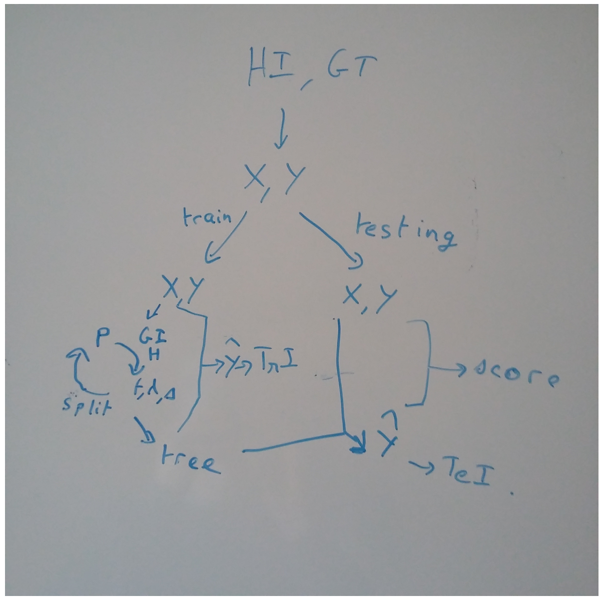

Figure 1: Training a Tree

Content 6
Extending splitting to the multiclass context
Training decision trees

1 Keywords

- Entropy of dataset

$$H = \sum_c P(Y=c)\log_2\left(\frac{1}{P(Y=C)}\right) = -\sum_c \frac{N_c}{N}\log_2\left(\frac{N_c}{N}\right) \text{ where } N_c = \sum_n \mathbf{1}(y_n = c) \tag{1}$$

- Gini index of the dataset

$$\text{GI} = \sum_{c\neq c'} P(Y=c)P(Y=c') = 1 - \sum_c (P(Y=c))^2 = 1 - \sum_c \left(\frac{N_c}{N}\right)^2 \tag{2}$$

- Information Gain of a split S_p at a node p (or mutual information)

$$\text{IG} = H(Y_p) - H(Y_p|S_p) \geq 0 \tag{3}$$

where S_p is the random variable associated to a split and Y_p is the random variable associated to the class distribution at node p.

- Random variable associated to a split

$$P(S=1) = \frac{1}{N}\sum_n \mathbf{1}(sx_{n,f} < s\lambda) \text{ and } P(S=0) = 1 - P(S=1) \tag{4}$$

- Conditional Entropy

$$H(Y_p|S_p) = P(S_p=1)H(Y_p|S_p=1) + P(S_p=0)H(Y_p|S_p=0) \tag{5}$$

$$H(Y_p|S_p) = P(S_p=1)\sum_c P(Y_p=c|S_p=1)\log_2\left(\frac{1}{P(Y_p=c|S_p=1)}\right) + P(S_p=0)\sum_c P(Y_p=c|S_p=0)\log_2\left(\frac{1}{P(Y_p=c|S_p=0)}\right) \tag{6}$$

- Confusion matrix of a split

$$C_{1,j} = \sum_n \mathbf{1}(s_n=1)\mathbf{1}(y_n=j) \text{ and } C_{2,j} = \sum_n \mathbf{1}(s_n=0)\mathbf{1}(y_n=j) \tag{7}$$

where $s_n = \mathbf{1}(sx_{n,f} < s\lambda)$

- Gini Gain of a split at node p

$$\text{GG}_p = \left(1 - \sum_c \left(P(Y = c)\right)^2\right) - P(S_p = 1)\left(1 - \sum_c \left(P(Y_p = c|S_p = 1)\right)^2\right) - P(S_p = 0)\left(1 - \sum_c \left(P(Y_p = c|S_p = 0)\right)^2\right) \tag{8}$$

GG_p can be negative.