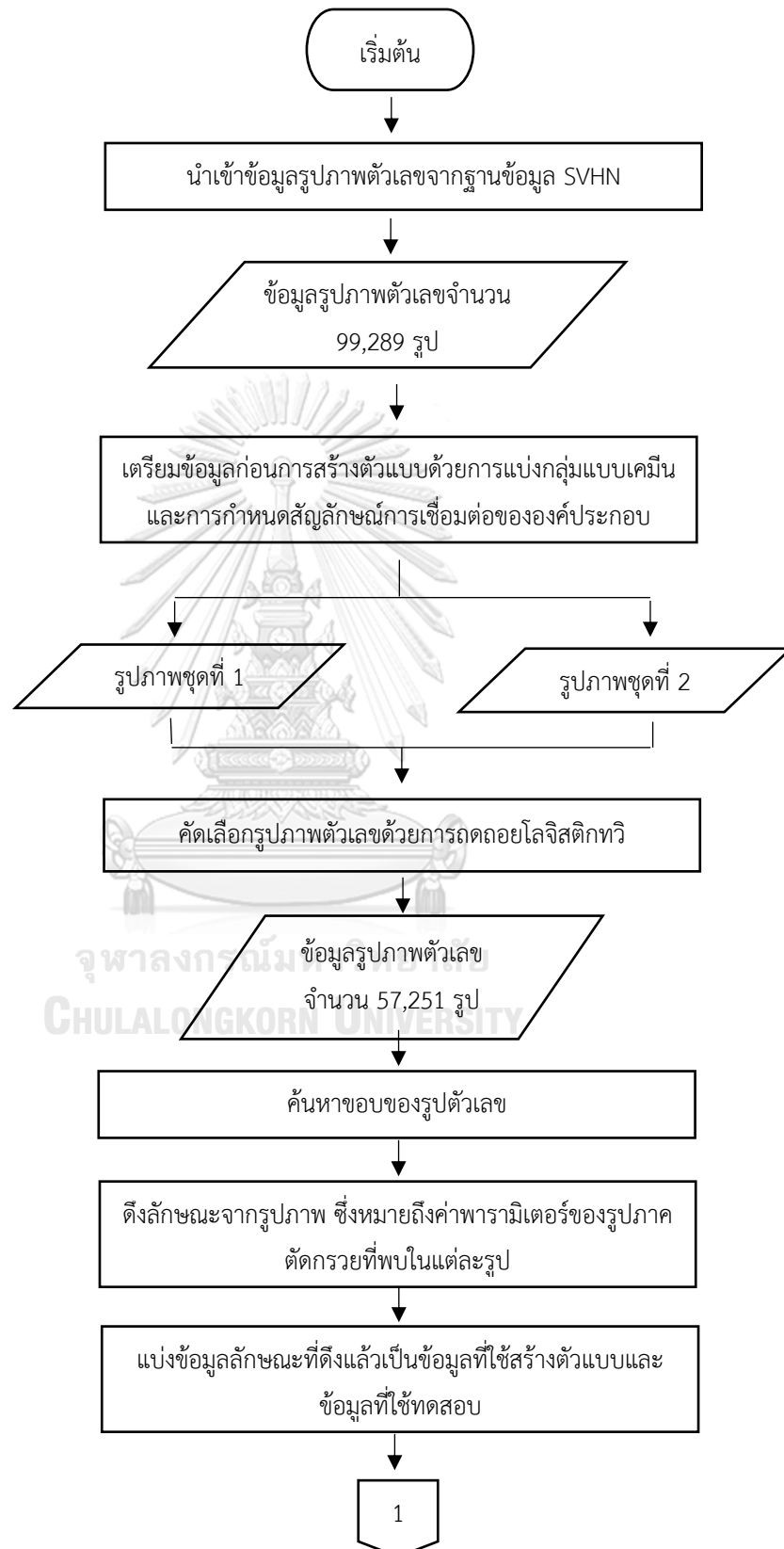
3.2.5.4 ทำเช่นนี้กับลักษณะทุกตัว

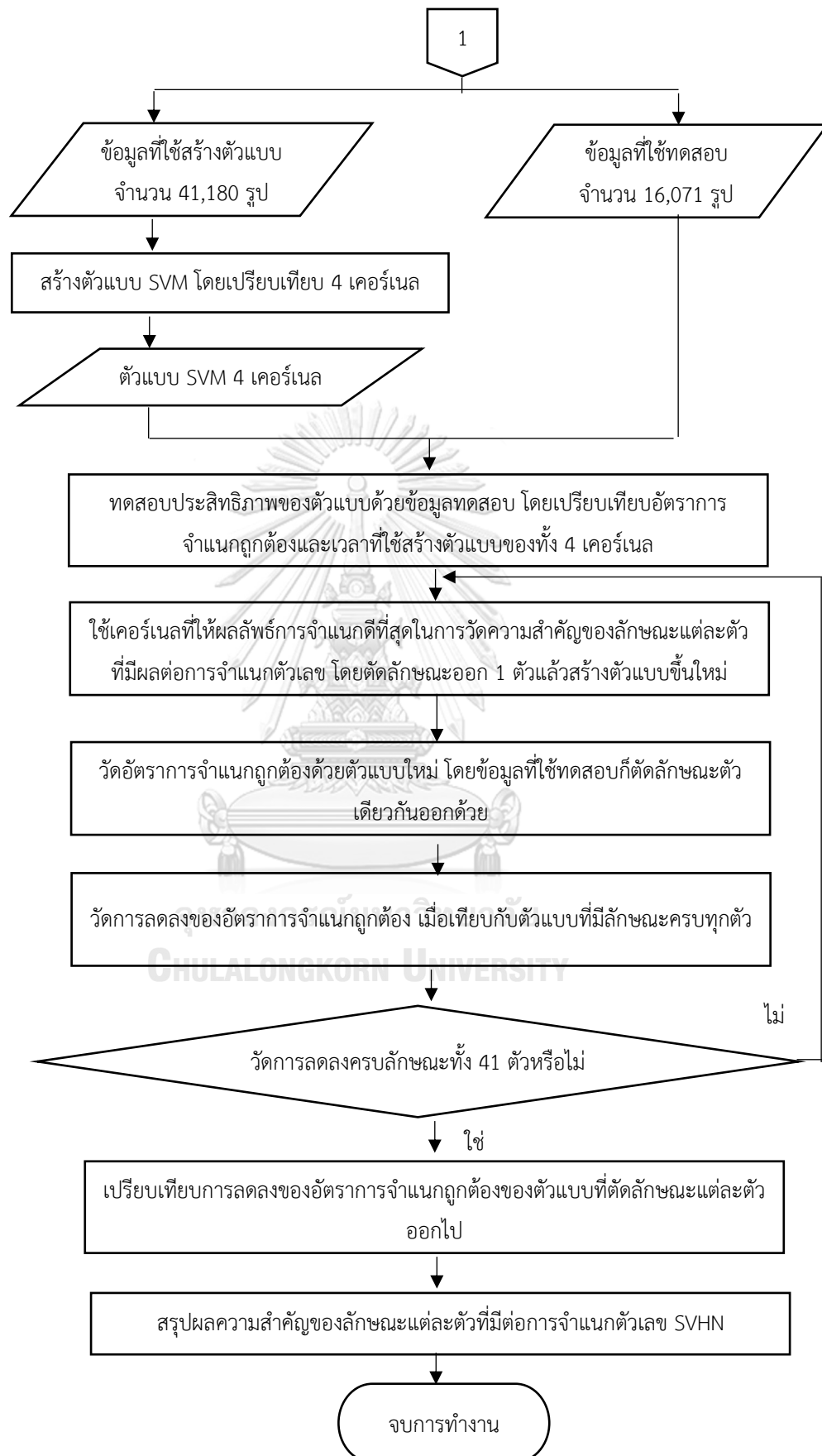
3.2.5.5 เปรียบเทียบอัตราการลดลงของอัตราการจำแนกถูกต้องของลักษณะแต่ละตัว รวมถึงวิเคราะห์และสรุปผลว่า ลักษณะใดบ้างที่มีความสำคัญต่อการจำแนกตัวเลข SVHN





### 3.3 ขั้นตอนการทำงานของโปรแกรม





## บทที่ 4

### ผลการวิจัย

งานวิจัยนี้ทดสอบประสิทธิภาพของการจำแนกตัวเลขจากฐานข้อมูล SVHN ด้วย SVM โดยใช้พารามิเตอร์ที่เกี่ยวข้องกับภาคตัดกรวยที่พบในรูปตัวเลขเป็นลักษณะในการสร้างตัวแบบ ผลการวิจัยจะแบ่งออกเป็น 3 ส่วน ได้แก่

- 1) ผลการปรับค่าพารามิเตอร์ของตัวแบบ SVM ทั้ง 4 เคอร์เนลด้วยวิธี 5-fold cross validation
- 2) ผลการเปรียบเทียบประสิทธิภาพของตัวแบบ SVM ในการจำแนกตัวเลข
- 3) ผลการเปรียบเทียบความสำคัญของลักษณะที่มีผลต่อการจำแนกตัวเลข

#### 4.1 ผลการปรับค่าพารามิเตอร์ของตัวแบบ SVM ทั้ง 4 เคอร์เนลด้วยวิธี 5-fold cross validation

จากการทำ tuning ค่าพารามิเตอร์ที่ใช้ในแต่ละเคอร์เนลของตัวแบบ SVM ด้วยวิธี 5-fold cross validation ได้ผลลัพธ์การ tuning ดังตารางที่ 4.1

ตารางที่ 4.1 ค่าพารามิเตอร์ที่ดีที่สุดของตัวแบบ SVM ทั้ง 4 เคอร์เนล

เคอร์เนล	$C$	$d$	$\gamma$	$\kappa$
เชิงเส้น	1	NA	NA	NA
พหุนาม	100	5	NA	NA
เรเดียล	10	NA	0.01	NA
ซิกมอยด์	1	NA	NA	0.01

#### 4.2 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบ SVM ในการจำแนกตัวเลข

จากการทดสอบประสิทธิภาพของตัวแบบโดยใช้ค่าพารามิเตอร์จากตารางที่ 4.1 ผลลัพธ์การจำแนกตัวเลขและเวลาที่ใช้สร้างตัวแบบแสดงดังตารางที่ 4.2

ตารางที่ 4.2 อัตราการจำแนกถูกต้องและเวลาที่ใช้ในการสร้างตัวแบบของ SVM ทั้ง 4 เคอร์เนล

เคอร์เนล	อัตราการจำแนกถูกต้อง (%)	เวลาที่ใช้ในการสร้างตัวแบบ* (นาที)
เชิงเส้น	63.42	40.73
พหุนาม	67.41	45.13
เรเดียล	72.17	22.36
ซิกมอยด์	56.26	28.44

\*ระบบประมวลผลที่ใช้ในงานวิจัยนี้คือ Intel® Core™ i7-6700HQ CPU@2.60GHz RAM 8.00GB

เคอร์เนลที่ให้ผลลัพธ์อัตราการจำแนกถูกต้องสูงสุด คือเคอร์เนลเรเดียล โดยมีอัตราการจำแนกถูกต้องเท่ากับ 72.17% ขณะที่เคอร์เนลซิกมอยด์ให้อัตราการจำแนกต่ำที่สุด เท่ากับ 56.26%

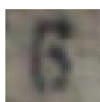
ในส่วนของเวลาที่ใช้ในการสร้างตัวแบบนั้น เคอร์เนลเรเดียลใช้เวลาในการสร้างตัวแบบน้อยที่สุด โดยใช้เวลาในสร้างตัวแบบเท่ากับ 22.36 นาที ขณะที่เคอร์เนลพหุนามใช้เวลาในการสร้างตัวแบบนานที่สุด เท่ากับ 45.13 นาที

เนื่องจากเคอร์เนลเรเดียลให้ผลลัพธ์ในการจำแนกดีที่สุดในแง่ของความถูกต้องและความเร็ว ดังนั้น ในการวิเคราะห์ต่อไปจะใช้เคอร์เนลเรเดียลเป็นตัวแบบหลัก Confusion matrix ของการจำแนกแต่ละตัวเลขสำหรับเคอร์เนลเรเดียลแสดงดังตารางที่ 4.3

ตารางที่ 4.3 Confusion Matrix สำหรับเคอร์เนลเรเดียล

		ตัวเลขที่ตัวแบบทำนาย (Predicted class)									
		1	2	3	4	5	6	7	8	9	0
ตัวเลขจริง (Actual class)	1	2265	79	58	56	26	25	54	36	27	43
	2	47	2053	202	43	67	15	48	42	39	21
	3	79	174	1124	36	135	27	38	50	145	35
	4	40	51	47	1467	17	47	15	16	20	20
	5	31	73	248	28	947	96	15	38	32	19
	6	23	35	61	51	107	839	6	43	14	71
	7	67	59	61	15	16	13	1020	6	9	5
	8	34	54	75	25	51	140	6	519	58	126
	9	23	77	65	11	42	27	11	36	672	95
	0	28	23	37	13	21	76	6	79	72	692

เมื่อพิจารณาผลการจำแนกตัวเลขของตัวแบบ พบว่า การจำแนกผิดที่สำคัญเกิดจากความคล้ายคลึงของรูปลักษณะของตัวเลขบางตัว เช่น จำแนกเลข 5 เป็นเลข 3 เป็นต้น ตัวอย่างรูปภาพตัวเลขที่ถูกจำแนกผิดแสดงดังภาพที่ 4.1 โดยตัวเลขทางฝั่งซ้ายของลูกศรคือตัวเลขจริง ตัวเลขด้านขวา คือตัวเลขที่ตัวแบบทำนาย



ก. เลข 5 -> 3

ข. เลข 6 -> 5

ค. เลข 7 -> 1

ง. เลข 8 -> 6

ภาพที่ 4.1 ภาพตัวอย่างการจำแนกผิด

สำหรับอัตราการจำแนกถูกต้องแยกตามแต่ละตัวเลข แสดงดังตารางที่ 4.4

ตารางที่ 4.4 อัตราการจำแนกถูกต้องแยกตามแต่ละตัวเลข สำหรับคอร์เนลเรเดียล

ตัวเลข	อัตราการจำแนกถูกต้อง (%)
1	84.86
2	79.67
3	60.99
4	84.31
5	62.02
6	67.12
7	80.25
8	47.70
9	63.46
0	66.09

#### 4.3 ผลการเปรียบเทียบความสำคัญของลักษณะที่มีผลต่อการจำแนกตัวเลข

ในการเปรียบเทียบความสำคัญของลักษณะที่มีผลต่อการจำแนกตัวเลขนั้น จะพิจารณา 2 แบบ ได้แก่ ความสำคัญของลักษณะแต่ละตัว และความสำคัญของกลุ่มของลักษณะ

ความสำคัญของลักษณะแต่ละตัววัดจากการลดลงของอัตราการจำแนกถูกต้องของตัวแบบที่มีลักษณะครบทุกตัวเมื่อเทียบกับอัตราการจำแนกถูกต้องของตัวแบบที่ไม่มีลักษณะนั้น ส่วนความสำคัญของกลุ่มของลักษณะวัดจากการลดลงของอัตราการจำแนกถูกต้องของตัวแบบที่มีลักษณะ

ครบทุกตัวเมื่อเทียบกับอัตราการจัดจำแนกถูกต้องของตัวแบบที่ไม่มีกลุ่มของลักษณะ โดยกลุ่มของลักษณะจะแบ่งตามรูปภาคตัดกรวยที่ค้นหา นั่นคือ เส้นตรง วงรี พาราโบลาแนวตั้ง และพาราโบลาแนวนอน ผลลัพธ์การจัดจำแนกถูกต้องที่ลดลงแสดงดังตารางที่ 4.5 และ 4.6



ตารางที่ 4.5 ลำดับความสำคัญของลักษณะ

อันดับ	ลักษณะ	อัตราการจัดอันดับที่ลดลง* (%)	อันดับ	ลักษณะ	อัตราการจัดอันดับที่ลดลง* (%)
1	OR ของวงรีแบบเต็มรูป	1.89	21	$y_0$ ของวงรีบนรูปภาพครึ่งบน	0.59
2	OR ของเส้นตรงแนวตั้ง	1.64	23	OR ของพาราโบลาแนวอนบนรูปภาพครึ่งบน	0.54
3	OR ของเส้นตรงแนวเฉียง	1.38	23	$b$ ของวงรีเต็มรูป	0.54
4	OR ของวงรีบนรูปภาพครึ่งล่าง	1.35	25	$a$ ของวงรีบนรูปภาพครึ่งล่าง	0.51
5	OR ของเส้นตรงแนวอน	1.28	25	$x_0$ ของวงรีบนรูปภาพครึ่งล่าง	0.51
6	OR ของวงรีบนรูปภาพครึ่งบน	1.22	27	$x_0$ ของวงรีบนรูปภาพครึ่งบน	0.47
7	$\rho$ ของเส้นตรงแนวตั้ง	1.21	28	$y_0$ ของวงรีเต็มรูป	0.40
8	$y_f$ ของพาราโบลาแนวอนบนรูปภาพครึ่งบน	1.12	29	$b$ ของวงรีบนรูปภาพครึ่งล่าง	0.37
9	$\rho$ ของเส้นตรงแนวเฉียง	0.97	30	ความยาวของเส้นตรงแนวตั้ง	0.32
10	$y_f$ ของพาราโบลาแนวอนบนรูปภาพครึ่งล่าง	0.87	31	$x_f$ ของพาราโบลาแนวตั้งบนรูปภาพครึ่งบน	0.31
11	OR ของพาราโบลาแนวตั้งบนรูปภาพครึ่งบน	0.86	32	$y_f$ ของพาราโบลาแนวตั้งบนรูปภาพครึ่งบน	0.29
12	$p$ ของพาราโบลาแนวอนบนรูปภาพครึ่งล่าง	0.84	32	$x_f$ ของพาราโบลาแนวตั้งบนรูปภาพครึ่งล่าง	0.29
13	$x_f$ ของพาราโบลาแนวอนบนรูปภาพครึ่งล่าง	0.80	34	$b$ ของวงรีบนรูปภาพครึ่งบน	0.27
14	$\rho$ ของเส้นตรงแนวอน	0.76	35	$x_0$ ของวงรีเต็มรูป	0.25
15	$\theta$ ของเส้นตรงแนวเฉียง	0.75	36	$p$ ของพาราโบลาแนวตั้งบนรูปภาพครึ่งบน	0.23
16	ความยาวของเส้นตรงแนวอน	0.73	37	$y_0$ ของวงรีบนรูปภาพครึ่งล่าง	0.22
17	OR ของพาราโบลาแนวตั้งบนรูปภาพครึ่งล่าง	0.71	38	$y_f$ ของพาราโบลาแนวตั้งบนรูปภาพครึ่งล่าง	0.14
18	$p$ ของพาราโบลาแนวอนบนรูปภาพครึ่งบน	0.68	39	$a$ ของวงรีเต็มรูป	0.13
19	OR ของพาราโบลาแนวอนบนรูปภาพครึ่งล่าง	0.64	40	$a$ ของวงรีบนรูปภาพครึ่งบน	0.11
20	$x_f$ ของพาราโบลาแนวอนบนรูปภาพครึ่งบน	0.61	41	$p$ ของพาราโบลาแนวตั้งบนรูปภาพครึ่งล่าง	0.03
21	ความยาวของเส้นตรงแนวเฉียง	0.59			

\*คำนวณจาก  $(0.7217 - \text{อัตราการจัดอันดับของตัวแบบเมื่อนำลักษณะนั้นออก})/0.7217$  โดย 0.7217 คืออัตราการจัดอันดับของตัวแบบ SVM เคอร์เนลเรเดียลที่มีทุกลักษณะ

ตารางที่ 4.6 ลำดับความสำคัญของรูปภาพตัดกรวยแต่ละรูป

อันดับ	กลุ่มของลักษณะ	อัตราการจำแนกถูกต้องที่ลดลง* (%)
1	เส้นตรง	10.24
2	วงรี	6.19
3	พาราโบลาแนวนอน	4.36
4	พาราโบลาแนวตั้ง	1.78

\*คำนวณจาก  $(0.7217 - \text{อัตราการจำแนกถูกต้องของตัวแบบเมื่อนำกลุ่มของลักษณะที่เกี่ยวข้องกับรูปภาพตัดกรวยนั้นออก})/0.7217$  โดย 0.7217 คืออัตราการจำแนกถูกต้องของตัวแบบ SVM เคอร์เนลเรเดียลที่มีทุกลักษณะ

จากตารางที่ 4.5 เมื่อเทียบลักษณะแต่ละตัวแล้ว อัตราการจำแนกถูกต้องที่ลดลงไม่แตกต่างกันมากนัก โดย 10 อันดับแรก เป็นลักษณะที่เกี่ยวข้องกับวงรี 3 ตัว เป็นลักษณะที่เกี่ยวข้องกับเส้นตรง 5 ตัว ขณะที่ในตารางที่ 4.6 เมื่อเทียบความสำคัญของรูปภาพตัดกรวยแต่ละรูปแล้ว อัตราการจำแนกถูกต้องที่ลดลงมีความแตกต่างกันอย่างชัดเจนมากกว่า โดยรูปเส้นตรงมีความสำคัญที่สุด อัตราการจำแนกถูกต้องลดลง 10.24% ขณะที่รูปพาราโบลาแนวตั้งมีความสำคัญน้อยที่สุด อัตราการจำแนกถูกต้องลดลง 1.78%

ใน ตารางที่ 4.7 แสดงลักษณะสำคัญที่สุด 3 อันดับแรกแยกตามแต่ละตัวเลข โดยวัดการลดลงของอัตราการจำแนกถูกต้องแยกตามแต่ละตัวเลขในตารางที่ 4.4 เมื่อเทียบกับอัตราการจำแนกถูกต้องแยกตามแต่ละตัวเลขเมื่อไม่มีลักษณะนั้น

ตารางที่ 4.7 ลำดับของลักษณะที่สำคัญที่สุด 3 อันดับแรกแยกตามแต่ละตัวเลข

ตัวเลข	ลักษณะอันดับ 1	ลักษณะอันดับ 2	ลักษณะอันดับ 3
1	OR ของเส้นตรงแนวตั้ง	$\rho$ ของเส้นตรงแนวตั้ง	OR ของวงรีรูปภาพครึ่งล่าง
2	OR ของเส้นตรงแนวตั้ง	OR ของพาราโบลาแนวตั้งบนรูปภาพครึ่งล่าง	OR ของเส้นตรงแนวนอน
3	$\rho$ ของเส้นตรงแนวตั้ง	OR ของเส้นตรงแนวเฉียง	$\theta$ ของเส้นตรงแนวเฉียง
4	$\rho$ ของเส้นตรงแนวเฉียง	OR ของวงรีรูปภาพครึ่งล่าง	$\rho$ ของเส้นตรงแนวตั้ง
5	OR ของเส้นตรงแนวเฉียง	$y_f$ ของพาราโบลาแนวนอนบนรูปภาพครึ่งบน	$p$ ของพาราโบลาแนวนอนบนรูปภาพครึ่งบน
6	OR ของวงรีรูปภาพครึ่งบน	OR ของวงรีเต็มรูป	$x_0$ ของวงรีรูปภาพครึ่งบน
7	OR ของเส้นตรงแนวนอน	OR ของเส้นตรงแนวเฉียง	OR ของเส้นตรงแนวตั้ง
8	OR ของวงรีรูปภาพครึ่งล่าง	$\rho$ ของเส้นตรงแนวตั้ง	$b$ ของวงรีรูปภาพครึ่งบน
9	OR ของวงรีเต็มรูป	$x_f$ ของพาราโบลาแนวนอนบนรูปภาพครึ่งล่าง	OR ของวงรีรูปภาพครึ่งล่าง
0	OR ของวงรีเต็มรูป	OR ของพาราโบลาแนวตั้งบนรูปภาพครึ่งบน	$y_0$ ของวงรีรูปภาพครึ่งล่าง



จะเห็นว่าลักษณะที่สำคัญที่สุด 3 อันดับแรกของแต่ละตัวเลข เป็นลักษณะที่เกี่ยวข้องกับเส้นตรง 14 ตัว วงรี 11 ตัว พาราโบลาแนวนอน 3 ตัว และพาราโบลาแนวตั้ง 2 ตัว ซึ่งสอดคล้องกับผลการเปรียบเทียบความสำคัญของรูปภาคตัดกรวยแต่ละรูปในตารางที่ 4.6 อย่างไรก็ตาม เมื่อพิจารณาบางตัวเลข พบว่า ความสำคัญของลักษณะบางตัวอาจเห็นได้ไม่ชัดเจนหากดูเพียงรูปลักษณ์ของตัวเลข ซึ่งอาจเกิดจากผลกระทบจากการค้นหาขอบของตัวเลข เช่น ลักษณะอันดับสองของเลข 5 คือ  $y_f$  ของพาราโบลาแนวนอนบนรูปภาพครึ่งบน, ลักษณะอันดับสามของเลข 1 คือ OR ของวงรีบนรูปภาพครึ่งล่าง เป็นต้น หรืออาจเกิดจากลักษณะของรูปภาพ SVHN ที่มีปัญหาเรื่องความบาง เช่น เลข 3 เป็นต้น แสดงตัวอย่างดังตารางที่ 4.8

ตารางที่ 4.8 ภาพตัวอย่างตัวเลขที่ได้รับผลกระทบจากการค้นหาขอบ

รูปตัวเลข	ผลกระทบต่อความสำคัญของลักษณะ
	ความบางของตัวเลขทำให้ส่วนที่ควรจะโค้งของเลข 3 กลายเป็นเส้นตรง
	ลักษณะของเลข 9 โดยรวมคล้ายวงรี
	ขอบของเลข 5 ในช่วงครึ่งบนเป็นพาราโบลาแนวนอน

## บทที่ 5

### สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้ ใช้ฐานข้อมูลตัวเลข SVHN ในการทดสอบการใช้พารามิเตอร์ที่เกี่ยวข้องกับรูปภาคตัดกรวยเป็นลักษณะในการจำแนกตัวเลข โดยประยุกต์ใช้การแปลงฮัฟในการค้นหารูปภาคตัดกรวยจากรูปตัวเลข และใช้ SVM เป็นตัวแบบในการจำแนกตัวเลข สามารถสรุปผลได้ดังนี้

#### 5.1 สรุปผลการวิจัย

ผลการวิจัยนี้ สามารถแบ่งการสรุปผลออกเป็น 3 ส่วน ได้แก่ ส่วนที่ 1 สรุปผลการจำแนกตัวเลขและเวลาที่ใช้ในการสร้างตัวแบบของตัวแบบ SVM แต่ละเคอร์เนล และส่วนที่ 2 สรุปผลความสำคัญของลักษณะแต่ละตัวและกลุ่มของลักษณะที่มีผลต่อการจำแนกตัวเลข และส่วนที่ 3 สรุปผลการใช้พารามิเตอร์จากภาคตัดกรวยเป็นลักษณะในการจำแนกตัวเลข SVHN เปรียบเทียบกับการใช้ลักษณะอื่น

##### ส่วนที่ 1 สรุปผลการจำแนกตัวเลขและเวลาที่ใช้ในการสร้างตัวแบบของตัวแบบ SVM แต่ละเคอร์เนล

เคอร์เนลที่มีประสิทธิภาพในการจำแนกสูงที่สุด คือเคอร์เนลเรเดียล โดยให้อัตราการจำแนกถูกต้องสูงที่สุดและใช้เวลาในการสร้างตัวแบบน้อยที่สุด ขณะที่เคอร์เนลชกมอยด์ให้อัตราการจำแนกต่ำที่สุด และเคอร์เนลพหุนามใช้เวลาในการสร้างตัวแบบนานที่สุด

เมื่อเปรียบเทียบผลลัพธ์ของเคอร์เนลเรเดียลกับเคอร์เนลเชิงเส้นแล้ว แสดงให้เห็นว่าลักษณะที่ใช้ในการสร้างตัวแบบนั้นไม่มีความเป็นเชิงเส้น (non-linear)

##### ส่วนที่ 2 สรุปผลความสำคัญของลักษณะแต่ละตัวและกลุ่มของลักษณะที่มีผลต่อการจำแนกตัวเลข

เมื่อเปรียบเทียบความสำคัญของลักษณะแต่ละตัวแล้ว อัตราการจำแนกถูกต้องที่ลดลงไม่แตกต่างกันมากนัก ในขณะที่เมื่อเปรียบเทียบความสำคัญของกลุ่มของลักษณะแล้ว อัตราการจำแนกถูกต้องที่ลดลงเห็นได้ชัดเจนมากกว่า โดยรูปภาคตัดกรวยที่มีความสำคัญที่สุดต่อการจำแนกตัวเลขคือเส้นตรง และรูปพาราโบลาแนวตั้งมีความสำคัญต่อการจำแนกตัวเลขน้อยที่สุด

การใช้ขอบของตัวเลขก่อนการดึง feature ด้วยการแปลงฮัฟส่งผลให้ลำดับความสำคัญของ feature ขึ้นกับลักษณะของขอบที่ปรากฏ อย่างไรก็ตาม การใช้ขอบมีข้อดีคือ ช่วยลดเวลาในการค้นหารูปร่าง (Davies, 1987) โดยเฉพาะอย่างยิ่ง รูปพาราโบลาและรูปวงรี ที่จำนวนมิติของ accumulator array เท่ากับ 3 มิติและ 4 มิติ ตามลำดับ ซึ่งจะใช้เวลาในการค้นหารูปร่างและเกิดความซ้ำซ้อนหากค้นหารูปร่างจากรูปตัวเลขธรรมดา

**ส่วนที่ 3** สรุปผลการใช้พารามิเตอร์จากภาคตัดกรวยเป็นลักษณะในการจำแนกตัวเลข SVHN เปรียบเทียบกับการใช้ลักษณะอื่น

เมื่อเปรียบเทียบลักษณะที่ใช้พารามิเตอร์จากภาคตัดกรวยกับงานวิจัยที่ใช้ลักษณะอื่นในการจำแนกตัวเลข SVHN เช่น Netzer และคณะ (2012) เป็นต้น พบว่า พารามิเตอร์จากภาคตัดกรวยให้ผลลัพธ์อัตราการจำแนกถูกต้องต่ำกว่าลักษณะอื่น ๆ (Netzer และคณะ (2012) เปรียบเทียบลักษณะ 4 แบบ ได้แก่ Histograms of Oriented Gradients, Weighted Direction Code Histogram, Stacked Sparse Auto-Encoders และ K-means-based features และใช้ SVM ในการจำแนกซึ่งได้อัตราการจำแนกถูกต้องเท่ากับ 85.0%, 63.3%, 89.7% และ 90.6% ตามลำดับ)

อย่างไรก็ตาม พารามิเตอร์จากภาคตัดกรวยเป็นลักษณะที่มีความซับซ้อนน้อยกว่า อีกทั้งจำนวนของลักษณะก็มีจำนวนน้อยกว่า (Histograms of Oriented Gradients มี 324 ลักษณะ K-means-based features มี 500 ลักษณะ ขณะที่พารามิเตอร์จากภาคตัดกรวยมี 41 ลักษณะ) แสดงให้เห็นว่า การจำแนกข้อมูลรูปภาพตัวเลข SVHN ให้ได้อัตราการจำแนกถูกต้องสูงชันนั้น จำเป็นต้องใช้ลักษณะที่มีความซับซ้อนมากขึ้น และมีจำนวนลักษณะเพิ่มมากขึ้น

## 5.2 ข้อเสนอแนะ

5.2.1 ในการวัดความสำคัญของลักษณะสำหรับ SVM สามารถใช้วิธี recursive backward elimination (Trziesiok, 2010) ซึ่งทดสอบโดยการตัดตัวแปรที่ไม่สำคัญ (ในที่นี้หมายถึง ตัวแปรที่ทำให้อัตราการจำแนกถูกต้องลดลงน้อยที่สุดเมื่อนำตัวแปรนั้นออกจากตัวแบบ) ออกทีละตัว ตัวแปรสุดท้ายที่เหลืออยู่จะเป็นลักษณะที่สำคัญที่สุด อย่างไรก็ตาม วิธีนี้จะใช้เวลาและทรัพยากรในการคำนวณมาก หากมีลักษณะจำนวนมาก

5.2.2 ขั้นตอนและเทคนิคการเตรียมข้อมูลรูปภาพก่อนสำหรับงานวิจัยนี้ยังมีข้อจำกัดในการใช้กับภาพถ่าย โดยเฉพาะอย่างยิ่ง ในสภาพแวดล้อมที่ไม่มีการควบคุมเรื่องแสงและเงา ซึ่งมีผลต่อความคมชัดของรูปตัวเลข ผลลัพธ์ที่ได้จากงานวิจัยนี้ จึงจำกัดเฉพาะกับขั้นตอนการ preprocess ที่ใช้

5.2.3 ในการระบุลักษณะ อาจกำหนดระดับ (threshold) ของ OR เพื่อระบุว่า มีหรือไม่มีรูปภาพตัดกรวยนั้นในรูปตัวเลข

5.2.4 ในกรณีที่ไม่ใช่ขอบของตัวเลขเป็นจุดเริ่มต้นในการดึงลักษณะอาจพิจารณาเส้นที่เหมาะสมที่จะสะท้อนรูปตัวเลขได้ชัดเจน เช่น เส้นที่อยู่กึ่งกลางของพิกเซลที่แสดงรูปตัวเลข

5.2.5 การใช้พารามิเตอร์จากรูปภาคตัดกรวยเป็นลักษณะในการสร้างตัวแบบจำแนกตัวเลขสามารถนำไปใช้กับฐานข้อมูลตัวเลขอื่น ๆ ได้ เช่น ฐานข้อมูล MNIST, ฐานข้อมูล USPS (United State Postal Service ฐานข้อมูลตัวเลขที่เขียนด้วยลายมือบนซองจดหมายในสหรัฐอเมริกา) เป็นต้น



## รายการอ้างอิง

- Aguado, A. S. (2012). *Feature Extraction and Image Processing for Computer Vision* (Vol. 3rd ed). Oxford: Academic Press.
- Chang, F., Chen, C. J., & Lu, C. J. (2004). A linear-time component-labeling algorithm using contour tracing technique. *Computer Vision and Image Understanding*, 93(2), 206-220.
- Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(3), 273-297.
- Duda, R. O., Hart, P. E., & Newman, W. (1972). Use of the Hough Transformation To Detect Lines and Curves in Pictures. *Communications of the ACM*, 15(1), 11.
- Garg, M., & Ahuja, D. (2013). A Novel Approach to Recognize the off-line Handwritten Numerals using MLP and SVM Classifiers. *International Journal of Computer Science & Engineering Technology*, 4(7), 953-958.
- Giuliodori, A., Lillo, R. E., & Peña, D. (2011). Handwritten digit classification. *DES - Working Papers. Statistics and Econometrics. WS*.
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The elements of statistical learning : data mining, inference, and prediction*: New York : Springer, c2001.
- Hough, P. V. C. (1962). METHOD AND MEANS FOR RECOGNIZING COMPLEX PATTERNS. In.
- Hsuan-tien, L., & Chih-jen, L. (2013). A Study on Sigmoid Kernels for SVM and the Training of non-PSD Kernels by SMO-type Methods. In. United States, North America.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning. [electronic resource] : with Applications in R*: New York, NY : Springer New York : Imprint: Springer, 2013.
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, Proc. IEEE*(11), 2278.

- Netzer, Y., Wang, T., Coates, A., Bissacco, R., Wu, B., & Ng, A. Y. (2012). Reading Digits in Natural Images with Unsupervised Feature Learning. In. United States, North America.
- Patel, I., Jagtap, V., & Kale, O. (2014). *A Survey on Feature Extraction Methods for Handwritten Digits Recognition* (Vol. 107).
- Prince, S. J. D. (2012). *Computer vision : models, learning, and inference*: New York : Cambridge University Press, 2012.
- Shapiro, L. G. (2001). *Computer vision*: Prentice Hall, 2001.
- Trzesiok, M. (2010). The importance of predictor variables for individual classes in SVM. In. Poland, Europe: Wydawnictwo Uniwersytetu Łódzkiego.



## ประวัติผู้เขียนวิทยานิพนธ์

นายพงศ์ศรัณย์ เทียมตะวัน เกิดวันที่ 18 มกราคม พ.ศ.2533 สำเร็จการศึกษาปริญญา  
บัญชีบัณฑิต (บช.บ.) วิชาเอกการบัญชี วิชาโทการธนาคารและการเงิน สาขาวิชาการบัญชี ภาค  
วิชาการบัญชี คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2554  
เข้าทำงานในบริษัท เคพีเอ็มจี ภูเก็ต สอนบัญชี จำกัด ในตำแหน่งผู้ช่วยผู้ตรวจสอบบัญชี เป็น  
ระยะเวลา 3 ปี และได้เข้าศึกษาต่อในหลักสูตรวิทยาศาสตรมหาบัณฑิต (วท.ม.) สาขาวิชาสถิติ  
ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2558

