Question 1

1. **Explain what is the bias-variance trade-off? Describe few techniques to reduce bias and variance respectively.**

   If the training model is simple and has less parameters then chances are that it is high bias and low variance. On the other hand, if the training model has large number of parameters then it's going to have high variance and low bias. So, the goal is to find the right/good balance without overfitting and underfitting the data. Hence, we perform this tradeoff where there is a tradeoff between bias and variance. An algorithm can't be more complex and less complex at the same time.

   Techniques to reduce variance:

   - Increasing Training set
   - Decreasing Features
   - Increasing Lambda

   Techniques to reduce bias:

   - Increasing Features
   - Perform feature engineering
   - Decreasing the alpha parameter of Regularization

2. **What is k-fold cross-validation? Why do we need it?**

   a k-fold cross-validation approach is in which cross validation is performed k different times, each time using a different partitioning of the data into training and validation sets, and the results are then averaged. Cross-validation approach is used to determine the number of iterations 'i' that yield the best performance on the validation set. The mean 'i' of these estimates for 'i' is then calculated, and a final run of backpropagation is performed training on all n examples for 'i' iterations, with no validation set.

   Reasons for using K-fold Cross-validation –

   - K-fold Cross-validation methods can be used to estimate an appropriate stopping point for gradient descent search and thus to minimize the risk of overfitting.
   - Using K-folds we are able to divide the data in equal parts so we end up using all of the data to train the algorithm
   - We get a better metric of algorithm performance because we are able to run to run the algorithm multiple times and compare the results

Question 2

1) Precision $= \dfrac{TP}{TP+FP} = \dfrac{50}{50+40} = \underline{0.555}$

2) Recall $= \dfrac{TP}{TP+FN} = \dfrac{50}{50+30} = \underline{0.625}$

3) F1 score $= \dfrac{2 \times \text{Precision} \times \text{recall}}{\text{Precision} + \text{Recall}}$

$= \dfrac{2 \times 0.555 \times 0.625}{0.555 + 0.625} = 0.\underline{587}$

QUESTION 3 (1)

Gain(Outlook) = E(playTennis) − E(Outlook, PlayTennis)

= 0.97 − 0.648 => 0.322

Gain(temp) = E(PlayTennis) − E(temp, Playtennis)

= 0.97 − 0.874 = 0.096

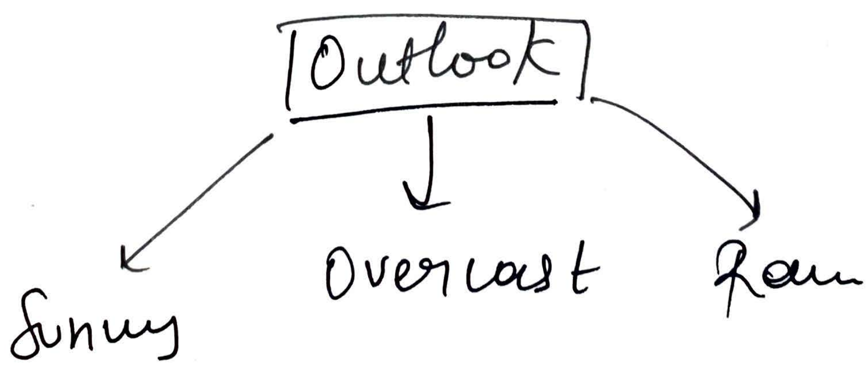
Gain(Humidity) = E(PlayTennis) − E(Humidity, Playtennis)

= 0.97 − 0.846 = 0.124

Gain(wind) = E(Playtennis) − E(wind, Playtennis)

= 0.97 − 0.795 = 0.195

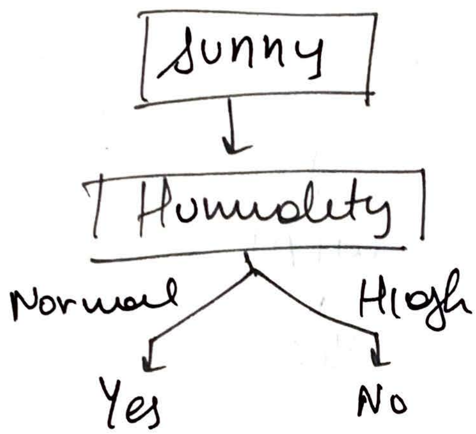
Gain for Outlook is greatest so it is chosen as root node



Similarly for Sunny,

Gain (humidity) = E(Playtennis) - E(humidity, Playtennis)
= 0

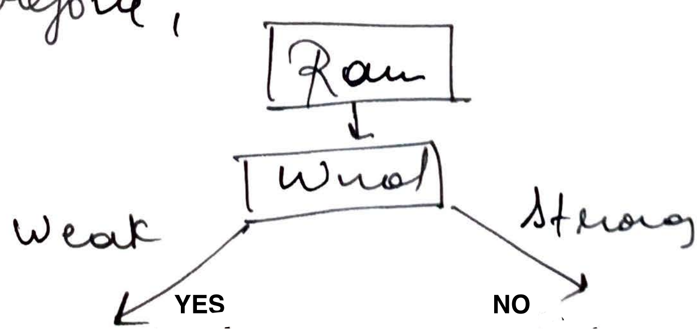
Since gain for humidity is 0, we will get that as leaf



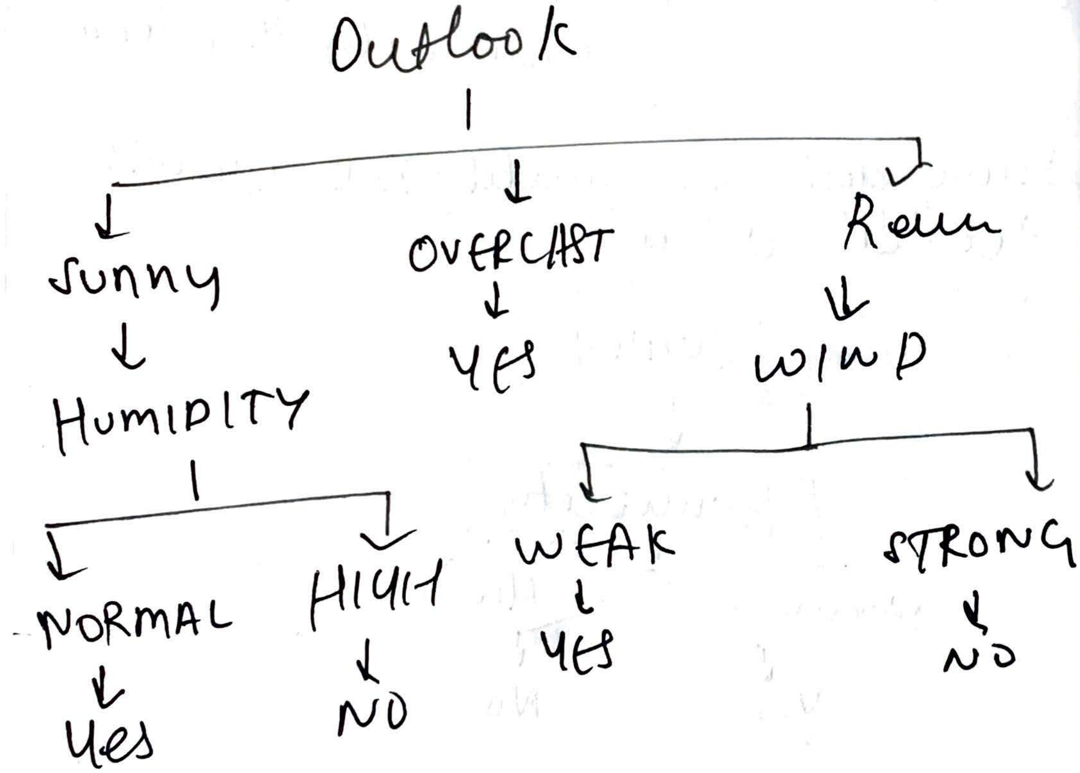
for Rainy

Gain(wind) = E (playtennis) - E(wind, Play tennis)
= 0

Therefore,



Final Output:-



QUESTION 3 (2)

sunny

$P_1 = 1 \quad n_1 = 3$

Overcast

$P_2 = 2 \quad n_2 = 0$

Rain

$P_3 = 3 \quad n_3 = 1$

$$\hat{p_i} = \frac{p}{p+n} |s_i| \qquad \hat{n_i} = \frac{n}{p+n} |s_i|$$

$$Q_1 = \frac{(P_1 - \hat{P}_1)^2}{\hat{p_i}} + \frac{(n_1 - \hat{n}_1)^2}{\hat{n}_1} = \underline{2.041}$$

$Q_2 = 1.333$

$Q_3 = 0.375$

$Q = Q_1 + Q_2 + Q_3$
$= 3.\underline{749}$

degree of freedom =
no of subsets - 1
$= 3 - 1$
$= 24$

Chi-Square Calculator

The chi-square distribution calculator makes it easy to compute cumulative probabilities, based on the chi-square statistic.

If anything is unclear, read the Frequently-Asked Questions or the Sample Problems. To learn more about the chi-square, read Stat Trek's tutorial on the chi-square distribution.

- Enter a value for degrees of freedom.
- Enter a value for one, and only one, of the remaining unshaded text boxes.
- Click the **Calculate** button to compute values for the other text boxes.

Degrees of freedom	2
Chi-square critical value (CV)	3.749
$P(X^2 < 3.749)$	0.85
$P(X^2 > 3.749)$	0.15

Calculate

QUESTION 4

Predicted events are :-

Classifier 1 : class 1 Classifier 2 : class 1

Classifier 3 : class 2

$\hat{P}(\omega_1 | d_{1,1}(x)=1) = \frac{40}{70}$ $\hat{P}(\omega_2 | d_{1,1}(x)=1) = \frac{30}{70}$

$\hat{P}(\omega_1 | d_{2,1}(x)=1) = \frac{20}{40}$ $\hat{P}(\omega_2 | d_{2,1}(x)=1) = \frac{20}{40}$

$\hat{P}(\omega_1 | d_{3,2}(x)=1) = \frac{0}{10}$ $\hat{P}(\omega_2 | d_{3,2}(x)=1) = \frac{10}{10}$

$\Rightarrow$ Class 1 : 0

$\Rightarrow$ Class 2 : 0.214

Hence we go with class 2