Lecture October 30

CART = classification and Regression Tree.

(y_i, x_i) $i = 0, \ldots n-1$

$x_i = (x_{i0}, x_{i1}, \ldots x_{ip-1})$

Regression

- partition into M regions $R_1 R_2 \ldots R_M$
- Model response as a constant c_m in each region

$$f(x) = \sum_{m=1}^{M} c_m I(x \in R_m)$$

$$\hat{c}_m = \text{ave}(y_i \mid x_i \in R_m)$$

- Minimize

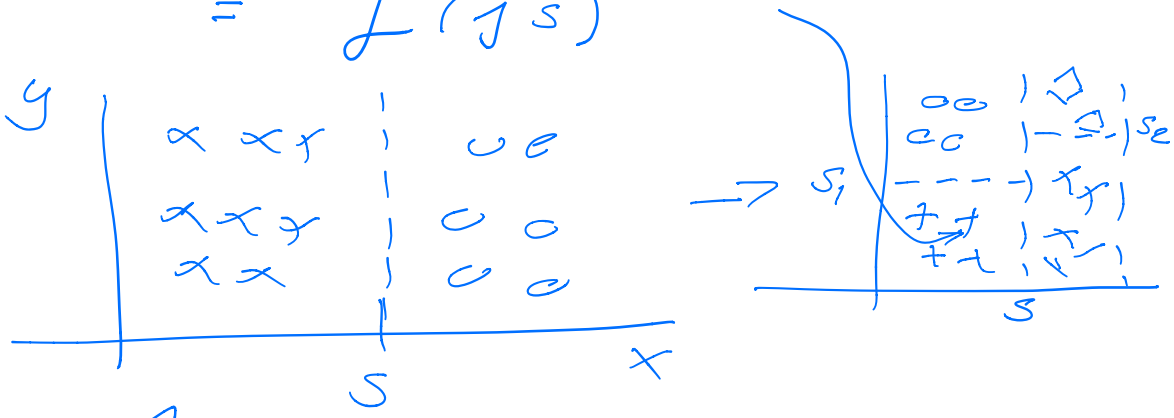
$$\sum (y_i - f(x_i))^2$$

- $R_1(j, s) = \{X \mid x_j \leq s\}$
- $R_2(j, s) = \{X \mid x_j > s\}$

- Minimize

$$\min_{j,s}\left[\min_{c_1}\sum_{x_i\in R_1(j,s)}(y_i-\hat{c}_1)^2 + \min_{c_2}\sum_{x_i\in R_2(j,s)}(y_i-\hat{c}_2)^2\right]$$

$$= J(j,s)$$



$$\hat{c}_1 = \text{ave}(y_i \mid x_i \in R_1(j,s))$$

$$\hat{c}_2 = \text{ave}(y_i \mid x_i \in R_2(j,s))$$

Compact CART

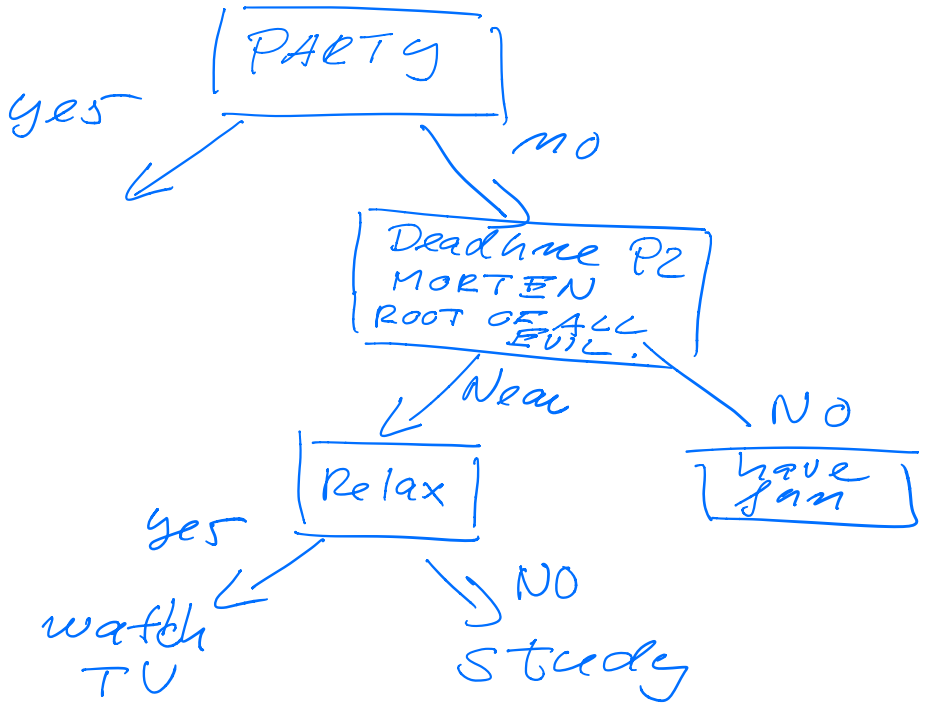
$$J(j,s) = \frac{N_{left}}{N} MSE_{left} + \frac{N_{right}}{N} MSE_{right}$$

Classification

$$J(j,s_j) = \frac{N_{left}}{N} G_{left} + \frac{N_{right}}{N} G_{right}$$

measures of impurity of the left/right subset.

G left/right ?



Define a node $-m-$
(Represents a Region R_m with N_m observations)

Proportion of class $-k-$ observations in node m

$$P_{mk} = \frac{1}{N_m} \sum_{x_i \in R_m} I(y_i = k)$$

Classify the observation

in node m of a tree

$$k(m) = \arg\max_k p_{mk}$$

Misclassification function

(i) $$\frac{1}{N_m} \sum_{i \in R_m} I(y_i \neq k(m)) = 1 - p_{mk}$$

(ii) Gini index (Default in Scikit-learn)

$$\sum_{k=1}^{K=\text{classes}} p_{mk}(1 - p_{mk})$$

(iii) Entropy

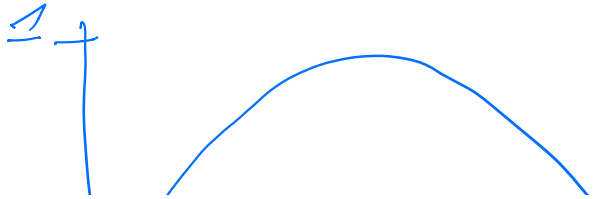
$$-\sum_{k=1} p_{mk} \log p_{mk}$$

what does the entropy represent?

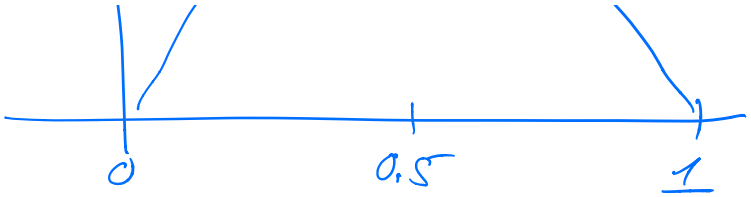
Binary case

$$\text{Entropy} = -\sum_i p_i \log_2 p_i$$

Define $0 \log_2 0 = 0$



positive
negative



if all samples are positive we don't get any extra information from knowing the value of the feature of any particular example, Entropy = 0

Gini impurity

one class in one node; purity, impurity = deviation from purity, N_i = # data points of class -i- (fraction)

- Also loops over the different features/classes.. and checks how many points belong to a specific class.
- For any feature k

$$G_k = \sum_{i=1}^{K=\#\text{classes}} \sum_{j \neq i} N_i N_j$$

$$\sum_i N_i = 1 \qquad \sum_{j \neq i} N_j = 1 - N_i$$

$$G_K = 1 - \sum_{i=1}^{K = \#\,classes} N_i^2$$