**Linear Regression**

**สูตรสำคัญ:**

* **ความชัน (m) :** 
* **จุดตัดแกน Y (c) :** 

**โจทย์ข้อที่ 1.1**

บริษัทขายไอศกรีมต้องการทำนายยอดขาย (ถ้วย) จากอุณหภูมิสูงสุดของวัน (องศาเซลเซียส) โดยมีข้อมูล 5 วันล่าสุดดังนี้

|  |  |
| --- | --- |
| **อุณหภูมิ (X)** | **ยอดขาย (Y)** |
| 25 | 150 |
| 30 | 200 |
| 32 | 230 |
| 28 | 180 |
| 35 | 250 |

**คำสั่ง:**

1. จงหาสมการ Linear Regression ***(y=mx+c)*** จากข้อมูลข้างต้น
2. ถ้าวันนี้อุณหภูมิ 33 องศาเซลเซียส คาดว่าจะขายไอศกรีมได้กี่ถ้วย?

**โจทย์ข้อที่ 1.2**

ฟิตเนสแห่งหนึ่งต้องการวิเคราะห์ความสัมพันธ์ระหว่างจำนวนชั่วโมงที่ลูกค้าออกกำลังกายต่อสัปดาห์ (X) กับน้ำหนักที่ลดลงในหนึ่งเดือน (กก.) (Y)

|  |  |
| --- | --- |
| **ชั่วโมง/สัปดาห์ (X)** | **น้ำหนักที่ลด (Y)** |
| 3 | 1.5 |
| 5 | 2.0 |
| 2 | 1.0 |
| 6 | 3.0 |
| 4 | 2.2 |
| 7 | 3.5 |

**คำสั่ง:**

1. จงหาสมการ Linear Regression
2. หากลูกค้าออกกำลังกาย 8 ชั่วโมง/สัปดาห์ คาดว่าน้ำหนักจะลดลงกี่กิโลกรัม?

**Decision Tree (Regression)**

**สูตรสำคัญ:**

* **Standard Deviation (SD) :** 
* **Standard Deviation Reduction (SDR) :** 

**โจทย์ข้อที่ 2.1**

ต้องการสร้างโมเดลทำนาย "ราคามือสอง" (Y, หน่วยเป็นพันบาท) ของสมาร์ทโฟน โดยพิจารณาจาก "อายุการใช้งาน (เดือน)" (X1)

|  |  |
| --- | --- |
| **อายุ (X1)** | **ราคา (Y)** |
| 6 | 18 |
| 12 | 14 |
| 24 | 9 |
| 8 | 17 |
| 18 | 11 |

**คำสั่ง:** จงหาการแบ่งครั้งแรก (First Split) ที่ดีที่สุด โดยคำนวณค่า Standard Deviation Reduction (SDR) ของทุกจุดแบ่งที่เป็นไปได้

**โจทย์ข้อที่ 2.2 (โจทย์ท้าทาย)**

บริษัทเกมต้องการสร้างโมเดลทำนาย "คะแนนในเกม" (Y) ของผู้เล่น โดยอ้างอิงจาก "ชั่วโมงที่เล่น" (X1) และ "เลเวลผู้เล่น" (X2) **เงื่อนไข:** หยุดแบ่ง Node (สร้าง Leaf) ก็ต่อเมื่อ Node นั้นมีข้อมูลน้อยกว่าหรือเท่ากับ 3 ชิ้น

|  |  |  |
| --- | --- | --- |
| **ชั่วโมงที่เล่น (X1)** | **เลเวลผู้เล่น (X2)** | **คะแนนในเกม (Y)** |
| 5 | 10 | 1200 |
| 15 | 25 | 3500 |
| 20 | 30 | 4500 |
| 2 | 5 | 500 |
| 8 | 15 | 1800 |
| 25 | 40 | 6000 |
| 12 | 20 | 2800 |
| 18 | 35 | 4000 |

**คำสั่ง:**

1. จงสร้าง Decision Tree จากข้อมูลทั้งหมดให้สมบูรณ์ตามขั้นตอน (แสดงการคำนวณเพื่อหาจุดแบ่งที่ดีที่สุดในแต่ละ Node)
2. วาดแผนผังต้นไม้ (Decision Tree) ที่สร้างเสร็จแล้ว
3. หากมีผู้เล่นใหม่ที่มี **ชั่วโมงที่เล่น 10 ชั่วโมง** และ **เลเวล 18** จงทำนายคะแนนของเขา

**K-Nearest Neighbors (K-NN)**

**สูตรสำคัญ:**

* **ระยะห่างแบบยูคลิด (Euclidean Distance) :** 

**โจทย์ข้อที่ 3.1**

นักวิเคราะห์สินเชื่อมีข้อมูลการอนุมัติสินเชื่อส่วนบุคคล โดยพิจารณาจาก "รายได้ต่อปี (แสนบาท)" (X1) และ "หนี้สินรวม (แสนบาท)" (X2)

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **รายได้ (X1)** | **หนี้สิน (X2)** | **ผลอนุมัติ (Y)** |
| P1 | 5 | 1 | อนุมัติ |
| P2 | 6 | 3 | อนุมัติ |
| P3 | 2 | 2 | ไม่อนุมัติ |
| P4 | 3 | 4 | ไม่อนุมัติ |
| P5 | 7 | 2 | อนุมัติ |
| P6 | 4 | 5 | ไม่อนุมัติ |

**คำสั่ง:** ลูกค้าใหม่ (P\_new) มี **รายได้ 6** แสนบาท และ **หนี้สิน 4** แสนบาท จงใช้ **K-NN (K=3)** ทำนายว่าลูกค้าคนนี้จะได้รับการอนุมัติหรือไม่?

**โจทย์ข้อที่ 3.2**

มหาวิทยาลัยแห่งหนึ่งใช้ข้อมูล "เกรดเฉลี่ยตอน ม.ปลาย" (X1) และ "คะแนนสอบเข้า" (X2) เพื่อคัดกรองนักศึกษาที่มีแนวโน้มจะ "เรียนต่อจนจบ" หรือ "ลาออก"

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **GPA (X1)** | **คะแนนสอบ (X2)** | **สถานะ (Y)** |
| S1 | 3.8 | 85 | เรียนจบ |
| S2 | 2.5 | 60 | ลาออก |
| S3 | 3.5 | 90 | เรียนจบ |
| S4 | 2.8 | 75 | ลาออก |
| S5 | 3.2 | 80 | เรียนจบ |
| S6 | 2.2 | 65 | ลาออก |
| S7 | 3.9 | 95 | เรียนจบ |

**คำสั่ง:** นักเรียนใหม่ (S\_new) มี **GPA 3.0** และ **คะแนนสอบ 70** จงใช้ **K-NN (K=5)** ทำนายสถานะของนักเรียนคนนี้

**4. Support Vector Machine (SVM)**

**โจทย์ข้อที่ 4.1**

มีข้อมูล 2 คลาส คือ A (สีฟ้า) และ B (สีแดง)

* **คลาส A:** P1(2, 5), P2(3, 2)
* **คลาส B:** P3(6, 4), P4(7, 7)

มีคนเสนอเส้นแบ่ง (Hyperplane) H1 คือเส้นแนวดิ่ง x=4.5**ผิดพลาด! ไม่ได้ระบุชื่อไฟล์**

**คำสั่ง:**

1. จงคำนวณหาระยะห่างจากทุกจุดไปยังเส้น H1
2. เส้น H1 มี Support Vectors คือจุดใดบ้าง? และมี Margin กว้างเท่าใด?
3. จงหาเส้นแบ่งที่ดีที่สุด (Optimal Hyperplane) และ Margin สูงสุดที่เป็นไปได้สำหรับข้อมูลชุดนี้

**โจทย์ข้อที่ 4.2**

จากข้อมูลชุดเดิมในข้อ 4.1 มีคนเสนอเส้นแบ่งใหม่ H2 คือ x+y−8=0**ผิดพลาด! ไม่ได้ระบุชื่อไฟล์**

**คำสั่ง:**

1. จงคำนวณหาระยะห่างจากทุกจุดไปยังเส้น H2
2. เส้น H2 มี Support Vectors คือจุดใดบ้าง และ Margin กว้างเท่าใด?
3. เปรียบเทียบกับผลลัพธ์ในข้อ 4.1 เส้น H2 เป็นเส้นแบ่งที่ดีที่สุดหรือไม่ เพราะอะไร?