Introduction to ML (CS771), Autumn 2018
Indian Institute of Technology Kanpur
Homework Assignment Number 1

Student Name: Lt Commander Rahul Raj
Roll Number: 18111053
Date: August 19, 2018

QUESTION 1

Sub Question (1) Misclassification Rate

Classification Error is one of the methods to measure the purity or homogenity of set of datapoints at a Node in Decision Tree. It is calculated as below,

$$Err(Node) = 1 - \max_{c \,\epsilon\, C}[P_c]$$

where P_c is the Probability of the Class 'c' in Node

Calculating misclassification error of Parent Node of given Tree A: -

$$\begin{aligned} Err(A_{Parent}) &= 1 - \max\ [P(A_{Parent})] \\ &= 1 - \max(P(A_{Left}), P(A_{Right})) \\ &= 1 - \max(P(400), P(400)) \\ &= 1 - \frac{1}{2} \\ &= 0.5 \end{aligned}$$

Similarly for Left Node and Right Node of Tree A

$$\begin{aligned} Err(A_{Left}) &= 1 - \max\ [P(A_{Left})] \\ &= 1 - \max\ (P(0), P(1)) \\ &= 1 - \frac{3}{4} \\ &= 0.25 \end{aligned} \qquad \begin{aligned} Err(A_{Right}) &= 1 - \max\ [P(A_{Right})] \\ &= 1 - \max\ (P(0), P(1)) \\ &= 1 - \frac{3}{4} \\ &= 0.25 \end{aligned}$$

The mislassification rate of **Tree A** is calculated as

$$\textbf{Misclassification Rate} = (P(A_{Left}) \times Err(A_{Left}) + P(A_{Right}) \times Err(A_{Right}))$$

$$\begin{aligned} &= (\frac{400}{800} \times 0.25) + (\frac{400}{800} \times 0.25) \\ &= \mathbf{0.25}\ \mathit{or}\ \mathbf{25\%} \end{aligned}$$

Similarly, the mislassification rate of **Tree B** is calculated as

$$\textbf{Misclassification Rate} = (P(B_{Left}) \times Err(B_{Left}) + P(B_{Right}) \times Err(B_{Right}))$$

$$\begin{aligned} &= (\frac{600}{800} \times \frac{1}{3}) + (\frac{200}{800} \times 0) \\ &= \mathbf{0.25}\ \mathit{or}\ \mathbf{25\%} \end{aligned}$$

Observations

From the above calculations, it is observed that the misclassificaiton rates of both Tree A and Tree B are same i.e 25%.

Conclusion

Growing the decision tree further will not yeild any good in improving/ minimizing the error rate.

Sub Question (2) Information Gain

Entropy is another way to measure the purity or homogenity of set of datapoints. High Entropy means less purity. Entropy is calculated as follows,

$$H(S) \ = \ -\sum_{c \ \epsilon \ C} P_c \times Log_2 P_c$$

where P_c is the probability of Class C

Information Gain of a node is the difference between Entropy before and after splitting datapoints of that node. In order to calculate the IG, the Entropy of Tree A need to be calculated as follows,

$$\begin{aligned} H(A_{Parent}) &= \ -\sum_{c \ \epsilon \ C} P_c \times Log_2 P_c \\ &= -\ (P(A_{Left}) \times \ Log_2 \ P(A_{Left}) \ + \ P(A_{Right}) \times \ Log_2 \ P(A_{Right})) \\ &= -\ (\frac{1}{2} \times Log \ \frac{1}{2} + \frac{1}{2} \times Log \ \frac{1}{2}) \\ &= 1 \end{aligned}$$

Similarly for Left Node and Right Node of Tree A

$$\begin{aligned} H(A_{Left}) &= \ -\sum_{c \ \epsilon \ C} P(A_{Left}) \times Log_2 P(A_{Left}) \\ &= -\ (\frac{3}{4} \times Log \ \frac{3}{4} + \frac{1}{4} \times Log \ \frac{1}{4}) \\ &= 0.81125 \end{aligned}$$

$$\begin{aligned} H(A_{Right}) &= \ -\sum_{c \ \epsilon \ C} P(A_{Right}) \times Log_2 P(A_{Right}) \\ &= -\ (\frac{1}{4} \times Log \ \frac{1}{4} + \frac{3}{4} \times Log \ \frac{3}{4}) \\ &= 0.81125 \end{aligned}$$

The Information Gain (IG$_A$) of **Tree A** calculated as

$$\textbf{Information Gain (IG}_A\textbf{)} = H(s) \ - \sum_{c \ \epsilon \ C} \frac{|H(s_c)|}{|H(s)|}) \times H(S_c)$$

$$\begin{aligned} &= 1 - ((\frac{400}{800} \times 0.81125) + (\frac{400}{800} \times 0.81125)) \\ &= \mathbf{0.18875} \end{aligned}$$

Similarly, Information Gain (IG$_B$) of **Tree B** is calculated as

$$\textbf{Information Gain (IG}_B\textbf{)} = H(s) \ - \sum_{c \ \epsilon \ C} \frac{|H(s_c)|}{|H(s)|}) \times H(S_c)$$

$$\begin{aligned} &= 0.81125 - ((\frac{600}{800} \times 0.917628) + (\frac{200}{800} \times 0)) \\ &= \mathbf{0.1230} \end{aligned}$$

Observations

From the above calculations, it is seen that the IG of Tree A is higher than the IG of Tree B. Information Gain is an indicator on change in purity of dataset while undertaking split at that point. More the IG, more the pure the dataset becomes.

Conclusion

Since $\text{IG}_A > \text{IG}_B$, we are preferring Tree A over Tree B for further growing the Decision Tree.

Sub Question (3) Different Answers for Sub Questions (1) and (2)

Observations

From the above, it is seen that the misclassification rate is same for Tree A and Tree B. This implies there is no advantage in further growing Decision Tree as there is no improvement in splitting either Tree A or Tree B.

However, while calculating the Entropy, we observed that the Entropy after split in both the cases have come down in both Tree A and Tree B. and the larger reduction happened in case of Tree A. This means that the Decision Tree can be further grown with Tree A.

Conclusion

If we are considering Misclassification Error at various nodes, and at some stage the same becomes equal, further analysis should be done with the help of Entropy method to find the further scope of growing the Decision Tree.

Introduction to ML (CS771), Autumn 2018
Indian Institute of Technology Kanpur
Homework Assignment Number 1

Student Name: Lt Commander Rahul Raj
Roll Number: 18111053
Date: August 19, 2018

QUESTION 2

Introduction to ML (CS771), Autumn 2018
Indian Institute of Technology Kanpur
Homework Assignment Number 1

Student Name: Lt Commander Rahul Raj
Roll Number: 18111053
Date: August 19, 2018

QUESTION 3

My solution to problem 3

Introduction to ML (CS771), Autumn 2018
Indian Institute of Technology Kanpur
Homework Assignment Number 1

Student Name: Lt Commander Rahul Raj
Roll Number: 18111053
Date: August 19, 2018

QUESTION 4

My solution to problem 4

Introduction to ML (CS771), Autumn 2018
Indian Institute of Technology Kanpur
Homework Assignment Number 1

Student Name: Lt Commander Rahul Raj
Roll Number: 18111053
Date: August 19, 2018

QUESTION 5

My solution to problem 5

Introduction to ML (CS771), Autumn 2018
Indian Institute of Technology Kanpur
Homework Assignment Number 1

Student Name: Lt Commander Rahul Raj
Roll Number: 18111053
Date: August 19, 2018

QUESTION 6

My solution to problem 6