Linear Regression

Q1. Suppose you have given a set of training examples:

$$\{(x_1,y_1),(x_2,y_2)\ldots,(x_n,y_n)\}$$

Find the equation of the line that best-fits the data / minimizes the squared error.

A: Estimate the square error:

$$J=\sum_{i=1}^{n}\left[y_i-f(x_i)\right]^2 \quad ; \; f() \rightarrow \text{linear Regression hypothesis}$$

$$f(x)=mx+c$$

$$\Rightarrow \boxed{J=\sum_{i=1}^{n}\left[y_i-mx_i-c\right]^2} \rightarrow \text{Minimize w.r.t } m \text{ and } c$$

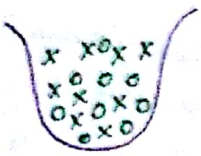
Tuning Parameters

$$\frac{\partial J}{\partial m}=0 \Rightarrow 2\sum_{i=1}^{n}\left[y_i-mx_i-c\right](-x_i)=0 \Rightarrow \boxed{\sum_{i=1}^{n}\left[y_i-mx_i-c\right]x_i=0} \quad (1)$$

$$\frac{\partial J}{\partial c}=0 \Rightarrow 2\sum_{i=1}^{n}\left[y_i-mx_i-c\right](-1)=0 \Rightarrow \boxed{\sum_{i=1}^{n}\left[y_i-mx_i-c\right]=0} \quad (2)$$

Put values of data points and calculate the values of m & c

Decision Tree :-



Data points $\in R^N$

N : Dimensionality of the feature space
or
N features of the input data

Steps : 1. Select a feature f_i

2. A point of split on that feature axis f_i so that, the subsets of data points going into the child nodes are the most homogeneous.

Method to measure Homogenity :

Entropy of a set :

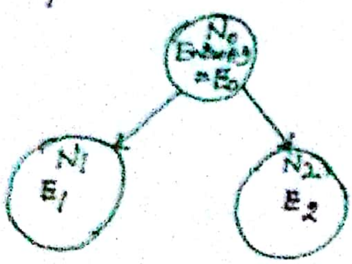
Say we have points from N different classes in a set. The probabilities/fractions of points of the different classes are $\{f_1, f_2, \cdots f_N\}$; such that $\sum_{i}^{N} f_i = 1$

Entropy of this set $\boxed{E=-\sum_{i=1}^{N} f_i \log_2 f_i}$.

Information Gain :-

Let N_0 be the initial data set and E_0 be the entropy. We choose a feature axis/vector and split the data set into two sets, $(N_1+N_2=N_0)$

$$\text{Information gain} = E_0-\left[\frac{N_1}{N_0}E_1+\frac{N_2}{N_0}E_2\right] \leftarrow \text{Quantify the reduction of Randomness}$$



Note : at every step of decision tree learning we chose the feature axis and splitting point on that axis, which maximized the information gain.