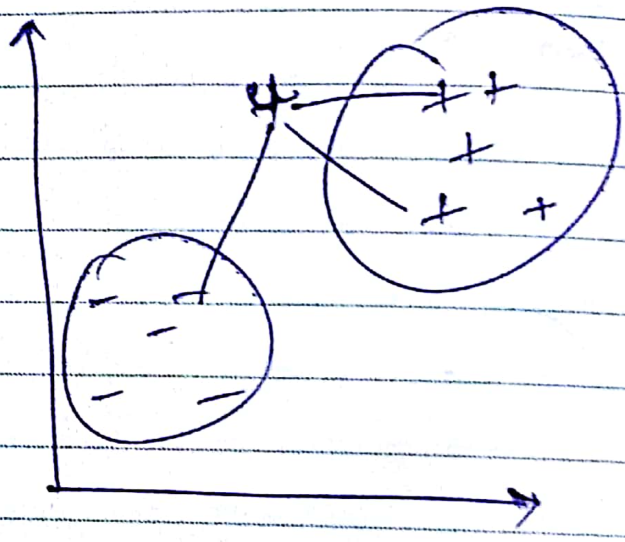
KNN

find euclidean dist between new features & existing ones.
The nearest neighbor' cluster is the ~~th~~ cluster of the new feature

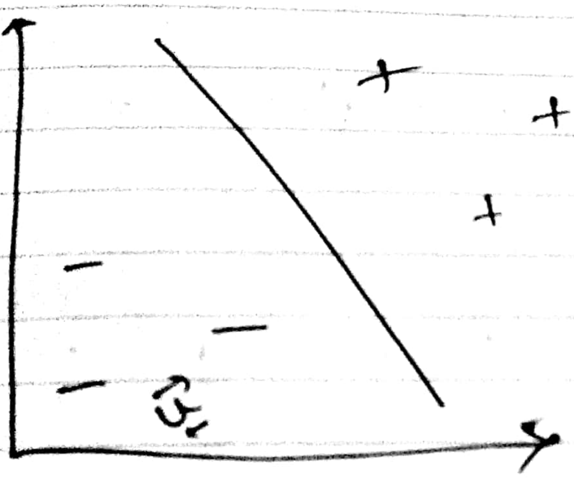


K=3
(find 3 nearest neighbors)

Since the nearest neighbors are +, +, -,

⇒ the new feature belongs to + cluster.

SVM (support vector machine)



$\vec{u} \cdot \vec{w} + b = y_i$ → if 'u' is on left, $y_i = -1$ else $y_i = 1$

$\vec{u}$ → unknown vector

$\vec{w}$ → vector that points ⊥ to the hyper plain

b → bias

Sunday 30

JANUARY						2011
S	M	T	W	T	F	S
30	31					1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29

$\vec{x}_{-sv} \rightarrow$ X Sub- minus support vector

$\vec{x}_{+sv} \rightarrow$ — " — plus — " —

$\vec{x}_{-sv} \cdot \vec{w} + b = -1 \rightarrow$ ①

$x_{+sv} \cdot \vec{w} + b = 1 \rightarrow$ ②

y_i (class)

$+ = +1$

$- = -1$

} ③

if y_i is '+' class $\rightarrow y_i = +1$

— " — '−' — " — $= -1$

+ class $\rightarrow$ $\vec{x_i} \cdot \vec{w} + b = 1 \rightarrow (4)$
$y_i = 1$

− class $\rightarrow$ $\vec{x_i} \cdot \vec{w} + b = -1 \rightarrow (5)$
$y_i = -1$

multiply (4) & (5) by y_i

or $y_i(\vec{x_i} \cdot \vec{w} + b) = (1) y_i$

$y_i(\vec{x_i} \cdot \vec{w} + b) = (-1) y_i$

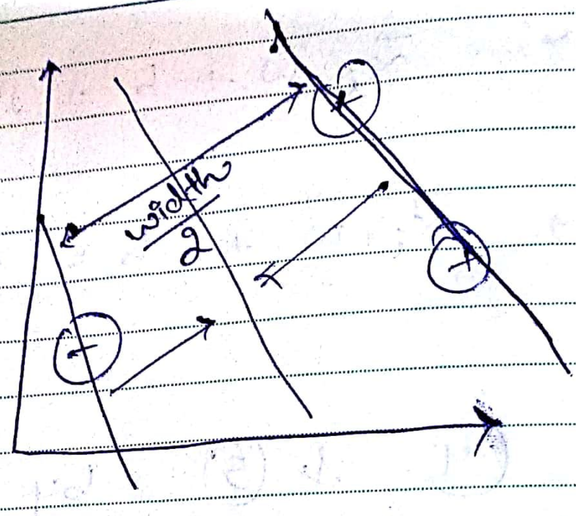
or $y_i(\vec{x_i} \cdot \vec{w} + b) = 1 \cdot 1 \quad (\because y_i = 1)$

$y_i(\vec{x_i} \cdot w + b) = (-1)(-1) \quad (\because y_i = -1)$ &

or (4) & (5) becomes (taking RHS to LHS)

or $y_i(\vec{x_i} \cdot \vec{w} + b) - 1 = 0$

$$\boxed{y_i(\vec{x_i} \cdot \vec{w} + b) - 1 = 0} \leftarrow (a)$$



$$\boxed{\text{width} = \left(\vec{X}_+ - \vec{X}_-\right) \cdot \frac{\vec{w}}{\|\vec{w}\|}} \rightarrow (1)$$

maximize

$\because \; X_+ = X_- = y_i\left(\vec{X}_i \cdot \vec{w} + b\right) = 1 \rightarrow (11)$

~~putting (11) in (1)~~

$\Rightarrow$ for $y_i = +1 \Rightarrow \vec{x}_+ \cdot \vec{w} = 1 - b$ (taking -1 & b to RHS)

for $y_i = -1 \Rightarrow \vec{x}_- \cdot \vec{w} = \cancel{(1+b)}$ (taking -1 & -b to RHS)

$= -(1-b)$

or ~~width~~ $\cancel{= [y_i(\vec{x}_i \cdot \vec{w} + b)] - 1}$

or width $= \cancel{\left\{ [y_i(\vec{x}_i \cdot \vec{w} + b) - 1] + [y_i(\vec{x}_i \cdot \vec{w} + b) - 1] \right\}} \cdot \frac{\vec{w}}{\|\vec{w}\|}$

width = ~~or~~ $\left\{ \cancel{[1(\vec{x}_i \cdot \vec{w} + b) - 1] + [-[1(\vec{x}_i \cdot \vec{w}) + b - 1]]} \right\} \cdot \frac{\vec{w}}{\|\vec{w}\|}$

~~or width~~ $= \cancel{\left[\vec{x}_i \cdot \vec{w} + b - 1 - [-\vec{x}_i \cdot \vec{w} - b] - 1 \right]} \cdot \frac{\vec{w}}{\|\vec{w}\|}$

or ~~width~~ $= \cancel{\frac{\vec{x}_+ \cdot \vec{w} - \vec{x}_- \cdot \vec{w}}{\|\vec{w}\|}}$

$= \cancel{1 - b - (1 - b}$

(for $y_i = 1$) $\vec{x}_+ \cdot \vec{w} = \cancel{y}\ 1 - b$

③

(for $y_i = -1$) $\vec{x}_- \cdot \vec{w} = \cancel{b+b}\ -(1+b)$

~~put ③ in ②~~

Expanding ①

$$\frac{\vec{x}_+ \cdot \vec{w} - \vec{x}_- \cdot \vec{w}}{\|\vec{w}\|} = \text{width} \rightarrow ④$$

⇒ put ③ in ④

$$\Rightarrow \cancel{\frac{1-b-[-(1+b)]}{\|w\|} = \text{width}}$$

$$\text{width} = \cancel{\frac{1-b+1-b}{\|w\|}}$$

width =

width

max

⇒

⇒ m

max $\frac{1}{\|w\|}$

∴ minim
will

$$\text{width} = \cancel{\frac{1-b + [\cancel{b-(1+b)}]}{\|w\|}}$$

$$= \frac{1-b+(1+b)}{\|w\|} \quad \left(\because \vec{x}\cdot\vec{w} = 1+b \text{ from eq } (a),\ \because y_i = -1\right)$$

$$\boxed{\text{width} = \frac{2}{\|w\|}}$$

maximise width

$\Rightarrow$ maximize $\rightarrow \dfrac{1}{\|w\|}$ (dropping constant (2))

$\Rightarrow$ minimize $\|w\|$

$$\max \frac{1}{\|w\|} \rightarrow \min \|w\| \rightarrow \boxed{\min \frac{1}{2}\|\vec{w}\|^2}$$

mathematically convinient

$\because$ minimizing $\frac{1}{2}\|w\|^2$ will maximize $\frac{1}{\|\vec{w}\|}$

$$L(w,b) = \frac{1}{2}\|\bar{w}\|^2 - \sum \alpha_i \left[y_i(\bar{w}.\bar{x}_i + b) - 1 \right] \rightarrow (b)$$

minimize | maximize

differentiating L wrt to all variables

$$\frac{dL}{dw} = \bar{w} - \sum \alpha_i y_i x_i = 0$$

$$\boxed{\Rightarrow \bar{w} = \sum \alpha_i y_i x_i} \rightarrow (c)$$

$$\frac{dL}{db} = -\sum \alpha_i y_i = 0$$

$$\boxed{\Rightarrow \sum \alpha_i y_i = 0} \rightarrow (d)$$

put (c) in (b)

// P.T.O

we're finding what this maximization depends on these sample vectors.

$$L = \frac{1}{2}\left[\left(\sum \alpha_i y_i \cdot x_i\right)\left(\sum \alpha_j y_j x_j\right)\right] -$$

$$\sum \alpha_i y_i x_i \cdot \left[\sum \alpha_j y_j x_j\right] - \underbrace{\sum \alpha_i y_i \cdot b}_{=0 \text{ (from d)}}$$

$$+ \sum \alpha_i$$

$$\boxed{L = \sum \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i \cdot y_j x_i \cdot x_j}$$

maximum

optimization depends only on (dot) product of pairs of samples

i.e., $(x_i \cdot x_j)$

$\sum \alpha_i y_i \underbrace{\vec{x_i} \cdot \vec{u}} + b \geq 0$ then +ve

dependent on dot product of 2 vectors

→ But for samples that are not linearly seperable,

then we need a transformation ~~to~~ that will take us from the space we're in to the space where thing are more convinient.

We'll that transformation → $\phi(\bar{x})$

~~$\phi(\bar{x}) \cdot \phi(\bar{y})$ to max~~

$\phi(x_i) \cdot \phi(x_j)$ to MAX

$\phi(x_i) \cdot \phi(u)$

$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$

↓

kernel function

Some of the popular kernels

(i) Linear kernels

$(\vec{u} \cdot \vec{v} + 1)^n$

(ii) $e^{-\frac{\|x_i - x_j\|}{\sigma}}$ Radial Basis Kernel uses exponential