以iris\_max3這張決策樹來解釋

一張含有 文字, 圖表, 方案, 字型 的圖片

自動產生的描述

1. 節點（Nodes）

每個方框代表一個節點，包含以下資訊：

• 特徵條件（如 x[2] <= 2.45）：決策樹根據此條件將數據分為左右子樹。

• 熵（Entropy）：衡量數據的不確定性，值越低，數據越純。

• 熵的公式為： \text{Entropy} = - \sum (p\_i \cdot \log\_2(p\_i))

• 0 表示純度最高（即該節點只有一種類別）。

• 樣本數量（Samples）：該節點包含的數據樣本總數。

• 類別分佈（Value）：該節點中各類別的樣本數量。例如，value = [40, 0, 0] 表示該節點包含 40 個屬於第一類的樣本。

2. 樹的結構

根節點（最上方節點）

• 條件：x[2] <= 2.45

• True（左子樹）：進入左邊分支。

• False（右子樹）：進入右邊分支。

• 熵：1.578，表示該節點的數據不確定性較高。

• 樣本數：105。

• 類別分佈：value = [40, 32, 33]，表示三種類別的分佈。

內部節點

• 每個內部節點基於一個特徵條件進行進一步分割。

• 分割目的是最大化每個分支的純度。

葉節點（終端節點）

• 熵為 0 的節點，例如 entropy = 0.0，表示該節點內的樣本完全屬於同一類別。

• 這些節點對應於最終的分類結果。

3. 如何解釋樹中的路徑

樹的每條路徑代表一個決策規則：

1. 從根節點開始，根據特徵條件進入左或右子樹。

2. 繼續沿路徑進行判斷，直到達到葉節點。

3. 葉節點的類別分佈（value）代表該條路徑上的分類結果。

範例：

• 如果樣本滿足條件 x[2] <= 2.45，則直接進入左側葉節點。

• 該葉節點有 40 個樣本，value = [40, 0, 0]，表示所有樣本都屬於第一類。

• 如果 x[2] > 2.45 且 x[3] <= 1.75，則進入右子樹第二層，再進行下一步判斷。

4. x[2] 是什麼？

• x[2] 指的是數據的第 2 個特徵（Feature）。

• 例如，在 Iris 資料集中，x[2] 對應於「花瓣長度（Petal Length）」。

• 特徵是決策樹用來進行分類或回歸的依據。