# Support Vector Machine

- Supervised ML algorithm.
- SVM can be used for both classification & regression. But more commonly used for classification.

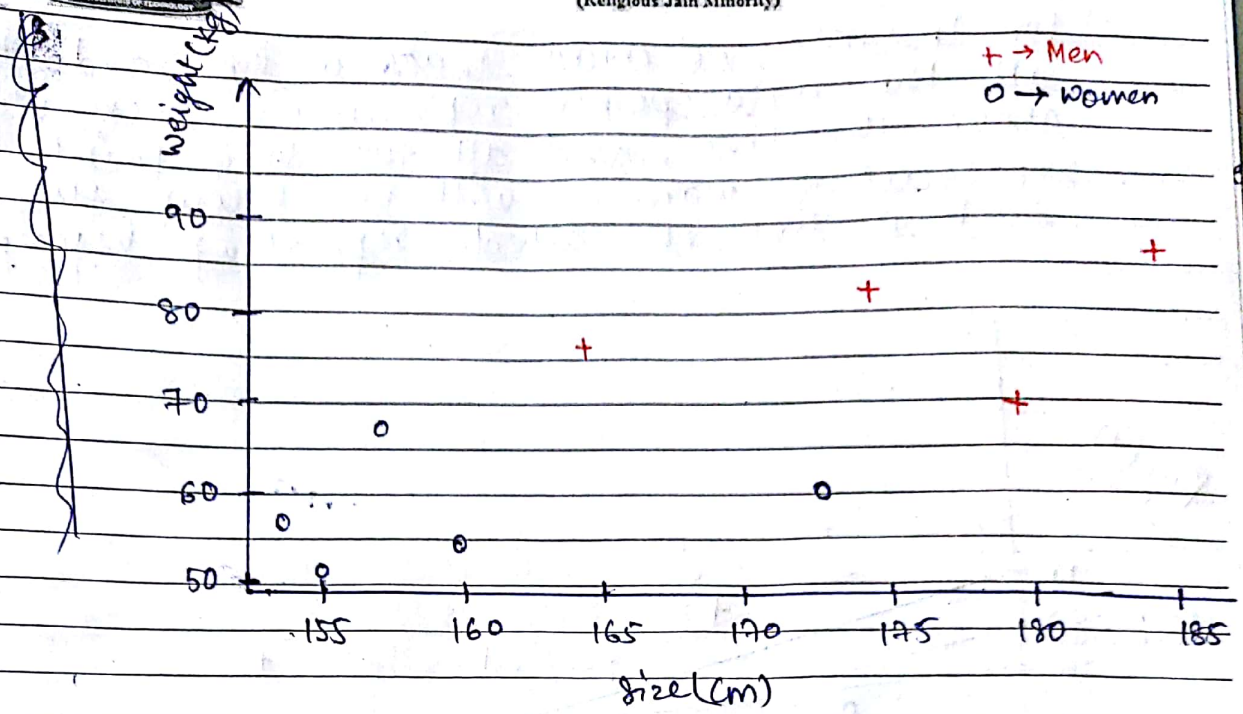
→ SVM tries to map an input space into an output space using a nonlinear mapping function $\phi$ such that, the data points become linearly seperable in the output space. When the points becomes linearly seperable then SVM discovers the optimal seperating hyperplane.

Linear SVM → {A model which assumes that data is linearly seperable}

→ The goal of an SVM is to find the optimal seperating hyperplane which maximizes the margin of the training data.

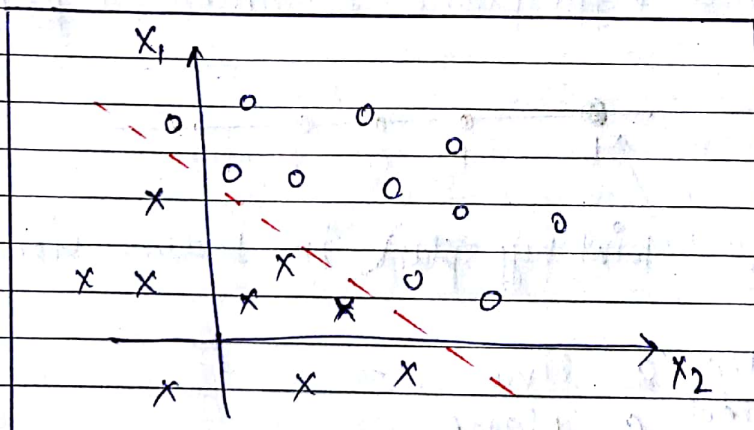
→ SVM is supervised learning algorithm which needs training data for finding the optimal seperating hyperplane.

Consider the training data as shown -




## Seperating hyperplanes →

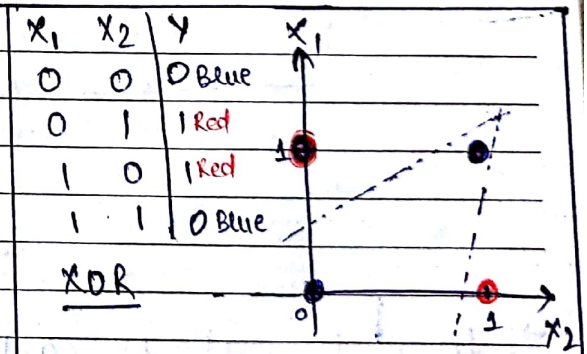
Just by looking at ~~data~~ the plot, we can see that it is possible to seperate the data.




Linearly seperable pattern

| $X_1$ | $X_2$ | Y |
|---|---|---|
| 0 | 0 | 0 Blue |
| 0 | 1 | 1 Red |
| 1 | 0 | 1 Red |
| 1 | 1 | 0 Blue |

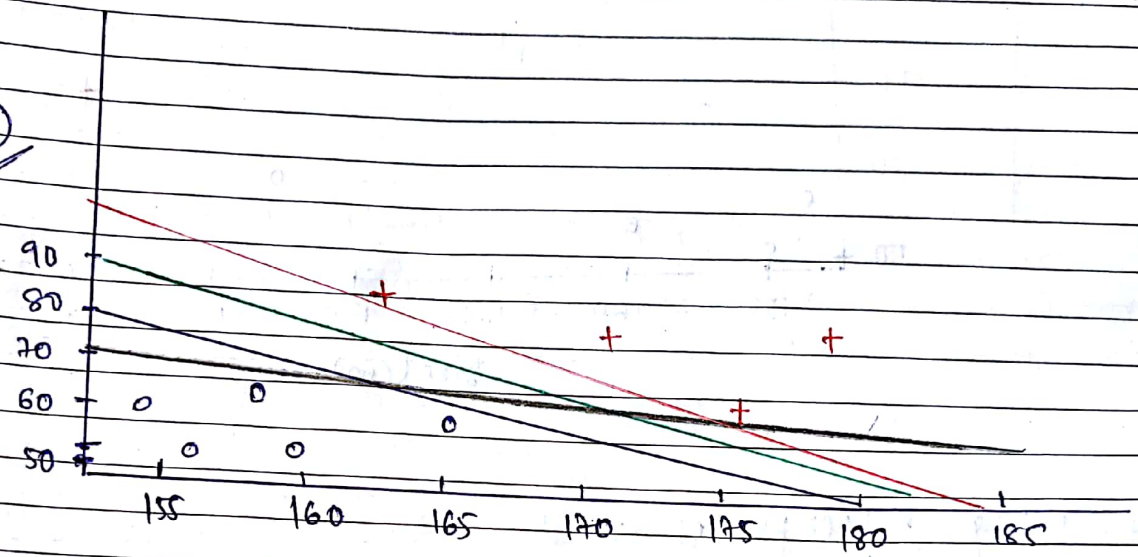
XOR




Linearly Inseperable pattern
(cannot be seperated with single line)

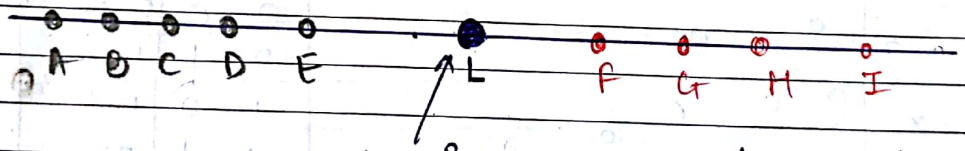
For Instance, we could trace a line and then all the data points representing men will be above the line, and all the data points representing women will be below the line. Such a line is called seperating hyperplane

Fig ①




SVM can work with any no. of dimensions.

→ In one-dimension, a hyperplane is called a point.




point hyperplane in 1-dimension

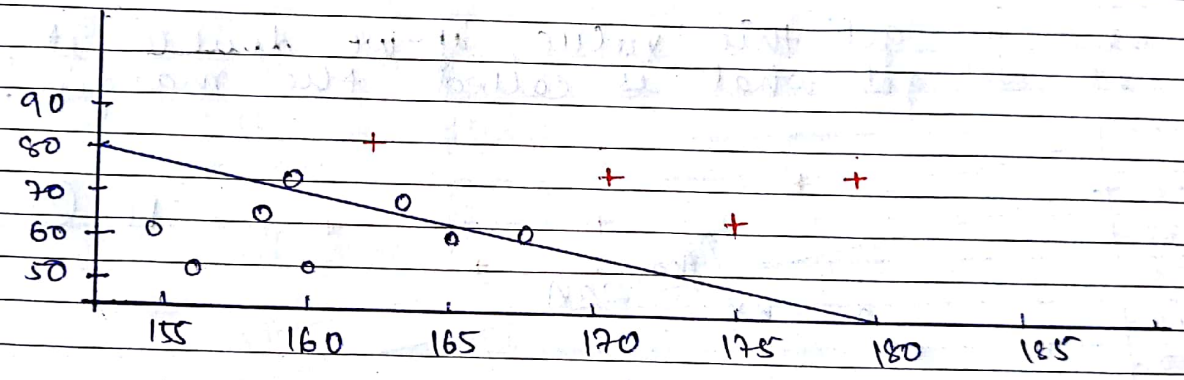
→ In 2-D, it is called a line.
→ In 3-D, it is called a plane.
→ In more-dimensions, we call it hyperplane.

Optimal Hyperplane →

The fact that you can find a seperating hyperplane does not mean it is the best one.

→ Figure ① shows that there can be several seperating hyperplanes. Each of them is valid as it successfully seperates the dataset with men on one side and women on the other side.

Suppose we select the blue hyperplane and use it to classify on real-life data, this hyperplane does not generalize well.




It fails to classify three womens correctly.

Intuitively, this can be because if a hyperplane is selected which is close to the data points of one class, then it might not generalize well.

→ Thus we need to select the hyperplane as far as possible from the data points of each category.

→ The green hyperplane classifies more accurately than blue hyperplane.

→ The objective of SVM is to find the optimal seperating hyperplane because

1) It correctly classifies the training data.
2) It generalize ~~well~~ better with unseen data.