SVM:-

Support vector Machine:-

* SVM comes under Supervised Learning
* SVM is used to solve both Regression & classification problems

① SVC → support vector classifier
② SVR → Support vector Regressor

① SVC:-
used to solve classification problem which has a categorical target.

Eg:- Diabetes, cancer...

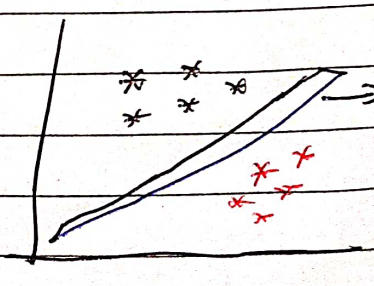
② SVR:-
used to solve regression problem which has continuous target

Goal of SVM:-

SVM is used to fit best fit line / straight line / hyperplane which seperates data into two class clearly.

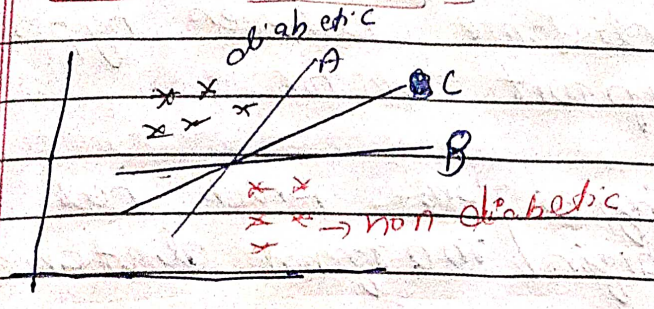
→ In simple words:-

"find hyperplane which seperates data into two class"



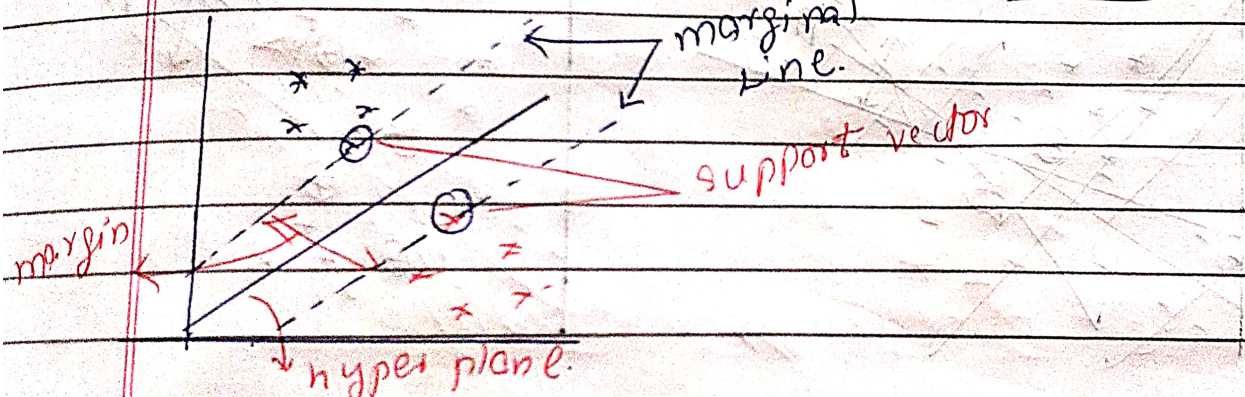
Hyperplane :- It is a straight line / decision boundary.

How SVM works :- Let us take diabetic data.



- There can be multiple hyperplanes which seperates data into two classes.
- In this example we can say that C is best hyperplane which seperate two classes.

→ How to select best hyper plane?



- Find point near to the hyperplane.
- Nearest points to hyperplane are called as support vectors.
- Draw a line near to the support vector
- The parallel lines drawn close to the support vectors are called as marginal line..

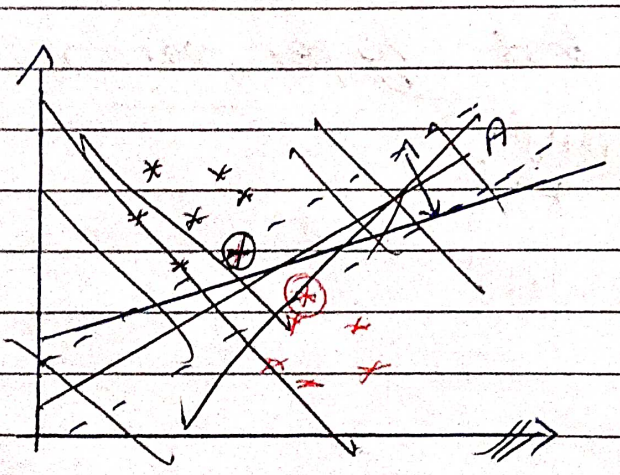
* find the distance b/w support vector & hyperplane.

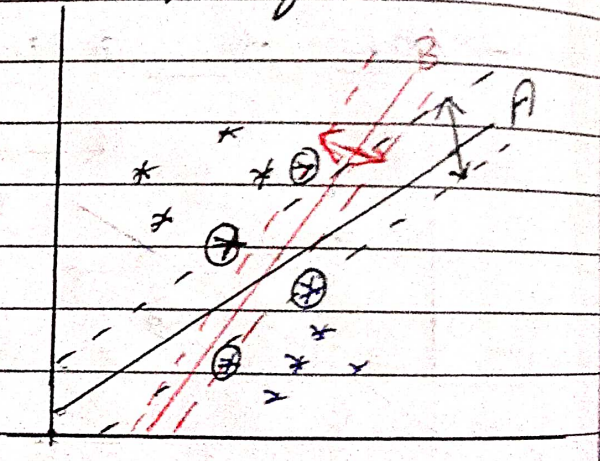
- **margin**:- the distance between hyperplane and support vector.

* select the hyperplane which has "maximum margin / marginal distance".

That means "select the hyperplane which has maximum distance b/w hyper-plane and support vector".

In short Goal of SVM:- maximize the margin.

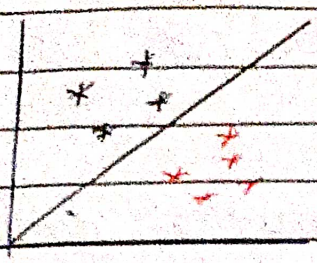




- select 'A' because It has maximum margin.

- This picture is in phone for reference

Linear data:



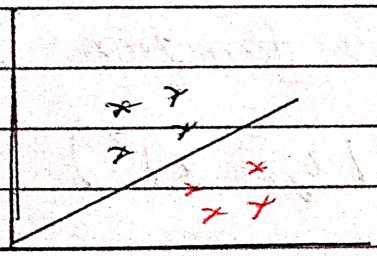
- If there is a clear seperation of classes we call it has Linear.

Non Linear data

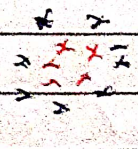
- If there is a over dapping of classes, we call it has non linear

Types of SVM:-

(1) Linear SVM:- If we can find a hyperplane which seperates data into two classes, we call it like Linear SVM



(2) Non linear SVM:- We can't find an hyperplane which seperates two classes.



1) How to overcome non linearity?

→ Non Linear svm uses Kernels to convert non linear data into linear

Kernels:-

Kernels are mathematical function that converts non linear data into linear data by converting lower dimension data into higher dimensions.

Types of Kernels:-

(1) Polynomial:-

Ex:- $y = x^2$, $y = x^3$, $z = x^2 + y^2$

(2) Sigmoid:-

$$ha(x) = \frac{1}{1+e^{-z}}$$

(3) Exponential:-

$y = e^x$ (or) $y = e^{4x}$

(4) RBF → Radial basis function

$$f(x, x_j) = Exp(-\gamma |x_j - x|^2)$$

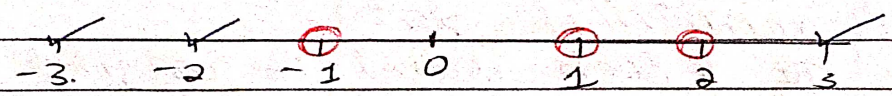
γ → Gamma tells us about how squared distance between two variables are varying.

Note:- SVM ignore outliers. no need to handle outliers.

(1)

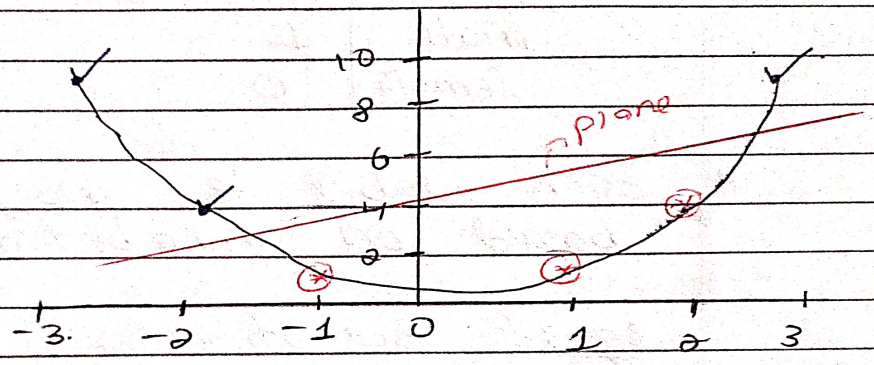
x	
-1	-ve
-2	+ve ✓
-3	+ve ✓
1	-ve
2	-ve
3	+ve ✓

Non linear data



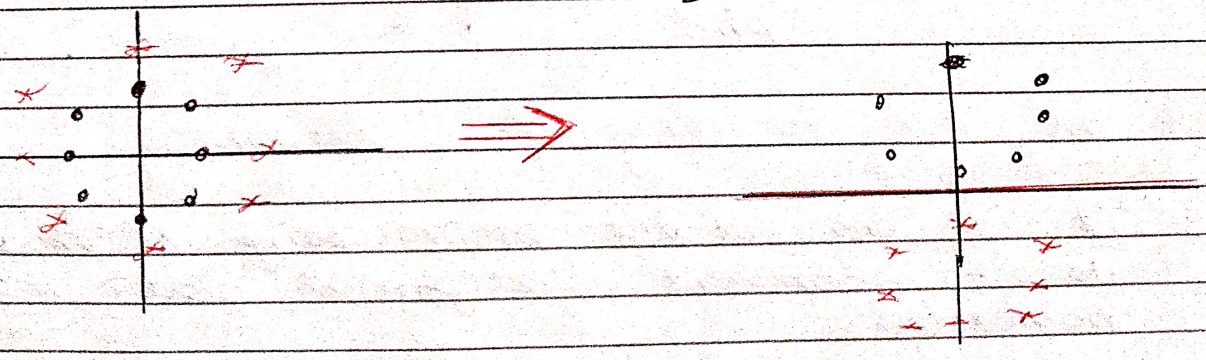
converting non linear to linear.

	x	$y = x^2$
-ve	-1	1
+ve	-2	4
+ve	-3	9
-ve	1	1
-ve	2	4
+ve	3	9.



(2)

$z = x^2 + y^2$



Converting categorical data into numerical data

(1) Label Encoder

- This is one of the preprocessing method used to convert categorical data into numerical data.
- Each label is assigned a unique integer based on alphabetical ordering.

Eg:-1)

Gender	Gender
male	1
female	0
male	1
female	0

Eg:-2

colur	Color
red	2
green	1
blue	0

Eg 3:-

Education level	
High school	0
Undergraduate	3
master	1
phd	2

(2) one hot Encoding / Get dummies

- It is one of the preprocessing method used to to convert categorical data into numerical.
- In one-hot encoding, each unique category in categorical variable is converted into a binary vector of zeros & ones.

- The length of the binary vector is equal to the number of unique categories in the variable.

- For each observation the corresponding element in the binary vector is set to 1 if the category matches that observation and 0 otherwise.

eg: 1

Gender	Male	Femal
male	1	0
female	0	1
male	1	0
female	0	1

eg 2:-

color	red	green	blue	orange
red	1	0	0	0
green	0	1	0	0
blue	0	0	1	0
orange	0	0	0	1

Why Label Encoder fails @ Limitations?

- Label Encoder result in high priority issue
- A label with high number will be considered as high priority than the label with low number.

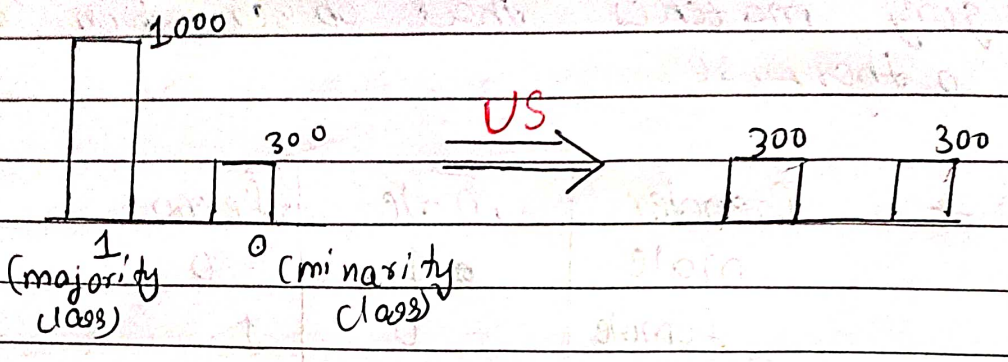
Eg:-

Result	
Pass	1
fail	0

Review	
good	1
bad	0

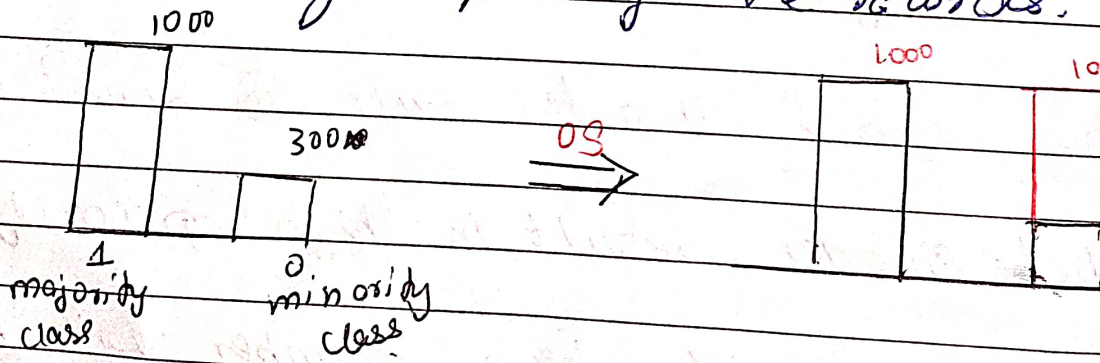
Balance data:-

① Under sampling:-
Reducing number of observation in majority class to match with minority class.



- failed :- Under sampling (US) fails because it results in loss of informations

② oversampling :-
Sampling technique which increases the number of observation in minority class to match with majority class by duplicating the records.



- oversampling fails :- It results in duplicates

(3) smote :-

- synthetic minority oversampling technique.
- It is a very popular method used to balance data.
- Smote is one of the oversampling technique which creates synthetic or artificial records in minority class to match with ~~majority class~~. majority class.

imp

- KNN ~~[redacted]~~ interpolation method to create artificial records.

Note:- Balancing is always done on training data.

Cross validation :- [recomended K=3 (or) K=5 (or) K=10]

- Cross validation is a method used to evaluate model at different datasets.
- K-Fold Cross validation
- It divides data into K sets of equal size.

Eg:- K=3 :-

test	train

train	test	train

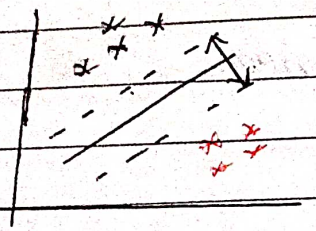
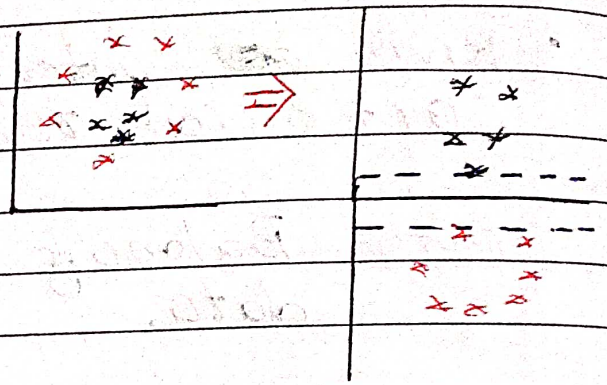
train	test

← All data →

Hyper parameters

- Learning parameters which helps in enhancing @ improving model performance.

→ Hyperparameters in SVM are
1) C
2) γ → RBF

Linear	Non Linear
• 	
• Margin seen in Linear data is called as hard margin	• Margin seen in non linear data is called as soft margin
• ~~No of misclassification will be seen in hard margin. [maximum margin]~~	• No of misclassification will be high in soft margin because the @ marginal distance will be less.
• No of misclassification will be ~~seen~~ less in hard margin because of maximum marginal distance	

(1) C → cost of misclassification

- 'C' is a learning parameter in SVM which controls the misclassification occured by soft margin.
- C adds penalty to the model to reduce the misclassification.

C is small :- ×

If 'C' is less then the penalty added will be less, so 'C' cannot be small.

(VVI) C is large :- ✓

If C is large then penalty added will be more, which reduces misclassification.

(2) γ - Gamma :-

Gamma is a learning parameter in SVM which control distance b/w observations

γ → is seen in RBF

$$f(x, x_j) = \exp\left(-\gamma (x - x_j)^2\right)$$

(VVI) γ → Gamma should be small.

Note:- C should be large & γ should be small.

- verbos = 2 → To show message
- confussion matrix → VVI

Hyperparametes tuning

- Hyperparametes tuning techniques are used to select best hyperparameter from the given list.

⇒ Grid search CV :-

It is one of the hyper parameter technique which train and Evaluate the model for Every combination of given parametes list.