

# การทำเหมืองข้อมูล เล่ม 2

## วิธีการและตัวแบบ

### Data Mining 2 : Methods and Models

รศ.สายชล สีนสมบูรณ์ทอง

ข้อสอบ

5

พ.ศ.

# DM

# DM DM DM

สำหรับ

สถิติ คณิตศาสตร์ การวิจัยการดำเนินงาน คอมพิวเตอร์  
คอมพิวเตอร์ธุรกิจ สารสนเทศ วิศวกรรมคอมพิวเตอร์ วิศวกรรม  
สารสนเทศ การวิจัย การเงิน การธนาคาร การตลาด  
การบริหารเทคโนโลยี การแพทย์ สาธารณสุข พืชสวน พืชไร่

เนื้อหาประกอบด้วย

- วิธีการลดมิติข้อมูล
- ตัวแบบการถดถอยเชิงเส้นอย่างง่าย
- ตัวแบบการถดถอยเชิงซ้อน
- ตัวแบบการถดถอยโลจิสติก
- การประมาณค่าหน่วยนิยม และโครงข่ายฮอปย์เชื่อม
- อัลกอริทึมเม็ทไฟนลุ่ม
- อัลกอริทึมการค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อย
- โปรแกรมในการทำเหมืองข้อมูล

# DM DM DM

# DM

ภาควิชาสถิติ

คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

# สารบัญ

	หน้า
<b>บทที่ 1 วิธีการลดมิติข้อมูล</b>	1
1.1 ความจำเป็นสำหรับการลดมิติข้อมูลในการทำเหมืองข้อมูล	1
1.2 การวิเคราะห์หองศ์ประกอบหลัก	2
1.3 การวิเคราะห์ปัจจัย	23
1.4 ส่วนประกอบที่ผู้ใช้กำหนด	31
1.5 สรุป	33
<b>แบบฝึกหัดที่ 1</b>	37
<b>เฉลยแบบฝึกหัดที่ 1</b>	42
<b>บทที่ 2 ตัวแบบการถดถอยเชิงเส้นอย่างง่าย</b>	45
2.1 ตัวอย่างการถดถอยเชิงเส้นอย่างง่าย	45
2.2 ค่าประมาณกำลังสองน้อยที่สุด	48
2.3 สัมประสิทธิ์การตัดสินใจ	52
2.4 ความคลาดเคลื่อนมาตรฐานของค่าประมาณ	58
2.5 สัมประสิทธิ์สหสัมพันธ์	60
2.6 ตารางวิเคราะห์ความแปรปรวน	62
2.7 ค่าออกกรอบ จุดตีเวยเรจที่มีค่ามาก และค่าสังเกตที่มีอิทธิพล	65
2.8 ตัวแบบการถดถอย	74
2.9 การอนุมานในการถดถอย	76
2.10 การตรวจสอบข้อสมมติเบื้องต้นของการถดถอย	86
2.11 ตัวอย่างข้อมูลเบสบอล	91
2.12 ตัวอย่างข้อมูลแคลิฟอร์เนีย	100
2.13 การแปลงข้อมูลให้เป็นเชิงเส้น	105
2.14 สรุป	112
<b>แบบฝึกหัดที่ 2</b>	116
<b>เฉลยแบบฝึกหัดที่ 2</b>	122

## สารบัญ (ต่อ)

	หน้า
<b>บทที่ 3 ตัวแบบการถดถอยเชิงซ้อน</b>	127
3.1 ตัวอย่างของการถดถอยเชิงซ้อน	127
3.2 ตัวแบบการถดถอยเชิงซ้อน	133
3.3 การอนุมานในการถดถอยเชิงซ้อน	135
3.4 การถดถอยที่มีตัวแปรทำนายเป็นข้อมูลเชิงกลุ่ม	141
3.5 ความสัมพันธ์กันของตัวแปรทำนาย	156
3.6 วิธีการเลือกตัวแปร	163
3.7 การประยุกต์ใช้วิธีการเลือกตัวแปร	168
3.8 สถิติ Mallows' $C_p$	174
3.9 เกณฑ์การเลือกตัวแปร	178
3.10 การใช้องค์ประกอบหลักเป็นตัวแปรทำนาย	189
3.11 สรุป	196
<b>แบบฝึกหัดที่ 3</b>	198
<b>เฉลยแบบฝึกหัดที่ 3</b>	204
<b>บทที่ 4 ตัวแบบการถดถอยโลจิสติก</b>	211
4.1 ตัวอย่างของการถดถอยโลจิสติก	211
4.2 การประมาณค่าความควรจะเป็นสูงสุด	214
4.3 การแปลผลการถดถอยโลจิสติก	215
4.4 การอนุมาน : ตัวแปรทำนายมีนัยสำคัญหรือไม่	216
4.5 การแปลความหมายตัวแบบการถดถอยโลจิสติก	219
4.6 ข้อสมมติเบื้องต้นของความเป็นเชิงเส้น	236
4.7 ปัญหาเรลส์เป็นศูนย์	240
4.8 ตัวแบบการถดถอยโลจิสติกเชิงซ้อน	242
4.9 เทอมอันดับที่สูงขึ้นเพื่อจัดการกับความไม่เป็นเชิงเส้น	248

## สารบัญ (ต่อ)

	หน้า
4.10 ความถูกต้องของตัวแบบการติดต่อใจจิตติก	257
4.11 สรุป	262
แบบฝึกหัดที่ 4	264
เฉลยแบบฝึกหัดที่ 4	270
<b>บทที่ 5 การประมาณค่านาอ็ฟเบย์และโครงข่ายเบย์เซียน</b>	<b>277</b>
5.1 วิธีเบย์เซียน	277
5.2 การจำแนกกลุ่มภายหลังสูงสุด	280
5.2.1 อัตราส่วนเชิดคภายหลัง	287
5.2.2 ความสมดุลกันของข้อมูล	289
5.3 การจำแนกกลุ่มนาอ็ฟเบย์	292
5.3.1 ตัวแปรทำนายเชิงตัวเลข	298
5.4 โครงข่ายความเชื่อของเบย์เซียน	303
5.4.1 ตัวอย่างการซื้อเสื้อผ้า	303
5.4.2 การใช้โครงข่ายเบย์เซียนเพื่อหาความน่าจะเป็น	306
5.5 สรุป	309
แบบฝึกหัดที่ 5	312
เฉลยแบบฝึกหัดที่ 5	314
<b>บทที่ 6 อัลกอริทึมเชิงพันธุกรรม</b>	<b>321</b>
6.1 อัลกอริทึมเชิงพันธุกรรมเบื้องต้น	321
6.2 กรอบในการทำงานเบื้องต้นของอัลกอริทึมเชิงพันธุกรรม	322
6.3 ตัวอย่างอย่างง่ายของอัลกอริทึมเชิงพันธุกรรม	323
6.4 การดัดแปลงและการเพิ่มขึ้น : การคัดเลือก	326
6.5 การดัดแปลงและการเพิ่มขึ้น : การผสมยีน	328

## สารบัญ (ต่อ)

	หน้า
6.6 อัลกอริทึมเชิงพันธุกรรมสำหรับตัวแปรค่าจริง	329
6.7 การใช้อัลกอริทึมเชิงพันธุกรรมเพื่อฝึกหัดโครงข่ายประสาท	331
6.8 อัลกอริทึมเชิงพันธุกรรมกับการเรียนแบบกระบวนการทางธรรมชาติเพื่อค้นหาคำตอบ	335
6.9 สรุป	343
แบบฝึกหัดที่ 6	346
เฉลยแบบฝึกหัดที่ 6	348
<b>บทที่ 7 อัลกอริทึมการค้นหาข้อมูลที่ปรากฏร่วมกันบ่อย</b>	349
7.1 อัลกอริทึม Apriori	349
7.2 อัลกอริทึม FP-Growth	355
แบบฝึกหัดที่ 7	363
เฉลยแบบฝึกหัดที่ 7	364
<b>บทที่ 8 โปรแกรมในการทำเหมืองข้อมูล</b>	365
8.1 การวิเคราะห์ข้อมูลโดยใช้องค์ประกอบหลัก	365
8.2 การวิเคราะห์ข้อมูลโดยใช้การถดถอยอย่างง่าย	369
8.3 การวิเคราะห์ข้อมูลโดยใช้การถดถอยเชิงซ้อน	373
8.4 การวิเคราะห์ข้อมูลโดยใช้การถดถอยโลจิสติก	376
8.5 การวิเคราะห์ข้อมูลโดยใช้นาอีฟเบย์	390
8.6 การวิเคราะห์ข้อมูลโดยใช้เบย์เนท	403
แบบฝึกหัดที่ 8	417
เฉลยแบบฝึกหัดที่ 8	418
<b>บรรณานุกรม</b>	420