Homework - 3

Aman Gupta.

Ques1 Bias - variance trade-off →

The ability of a model to match the training data well (low bais) and its ability to genralize well to new, unseen data are Bias - variance trade-off.

the difference blw a model's expected output and actual output is reffred as the bias. high bais indicates that the model is underfitted because it is oversimplified.

Variance:→ It describes how differently the model predicts outcomes for various training groups. High variance indicates an overfitted model that is too complicated and has learned to account for train data noise.

To reduce the bias.

1. Increase model complexity.
2. Add more features.
3. Modifying the model architecture will enable us to better tailor it to the particular issue we are attempting to address.

To reduce variance

→ Regularization

→ Dropout

→ Ensembling

Ques 2

	Class 1	Class 2
Class 1	50	30
Class 2	40	60

TP → 50, False Positive (FP) = 30,

FN => 40, True Negative = 60

1) Precision = $\frac{TP}{TP+FP}$ => $\frac{50}{50+30} = \frac{5}{8} = 0.625$

2) Recall = $\frac{TP}{TP+FN} = \frac{50}{50+40} = \frac{5}{9} = 0.555$

3) F_1 score = $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{5/8 \cdot 5/9}{5/8 + 5/9}$

$= 2 \times \frac{25}{85} = 0.558$

Ques 3 $\text{Entropy} = \frac{-p}{p+n} \log_2 \left(\frac{p}{p+n}\right) - \frac{n}{p+n} \log_2 \left(\frac{n}{p+n}\right)$

Average Information $I = \sum \frac{p_i + n_i}{p+n} \text{Entropy}(A)$

Information Gain = Entropy (S) - I

$p \Rightarrow 6 \quad n = 4$

$$\text{Entropy} = \frac{-6}{10} \log_2\left(\frac{6}{10}\right) - \frac{4}{10} \log_2\left(\frac{4}{10}\right) \Rightarrow 0.97095$$

Outlook.

	Yes	No	Total.
sunny	1	3	4
overcast	2	0	2
Rain	3	1	4
total	6	4	10

$$I(S, \text{outlook}) = \sum \frac{p_i + n_i}{p+n} \text{Entropy}(\text{outlook})$$

$$= \frac{4}{10}\left(-\frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{3}{4}\log_2\left(\frac{3}{4}\right)\right) + \frac{2}{10}\left(-\frac{2}{2}\log_2(1)\right) + \frac{4}{10}\left(-\frac{3}{4}\log_2\left(\frac{3}{4}\right) + \frac{1}{4}\log_2\left(\frac{1}{4}\right)\right)$$

$$\Rightarrow 0.64902\ldots$$

$$I(S, \text{temp}) = \left(-\frac{3}{10}\left(-\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \left(\left(\frac{2}{3}\log_2\left(\frac{2}{3}\right)\right)\right)\right) + \frac{4}{10}\left(-\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right)\right) + \frac{3}{10}\left(-\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right)$$

$$\Rightarrow 1$$

$$I(S, \text{Humidity}) \Rightarrow \frac{5}{10}\left(-\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \left(\frac{3}{5}\log_2\left(\frac{3}{5}\right)\right) + \frac{5}{10}\left(-\frac{4}{5}\log_2\left(\frac{4}{5}\right) - \frac{1}{5}\log_2\left(\frac{1}{5}\right)\right) \Rightarrow 0.8464$$

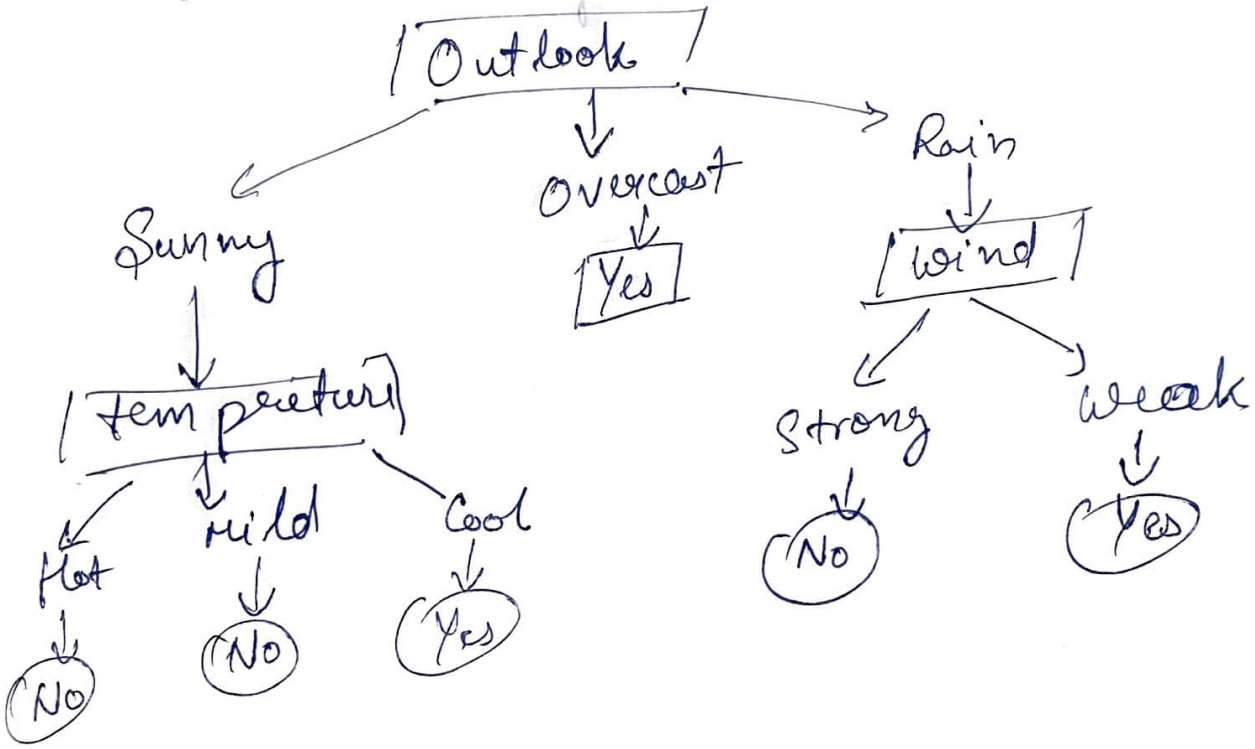
$$I(S, windy) = \left(\frac{7}{10}\left(-\frac{5}{7}\log_2\left(\frac{5}{7}\right) - \frac{2}{7}\log_2\left(\frac{2}{7}\right)\right)\right.$$
$$\left. + \frac{3}{10}\left(-\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log\left(\frac{2}{3}\right)\right)\right)$$

$$\Rightarrow 0.87867.$$

Information gain.

Gain (Outlook) = 0.321
Gain (Temperature) = 0.029
Gain (Humidity) = 0.124511
Gain (windy) = 0.091277.

Final Decision Tree.



Ques 4

For classifier 1.

$$P(Class1|x) = (40/(40+30)) \times (20/(20+20)) \times (50/(50+40)) = 0.137$$

Ans is 0.137 for a Part.

(b) Classifier -2.

$P(\text{class } 1/x) = (20/(20+20) \times 50/(50+40)) \times (0/(0+0))$

$= 0$ for class 1

(c) Classifier - 3

$P(\text{class } 1/x) = (40/(40+10)) \times (30/(30+20)) \times 0/(0+10)) = 0$

$P(\text{class } 2/x) = 1 - P(\text{class } 1/x)$

$= 1$ for class 2 Ans.