Q1) The bias-variance tradeoff describes the issue in many machine learning problems between minimizing bias and variance, which inherently affect each other. The bias problem is when in a classification machine learning problem the hypothesis space made available does not include sufficient hypotheses. That is, a high bias would give a large error when a particular classification model is underfitting. On the other hand, the variance problem is when in a classification machine learning problem, the hypothesis space made available is too large for the training data and the selected hypothesis may not be accurate on new data. That is, a high variance may tend to overfit data. As evident by definition, the bias-variance tradeoff is the conflict between complex and simple classification algorithms, and it is impossible to have no bias and no variance, but instead an algorithm should have be in the middle ground. One example of how to lower bias and variance is by including a penalty term to loss function, which will discourage the model from assigning too much importance to any one specific feature. To reduce bias specifically, we can increase the model complexity to reduce underfitting. To reduce variance, we can use techniques like resampling such as in a random forest.

Here, the bias can stay the same, as we average many trees. Other ensemble methods like bagging, boosting, and stacking can help in reducing the variance and bias.

2) Precision: $TP/(TP+FP)$

Recall: $r = TP/(TP+FN)$

F_1: $2rP/(P+r)$

Class 1: TP = 50
FP = 40
FN = 30

Class 2: TP = 60
FP = 30
FN = 40

Precision: Class 1 = $50/90 = 0.5556$

Class 2 = $60/90 = 0.6667$

Recall: Class 1: $50/80 = 0.625$

Class 2: $60/100 = 0.6$

$F_1 =$ Class 1: $\dfrac{2 \cdot \frac{5}{9} \cdot \frac{5}{8}}{\frac{5}{9} + \frac{5}{8}} = 0.5884$

$F_2 =$ Class 2: $\dfrac{2 \cdot \frac{6}{9} \cdot \frac{6}{10}}{\frac{6}{9} + \frac{6}{10}} = 0.6316$

3) Entropy(6,4)

* using base 10

$$= \frac{4}{10}\log\frac{4}{10} - \frac{6}{10}\log\frac{6}{10} = 0.2922852532$$

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes

Outlook

Sunny: 3 No, 1 Yes $= \frac{-3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4} = 0.244219503$

Overcast: 0 No, 2 Yes = 0

Rain: 1 No, 3 Yes $= \frac{-1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4} = 0.244219503$

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes

Humidity

High: 3 No, 2 Yes $= \frac{-3}{5}\log\frac{3}{5} - \frac{2}{5}\log\frac{2}{5} = 0.29228525 32$

Normal: 1 No, 4 Yes $= \frac{-1}{5}\log\frac{1}{5} - \frac{4}{5}\log\frac{4}{5} = 0.217322 0113$

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes

Temperature

Hot: 2 No 1 Yes $= \frac{-2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}$

Mild: 1 No 2 Yes $= \frac{-1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3}$

Cool: 1 No 3 Yes $= \frac{-1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4}$

Wind

Weak: 2 No 5 Yes $= \frac{-2}{7}\log\frac{2}{7} - \frac{5}{7}\log\frac{5}{7}$

Strong: 2 No 1 Yes $= \frac{-2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}$

W = 0.2598 25181

S = 0.2764345909

Hot: 0.2764345 901

Mild: 0.2764345 909

Cool: 0.244 219 0503

Outlook: $0.2922852532 - \left(\frac{4}{10} \times .2442190503 + \frac{4}{10} \times .2442190503\right)$

Temp: $0.2922852532 - \left(\frac{3}{10} \times 0.2764345909 + \frac{3}{10} \times 0.2764345909 + \frac{4}{10} \times 0.2442190503\right)$

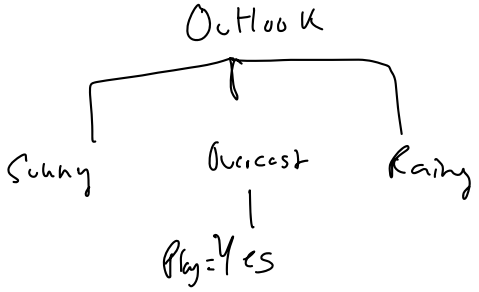
Humidity: $0.2922852532 - \left(\frac{5}{10} \times 0.2922852532 + \frac{5}{10} \times 0.2173220113\right)$

Wind: $0.2922852532 - \left(\frac{7}{10} \times 0.2598 25181 + \frac{3}{10} \times 0.2764345909\right)$

O = 0.0969 H = 0.0375

T = 0.0287 W = 0.0275

Outlook has highest gain.



Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes

Sunny

Temp

Hot : 2 No 0 yes = 0

Mild : 1 No 0 yes = 0

Cool : 1 yes 0 No = 0

$= 0.244\,219\,0503$

Humidity

High = 3 No 0 yes = 0 $= 0.244\,219\,0503$

Normal = 1 yes 0 No = 0

Wind

Weak : 2 No 1 yes. $= -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} = 0.27643\,45909$

Strong : 1 No 0 yes = 0

$= 0.244\,219\,0503 - \left(\frac{3}{4} \times 0.27643\,45909\right) = 0.036\,893\,10712$

Sunny, Temp

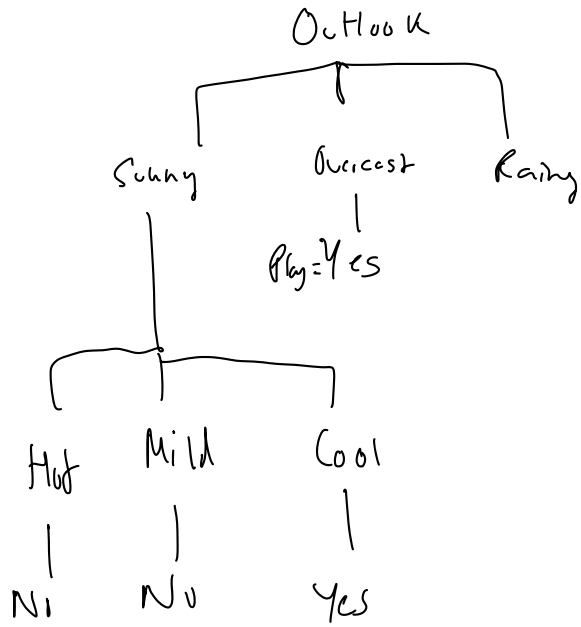
Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes

Temp

Hot = No

Mild = No

Cool = yes



Rainy

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes

Temp

Mild : 2 yes 0 No = 0

Cool : 1 yes 1 No = $-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}$

$= 0.301\,029\,9957$

Humidity

High : 1 yes, 0 no = 0

Normal : 2 yes 1 No

High = 0

$= -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}$

$= 0.27643\,45909$

Wind

Weak : 3 yes, 0 No = 0

Strong : 1 No, 0 yes = 0

Temp : $\frac{2}{4} \times 0.301\,029\,9957 = 0.150\,514\,9978$

Humidity = $\frac{3}{4} \times 0.27643\,45909 = 0.207\,325\,9432$

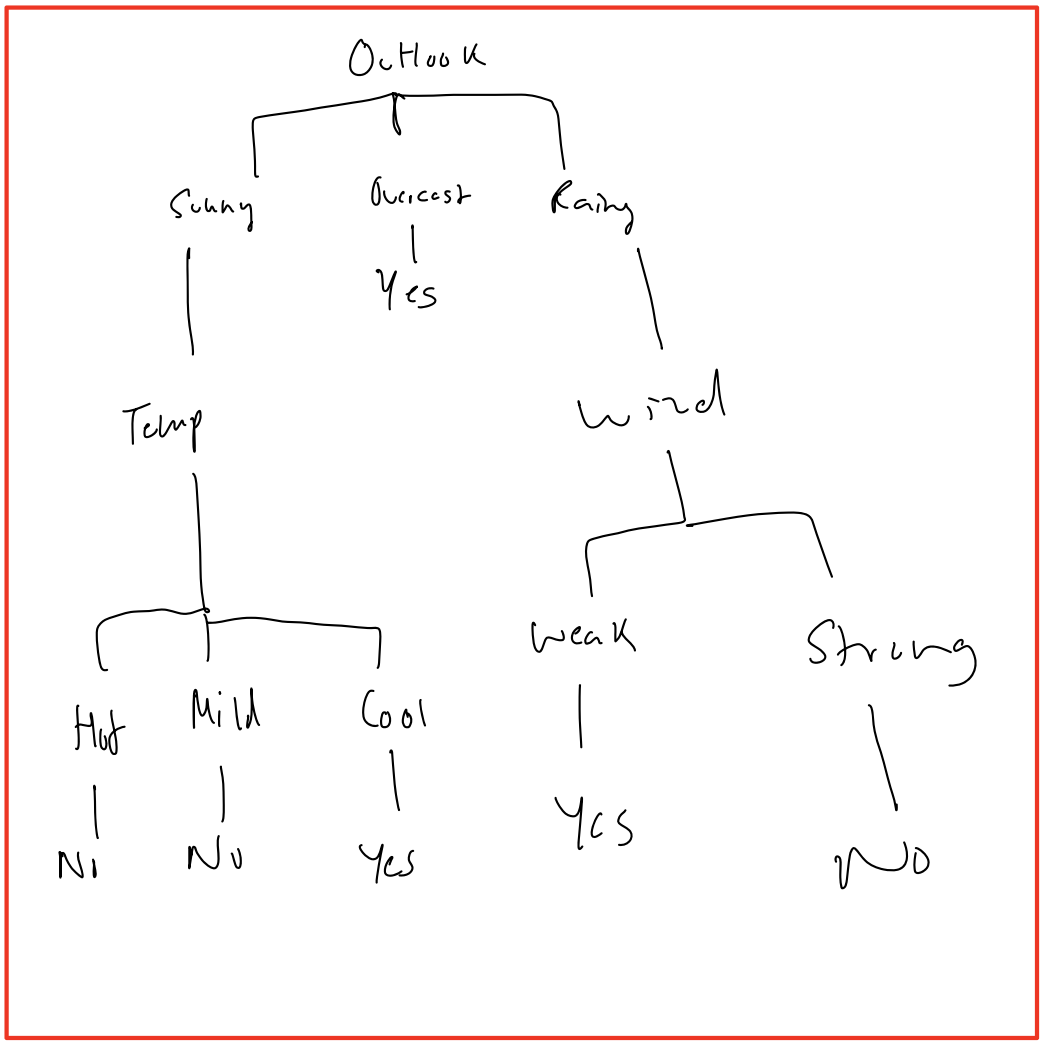
$.2442190503 - 0.1505149978 = 0.0937040525$ ⋮ $.2442190503 - 0.2073259432 = 0.0368931071$

Wind = 0.2442190503

↓

Weak = yes

strong = No



4) Classifier 1

$P(\text{Class 1} \mid \text{Pred}) = \frac{TP}{\text{Class 1 Pred}} = \frac{4}{7}$

$P(\text{Class 2} \mid \text{Pred}) = \frac{TN}{\text{Class 2 Pred}} = \frac{2}{3}$

Classifier 2

$P(\text{Class 1} \mid \text{Pred}) = \frac{1}{2}$

$P(\text{Class 2} \mid \text{Pred}) = \frac{1}{2}$

Classifier 3

$P(\text{Class 1} \mid \text{Pred}) = \frac{5}{9}$

$P(\text{Class 2} \mid \text{Pred}) = 1$

$$P(C1) = \frac{4}{7} \times \frac{1}{2} \times \frac{5}{9}$$

$$= \frac{10}{63}$$

$$P(C2) = \frac{2}{3} \times \frac{1}{2} \times 1 = \frac{2}{6}$$

$$= \frac{1}{3}$$

The final decision would be **class 2**.