6252300462 นายพงศกร สมอไทย

1.(10 คะแนน) จงหาอัตราความถูกต้อง (มีค่าตั้งแต่ 0 ถึง 1 ตอบเป็นทศนิยม 3 ตำแหน่ง) ในการใช้งานชั้นตอนวิธี k-nearest neighbors ในการจำแนกชุดข้อมูลต่อไปนี้ทุกค่า k ที่เป็นไปได้ กำหนดให้ใช้ข้อมูล id ที่ 1-8 เป็นชุดฝึกฝนและข้อมูล id ที่ 9-16 เป็นชุดทดสอบ และใช้ squared Euclidean distance ในการหาระยะห่างและการให้เปรียบเทียบตามลำดับตัวอักษร A-Z

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id | A1 | A2 | A3 | A4 | target |
| 1 | 1 | 1 | 4 | 4 | B |
| 2 | 3 | 0 | 5 | 1 | C |
| 3 | 1 | 0 | 0 | 5 | B |
| 4 | 2 | 5 | 0 | 1 | C |
| 5 | 2 | 1 | 1 | 2 | D |
| 6 | 4 | 3 | 3 | 3 | A |
| 7 | 0 | 1 | 0 | 2 | D |
| 8 | 3 | 4 | 4 | 4 | A |
| 9 | 4 | 1 | 4 | 3 | D |
| 10 | 3 | 1 | 3 | 3 | A |
| 11 | 2 | 1 | 2 | 1 | A |
| 12 | 3 | 2 | 3 | 4 | D |
| 13 | 1 | 5 | 3 | 1 | B |
| 14 | 4 | 2 | 0 | 1 | C |
| 15 | 1 | 1 | 4 | 3 | B |
| 16 | 3 | 0 | 0 | 0 | C |

6252300462 นายพงศกร สมอไทย

2.(5 คะแนน) จงแปลงชุดข้อมูลทั้งหมดในข้อที่ 1 เป็นรหัสไบนารีด้วยวิธี Locality Sensitive Hashing (LSH) โดยใช้เวกเตอร์การแปลงดังต่อไปนี้ กำหนดให้รหัสมีค่า 1 เมื่อผลคูณภายในของข้อมูลและเวกเตอร์การแปลงมีค่าไม่ติดลบนอกนั้นรหัสมีค่าเป็น 0

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| id | C1 | C2 | C3 | C4 |
| 1 | 5 | 5 | 1 | -1 |
| 2 | 4 | 4 | 5 | 5 |
| 3 | 3 | -2 | 5 | 1 |

6252300462 นายพงศกร สมอไทย

3.(15 คะแนน) จงแสดงวิธีการสร้างต้นไม้ตัดสินใจด้วยขั้นตอนวิธี ID3 จากชุดข้อมูลต่อไปนี้ (ใช้ log ฐาน 2)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| id | A1 | A2 | A3 | target |
| 1 | 4 | 1 | 3 | - |
| 2 | 1 | 1 | 1 | - |
| 3 | 3 | 2 | 1 | - |
| 4 | 2 | 4 | 3 | + |
| 5 | 1 | 1 | 3 | - |
| 6 | 3 | 4 | 1 | + |
| 7 | 4 | 1 | 3 | - |
| 8 | 2 | 3 | 1 | - |
| 9 | 1 | 2 | 2 | + |
| 10 | 1 | 2 | 3 | + |
| 11 | 1 | 3 | 1 | + |
| 12 | 3 | 2 | 3 | + |