Support Vector Machine

06 June 2023 07:30 AM

Actual data

2 m Rows → 50 -100 coloums

4-5 algo

- Linear model
- KNN
- DT
- RF
- Ada Boost

Naive Bayes →

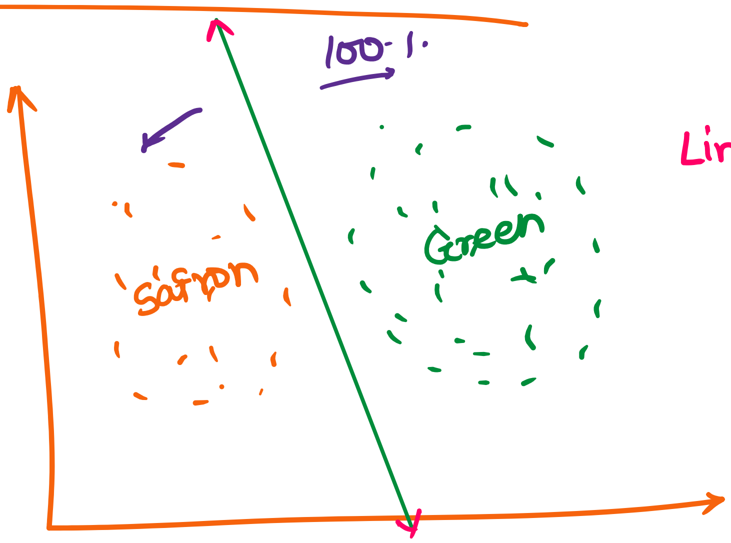
10K Rows and 5K columns

small data size

10K and 20K - 50K

Linear models and KNN —

good accuracy

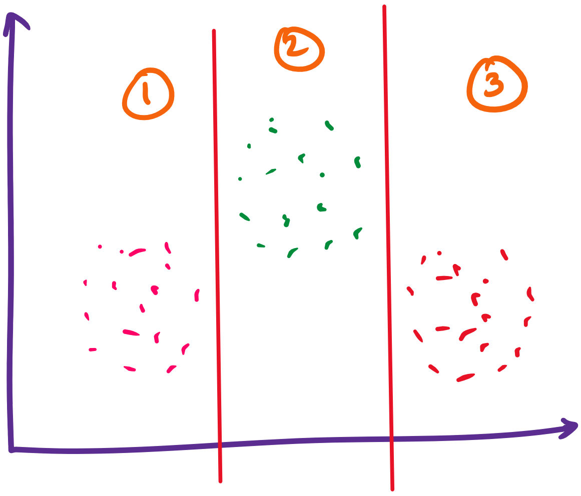


Linearly separable data.

Logistic reg.

✓ $= \frac{1}{1+e^{y}}$

0 to 1



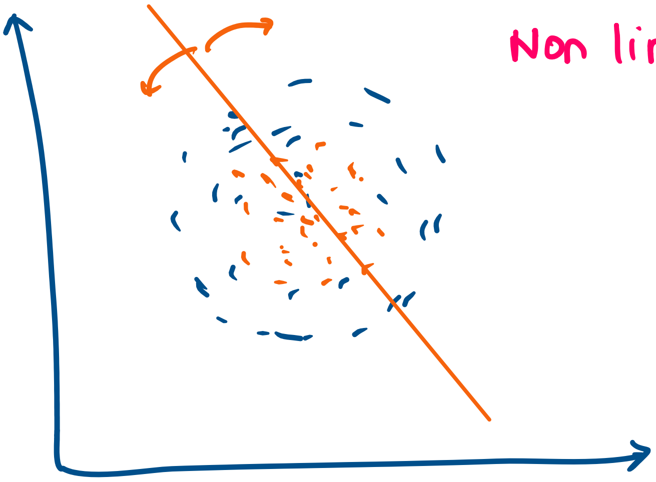
Linear line

✓ Linear model

✓ KNN

✓ DT

✓ Ensemble



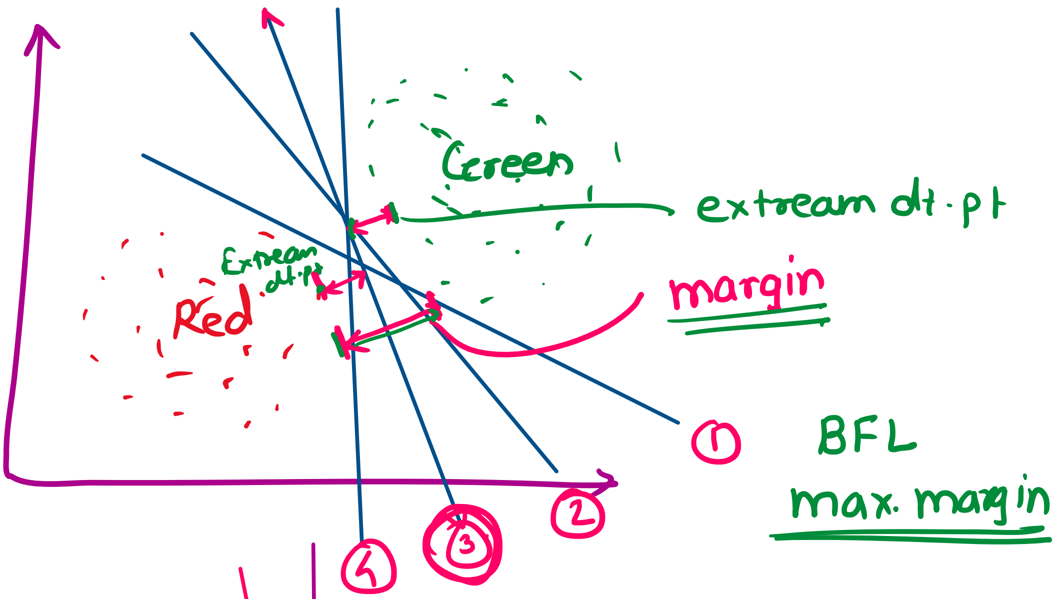
Non linearly separable

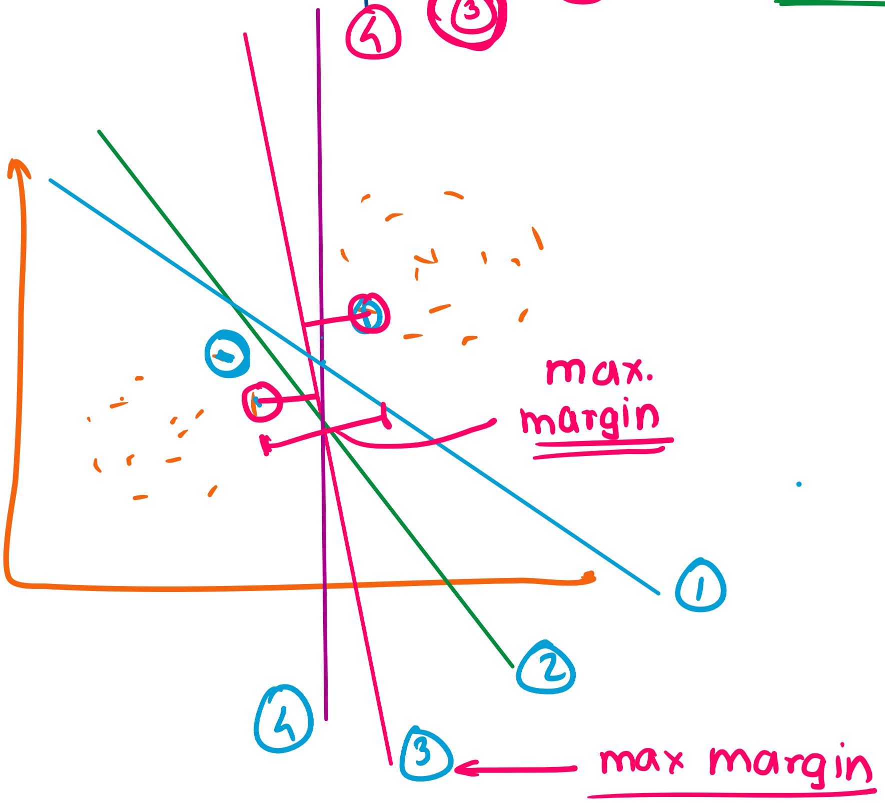
SVM will work.

SVM → support Vector Machine

handle both linearly separable and non linearly separable data

① linearly separable data





Goal of SVM ⟶ decision boundary

✓ 'Hyperplane'

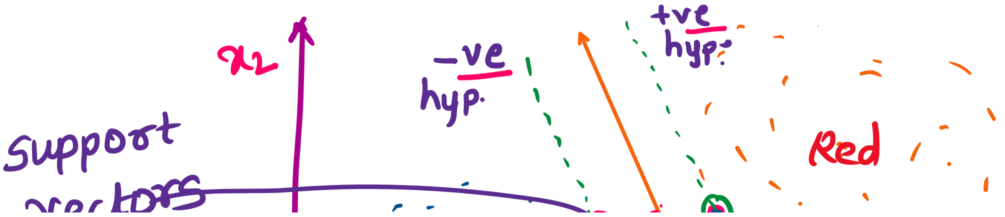
To find best hyperplane

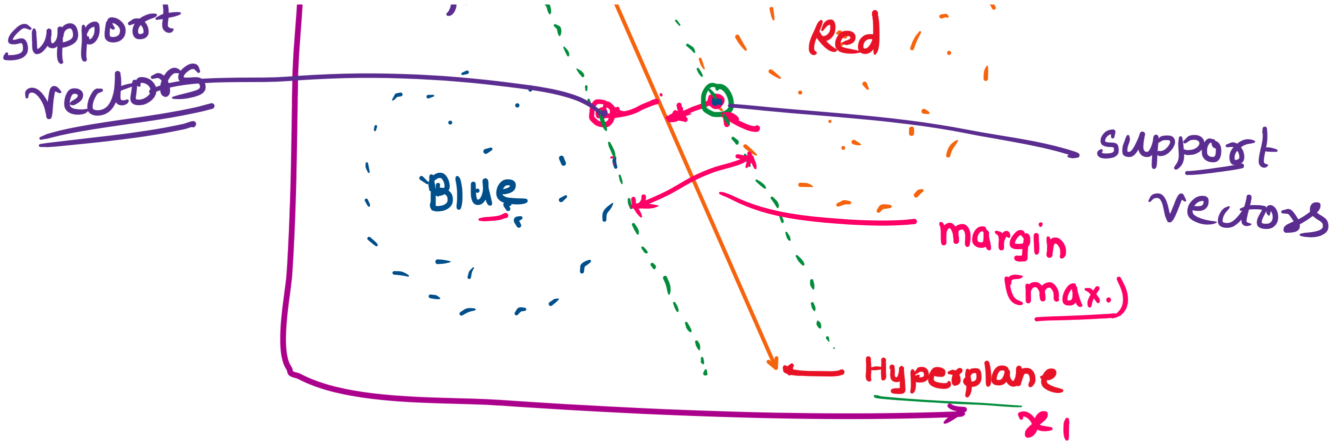
Terminology in SVM

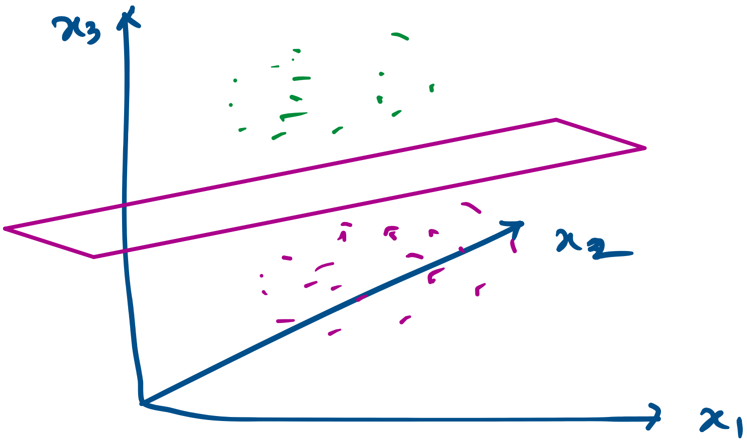
2-features →

1D - hyperplane

Straight line

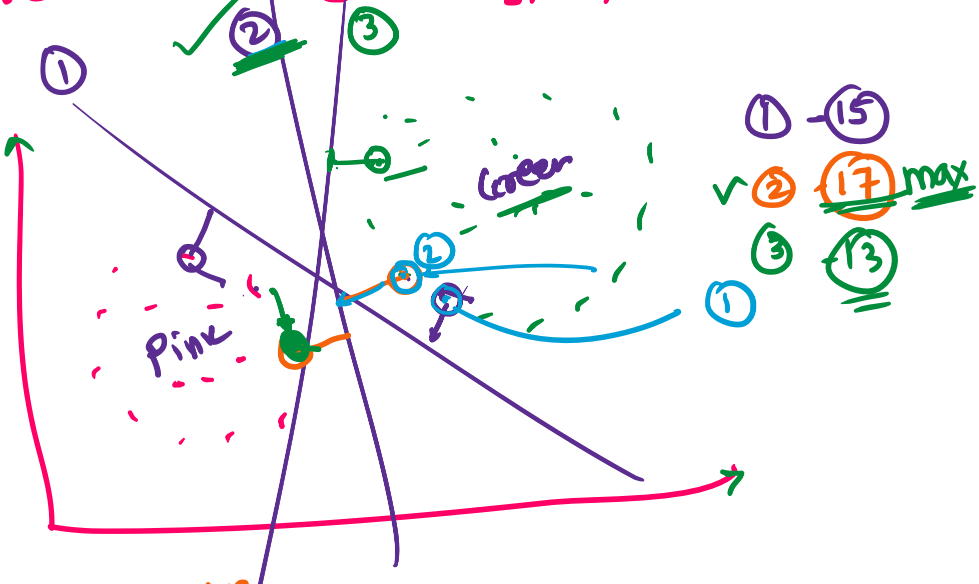


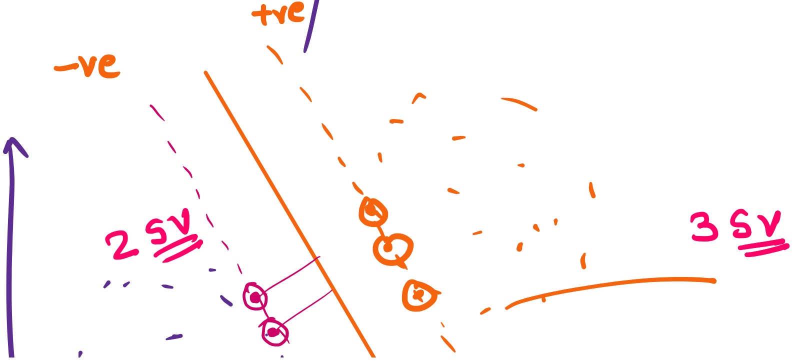


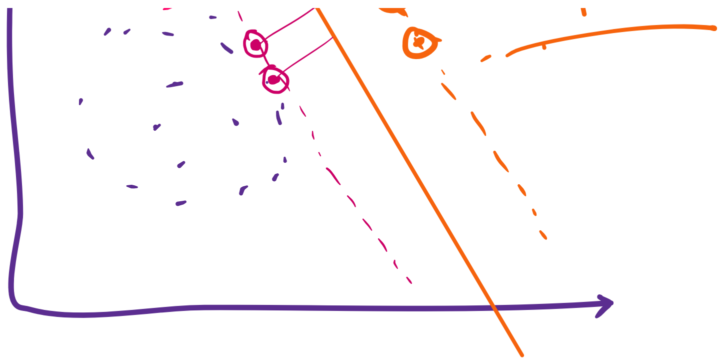


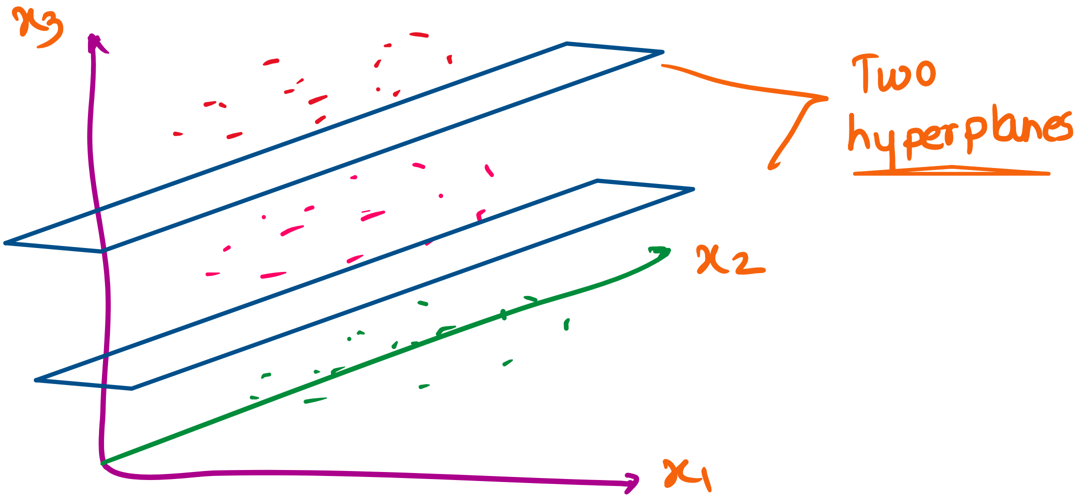
✓ 3-features → 2D-hyperplan

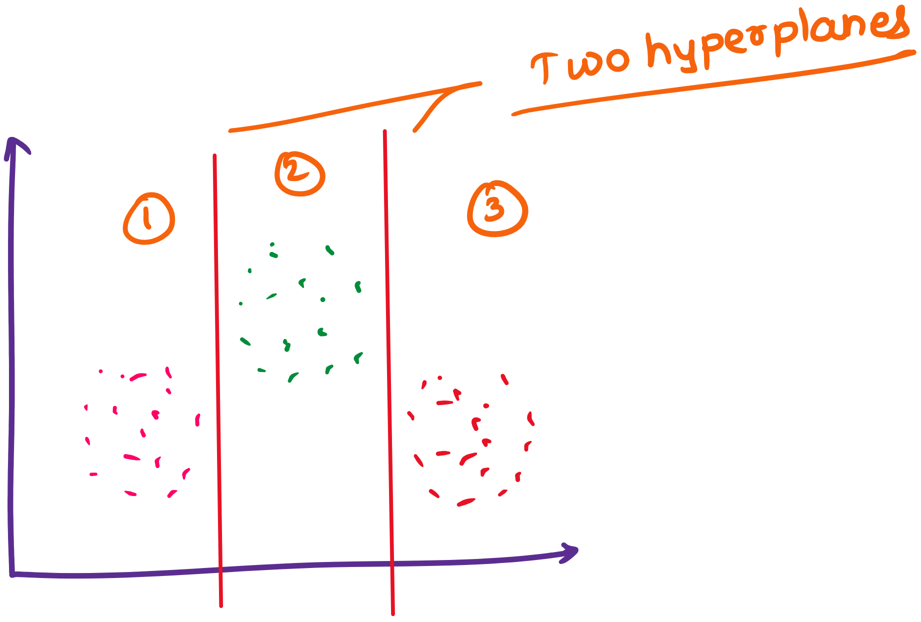
4-features ——— 3-D hyperplane



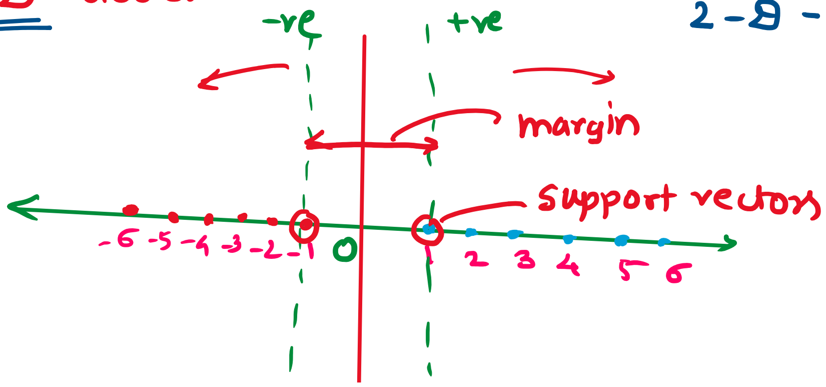








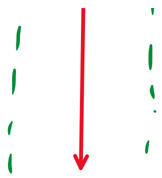
① For 1-D data



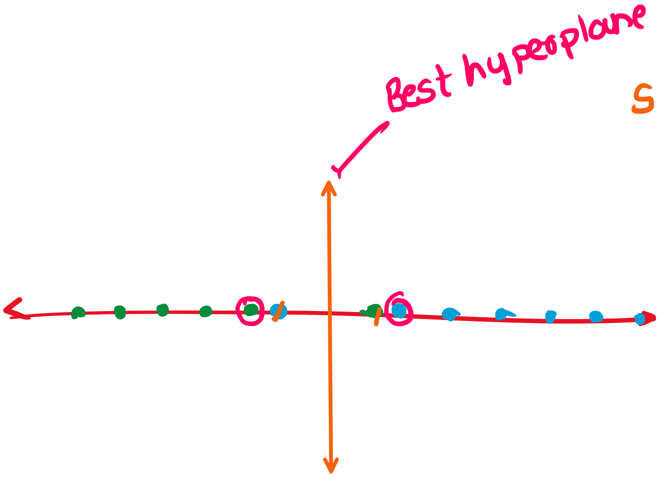
1-D — 1 Line

2-D - x-y-plane

3D -



② case ②

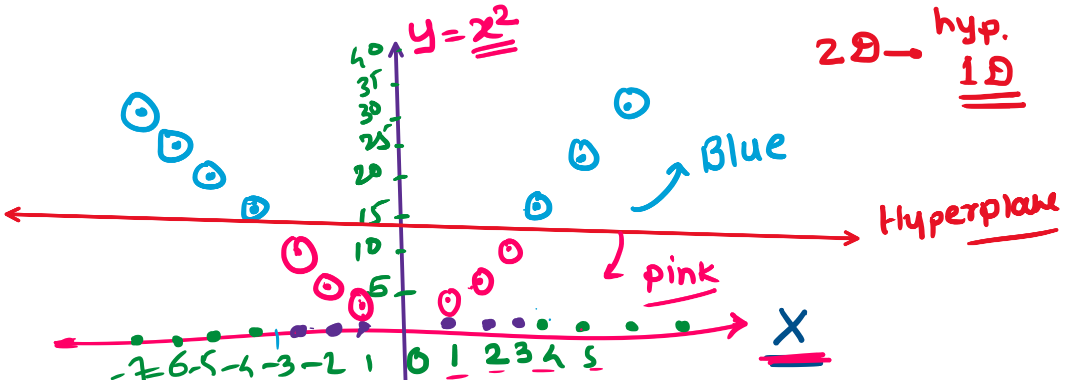


skip two missclassified data points

draw hyp. on remaining correctly classified dt. pts.

case - 3

2D → hyp.
1D



SVM

Kernel

convert Lower D to Higher D

1D → 2-D

Polynomial $f(x) = x^2$ Assume $y = f(x)$

Kernel → select best function

Best kernel

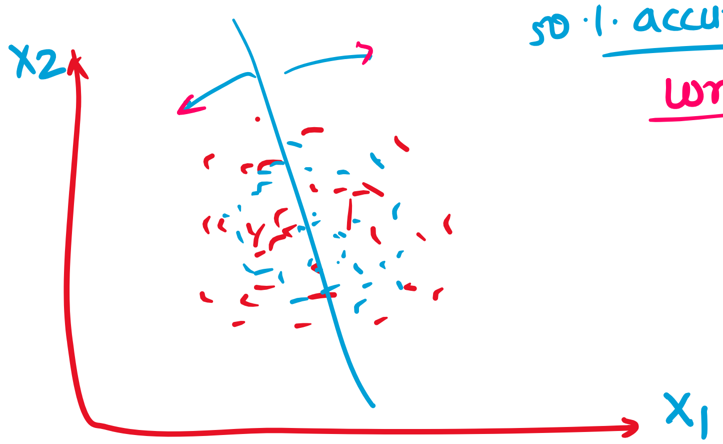
by using hyperparameter tunning

Kernels

1. polynomial
2. RBF → Radial Basis Function (By default)
3. Sigmoid kernel
4. Linear kernel

50 ·/· accuracy

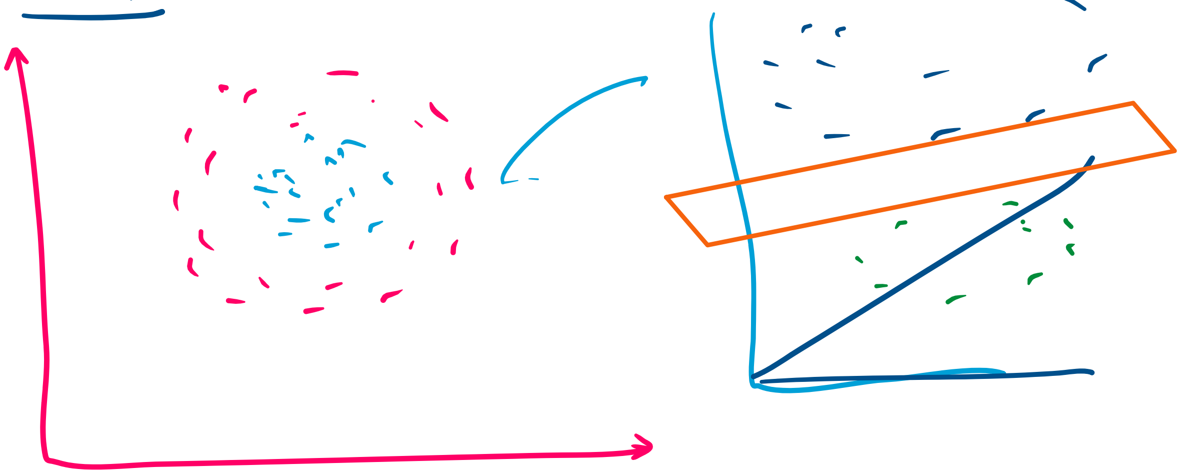
lowest accuracy

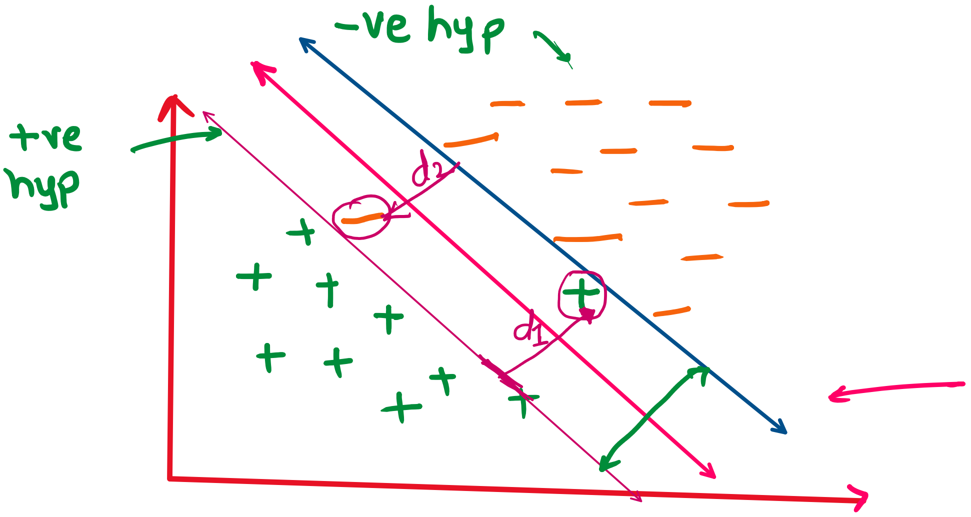


50 % accuracy

wrost accuracy

some pattern





SVM Errors = Marginal Error + Classification Error

High Margine - clear & distinct classification

less|min error

classification Error :-

$\Sigma(d_1 + d_2)$

$C \cdot \sum_{i=1}^{n} \xi$ (zeta)

SVM Errors = Marginal Error + Classification Error

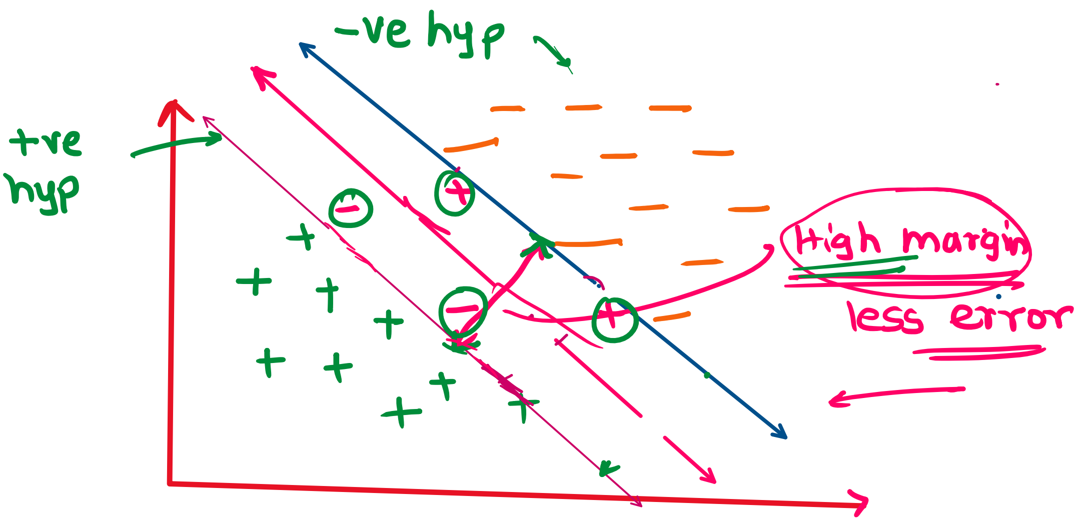
Marginal Error → concern

Error → min

confunsion Matrix

FP — precision

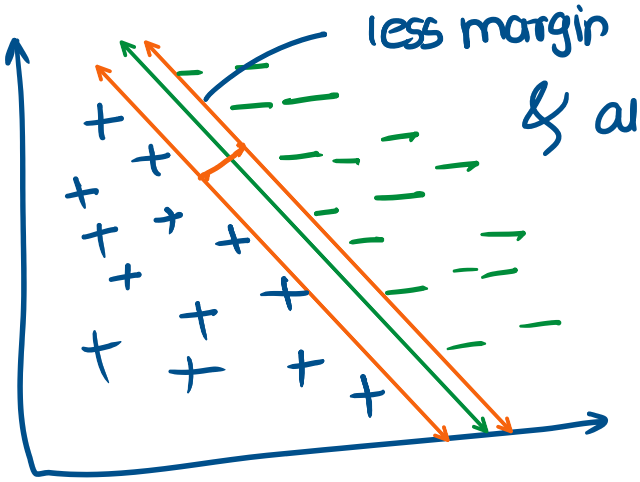
FN — Recall



✓ Soft Margin

SVM Errors = Marginal Error + Classification Error (min. classification Error)

Error → Concern



& all the data points are correctly classified

Hard Margin

Svm Errors = Marginal Error + Classification Error

(zeta)

$C \sum_{i=1}^{n} \{$

C value is 1000

Focusing on

Too high

Fousing error 1000 Classification Error

High

Default value of C = 1.0