KNN - For two groups

1) Determine the distance between the new observation and all data points in the training set

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

2) Sort the distances

3) Identify the k closest neighbors.

4) Determine the class of the new observation

$$\text{Accuracy} = \frac{\text{No of correct predictions}}{\text{Total No of predictions}}$$

Optimal value of k is

$$k = \sqrt{n}$$

More than two groups

1) Standarize the data

$$Z_i = \frac{X_i - \bar{X}}{SD}$$

$$SD = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}}$$

$\bar{X} = 0$
$S.D = 1$

→ should be for all variables

(2)

2) Determine the distance b/w new observation & all data points

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 + \cdots}$$

3) Sort the distance

4) Identify k closest neighbour

$$k = \sqrt{n}$$

5) Determine the class of new observation