Machine Learning HW2

Ritika Nigam, UIN: 633002197

March 8, 2024

1 Problem 1: Information Gain

1.1 Conditional Entropy of X1 and X2

$H(Y|X1) = -P(X1 = 0)\sum_{yi=1}^{n=3} P(Y = yi|X1 = 0)\log_2 P(Y = yi|X1 = 0)$
$-P(X1 = 1)\sum_{yi=1}^{n=3} P(Y = yi|X1 = 1)\log_2 P(Y = yi|X1 = 1)$

$= -[\frac{2}{6}(\frac{0}{2}\log_2\frac{0}{2} + \frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}) + \frac{4}{6}(\frac{2}{4}\log_2\frac{2}{4} + \frac{1}{4}\log_2\frac{1}{4} + \frac{1}{4}\log_2\frac{1}{4})]$

$= -(-\frac{2}{6} - 1)$

$= \frac{4}{3}$

$H(Y|X2) = -P(X2 = 0)\sum_{yi=1}^{n=3} P(Y = yi|X2 = 0)\log_2 P(Y = yi|X2 = 0)$
$-P(X2 = 1)\sum_{yi=1}^{n=3} P(Y = yi|X2 = 1)\log_2 P(Y = yi|X2 = 1)$

$= -[\frac{3}{6}(\frac{0}{3}\log_2\frac{0}{3} + \frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}) + \frac{3}{6}(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3} + \frac{0}{3}\log_2\frac{0}{3})]$

$= -(-\frac{2}{3} - \frac{19}{12})$

$= \frac{11}{12}$

$H(Y|X1) = \frac{4}{3}$ and $H(Y|X2) = \frac{11}{12}$

1.2 Information gain if we split based on 1) X1 or 2) X2.

$H(Y) = -\sum_{yi=1}^{n=3} P(Y = yi)\log_2 P(Y = yi)$

$= -[\frac{2}{6}\log_2\frac{2}{6} + \frac{2}{6}\log_2\frac{2}{6} + \frac{2}{6}\log_2\frac{2}{6}]$

$= -[\log_2\frac{1}{3}]$

$= \log_2 3 \approx \frac{19}{12}$

$IG(X1) = H(Y)$ - $H(Y|X1) = \frac{19}{12} - \frac{4}{3} \approx \frac{1}{4}$

$IG(X2) = H(Y)$ - $H(Y|X2) = \frac{19}{12} - \frac{11}{12} \approx \frac{2}{3}$

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

Since the information gain of X2 is greater than X1's information gain, Attribute X2 is used for the first split.

[redacted]

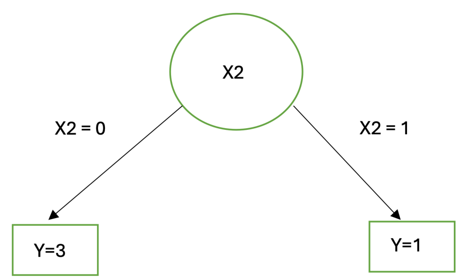

Figure 1: Decision Tree

1.4 Conduct classification for the test example X1 = 0 and X2 = 1.

If we consider the test sample with X1 = 0 and X2 = 1 then this decision tree will classify Y=1 because X2=1 in this case.