แบบฝึกหัด

1. กำหนดแอทริบิวต์ดังต่อไปนี้ จงจำแนกประเภทของแอทริบิวต์ว่าเป็นแอทริบิวต์เชิงคุณภาพ (qualitative) ชนิดใด (nominal, ordinal) หรือเป็นแอทริบิวต์เชิงปริมาณ (quantitative) ชนิดใด (interval, ratio) นอกจากนี้ให้จำแนกด้วยว่าแอทริบิวต์แต่ละตัวมีชนิดเป็น Binary, Discrete, หรือ Continuous
   (ก) เวลาในรูปแบบ AM หรือ PM
   (ข) ความสว่างที่วัดโดยมิเตอร์วัดแสง
   (ค) ความสว่างที่วัดจากความรู้สึกของมนุษย์

(ง) มุมที่วัดเป็นองศาระหว่าง 0 ถึง 360

(จ) เหรียญทอง เหรียญเงิน เหรียญทองแดง ของกีฬาโอลิมปิก

(ฉ) ความสูงจากระดับน้ำทะเล

(ช) จำนวนผู้ป่วยในโรงพยาบาลแห่งหนึ่ง

(ซ) เลข ISBN ของหนังสือ

(ฌ) ความหนาแน่นของสสารหน่วยเป็น กรัมต่อลูกบาศก์เซนติเมตร

(ญ) ระยะทางจากจุดศูนย์กลางของวิทยาเขตหน่วยเป็นเมตร

(ฎ) ชั้นยศของกองทัพ

(ฏ) ความสามารถในการส่งผ่านคลื่นแสง: opaque translucent transparent

2. จงยกตัวอย่างสถาณการณ์ที่ identification numbers (รหัสประจำตัว) น่าจะมีประโยชน์สำหรับการทำนาย
3. ปริมาณใดต่อไปนี้ที่มีคุณสมบัติ spatial autocorrelation : daily rainfall หรือ daily temperature และทำไมจึงเป็นเช่นนั้น
4. โปรแกรมเมอร์คนหนึ่งได้ออกแบบอัลกอริทึม k-nearest neighbors ดังนี้

Algorithm 2.1 Algorithm for finding K nearest neighbors.

1: **for** $i = 1$ to *number of data objects* **do**
2: Find the distances of the i^{th} object to all other objects.
3: Sort these distances in decreasing order.
(Keep track of which object is associated with each distance.)
4: **return** the objects associated with the first K distances of the sorted list
5: **end for**

(ก) จงอภิปรายว่าจะมีปัญหาอะไรเกิดขึ้นได้บ้างกับอัลกอริทึมนี้ ถ้าดาต้าเซตมีข้อมูลซ้ำ (duplicates)

(ข) จงเสนอวิธีการแก้ไขปัญหาที่เกิดจากการมีข้อมูลซ้ำซ้อนในดาต้าเซต

5. คำนวณค่า cosine, correlation, Jaccard และ Euclidean distance ของ ดาต้าอ็อบเจ็กต์ x และ y ดังต่อไปนี้

(ก) x = (1, 1, 1, 1), y = (2, 2, 2, 2)

(ข) x = (0, 1, 0, 1), y = (1, 0, 1, 0)

(ค) x = (0, -1, 0, 1), y = (1, 0, -1, 0)

6. คำนวณค่า Mutual information ของดาต้าอ็อบเจ็กต์ x และ y ดังต่อไปนี้

(ก) x = (-7, -2, 1, 0, 1, 2), y = (9, 4, 1, 0, 4, 1)

(ข) x = (1,1,1,1), y = (2,2,2,2)

เอกสารอ้างอิง

[1] Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar. “Introduction to Data Mining”. Pearson, 2nd edition, 2018.