

การพัฒนาศูนย์วัดความสำคัญของตัวแปรและเกณฑ์จุดตัดในการคัดเลือกตัวแปร
สำหรับการจำแนกประเภทการเป็นมะเร็งลำไส้ใหญ่

นพมาศ อัครจันทโชติ

ดุษฎีนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต

สาขาวิชาการวิจัยและสถิติทางวิทยาการปัญญา

วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา

กุมภาพันธ์ 2557

ลิขสิทธิ์เป็นของมหาวิทยาลัยบูรพา

คณะกรรมการควบคุมคุณิพนธ์และคณะกรรมการสอบคุณิพนธ์ ได้พิจารณา
คุณิพนธ์ของ นพมาศ อัครจันทโชติ ฉบับนี้แล้ว เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตาม
หลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาการวิจัยและสถิติทางวิทยาการปัญญา ของมหาวิทยาลัยบูรพาได้

คณะกรรมการควบคุมคุณิพนธ์


.....อาจารย์ที่ปรึกษาหลัก
(รองศาสตราจารย์ ดร.เสรี ชัดแฉ้ม)


.....อาจารย์ที่ปรึกษาร่วม
(ผู้ช่วยศาสตราจารย์ ดร.กิตาการ สายธนู)

คณะกรรมการสอบคุณิพนธ์


.....ประธาน
(ผู้ช่วยศาสตราจารย์ ดร.จุฑาธิป ตันสถิตย์)

.....กรรมการ
(รองศาสตราจารย์ ดร.เสรี ชัดแฉ้ม)


.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.กิตาการ สายธนู)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สุชาดา กรเพชรปานิ)

วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญาอนุมัติให้รับคุณิพนธ์ฉบับนี้เป็นส่วนหนึ่งของ
การศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาการวิจัยและสถิติทางวิทยาการปัญญาของ
มหาวิทยาลัยบูรพา


.....คณบดีวิทยาลัยวิทยาการวิจัย
และวิทยาการปัญญา

(ผู้ช่วยศาสตราจารย์ ดร.สุชาดา กรเพชรปานิ)
วันที่ ๒๘ เดือน กุมภาพันธ์ พ.ศ. ๒๕๕๗

ประกาศคุณูปการ

ดุขฉุฉนัฟนัฟฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี ด้วยความกรุณาจาก รองศาสตราจารย์ ดร.เสรี ชัดแฉัฒ อาจารย์ที่ปรึกษาหลัก และผู้ช่วยศาสตราจารย์ ดร.กิตาการ สายธนู อาจารย์ที่ปรึกษาร่วม ที่กรุณาให้คำปรึกษาแนะนำแนวทางที่ถูกต้อง ตลอดจนแก้ไขข้อบกพร่องต่าง ๆ ด้วยความละเอียดถี่ถ้วนและเอาใจใส่ด้วยดีเสมอมา ผู้วิจัยรู้สึกซาบซึ้งเป็นอย่างยิ่ง จึงขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ผู้วิจัยขอขอบคุณ มหาวิทยาลัยหัวเฉียวเฉลิมพระเกียรติ ที่สนับสนุนทุนการศึกษาเพื่อศึกษาต่อในระดับปริญญาเอก และขอขอบคุณ เพื่อน ๆ พี่ ๆ และน้อง ๆ ทุกคนที่สนับสนุนผู้วิจัยทั้งด้านวิชาการและการเป็นกำลังใจแก่ผู้วิจัย

ผู้วิจัยขอขอบคุณผู้ช่วยศาสตราจารย์ ดร.จุฑาธิป ตัณสถิตย์ และผู้ช่วยศาสตราจารย์ ดร.สุชาดา กรเพชรปาณี ที่กรุณาเป็นประธานและกรรมการสอบดุขฉุฉนัฟนัฟ พร้อมทั้งให้คำแนะนำในการปรับแก้ดุขฉุฉนัฟนัฟ และขอขอบคุณ ดร.ปรัชญา แก้วแก่น ที่ให้คำแนะนำในเรื่องอื่นแก่ผู้วิจัย ทำให้ดุขฉุฉนัฟนัฟนี้มีความสมบูรณ์ยิ่งขึ้น

คุณค่าและประโยชน์ของดุขฉุฉนัฟนัฟฉบับนี้ ผู้วิจัยขอมอบเป็นกตัญญูตเวทิตาแด่บุพการี บุรพจารย์ และผู้มีพระคุณทุกท่านทั้งในอดีตและปัจจุบัน ที่ทำให้ข้าพเจ้าเป็นผู้มีการศึกษาและประสบความสำเร็จมาจนตราบนานเท่านานนี้

นพมาศ อัครจันทโชติ

53810233: สาขาวิชา: การวิจัยและสถิติทางวิทยาการปัญญา

ปร.ด. (การวิจัยและสถิติทางวิทยาการปัญญา)

คำสำคัญ: การคัดเลือกตัวแปร/ ดัชนีวัดความสำคัญของตัวแปร/ เกณฑ์จุดตัด/ วิธีการสองน้อยที่สุดบางส่วน/ การจำแนกประเภทการเป็นมะเร็งลำไส้ใหญ่

นพมาศ อัครจันทโชติ : การพัฒนาดัชนีวัดความสำคัญของตัวแปรและเกณฑ์จุดตัดในการคัดเลือกตัวแปร สำหรับการจำแนกประเภทการเป็นมะเร็งลำไส้ใหญ่ (THE DEVELOPMENT OF VARIABLE IMPORTANCE INDEX AND CUTOFF THRESHOLD CRITERION IN VARIABLE SELECTION FOR COLON CANCER CLASSIFICATION) คณะกรรมการควบคุมดัชนีพนธ์: เสรี ชัดรัมย์, ค.ด., กิตาการ สายธนู, Ph.D. 174 หน้า. ปี พ.ศ. 2557.

การวิจัยนี้มีวัตถุประสงค์เพื่อ พัฒนาดัชนีวัดความสำคัญของตัวแปรและเกณฑ์จุดตัดในการคัดเลือกตัวแปรบนพื้นฐานวิธีการสองน้อยที่สุดบางส่วน สำหรับการจำแนกประเภทข้อมูล โดยดัชนีวัดความสำคัญของตัวแปรที่พัฒนามีชื่อว่า VIIC (Variable Importance Index for Classification) และเกณฑ์จุดตัดที่พัฒนามีชื่อว่า เกณฑ์จุดตัดของแผนภาพกล่อง (Boxplot Cutoff Threshold หรือเกณฑ์จุดตัด BCT) ดัชนีและเกณฑ์จุดตัดที่พัฒนาขึ้นนี้ นำมาสร้างวิธีการคัดเลือกตัวแปร 2 วิธี คือ วิธี VIP-BCT ซึ่งเป็นวิธีที่ใช้ดัชนีเดิมคือ Variable Influence on the Projection (VIP) และเกณฑ์จุดตัดใหม่ BCT และวิธี VIIC-BCT ซึ่งเป็นวิธีที่ใช้ดัชนีใหม่ VIIC และเกณฑ์จุดตัดใหม่ BCT โดยเปรียบเทียบกับวิธีการคัดเลือกตัวแปรแบบดั้งเดิม (VIP-1) ที่ใช้ดัชนี VIP ที่มีเกณฑ์จุดตัด 1.0 ภายใต้งานจำลองของ 4 พารามิเตอร์ คือ 1) เปอร์เซนต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล 2) ความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง 3) แบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้อง และ 4) ขนาดตัวอย่าง โดยนำวิธีการคัดเลือกตัวแปรทั้ง 3 วิธีนี้ไปศึกษาความสามารถของยีนที่ได้จากการคัดเลือกตัวแปร ในการจำแนกประเภทเนื้อเยื่อมะเร็งลำไส้ใหญ่ และเนื้อเยื่อปกติ ผลการศึกษาปรากฏว่า

$$1. \text{ ดัชนีวัดความสำคัญของตัวแปรที่พัฒนาคือ } VIIC_j = \sum_{o=1}^h ACR_o \cdot (w_{oj} / \|w_o\|)^2 \text{ โดยที่ } ACR_o \text{ แทน}$$

Adjusted Count R^2 ขององค์ประกอบ t_o และ w_{oj} แทนน้ำหนักของตัวแปรทำนายที่ j ในองค์ประกอบ t_o ส่วนเกณฑ์จุดตัดที่พัฒนาคือ $BCT = Q_3 + 1.5 \times IQR$ โดยที่ Q_3 แทน ควอร์ไทล์ที่ 3 ของค่าดัชนีของตัวแปรบวกรวม และ IQR แทนพิสัยระหว่างควอร์ไทล์

2. วิธีการคัดเลือกตัวแปรแบบ VIP-BCT มีความแม่นยำในการคัดเลือกตัวแปรมากกว่าวิธีดั้งเดิม (VIP-1) ในสถานการณ์ส่วนใหญ่เมื่อเปอร์เซนต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล เท่ากับ 1, 3 และ 5 ส่วนวิธีการคัดเลือกตัวแปรแบบ VIIC-BCT มีความแม่นยำในการคัดเลือกตัวแปรมากกว่าวิธีดั้งเดิม (VIP-1) ในสถานการณ์ส่วนใหญ่เมื่อเปอร์เซนต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล เท่ากับ 1, 3 และ 5 และ ความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง เท่ากับ 1

3. ความแม่นยำโดยเฉลี่ยในการจำแนกประเภทเนื้อเยื่อมะเร็งลำไส้ใหญ่และเนื้อเยื่อปกติ โดยใช้ยีนที่ได้รับการคัดเลือกด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT เท่ากับ 82.02% 82.71% และ 83.21% ตามลำดับ เมื่อจำแนกประเภทด้วยวิธีการวิเคราะห์การจำแนกประเภทของวิธีการสองน้อยที่สุดบางส่วน

53810233: MAJOR: RESEARCH AND STATISTICS IN COGNITIVE SCIENCE

Ph.D. (RESEARCH AND STATISTICS IN COGNITIVE SCIENCES)

KEYWORDS: VARIABLE SELECTION/ VARIABLE IMPORTANCE INDEX/ CUTOFF

THRESHOLD CRITERION/ PARTIAL LEAST SQUARES/ COLON CANCER
CLASSIFICATION

NOPPAMAS AKARACHANTACHOTE: THE DEVELOPMENT OF VARIABLE
IMPORTANCE INDEX AND CUTOFF THRESHOLD CRITERION IN VARIABLE SELECTION FOR
COLON CANCER CLASSIFICATION. ADVISORY COMMITTEE: SEREE CHADCHAM, Ph.D.,
KIDAKAN SAITHANU, Ph.D., 174 P. 2014.

The objective of this research was to develop a variable importance index and cutoff threshold criterion for variable selection based on partial least squares for classification. The proposed variable importance index was Variable Importance Index for Classification, VIIC, and the proposed cutoff threshold criterion was Boxplot Cutoff Threshold, BCT. Both VIIC and BCT were applied to generate a new variable selection method, VIIC-BCT. The classical index, Variable Influence on the Projection (VIP), and the BCT formed the basis for a new variable selection method, VIP-BCT. These two methods were compared to the classical method VIP-1, which used the VIP as an index and constant value of 1.0 as a cutoff threshold criterion, through situation simulation. There were 180 situations generated based on four parameters: 1) Percentage of the number of relevant variables, 2) Magnitude of mean difference of relevant variables between two groups, 3) Pattern of correlation between relevant variables, and 4) sample size. The variable selection methods were applied to select genes that can classify colon cancer tissues from normal tissues. Additionally, accuracies of the classification were studied.

The research results were as follows:

1. The proposed variable importance index was $VIIC_j = \sum_{a=1}^h ACR_a \cdot (w_{aj} / \|w_a\|)^2$ where

ACR_a is the adjusted count R^2 of component t_a , and w_{aj} is weight of the j^{th} predictor variable in component t_a . The proposed cutoff threshold criterion was $BCT = Q_3 + 1.5 \times IQR$ where Q_3 is upper quartile of the variable importance index of noise variables, and the IQR is interquartile range.

2. The VIP-BCT method was superior to VIP-1 in identifying relevant variables in most of the situations when the percentage of the number of relevant variables was 1, 3 and 5. The VIIC-BCT method was superior to VIP-1 in most of the situations when the percentage of the number of relevant variables was 1, 3 and 5, and the magnitude of mean difference of relevant variables between two groups was 1.

3. The balanced accuracy of classifying colon cancer tissues against normal tissues when applied to the three variable selection methods, VIP-1, VIP-BCT, and VIIC-BCT were 82.02%, 82.71%, and 83.21%, respectively, using partial least squares discriminant analysis as a classifier.

สารบัญ

	หน้า
สารบัญ	ฉ
สารบัญตาราง	ซ
สารบัญภาพ	๗
บทที่	
1 บทนำ.....	1
ความเป็นมาและความสำคัญของปัญหา.....	1
วัตถุประสงค์ของการวิจัย.....	5
กรอบแนวคิดในการวิจัย.....	5
ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย.....	10
ขอบเขตของการวิจัย.....	10
นิยามศัพท์เฉพาะ.....	15
2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	18
ตอนที่ 1 การคัดเลือกตัวแปร	18
ตอนที่ 2 วิธีกำลังสองน้อยที่สุดบางส่วน	22
ตอนที่ 3 การจำแนกประเภท	30
ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง.....	36
3 วิธีดำเนินการวิจัย.....	42
ตอนที่ 1 การพัฒนาดัชนีวัดความสำคัญของตัวแปรและเกณฑ์จุดตัดในการคัดเลือกตัวแปร	42
ตอนที่ 2 การเปรียบเทียบการคัดเลือกตัวแปร.....	44
ตอนที่ 3 การศึกษาความสามารถของยีนที่ได้จากการคัดเลือกตัวแปรในการจำแนกประเภทเนื้อเยื่อมะเร็งลำไส้ใหญ่ และเนื้อเยื่อปกติ.....	49
4 ผลการวิเคราะห์ข้อมูล.....	52
ตอนที่ 1 ผลการพัฒนาดัชนีวัดความสำคัญของตัวแปรและเกณฑ์จุดตัดในการคัดเลือกตัวแปร.....	52
ตอนที่ 2 ผลการเปรียบเทียบการคัดเลือกตัวแปร.....	54
ตอนที่ 3 ผลการศึกษาความสามารถของยีนที่ได้จากการคัดเลือกตัวแปรในการจำแนกประเภทเนื้อเยื่อมะเร็งลำไส้ใหญ่ และเนื้อเยื่อปกติ.....	102

สารบัญ (ต่อ)

	หน้า
5 สรุปและอภิปรายผล.....	114
สรุปผลการวิจัย.....	114
อภิปรายผล.....	116
ข้อเสนอแนะ.....	118
บรรณานุกรม.....	121
ภาคผนวก.....	129
ภาคผนวก ก ชื่อสินค้าที่ได้รับการคัดเลือก.....	130
ภาคผนวก ข ผลการเปรียบเทียบค่าเฉลี่ย Sen และ Spe ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปรด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตาม Prel, Mdif, Σ และ n	148
ภาคผนวก ค เกณฑ์จุดตัดเฉลี่ยสำหรับดัชนี VIP เมื่อใช้วิธีการคัดเลือกตัวแปร VIP-BCT (BCT_{VIP}) โดยจำแนกตาม Prel, Mdif, Σ และ n	157
ภาคผนวก ง การเปรียบเทียบค่าความเบ้ของ VIP_{noise} และ $VIIC_{noise}$ เมื่อคัดเลือกตัวแปร ด้วยวิธี VIP-BCT และวิธี VIIC-BCT.....	159
ภาคผนวก จ โปรแกรมที่ใช้ในงานวิจัย.....	163
ภาคผนวก ฉ ตัวอย่างข้อมูลดิบ.....	171
ประวัติย่อของผู้วิจัย.....	174

สารบัญตาราง

ตารางที่	หน้า
1 ผลการจำแนกประเภทตัวแปร.....	13
2 ผลการจำแนกประเภทการเป็นมะเร็งลำไส้ใหญ่.....	14
3 ผลการจำแนกประเภทของหน่วยตัวอย่าง.....	33
4 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปรด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ในกรณีเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 1% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบไม่สัมพันธ์กัน ($\Sigma = \Sigma_1$)	59
5 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปรด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ในกรณีเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 1% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบสัมพันธ์กันที่ระดับสัมประสิทธิ์สหสัมพันธ์คงที่ 0.5 ($\Sigma = \Sigma_2$).....	60
6 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปรด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ในกรณีเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 1% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบสัมพันธ์กันที่ระดับสัมประสิทธิ์สหสัมพันธ์คงที่ 0.9 ($\Sigma = \Sigma_3$).....	62
7 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปรด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ในกรณีเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 1% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบสัมพันธ์กันที่ระดับสัมประสิทธิ์สหสัมพันธ์ไม่คงที่ ซึ่งกำหนดสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรทำนายที่ i กับ j เท่ากับ $(0.5)^{ i-j }$ ($\Sigma = \Sigma_4$).....	64

สารบัญตาราง (ต่อ)

ตารางที่	หน้า
8 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปร ด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของ ค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ใน กรณิเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 1% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบสัมพันธ์กันที่ระดับ สัมประสิทธิ์สหสัมพันธ์ไม่คงที่ ซึ่งกำหนดสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรทำนาย ที่ i กับ j เท่ากับ $(0.9)^{ i-j }$ ($\Sigma = \Sigma_5$).....	66
9 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปร ด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของ ค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ใน กรณิเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 3% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบไม่สัมพันธ์กัน ($\Sigma = \Sigma_1$).....	67
10 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปร ด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของ ค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ใน กรณิเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 3% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบสัมพันธ์กันที่ระดับ สัมประสิทธิ์สหสัมพันธ์คงที่ 0.5 ($\Sigma = \Sigma_2$).....	69
11 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปร ด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของ ค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ใน กรณิเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 3% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบสัมพันธ์กันที่ระดับ สัมประสิทธิ์สหสัมพันธ์คงที่ 0.9 ($\Sigma = \Sigma_3$).....	71

สารบัญตาราง (ต่อ)

ตารางที่	หน้า
12 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปร ด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของ ค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ใน กรณิเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 3% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบสัมพันธ์กันที่ระดับ สัมประสิทธิ์สหสัมพันธ์ไม่คงที่ ซึ่งกำหนดสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรทำนาย ที่ i กับ j เท่ากับ $(0.5)^{ i-j }$ ($\Sigma = \Sigma_4$).....	73
13 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปร ด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของ ค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ใน กรณิเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 3% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบสัมพันธ์กันที่ระดับ สัมประสิทธิ์สหสัมพันธ์ไม่คงที่ ซึ่งกำหนดสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรทำนาย ที่ i กับ j เท่ากับ $(0.9)^{ i-j }$ ($\Sigma = \Sigma_5$).....	75
14 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปร ด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของ ค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ใน กรณิเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 5% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบไม่สัมพันธ์กัน ($\Sigma = \Sigma_1$).....	77
15 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปร ด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของ ค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ใน กรณิเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 5% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบสัมพันธ์กันที่ระดับ สัมประสิทธิ์สหสัมพันธ์คงที่ 0.5 ($\Sigma = \Sigma_2$).....	79

สารบัญตาราง (ต่อ)

ตารางที่	หน้า
16 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปร ด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของ ค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ใน กรณิเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 5% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบสัมพันธ์กันที่ระดับ สัมประสิทธิ์สหสัมพันธ์คงที่ 0.9 ($\Sigma = \Sigma_3$).....	81
17 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปร ด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของ ค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ใน กรณิเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 5% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบสัมพันธ์กันที่ระดับ สัมประสิทธิ์สหสัมพันธ์ไม่คงที่ ซึ่งกำหนดสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรทำนาย ที่ i กับ j เท่ากับ $(0.5)^{ i-j }$ ($\Sigma = \Sigma_4$).....	83
18 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปร ด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของ ค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ใน กรณิเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 5% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบสัมพันธ์กันที่ระดับ สัมประสิทธิ์สหสัมพันธ์ไม่คงที่ ซึ่งกำหนดสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรทำนาย ที่ i กับ j เท่ากับ $(0.9)^{ i-j }$ ($\Sigma = \Sigma_5$).....	85
19 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปร ด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของ ค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ใน กรณิเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 10% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบไม่สัมพันธ์กัน ($\Sigma = \Sigma_1$).....	87

สารบัญตาราง (ต่อ)

ตารางที่	หน้า
20 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปร ด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของ ค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ใน กรณิเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 10% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบสัมพันธ์กันที่ระดับ สัมประสิทธิ์สหสัมพันธ์คงที่ 0.5 ($\Sigma = \Sigma_2$).....	89
21 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปร ด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของ ค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ใน กรณิเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 10% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบสัมพันธ์กันที่ระดับ สัมประสิทธิ์สหสัมพันธ์คงที่ 0.9 ($\Sigma = \Sigma_3$).....	91
22 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปร ด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของ ค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ใน กรณิเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 10% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบสัมพันธ์กันที่ระดับ สัมประสิทธิ์สหสัมพันธ์ไม่คงที่ ซึ่งกำหนดสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร ทำนายที่ i กับ j เท่ากับ $(0.5)^{ i-j }$ ($\Sigma = \Sigma_4$).....	93
23 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปร ด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามความแตกต่างของ ค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) และขนาดตัวอย่าง (n) ใน กรณิเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) เท่ากับ 10% และแบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องอยู่ในรูปแบบสัมพันธ์กันที่ระดับ สัมประสิทธิ์สหสัมพันธ์ไม่คงที่ ซึ่งกำหนดสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร ทำนายที่ i กับ j เท่ากับ $(0.9)^{ i-j }$ ($\Sigma = \Sigma_5$).....	95

สารบัญตาราง (ต่อ)

ตารางที่	หน้า
24 ผลการเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปรด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตามเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล (Prel) ความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง (Mdif) แบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้อง (Σ) และขนาดตัวอย่าง (n).....	97
25 จำนวนและเปอร์เซ็นต์ของเนื้อเยื่อมะเร็งลำไส้ใหญ่ และเนื้อเยื่อปกติ.....	102
26 ค่าสถิติพื้นฐานของค่าระดับการส่งออกของยีน 2,000 ยีน.....	102
27 ยีนที่ได้รับการคัดเลือกจากวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT	105
28 สถิติพื้นฐานของค่า BAC ของการจำแนกประเภทการเป็นมะเร็งลำไส้ใหญ่ ตลอดการแบ่งข้อมูลซ้ำ 200 รอบ สำหรับวิธีการคัดเลือกตัวแปรแต่ละวิธี.....	111
29 ค่าเฉลี่ยของ BAC และการทดสอบ z	112
30 รายละเอียดของชื่อยีนที่ได้รับการคัดเลือก.....	131
31 ผลการเปรียบเทียบค่าเฉลี่ย Sen ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปรด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตาม Prel, Mdif, Σ และ n	149
32 ผลการเปรียบเทียบค่าเฉลี่ย Spe ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปรด้วยวิธี VIP-1, VIP-BCT และ VIIC-BCT โดยจำแนกตาม Prel, Mdif, Σ และ n	153
33 เกณฑ์จุดตัดเฉลี่ยสำหรับดัชนี VIP เมื่อใช้วิธีการคัดเลือกตัวแปร VIP-BCT (BCT_{VIP}) โดยจำแนกตาม Prel, Mdif, Σ และ n	158
34 ผลการเปรียบเทียบค่าความเบ้เฉลี่ยของ VIP_{noise} และ $VIIC_{noise}$ เมื่อคัดเลือกตัวแปรด้วยวิธี VIP-BCT และวิธี VIIC-BCT โดยจำแนกตาม Prel, Mdif, Σ และ n	160
35 ตัวอย่างข้อมูลดิบของข้อมูลไมโครอาร์เรย์เกี่ยวกับมะเร็งลำไส้ใหญ่ที่ใช้ในงานวิจัย	172
36 ตัวอย่างข้อมูลดิบของค่า BAC, Sen และ Spe ของการจำแนกประเภทการเป็นมะเร็งลำไส้ใหญ่เมื่อใช้ยีนที่ได้จากการคัดเลือกตัวแปรด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT และจำแนกประเภทด้วยวิธี PLS-DA	173

สารบัญภาพ

ภาพที่	หน้า
1 กรอบแนวคิดการพัฒนาศูนย์วัดความสำคัญของตัวแปรและเกณฑ์จุดตัดในการคัดเลือกตัวแปร สำหรับการจำแนกประเภทข้อมูล.....	8
2 กรอบแนวคิดของการวิจัยในภาพรวม.....	9
3 การเปรียบเทียบแนวคิดของวิธีฟิเตอร์ วิธีเรปเปอร์ และ วิธีฝังตัว.....	20
4 ขั้นตอนการสร้างตัวแบบของ PLS-R.....	27
5 การทำ 5-fold Cross-Validation	28
6 แนวทางการสร้างตัวแบบการจำแนกประเภท.....	32
7 ขั้นตอนการทำนายกลุ่มจากตัวแบบ PLS-DA	35
8 แผนผังการพัฒนาศูนย์วัดความสำคัญของตัวแปรและเกณฑ์จุดตัดในการคัดเลือกตัวแปร สำหรับการจำแนกประเภทข้อมูล.....	44
9 แผนผังการจำลองข้อมูลเพื่อเปรียบเทียบการคัดเลือกตัวแปร.....	48
10 แผนผังการศึกษาความสามารถของยีนที่ได้จากการคัดเลือกตัวแปรในการจำแนกประเภทเนื้อเยื่อมะเร็งลำไส้ใหญ่ และเนื้อเยื่อปกติ.....	51
11 การเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปรด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT ในกรณี $Prel=1$ และ $\Sigma = \Sigma_1$	60
12 การเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปรด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT ในกรณี $Prel=1$ และ $\Sigma = \Sigma_2$	61
13 การเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปรด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT ในกรณี $Prel=1$ และ $\Sigma = \Sigma_3$	63
14 การเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปรด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT ในกรณี $Prel=1$ และ $\Sigma = \Sigma_4$	65
15 การเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปรด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT ในกรณี $Prel=1$ และ $\Sigma = \Sigma_5$	67
16 การเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปรด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT ในกรณี $Prel=3$ และ $\Sigma = \Sigma_1$	68
17 การเปรียบเทียบค่าเฉลี่ย BAC ของการจำแนกประเภทตัวแปร เมื่อคัดเลือกตัวแปรด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT ในกรณี $Prel=3$ และ $\Sigma = \Sigma_2$	70

สารบัญภาพ (ต่อ)

ภาพที่

หน้า

- 33 ตัวอย่างฮิสโทแกรมของ VIP_{noise} และ $VII C_{noise}$ สำหรับสถานการณ์ที่ $Prel=10$,
 $Mdif=1$, $n=70$ และ $\Sigma = \Sigma_3$ 162

มหาวิทยาลัยบูรพา
 Burapha University