

Feature extraction

Many of our model in machine learning have large number of features (independent variables). This makes our algorithm less efficient as we have to account for so many such variable. To solve this, we do feature extraction.

In feature extraction we combine of the similar features together to reduce the number of features.

Feature scaling

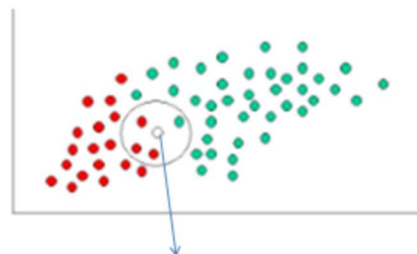
We do not consider units in machine learning so at times certain features tend to dominate others because of their large magnitude, like if we have weight and height as features and the values are like 5000g and 2m then weight tends to largely influence the outcome but it should not so we scale it down by certain factor. This is called feature scaling.

Naïve Bayes' theorem

The theorem is same as the Bayes' theorem but we assume that the features are independent of each other.

This is used to solve classification problems.

Suppose we have given set of green and red dots and we want ot predict the porbabilty of a new dot being red/green, this can be done by the theorem as



LIKELIHOOD OF THIS BEING RED = (NO. OF RED IN THE VICINITY / TOTAL NO OF RED)

$$\text{Posterior Probability of GREEN} = \text{Prior Probability of GREEN} \times \text{Likelihood of GREEN} = \frac{40}{60} \times \frac{1}{40}$$

$$\text{Posterior Probability of RED} = \text{Prior Probability of RED} \times \text{Likelihood of RED} = \frac{20}{60} \times \frac{3}{20}$$