**เฉลยใบงาน: Advanced Topics in Computer Engineering**

**ข้อที่ 1: K-Means Clustering (คำนวณ) (20 คะแนน)**

**ขั้นตอนที่ 1: การจัดสรร (Assignment Step) - รอบที่ 1** คำนวณระยะทางแบบยุคลิด (Euclidean Distance) จากทุกจุดไปยัง Centroid เริ่มต้น C1(2,2) และ C2(8,7)

รูปภาพประกอบด้วย ข้อความ, ภาพหน้าจอ, ตัวอักษร

เนื้อหาที่สร้างโดย AI อาจไม่ถูกต้อง

**ผลการจัดกลุ่มรอบที่ 1:**

* **กลุ่ม 1:** {P1(2,2), P2(3,2), P5(3,4)}
* **กลุ่ม 2:** {P3(8,7), P4(7,8)}

**ขั้นตอนที่ 2: การปรับปรุง (Update Step) - รอบที่ 1** คำนวณตำแหน่ง Centroid ใหม่จากค่าเฉลี่ยของสมาชิกในแต่ละกลุ่ม

* **Centroid ใหม่ C1':**
  + ค่าเฉลี่ยแกน x = (2 + 3 + 3) / 3 = 8 / 3 ≈ 2.67
  + ค่าเฉลี่ยแกน y = (2 + 2 + 4) / 3 = 8 / 3 ≈ 2.67
  + **C1' = (2.67, 2.67)**
* **Centroid ใหม่ C2':**
  + ค่าเฉลี่ยแกน x = (8 + 7) / 2 = 15 / 2 = 7.5
  + ค่าเฉลี่ยแกน y = (7 + 8) / 2 = 15 / 2 = 7.5
  + **C2' = (7.5, 7.5)**

**คำตอบ:** หลังจากอัลกอริทึมทำงานครบ 1 รอบ ตำแหน่งของ Centroid ใหม่คือ **C1' = (2.67, 2.67)** และ **C2' = (7.5, 7.5)**

**ข้อที่ 2: Principal Component Analysis (PCA) (คำนวณและวิเคราะห์) (30 คะแนน)**

**2.1) คำนวณหาค่าเฉลี่ยและปรับข้อมูล (Center the Data)**

* **ค่าเฉลี่ยแกน x:** (2 + 3 + 5 + 6) / 4 = 16 / 4 = **4.0**
* **ค่าเฉลี่ยแกน y:** (3 + 5 + 6 + 7) / 4 = 21 / 4 = **5.25**
* **ข้อมูลที่ปรับแล้ว (Centered Data):**
  + A' = (2 - 4.0, 3 - 5.25) = (-2, -2.25)
  + B' = (3 - 4.0, 5 - 5.25) = (-1, -0.25)
  + C' = (5 - 4.0, 6 - 5.25) = (1, 0.75)
  + D' = (6 - 4.0, 7 - 5.25) = (2, 1.75)

**2.2) คำนวณหา Covariance Matrix**

* **Var(x) = Cov(x,x):** ((-2)² + (-1)² + 1² + 2²) / (4-1) = (4 + 1 + 1 + 4) / 3 = 10 / 3 ≈ **3.33**
* **Cov(x,y):** ((-2)(-2.25) + (-1)(-0.25) + (1)(0.75) + (2)(1.75)) / (4-1) = (4.5 + 0.25 + 0.75 + 3.5) / 3 = 9 / 3 = **3.0**
* **Var(y) = Cov(y,y):** ((-2.25)² + (-0.25)² + (0.75)² + (1.75)²) / (4-1) = (5.0625 + 0.0625 + 0.5625 + 3.0625) / 3 = 8.75 / 3 ≈ **2.92**

รูปภาพประกอบด้วย ข้อความ, ตัวอักษร, ภาพหน้าจอ, จำนวน

เนื้อหาที่สร้างโดย AI อาจไม่ถูกต้อง

2.3) คำนวณหา Eigenvector ของ PC1

รูปภาพประกอบด้วย ข้อความ, ภาพหน้าจอ, ตัวอักษร, จำนวน

เนื้อหาที่สร้างโดย AI อาจไม่ถูกต้อง

จากสมการแถวแรก: -2.8x + 3.0y = 0 --> y = (2.8/3.0)x ≈ 0.93x ถ้าให้ x=1, จะได้ y=0.93 --> เวกเตอร์ v = [1, 0.93]

**ทำ Normalization:**



* x\_unit = 1 / 1.365 ≈ 0.732
* y\_unit = 0.93 / 1.365 ≈ 0.681
* **Eigenvector (PC1) ≈ [0.73, 0.68]**

**2.4) แปลงข้อมูลไปยังมิติใหม่ (1D)** นำข้อมูลที่ปรับแล้ว (Centered Data) มา dot product กับ PC1

* **A\_new:** (-2 \* 0.73) + (-2.25 \* 0.68) = -1.46 - 1.53 = **-2.99**
* **B\_new:** (-1 \* 0.73) + (-0.25 \* 0.68) = -0.73 - 0.17 = **-0.90**
* **C\_new:** (1 \* 0.73) + (0.75 \* 0.68) = 0.73 + 0.51 = **1.24**
* **D\_new:** (2 \* 0.73) + (1.75 \* 0.68) = 1.46 + 1.19 = **2.65**

**ข้อที่ 3: Artificial Neural Networks (ANN) (วิเคราะห์) (20 คะแนน)**

**3.1) Input Layer:**

* **วิธีคำนวณ:** เนื่องจาก ANN แบบดั้งเดิมไม่สามารถประมวลผลข้อมูล 2 มิติ (ภาพ) ได้โดยตรง เราจึงต้องทำการ "คลี่" (Flatten) ภาพให้เป็นเวกเตอร์ 1 มิติ จำนวนเซลล์ประสาทจะเท่ากับจำนวนพิกเซลทั้งหมด จำนวนเซลล์ = ความกว้าง x ความสูง x จำนวนช่องสี (channels) จำนวนเซลล์ = 64 x 64 x 3 = 12,288
* **คำตอบ:** Input Layer ควรมี **12,288 เซลล์**

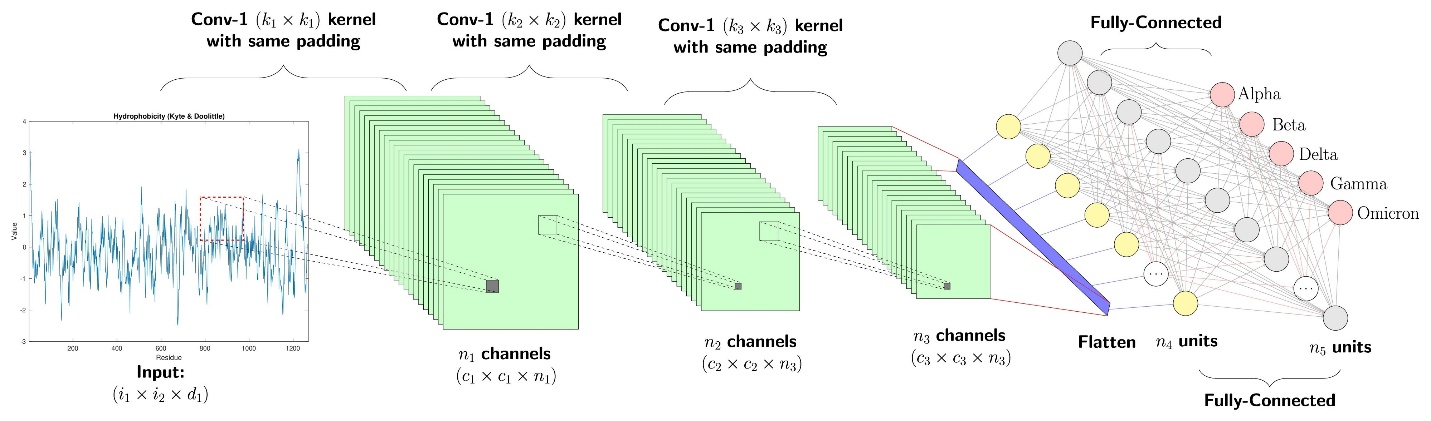
**3.2) Output Layer:**

* **จำนวนเซลล์:** สำหรับปัญหาการจำแนก 2 กลุ่ม (Binary Classification) เราใช้เพียง **1 เซลล์** ก็เพียงพอ
* **Activation Function:** ควรใช้ **Sigmoid** ฟังก์ชัน
* **เหตุผล:** ฟังก์ชัน Sigmoid จะให้ผลลัพธ์ออกมาเป็นค่าระหว่าง 0 ถึง 1 ซึ่งสามารถตีความได้ว่าเป็น "ความน่าจะเป็น" (Probability) ที่ภาพนั้นจะเป็นหนึ่งในสองคลาส เช่น ถ้าค่าเข้าใกล้ 1 หมายถึง "สุนัข" และถ้าค่าเข้าใกล้ 0 หมายถึง "แมว"

**ข้อที่ 4: Computer Vision (วิเคราะห์) (15 คะแนน)**

CNNs สามารถจดจำวัตถุได้ดีกว่า MLP แบบดั้งเดิมในสภาวะที่หลากหลาย เพราะสถาปัตยกรรมของ CNN ถูกออกแบบมาเพื่อจัดการกับข้อมูลเชิงพื้นที่ (Spatial Data) โดยเฉพาะ ผ่านองค์ประกอบสำคัญดังนี้:

1. **Convolutional Layer:** ทำหน้าที่เป็น "ตัวตรวจจับคุณลักษณะ" (Feature Detector) โดยใช้ฟิลเตอร์ (Filter/Kernel) เลื่อนไปทั่วภาพเพื่อตรวจจับลักษณะเฉพาะ เช่น เส้นขอบ, มุม, หรือพื้นผิว การที่ฟิลเตอร์ตัวเดียวกันถูกใช้ซ้ำทั่วทั้งภาพ ทำให้ CNN สามารถตรวจจับคุณลักษณะได้ไม่ว่ามันจะปรากฏอยู่ส่วนไหนของภาพก็ตาม (คุณสมบัตินี้เรียกว่า **Translation Invariance**) ซึ่งแก้ปัญหาเรื่อง **มุมที่ต่างกัน** ได้ดี
2. **Pooling Layer:** ทำหน้าที่ลดขนาดของ Feature Map (Downsampling) โดยสรุปข้อมูลในบริเวณเล็กๆ ให้เหลือค่าเดียว (เช่น Max Pooling จะเลือกค่าที่สูงสุด) กระบวนการนี้ทำให้แบบจำลองทนทานต่อการเปลี่ยนแปลงเล็กๆ น้อยๆ เช่น การเลื่อนตำแหน่งหรือการบิดเบี้ยวของวัตถุ และยังช่วยให้สามารถจดจำวัตถุได้แม้จะมี **ขนาดในภาพไม่เท่ากัน** (Scale Invariance)



ในทางตรงกันข้าม MLP จะทำการคลี่ภาพ (Flatten) เป็นเวกเตอร์ 1 มิติก่อนประมวลผล ทำให้ข้อมูลความสัมพันธ์เชิงพื้นที่ระหว่างพิกเซลข้างเคียงสูญหายไปทั้งหมด จึงไม่สามารถเรียนรู้คุณลักษณะเหล่านี้ได้อย่างมีประสิทธิภาพ

**ข้อที่ 5: AI Ethics (คิดวิเคราะห์) (15 คะแนน)**

**5.1) ปัญหาและสาเหตุ:**

* **ชื่อปัญหา:** ปัญหานี้เรียกว่า **อคติของอัลกอริทึม (Algorithmic Bias)** หรือ **อคติใน AI (AI Bias)**
* **สาเหตุที่เป็นไปได้:** สาเหตุหลักน่าจะมาจาก **อคติในข้อมูลที่ใช้ฝึก (Data Bias)** กล่าวคือ ข้อมูลประวัติการรับสมัครงานในอดีตที่ใช้สอน AI อาจมีจำนวนผู้ชายที่ได้รับการคัดเลือกในตำแหน่งวิศวกรมากกว่าผู้หญิงอย่างมีนัยสำคัญ ทำให้ AI "เรียนรู้" ว่าเพศชายเป็นปัจจัยหนึ่งที่ส่งผลต่อการได้รับคัดเลือก และนำอคตินั้นมาใช้ในการตัดสินใจกับข้อมูลใหม่

**5.2) แนวทางการแก้ไข:**

1. **การตรวจสอบและปรับปรุงชุดข้อมูล (Data Auditing and Preprocessing):** ก่อนนำข้อมูลไปฝึกสอน ควรมีการตรวจสอบเพื่อหาอคติที่อาจแฝงอยู่ หากพบว่าข้อมูลไม่สมดุล (เช่น มีข้อมูลผู้สมัครชายมากกว่าหญิง) ควรทำการแก้ไข เช่น การสุ่มตัวอย่างข้อมูลจากกลุ่มส่วนน้อยเพิ่มขึ้น (Oversampling) หรือการลดข้อมูลจากกลุ่มส่วนใหญ่ลง (Undersampling) เพื่อให้ AI เรียนรู้จากข้อมูลที่เป็นตัวแทนของทุกกลุ่มอย่างเท่าเทียม
2. **การใช้เทคนิค Fairness-aware Machine Learning:** นำเมตริกวัด "ความเป็นธรรม" (Fairness Metrics) มาใช้ในระหว่างการประเมินผลโมเดล เพื่อตรวจสอบว่าโมเดลให้ผลลัพธ์ที่แตกต่างกันอย่างมีนัยสำคัญระหว่างกลุ่มประชากรต่างๆ (เช่น ชาย-หญิง) หรือไม่ และอาจปรับปรุงกระบวนการเรียนรู้ของโมเดลให้พยายามลดอคติดังกล่าวโดยตรง
3. **การใช้มนุษย์ในกระบวนการตัดสินใจ (Human-in-the-loop):** ไม่ควรให้ AI ตัดสินใจคัดเลือกผู้สมัครโดยอัตโนมัติ 100% แต่ควรใช้เป็นเครื่องมือ "ช่วยแนะนำ" หรือ "คัดกรองเบื้องต้น" เท่านั้น โดยให้มีเจ้าหน้าที่สรรหาบุคลากร (ที่เป็นมนุษย์) เป็นผู้พิจารณาและตัดสินใจในขั้นตอนสุดท้ายเสมอ เพื่อป้องกันการตัดสินใจที่ผิดพลาดจากอคติของ AI