Casey Pei
CSCE 421-500
Texas A&M University

March 9, 2024

HOMEWORK 2 — Tree-based Models

1 Math Questions

1.1 Information Gain (20 points)

NOTE: This is not a programming assignment, so you may NOT use programming tools to help solve this problem. Show your work.

.

Suppose you are given 6 training points as seen below, for a classification problem with two binary attributes X_1 and X_2 and three classes $Y \in 1, 2, 3$. You will use a decision tree learner based on information gain

X_1	X_2	Y
1	1	1
1	1	1
1	1	2
1	0	3
0	0	2
0	0	3

1. Calculate the conditional entropy for both X_1 and X_2.

   First, I will calculate the conditional probabilities for X_1 and X_2 on Y.

$P(X_1\|Y)$	$P(X_2\|Y)$
$P(X_1 = 1 \mid Y = 1) = \frac{2}{3}$	$P(X_2 = 1 \mid Y = 1) = \frac{3}{3}$
$P(X_1 = 0 \mid Y = 1) = \frac{1}{3}$	$P(X_2 = 0 \mid Y = 1) = \frac{0}{3}$
$P(X_1 = 1 \mid Y = 2) = \frac{1}{1}$	$P(X_2 = 1 \mid Y = 2) = \frac{0}{1}$
$P(X_1 = 0 \mid Y = 2) = \frac{0}{1}$	$P(X_2 = 0 \mid Y = 2) = \frac{1}{1}$
$P(X_1 = 1 \mid Y = 3) = \frac{0}{1}$	$P(X_2 = 1 \mid Y = 3) = \frac{0}{1}$
$P(X_1 = 0 \mid Y = 3) = \frac{1}{1}$	$P(X_2 = 0 \mid Y = 3) = \frac{1}{1}$

   Secondly, we compute the entropy $H(Y)$

$$\begin{aligned}
H(Y) &= \sum_{y_i=1}^{n=3} P(Y = y_i) log_2 P(Y = y_i) \\
&= \sum_{y_i=1}^{n=3} \frac{1}{3} log_2(\frac{1}{3}) \\
&= log_2(3) \\
&= 1.58496
\end{aligned}$$

Next we compute the conditional entropy for the X_1 split:

$$\begin{aligned}
H(X_1|Y) &= -\sum P(X_1|Y) \times log_2(P(X_1|Y)) \\
&= -\left[\frac{2}{3}log_2\left(\frac{2}{3}\right) + \frac{1}{3}log_2\left(\frac{1}{3}\right) + \frac{1}{1}log_2\left(\frac{1}{1}\right) + \frac{0}{1}log_2\left(\frac{0}{1}\right) + \frac{0}{1}log_2\left(\frac{0}{1}\right) + \frac{1}{1}log_2\left(\frac{1}{1}\right)\right] \\
&= -\left(-\frac{2}{6} - 1\right) \\
&= \frac{4}{3}
\end{aligned}$$

Finally we compute the conditional entropy for the X_2 split:

$$\begin{aligned}
H(X_2|Y) &= -\sum P(X_2|Y) \times log_2(P(X_2|Y)) \\
&= -\left[\frac{3}{3}log_2\left(\frac{3}{3}\right) + \frac{0}{3}log_2\left(\frac{0}{3}\right) + \frac{0}{1}log_2\left(\frac{0}{1}\right) + \frac{1}{1}log_2\left(\frac{1}{1}\right) + \frac{0}{1}log_2\left(\frac{0}{1}\right) + \frac{1}{1}log_2\left(\frac{1}{1}\right)\right] \\
&= -\left(-\frac{2}{3} - \frac{19}{12}\right) \\
&= \frac{11}{12}
\end{aligned}$$

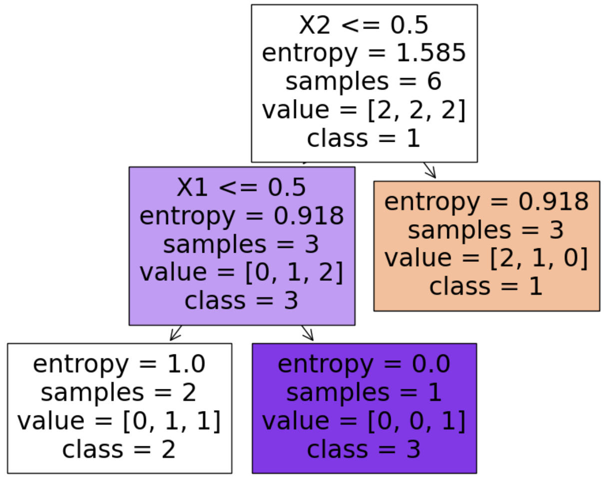
2. Calculate the information gain if we split based on 1) X_1 or 2) X_2

$$\begin{aligned}
IG(X_1) &= H(Y) - H(Y|X_1) \\
&= 1.58496 - \frac{4}{3} \\
&= 0.251629
\end{aligned}$$

$$\begin{aligned}
IG(X_2) &= H(Y) - H(Y|X_2) \\
&= 1.58496 - \frac{11}{12} \\
&= 0.6682958
\end{aligned}$$

3. Report which attribute is used for the first split. Draw the decision tree using this split.

   Since the information gain for splitting based on X_2 (0.668) is greater than the information gain for splitting based on X_1 (0.251), the first split in the decision tree will be based on X_2.



4. Conduct classification for the test example $X_1 = 0$ and $X_2 = 1$.

   Following the above decision tree, we predict that $Y = 1$ given $X_1 = 0$ and $X_2 = 1$.

2 Programming Questions

Answers for these are located in the attached Casey_Pei_HW2.ipynb file.

Submitted by Casey Pei on March 9, 2024.