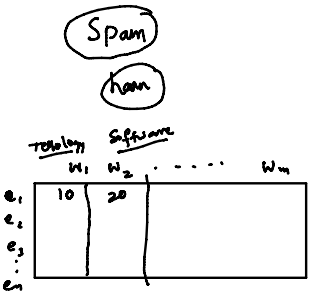
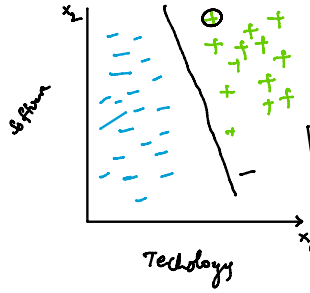
SVM-1

20 December 2022 20:42

Support Vector Machines



Logistic Regression

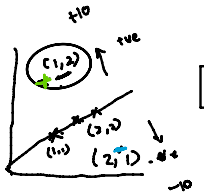


$w_0 + w_1 x_1 + w_2 x_2 = 0$ (eq 1)

$ax + by + c = 0$ ← line

$by = -c - ax$

$y = \frac{-c}{b} - \frac{a}{b} x$



$y = x$ $c = 0$ $m = 1$

$m = \frac{(y_2 - y_1)}{x_2 - x_1} = \frac{2-1}{2-1} = 1$

$y - x = 0$ ✓

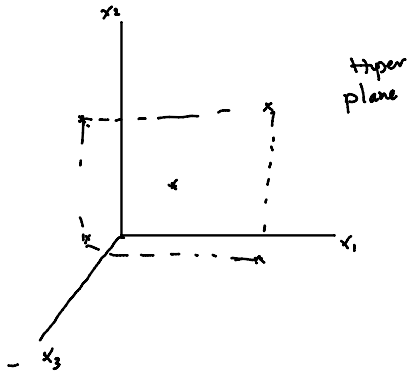
$\left(\frac{1}{1+e^{-x}}\right) \sim (0-1)$

$y - x = 0$ $(2,2)$ $(1,1)$ ✓

$(1,2)$ $y - x > 0$ $2 - 1 = 1$ ✓

$(2,1)$ $y - x < 0$ $1 - 2 = -1$

3-dimensional



hyper plane

Hyperplane

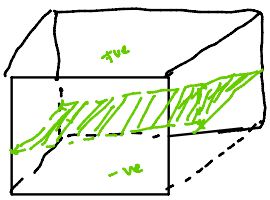
$ax + by + cz + d = 0$

$w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3 = 0$ ✓ Sigmoid $p = 0.5$

$w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3 > 0$ +ve $p > 0.5$

$w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3 < 0$ -ve < 0.5

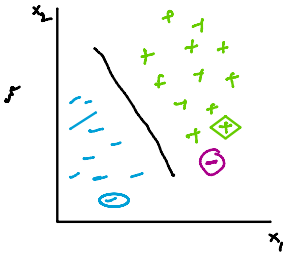
(10 input)



if $\sum_{i=1}^{p} w_i \cdot x_i + w_0 = 0$ ← hyperplane

$y \in \{0, 1\}$

$y \in \{+1 - \text{Pos}, -1 - \text{Neg}\}$



$w_0 + w_1 x_1 = 0$ ✓

$w_0 + w_1 x_1 > 0$

$w_0 + w_1 x_1 < 0$

$-1 \times -ve \; (+ve)$

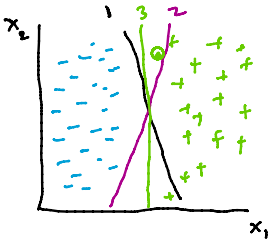
$y \in \{+1 \text{ Pos}, -1 \text{ Neg}\}$

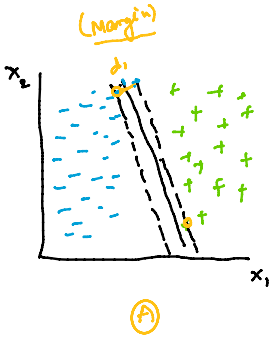
$y_i (w_0 + w_1 x_1) > 0$ (Correctly classified)

y_i

$-1 \; (-ve) \quad +ve \leq 0$ (Wrongly classified)

Margin classifier

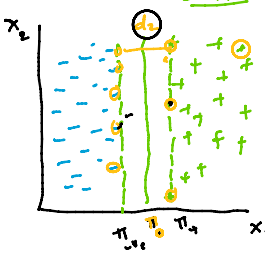




(A)

"Maximum Margin classifier"

'Support Vectors' ✓



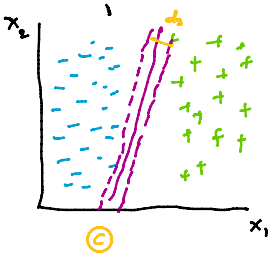
(Generalization)

(B)

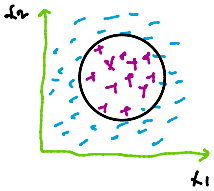
$a_1 x + b_1 y + c_2 + d = 0$

$a x + b y + c_2 + d = 0$

$a_1 x + b_1 y + c_2 + d = 0$



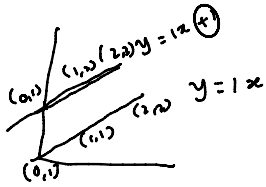
(C)



Objective

To find a line that can have Maximum margin?

Max (d) → or Minimize $\left(\frac{1}{d}\right)$

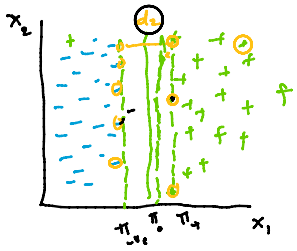


(Maximize the margin)

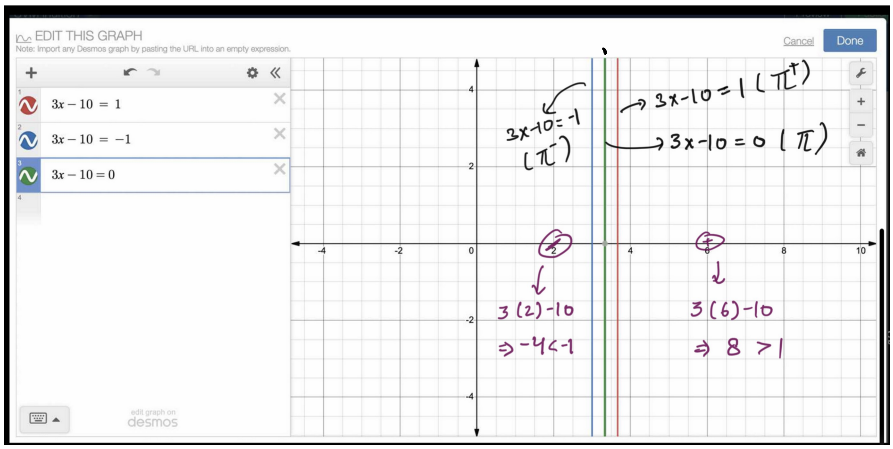
Hard Margin ✓ (Maximum Margin classifier)

Min $\left(\frac{1}{d}\right)$

Max (d)



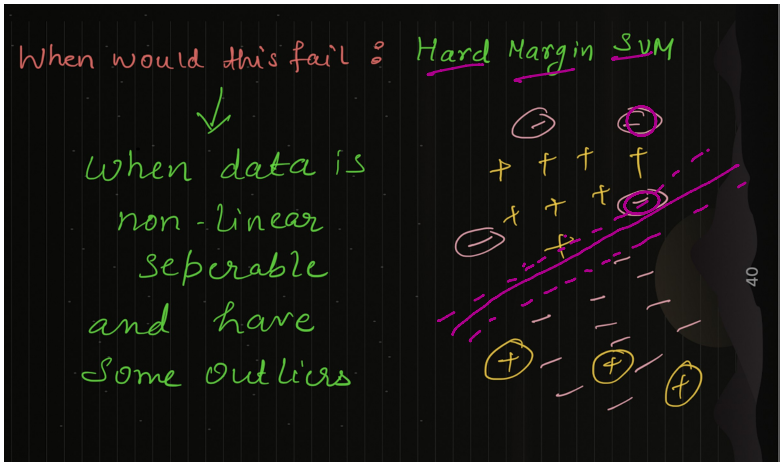
Driven by the Support vectors that are nearest and no Misclassification



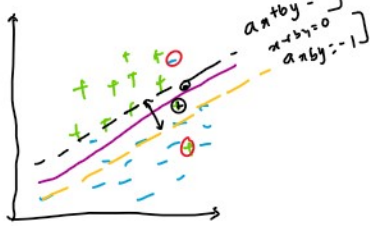
$ax + by = 0$

$ax + by = 1$ (K)

$ax + by = -1$ (-K)



Soft Margin classifier



0-1

[0-1]

(ϵ_i)

hard → soft margin

Obj

Max d + min (ϵ_i) ?

Cost = $\min\left(\frac{1}{d}\right) + \min(\epsilon_i)$

[Regularization
Penalization]

how much error is acceptable?

Cost = $\min\left(\frac{1}{d}\right) + C * \min(\epsilon_i)$

C. hyper meter

$C = 10$

Cost = $\min\left(\frac{1}{d}\right) + 10 \cdot (\epsilon_i)$ ↓

+ (20)

+ 10 (1) = 10 ↓ ↑

ϵ_i (1, 2)

1, 1

ϵ_i (1)

High value 'C' ← Overfitting

ϵ_i (10) looks wrong

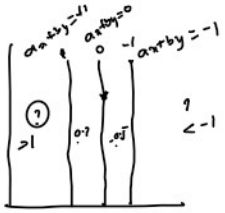
$C = 1$

+ 1 (10) ↓

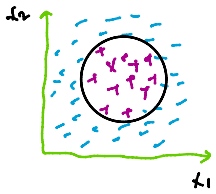
10 ✓ (Underfitting)

(1 - 10)

Break 10:29 pm



Non-linear ✗



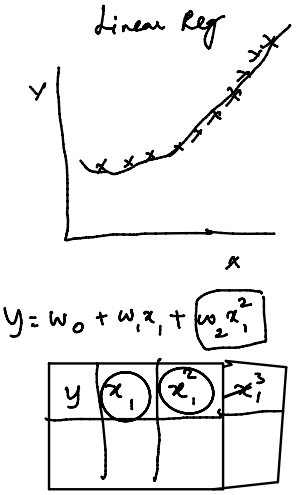
✓

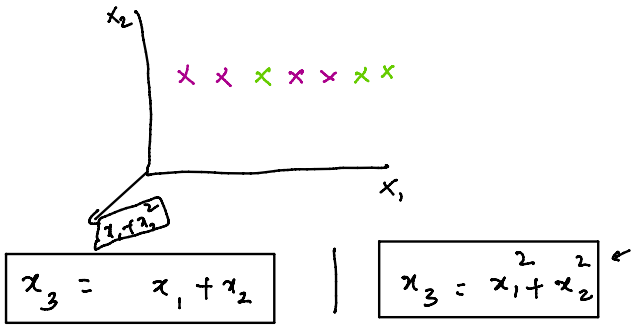
Kernels - (Mathematical functions)

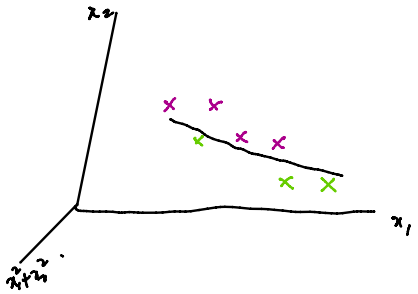
Map $d \longrightarrow d'$ where $d' >> d$

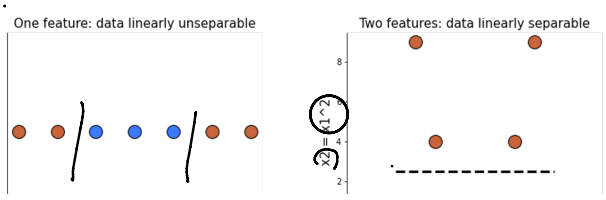
3 → (5)

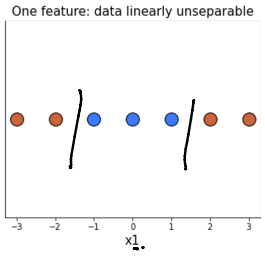
5 >> 3

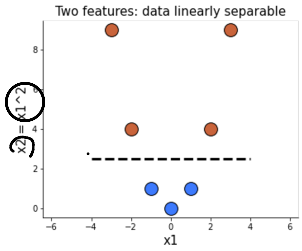








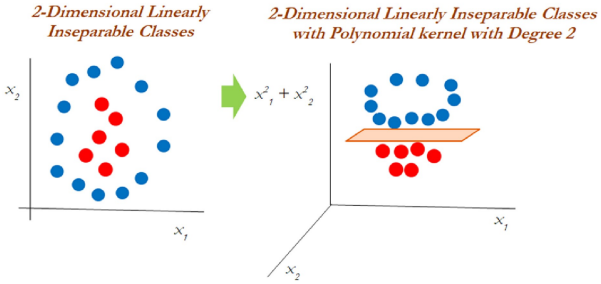




- Linear →
- polynomial
- RBF (Radial Basis function)

'Sigmoid'

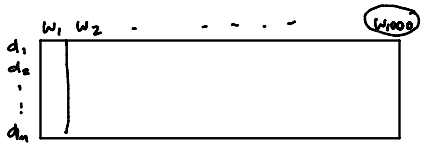
Map (d) → (d) ← find a linear Separation in (higher dimensional)



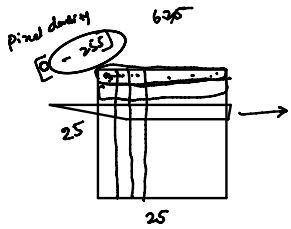
'kernel trick'

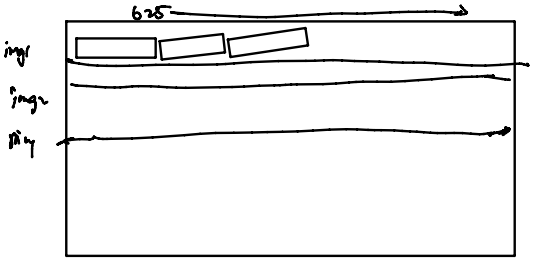
Spam

Text data



↗ high dimensional space

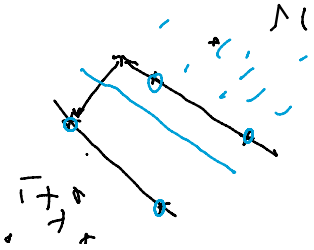


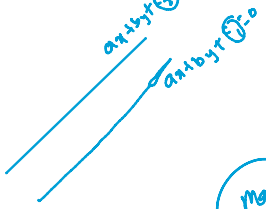


SVM

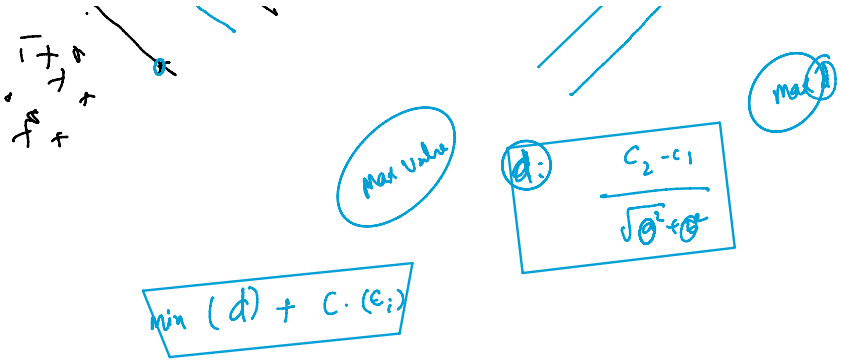
10000

$\wedge(x_1, x_2)$ or $\wedge(x_1, x_2 \ldots x_{10000})$





max



Advantages of SVM over other algorithms.

- less data (500 rows 400 features columns)

$n = 500$, features $= 400$

Classification model ? → (400) × — atleast 100 training examples

$(400) \sim (400 \times 100)$

Bigdata $\{n : 10,000\}$

{Healthcare data} (Rare disease)

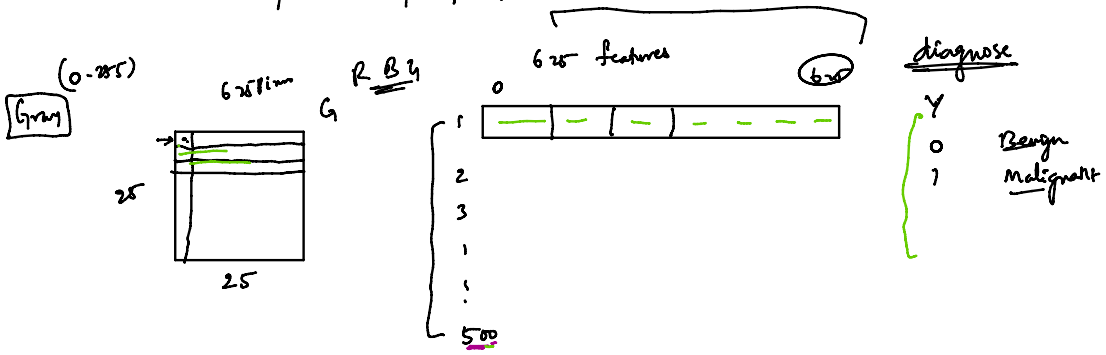
X rays, Scans

1000 features

250 Patients

{500} people

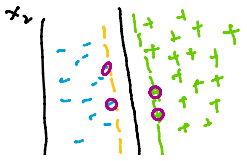
300 pt Healthy people

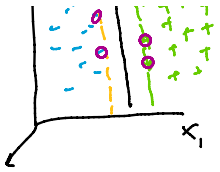


DL

Classified

SVM.





$P(S/w) =$

likelihood (Prob)

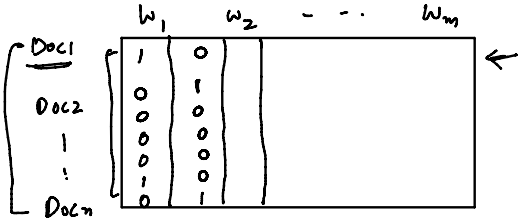
Sparse data

Text → Matrix

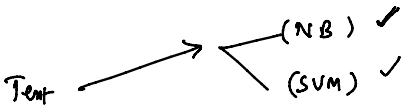
(NB) ←
(SVM) ←

DTM

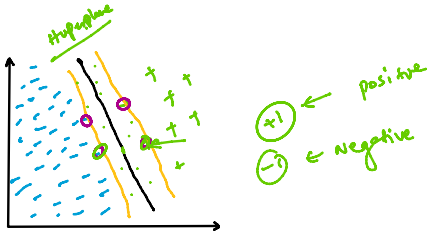
Text



L.R } ⊗
D.T
ensembles ⊗

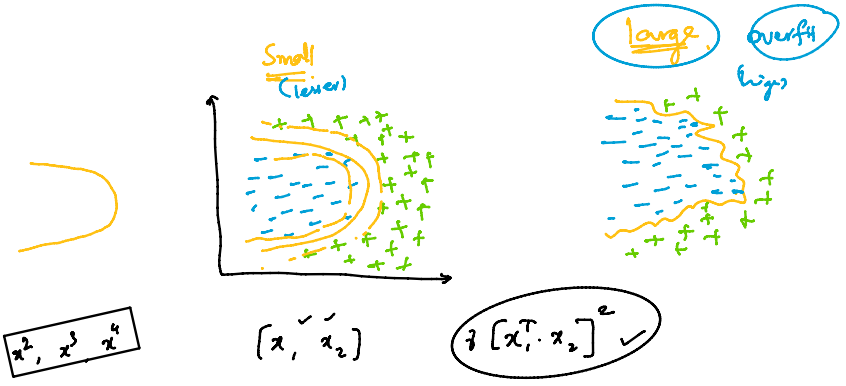


⊕ Class Imbalance. "a" will it effect SVM ?



gamma (?) ← Non linear Kernels, (poly, RBF, Sigmoid)

? - does → how smooth the Seperation is.



$x \rightarrow [x^2, x^3, x^4]$

$[x_1, x_2]$

$? [x_1^T \cdot x_2]^2$ ✓

Optimization

Constrait optimization

$$\text{Linear:} \quad K(\mathbf{a}, \mathbf{b}) = \mathbf{a}^\top \mathbf{b}$$

$$\text{Polynomial:} \quad K(\mathbf{a}, \mathbf{b}) = (\gamma \mathbf{a}^\top \mathbf{b} + r)^d$$

$$\text{Gaussian RBF:} \quad K(\mathbf{a}, \mathbf{b}) = \exp\left(-\gamma \|\mathbf{a} - \mathbf{b}\|^2\right)$$

$$\text{Sigmoid:} \quad K(\mathbf{a}, \mathbf{b}) = \tanh\left(\gamma \mathbf{a}^\top \mathbf{b} + r\right)$$

Vladimir vanpik

[Primal form, Dual form]