

A Assignment - 3

Q1

Decision tree structure :-

Root node (first split)

Feature : x_2

Threshold : b logic if $x_2 > b$ all data points are crossed. If $x_2 \leq b$ we have a mix of dots and crosses that require another split.

Second Split

Second split

feature: $-x_1 \leq a$

threshold: $-a$

logic looking at $x_2 \leq b$

if $x_1 \leq a$ all dots

otherwise miss all cross

Leaf 1: Assign Class +1 (Points where $x_2 > b$)

Leaf 2: Assign Class -1 ($x_2 \leq b$ & $x_1 \leq a$)

Leaf 3: Assign Class +1 ($x_2 \leq b$ & $x_1 > a$)

Q2. A random forest is an ensemble of decision trees.

Forest because many trees \rightarrow diverse

Random due to Bootstrapped sample, Random subsets of features at each split.

Randomness helps in reducing correlation b/w trees lowering variance and improving generalization

Q3. An ensemble method combines predictions from multiple models to produce final prediction.

Motivation: it reduces biasness

Reduce variance

Improves accuracy

Q4 confusion Matrix

① TP: 180
TN: 230

FP: 70
FN: 20

② Performance metrics

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{total}} = \frac{180 + 230}{1000} = 0.91$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{180}{250} = 0.72$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{180}{200} = 0.90$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{230}{800} = 0.7125$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times 0.72 \times 0.90}{0.72 + 0.90} = 0.80$$

③ In this case Recall would be prioritize

High recall ensures that as many sick individual as possible are flagged, even if some false call in case of a disease detection.

④ entries likely to increase would be TP & FP
lowering the threshold means the model becomes "less strict" allowing both more no. of

patient as positive and will naturally catch more 'true patients'.

5. Yes, two classifier can have same accuracy but very different confusion matrix.

Classifier A :- High TP & High FP

Classifier B :- High TN & High FN

Their accuracy can be same as the accuracy does not distinguish b/w TP & TN.

in Classifier A

Recall increases ~~Precision~~.

Precision & Specificity decreases

in Classifier B opposite of it happens