

EGCO 425 Data Mining (T2/2019)

SO 1 An ability to identify, formulate, and solve complex engineering problems by applying principles of engineering, science, and mathematics

PI 1.1 Identify and formulate engineering problems

PI 1.2 Solve engineering problems by applying mathematics and engineering knowledge

Passing criterion $\geq 70\%$

PI 1.1 Identify and formulate engineering problems

Assessment 1.1.1 Midterm exam : identify effects of parameter setups

Attainability = 24 out of 26 students (92%)

3.6 MinPts ในอัลกอริทึม DBSCAN มีบทบาทในการทำ clustering อย่างไร ถ้าพารามิเตอร์ตัวนี้มีค่าสูงเกินไปจะมีผลอย่างไร และถ้ามีค่าต่ำเกินไปจะมีผลอย่างไร (3 คะแนน)

MinPts = เกณฑ์ที่บอกว่าจุด p ใดๆ เป็น core point หรือไม่ (ภายในรัศมี r จากที่มี p เป็นจุดศูนย์กลาง ต้องมีสมาชิกอย่างน้อย MinPts)

สูงเกินไป : p อาจไม่ผ่านเกณฑ์ กลายเป็น noise ทั้งที่ไม่ควรเป็น (1)

ต่ำเกินไป : มี noise ผ่านเกณฑ์ กลายเป็น core point (1)

ABET : course objective 1, SO 1 → pass at 1.5 points

Average attainability = 92%

Student performance

	Student ID	Assessment 1.1.1 Pass at 1.5 points
1	5913200	3.0
2	5913376	3.0
3	5913386	3.0
4	6013110	1.5
5	6013111	3.0
6	6013112	3.0
7	6013113	3.0
8	6013115	3.0
9	6013117	3.0
10	6013118	0.0
11	6013120	2.5
12	6013121	2.0
13	6013122	3.0
14	6013123	3.0
15	6013124	3.0
16	6013125	2.5
17	6013130	0.5
18	6013287	3.0
19	6013288	3.0
20	6013289	3.0
21	6013290	2.5
22	6013291	2.0
23	6013293	3.0
24	6013294	3.0
25	6013300	3.0
26	6013301	3.0

PI 1.2 Solve engineering problems by applying mathematics and engineering knowledge

Assessment 1.2.1 Midterm exam : apply math & stat methods to analyze data
 Attainability = 22 out of 26 students (85%)

1.3 ต้องการ discretize ตัวแปร score โดยแบ่งค่าของตัวแปรเป็น 2 bin

แบ่งแบบที่ 1	Bin 1: score ≤ 19						Bin 2: score > 19								
Score	10	10	12	14	15	18	20	22	23	23	25	28	29	29	30
Class	M	F	F	M	M	M	F	M	M	M	M	M	F	F	M

แบ่งแบบที่ 2	Bin 1: score ≤ 21							Bin 2: score > 21							
Score	10	10	12	14	15	18	20	22	23	23	25	28	29	29	30
Class	M	F	F	M	M	M	F	M	M	M	M	M	F	F	M

ค่า entropy หลังแบ่งข้อมูลแต่ละแบบจะเป็นเท่าไร และวิธีแบ่งข้อมูลแบบไหนดีกว่ากัน (3 คะแนน)

case 1: $\frac{6}{15} \left[-\frac{4}{6} \log \frac{4}{6} - \frac{2}{6} \log \frac{2}{6} \right] + \frac{9}{15} \left[-\frac{6}{9} \log \frac{6}{9} - \frac{3}{9} \log \frac{3}{9} \right] = \frac{6}{15} (0.9183) + \frac{9}{15} (0.9183) = 0.9183$

case 2: $\frac{7}{15} \left[-\frac{4}{7} \log \frac{4}{7} - \frac{3}{7} \log \frac{3}{7} \right] + \frac{8}{15} \left[-\frac{6}{8} \log \frac{6}{8} - \frac{2}{8} \log \frac{2}{8} \right] = \frac{7}{15} (0.9852) + \frac{8}{15} (0.8113) = 0.8924$

แบบที่ 2 ดีกว่า เพราะให้ค่า entropy ต่ำกว่า

ABET : course objective 2, SO 1 → pass at 1.5 points

Average attainability = 85%

Student performance

	Student ID	Assessment 1.2.1 Pass at 1.5 points
1	5913200	3.0
2	5913376	3.0
3	5913386	3.0
4	6013110	3.0
5	6013111	3.0
6	6013112	2.5
7	6013113	3.0
8	6013115	3.0
9	6013117	3.0
10	6013118	1.0
11	6013120	3.0
12	6013121	1.5
13	6013122	3.0
14	6013123	3.0
15	6013124	3.0
16	6013125	3.0
17	6013130	2.5
18	6013287	3.0
19	6013288	0.0
20	6013289	2.0
21	6013290	3.0
22	6013291	3.0
23	6013293	3.0
24	6013294	1.0
25	6013300	3.0
26	6013301	1.0

SO 6 An ability to develop and conduct appropriate experimentation, analyze and interpret data, and use engineering judgment to draw conclusions

PI 6.1 Develop and conduct experimentation

PI 6.2 Analyze and interpret data, and use engineering judgment to draw conclusions

Passing criterion $\geq 70\%$

PI 6.1 Develop and conduct experimentation

Assessment 6.1.1 Group project 2. Apply clustering methods to wine data (<https://bit.ly/2T6me4P>) and write a report. This assessment focuses on data preparation, algorithm selection, and parameter setups (marking criteria 1.2, 2.1, 3.1 – pass at 15 out of 20 points).

Attainability = 26 out of 26 students (100%)

Average attainability = 100%

Student performance

	Student ID	Assessment 6.1.1 Pass at 15 points
1	5913200	20.0
2	5913376	20.0
3	5913386	20.0
4	6013110	16.0
5	6013111	16.0
6	6013112	17.0
7	6013113	20.0
8	6013115	20.0
9	6013117	20.0
10	6013118	20.0
11	6013120	20.0
12	6013121	17.0
13	6013122	20.0
14	6013123	20.0
15	6013124	17.0
16	6013125	20.0
17	6013130	17.0
18	6013287	20.0
19	6013288	17.0
20	6013289	18.0
21	6013290	18.0
22	6013291	16.0
23	6013293	20.0
24	6013294	20.0
25	6013300	20.0
26	6013301	20.0

PI 6.2 Analyze and interpret data, and use engineering judgment to draw conclusions

Assessment 6.2.1 Group project 2. Apply clustering methods to wine data (<https://bit.ly/2T6me4P>) and write a report. This assessment focuses on result presentation, discussion, and conclusion (marking criteria 2.2, 3.2, 4 – pass at 35 out of 55 points).

Attainability = 18 out of 26 students (69%)

Assessment 6.2.2 Final exam : evaluate and compare the performance of 2 classifiers
Attainability = 24 out of 26 students (92%)

7. ผลจากการจำแนกรูปถ่ายผลไม้ 3 ชนิดคือ กล้วย, สับปะรด, และสตรอเบอรี่ ด้วย classifier 2 ตัวมีดังนี้

Classifier 1

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.933	0.070	0.778	0.933	0.848	0.987	banana
0.833	0.133	0.556	0.833	0.667	0.974	pineapple
0.733	0.111	0.917	0.733	0.815	0.916	strawberry
0.792	0.106	0.828	0.792	0.797	0.924	Weighted Avg.

Classifier 2

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.800	0.070	0.750	0.800	0.774	0.961	banana
0.833	0.033	0.833	0.833	0.833	0.997	pineapple
0.867	0.185	0.886	0.867	0.876	0.895	strawberry
0.847	0.136	0.849	0.847	0.848	0.926	Weighted Avg.

ให้เลือกตัววัดเพียง 2 ตัวจากตาราง ที่จะใช้เปรียบเทียบความสามารถในการจำแนกคลาสแต่ละคลาส (per-class performance) โดย classifier แต่ละตัว อธิบายเหตุผลที่เลือกตัววัดแต่ละตัว บอกด้วยว่าตัววัดนั้นๆ วัดอะไร

และจากตัววัดที่เลือกมา classifier ไດจำแนกคลาสกล้วยได้ดีกว่ากัน, classifier ไດจำแนกคลาสสับปะรดได้ดีกว่ากัน, และ classifier ไດจำแนกคลาสสตรอเบอรี่ได้ดีกว่ากัน (7 คะแนน)

- อธิบายตัววัด 2 ตัว บอกด้วยว่าค่าสูงแปลว่าดี หรือต่ำแปลว่าดี 4
- กล้วย (classifier 1) 1
- สับปะรด (classifier 2) 1
- สตรอเบอรี่ (classifier 2) 1

ABET : course objective 4, SO 6 → pass at 4 points

Average attainability = $(69 + 92)/2 = 80\%$

Student performance

	Student ID	Assessment 6.2.1 Pass at 35 points	Assessment 6.2.2 Pass at 4 points
1	5913200	53.0	6.0
2	5913376	53.0	6.5
3	5913386	53.0	6.5
4	6013110	32.0	6.0
5	6013111	32.0	4.0
6	6013112	51.0	6.0
7	6013113	55.0	7.0
8	6013115	55.0	7.0
9	6013117	26.0	5.5
10	6013118	26.0	4.5
11	6013120	40.0	5.0
12	6013121	51.0	5.0
13	6013122	38.0	6.0
14	6013123	40.0	4.5
15	6013124	47.0	6.5
16	6013125	51.0	5.5
17	6013130	51.0	3.5
18	6013287	38.0	5.0
19	6013288	47.0	4.5
20	6013289	26.0	6.5
21	6013290	26.0	3.5
22	6013291	32.0	4.0
23	6013293	26.0	6.5
24	6013294	55.0	5.0
25	6013300	51.0	6.5
26	6013301	51.0	6.5

SO 7 An ability to acquire and apply new knowledge as needed, using appropriate learning strategies**PI 7.2 Use appropriate sources of knowledge****Passing criterion $\geq 70\%$**

PI 7.2 Use appropriate sources of knowledge

Assessment 7.2.1 Group project 2. Apply clustering methods to wine data (<https://bit.ly/2T6me4P>) and write a report. This assessment focuses on searching relevant information and background about wine quality (marking criteria 1.1 – pass at 6 out of 10 points)

Attainability = 23 out of 26 students (88%)

Average attainability = 88%**Student performance**

	Student ID	Assessment 7.2.1 Pass at 6 points
1	5913200	10.0
2	5913376	10.0
3	5913386	10.0
4	6013110	5.0
5	6013111	5.0
6	6013112	9.0
7	6013113	10.0
8	6013115	10.0
9	6013117	7.0
10	6013118	7.0
11	6013120	10.0
12	6013121	9.0
13	6013122	7.0
14	6013123	10.0
15	6013124	10.0
16	6013125	9.0
17	6013130	9.0
18	6013287	7.0
19	6013288	10.0
20	6013289	7.0
21	6013290	7.0
22	6013291	5.0
23	6013293	7.0
24	6013294	10.0
25	6013300	9.0
26	6013301	9.0

Summary

	PI	Attainable ($\geq 70\%$)	Reasons	Remedial Actions	Action Plan	Measurements
SO1	1.1	Yes				
	1.2	Yes				
SO6	6.1	Yes				
	6.2	Yes	<ul style="list-style-type: none"> Students could interpret clustering results. But some struggled to discuss them and draw conclusions in the context of the given domain (wine quality) 	<ul style="list-style-type: none"> More case studies, with more focus on discussion and conclusion of results 	Next year	Next year
SO7	7.2	Yes				