Machine Learning Exercises (chapter 18)

Mohammad Javad Abbas pour

18.2.1

We can run the ID3 Algorithm and show that the error of the decision tree algorithm is at least $\frac{1}{4}$. At first we compute information gains:

$$H(Y) = \left(-\frac{1}{2}\right)\log\left(\frac{1}{2}\right) - \left(\frac{1}{2}\right)\log\left(\frac{1}{2}\right) = 1$$

$$IG(X_1) = H(Y) - H(Y|X_1) = 1 - \frac{3}{4}\left(-\frac{2}{3}\log\left(\frac{2}{3}\right) - \frac{1}{3}\log\left(\frac{1}{3}\right)\right) \Rightarrow IG(X_1) > 0$$

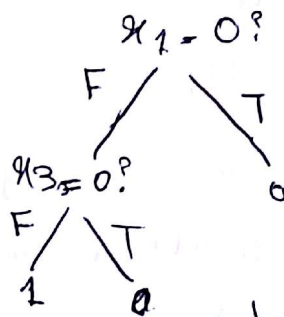$$IG(X_2) = H(Y) - H(Y|X_2) = 1 - \left(\frac{1}{2}(-1) + \frac{1}{2}(-1)\right) = 0$$

$$IG(X_3) = H(Y) - H(Y|X_3) = 0 \longrightarrow \text{So we choose } X_1 = 0.$$

$\xrightarrow[\text{Left}]{\text{node}}$ $H(Y) = -\frac{2}{3}\log\left(\frac{2}{3}\right) - \frac{1}{3}\log\left(\frac{1}{3}\right)$

$$IG(X_2) = H(Y) - \frac{2}{3}, \quad IG(X_3) = H(Y) - \frac{2}{3} \longrightarrow \text{So we can choose left node or right.}$$
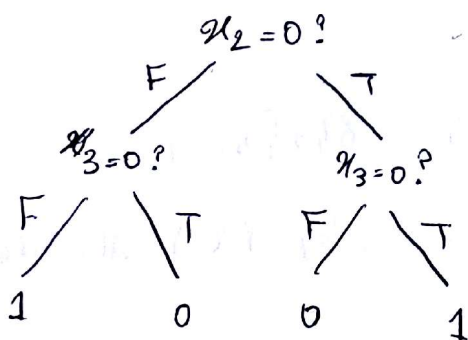
$\xrightarrow[\text{node}]{\text{right}}$

> In this case Error point is
> $((1,0,0),1)$, so error is $\frac{1}{4}$.

$\downarrow$
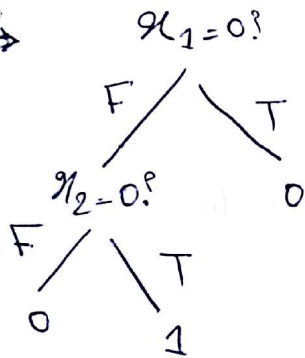
In this case Error point is

$((1,1,1),1)$, so error is $\frac{1}{4}$.

hence we observe that ony tree with 2 depth error is $\frac{1}{4}$.

18.2.2

X_2, X_3 describe the lables completly so we can ignor X_1 to make decision tree.