Basic Terminology in Classification Algorithms

- **Classifier:** An algorithm that maps the input data to a specific category.
- **Classification model:** A classification model tries to draw some conclusion from the input values given for training. It will predict the class labels/categories for the new data.
- **Feature:** A feature is an individual measurable property of a phenomenon being observed.
- **Binary Classification:** Classification task with two possible outcomes. **Eg: Gender classification (Male / Female)**
- **Multi-class classification:** Classification with more than two classes. In multi-class classification, each sample is assigned to one and only one target label. **Eg: An animal can be a cat or dog but not both at the same time.**
- **Multi-label classification:** Classification task where each sample is mapped to a set of target labels (more than one class). **Eg: A news article can be about sports, a person, and location at the same time.**

Applications of Classification Algorithms

- Email spam classification
- Bank customers loan pay willingness prediction.
- Cancer tumor cells identification.
- Sentiment analysis
- Drugs classification
- Facial key points detection

Naïve Bayes Classification

- Based on Bayes theorem
- Assumption
  - Presence of one evidence is independent of other other evidence /feature (naïve)
  - equal contribution to the outcome.
- Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
- **It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.**
- Some popular examples of Naïve Bayes Algorithm are **spam filtration, Sentimental analysis, and classifying articles**.

Why is it called Naïve Bayes?

LIVE
shrutinewar98@gmail.com
9654220923

- **Naïve**: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other.
- **Bayes**: It is called Bayes because it depends on the principle of Bayes' Theorem.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where,

- **P(A|B) is Posterior probability**: Probability of hypothesis A on the observed event B.
- **P(B|A) is Likelihood probability**: Probability of the evidence given that the probability of a hypothesis is true.
- **P(A) is Prior Probability**: Probability of hypothesis before observing the evidence.
- **P(B) is Marginal Probability**: Probability of Evidence.

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

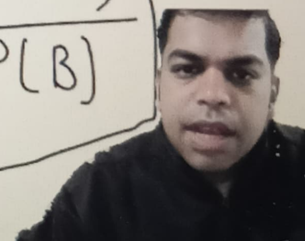
$$P(A \text{ and } B) = P(B)\,P(A|B)$$

Similarly

$$P(B \text{ and } A) = P(B/A)\,P(A)$$

$$P(B)\,P(A|B) = P(B|A)\,P(A)$$

$$\boxed{P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}}$$



Working of Naïve Bayes' Classifier:

- Convert the given dataset into frequency tables.
- Generate Likelihood table by finding the probabilities of given features.
- Now, use Bayes theorem to calculate the posterior probability.

x_1	x_2	x_3	y
a_1	b_1	c_1	Yes
a_2	b_1	c_2	No
a_1	b_2	c_2	NO

$x_1, x_2, x_3 \Rightarrow$ Independent

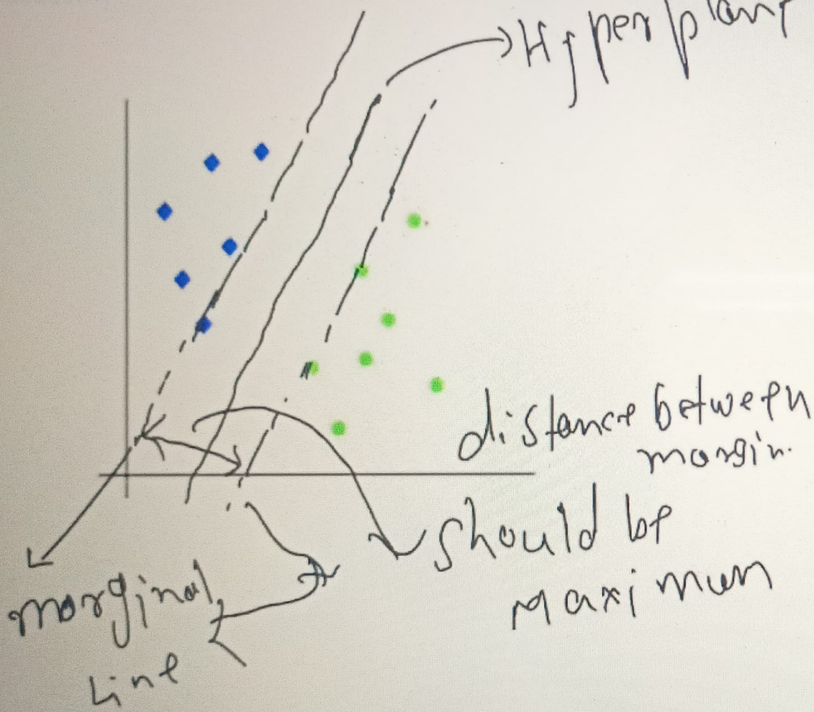
y = dependent

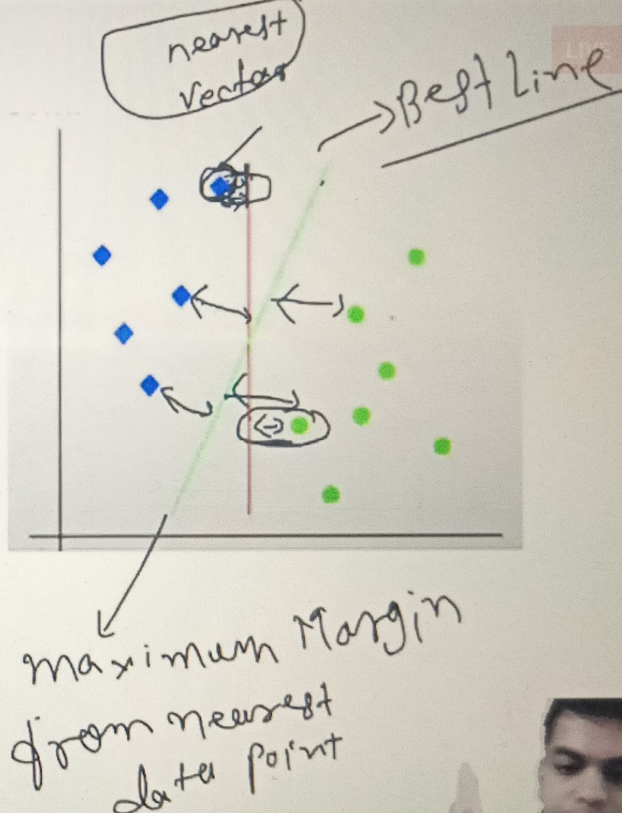
No. of class = 2 (Yes & No)

feature = 3 (x_1, x_2, x_3)

SVM

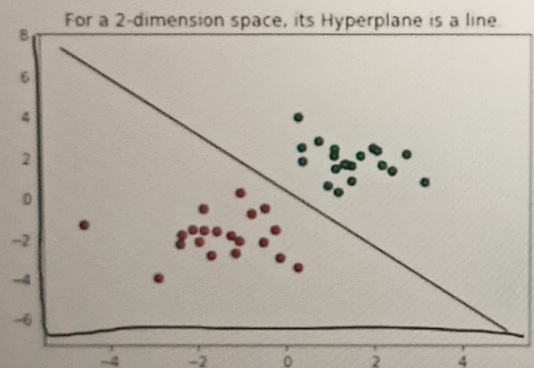
A **Support Vector Machine (SVM)** is a supervised machine learning algorithm which can be used for both classification and regression problems.

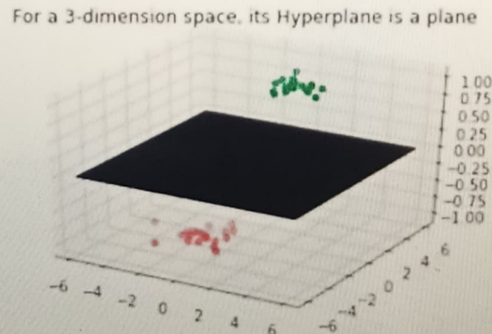

Hyperplane
distance between margin
should be maximum
marginal line


nearest vector
Best Line
maximum Margin from nearest data point

So what is a **Hyperplane**?

- Hyperplane is an (n minus 1)-dimensional subspace for an n-dimensional space
- For a 2-dimension space, its hyperplane will be 1-dimension, which is just a line.
- For a 3-dimension space, its hyperplane will be 2-dimension, which is a plane that slice the cube.





- Any Hyperplane can be written mathematically as below:

$$\beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \ldots + \beta_n * x_n = 0$$

- For a 2-dimensional space, the Hyperplane, which is the line.

$$\beta_0 + \beta_1 * x_1 + \beta_2 * x_2 = 0$$

$a_1 x + b_1 y_1$ $= y = mx + c$

- Any Hyperplane can be written mathematically as below:

$ax + by + c = 0$

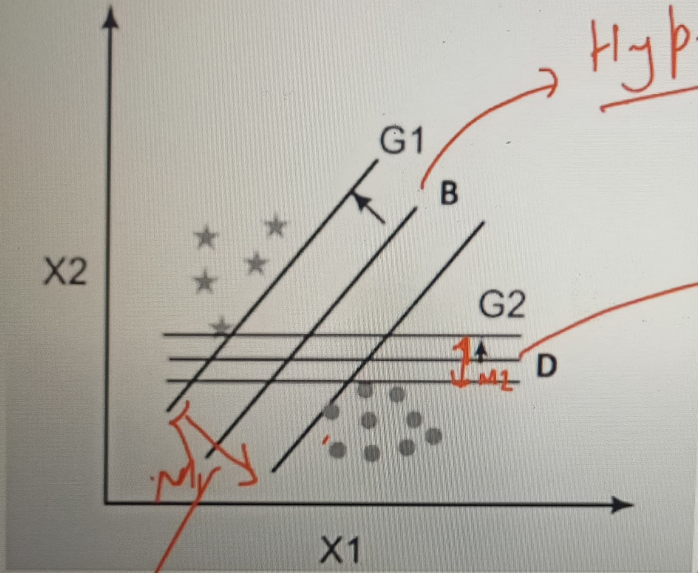
$$\beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \ldots + \beta_n * x_n = 0$$

$y = mx + c$

- For a 2-dimensional space, the Hyperplane, which is the line.

$$\beta_0 + \beta_1 * x_1 + \beta_2 * x_2 = 0$$

unacademy



Hyperplane

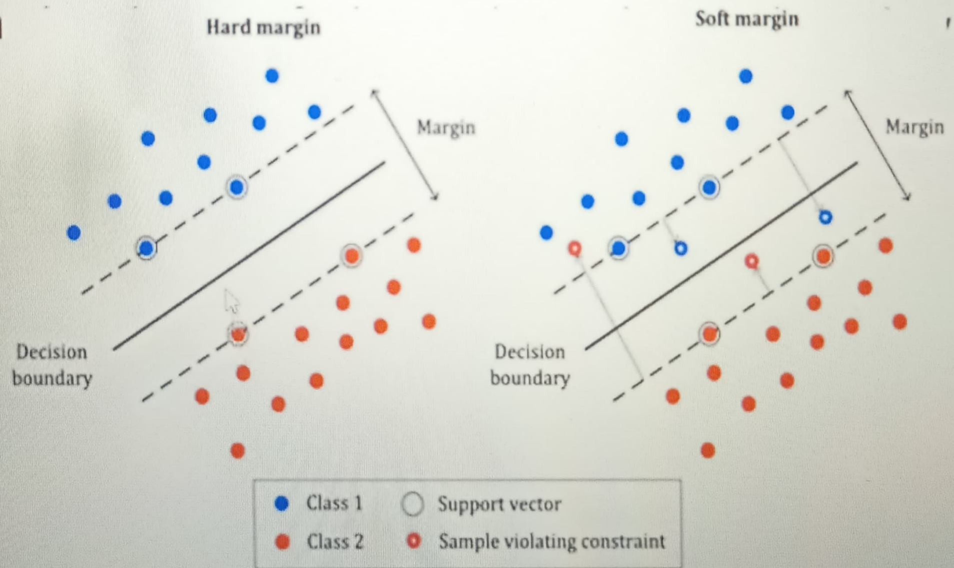
Hyperplane

maximum margin

$M_1 > M_2$

Soft Margin Vs. Hard Margin

- **Soft Margin:** try to find a line to separate, but tolerate one or few misclassified dots (e.g. the dots circled in red)
- **Kerı**



unacademy

Comparison with other classifiers

- SVM can be used to separate linear as well as non-linear space (kernel trick)
- Due to maximum margin on both sides it is less likely to result in over-fitting
- Can work with even smaller datasets
- Complexity is linear

unacademy

- ***Degree of tolerance***
  How much tolerance(soft) we want to give when finding the decision boundary is an important hyper-parameter for the SVM (both linear and nonlinear solutions).
- It is represented as the penalty term — 'C'.
- The bigger the C, the more penalty SVM gets when it makes misclassification..