⇒ SVM

1) Maximal Margin classifier (Constrained optimization problem)

$$\underset{\beta_0, \beta_1, \beta_2, \ldots \beta_p, M}{\text{Maximize}} \; M \quad \text{—— ①}$$

Subject to

$$\sum_{j=1}^{p} \beta_j^2 = 1 \quad \text{—— ②}$$

* similar to l2 regularization

$$y^i(\beta_0 + \beta_1 x_1 + \beta_2 x_2 \cdots \beta_p x_p) \geq M \quad \text{— ③}$$

for all i = 1 to n

$$d = \frac{|\beta_0 + \beta_1 x_1 + \beta_2 x_2|}{\sqrt{\beta_1^2 + \beta_2^2}}$$

eg: Three lines (hyperplane)

$2x_1 + 3x_2 - 5 = 0$

$-x_1 + 4x_2 + 7 = 0$

$5x_1 - 12x_2 + 10 = 0$

Dataset

x_1	x_2	y
3	4	1
2	3	1
1	-1	-1
-2	1	-1

$$\cancel{M1 = -5 + 2(3) + 3(4)}$$
$$\cancel{= -5 + 6 + 12}$$
$$\cancel{= 13}$$

$$M1 = \frac{|-5 + 2(3) + 3(4)|}{\sqrt{2^2 + 3^2}}$$
$$= \frac{13}{\sqrt{13}}$$

$$M2 = \frac{-5 + 2(2) + 3(3)}{\sqrt{4+9}}$$
$$= \frac{-5 + 4 + 9}{\sqrt{13}}$$
$$= \frac{8}{\sqrt{13}}$$

$$M3 = \frac{|-5 + 2(1) + 3(-1)|}{\sqrt{13}}$$
$$= \frac{|-5 + 2 - 3|}{\sqrt{13}}$$
$$= \frac{+6}{\sqrt{13}}$$

$$M4 = \frac{-5 + 2(-2) + 3(1)}{\sqrt{13}}$$
$$= -5 - 4 + 3$$
$$= \frac{+6}{\sqrt{13}}$$

$$\min = 6/\sqrt{13} = 1.664$$

=> Hyperplane - 2

$$M1 = \frac{+7 - 1(3) + 4(4)}{\sqrt{1^2 + 4^2}}$$
$$= \frac{7 - 3 + 16}{\sqrt{17}}$$
$$= \frac{20}{\sqrt{17}}$$

$$M2 = \frac{+7 - 1(2) + 4(3)}{\sqrt{17}}$$
$$= \frac{7 - 2 + 12}{\sqrt{17}}$$
$$= \frac{17}{\sqrt{17}}$$

$$M3 = \frac{+7 - 1(1) + \overset{4}{\cancel{3}}(-1)}{\sqrt{17}}$$
$$= \frac{+7 - 1 - \overset{4}{\cancel{3}}}{\sqrt{17}} = \frac{2}{\sqrt{17}}$$

$$M4 = \frac{+7 - 1(-2) \; \overset{4}{\cancel{3}}(1)}{\sqrt{17}}$$
$$= \frac{7 + 2 + 4}{\sqrt{17}} = \frac{13}{\sqrt{17}}$$

$$\min = \frac{2}{\sqrt{17}} = 0.485$$

Hyperplane - 3

$$d_1 = \frac{10 + 5(3) - 12(4)}{\sqrt{5^2 + 12^2}}$$

$$= \frac{10 + 15 - 48}{\sqrt{169}}$$

$$= \frac{23}{13}$$

$$d_2 = \frac{10 + 5(2) - 12(3)}{13}$$

$$= \frac{10 + 10 - 36}{13}$$

$$= \frac{16}{13}$$

$$d_3 = \frac{10 + 5(1) - 12(-1)}{13}$$

$$= \frac{10 + 5 + 12}{13}$$

$$= \frac{27}{13}$$

$$d_4 = \frac{10 + 5(-2) - 12(1)}{13}$$

$$= \frac{10 - 10 - 12}{13}$$

$$= \frac{12}{13}$$

$$\min = \frac{12}{13} = 0.92$$

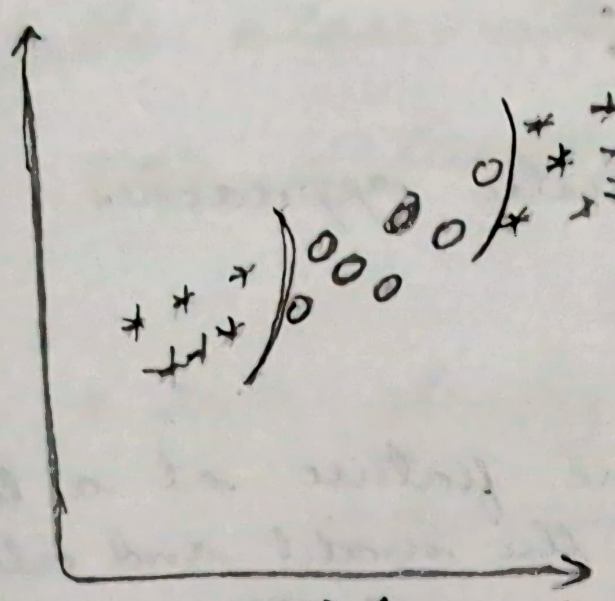
$$M = \max\left\{\frac{6}{\sqrt{13}}, \frac{2}{\sqrt{17}}, \frac{12}{13}\right\}$$

$$= \frac{6}{\sqrt{13}}$$

$$= 1.664$$

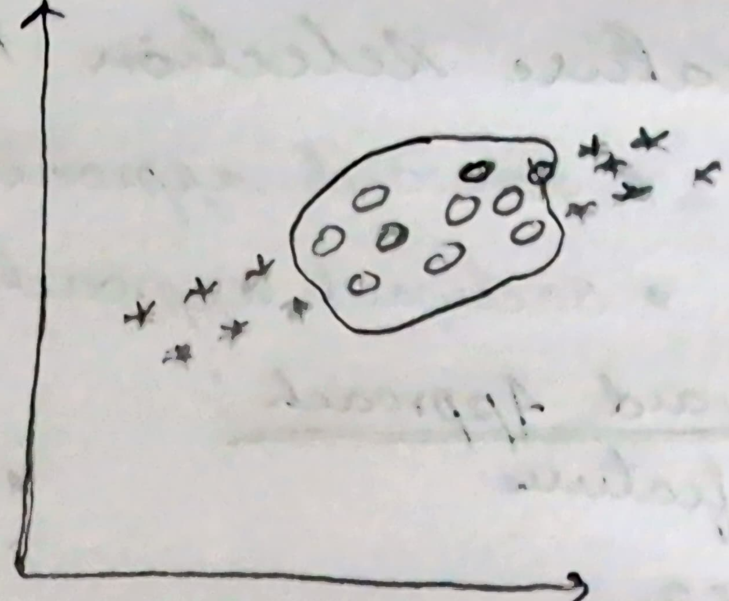
Revise

1) Projection
2) Norm of vector
3) Vector addition
4) Orthogonality
5) Subspaces



Polynomial Kernel

Tries to fit in curves in the model instead of a linear line.



RBF Kernel

Creates a very distinct boundary around one type of dataset.

One (vs) Rest

Eg:- $y = \{Red, Green, Blue\}$

Model 1 :- Red vs others

Model 2 :- Green vs others

Model 3 :- Blue vs others

x_1	x_2	y
1	0	Red
2	1	Green
3	5	Blue
4	2	Red

will be labelled as others

* Each model will give us a score

Model 1 = 0.8

Model 2 = 0.9

Model 3 = 0.7

* Then we select model with highest score.
* Probability score of the test sample will correspond to the positive sample in the model.
* ie, when we pass a test data to the model, if the sigmoid function value of the model is equal to value of the test data sigmoid function value, then the test data label will be same as model positive label

⇒ Evaluation Metrics Example :-

Observation	Actual label	Predicted Score	Predicted label
1	1	0.85	1
2	0	0.60	1 * FP
3	1	0.70	1
4	1	0.40	0 * FN
5	0	0.55	1 * FP
6	1	0.50	1
7	0	0.65	1 * FP
8	0	0.35	0
9	1	0.60	1
10	0	0.20	0

Predict

1) Accuracy
2) Precision
3) Sensitivity / True positive or Recall
4) Specificity
5) False positive
6) F1 score.
7) ROC plot
8) AUC.

	Original (+)	Original (-)
Predicted (+ve)	4	3
Predicted (-)	1	2

1) Accuracy : $\cancel{\dfrac{\text{Total (+) + Total (-)}}{\text{Total}}}$ $\dfrac{\text{True (+) + True (-)}}{\text{Total}}$

$= \dfrac{4+2}{10} = \dfrac{6}{10}$

2) Precision = $\dfrac{\text{True (+)}}{\text{True (+) + False (+)}} = \dfrac{4}{4+3} = \dfrac{4}{7}$

3) Sensitivity = $\dfrac{\text{True (+)}}{\text{True (+) + False (-)}} = \dfrac{4}{4+1} = \dfrac{4}{5} = \boxed{0.80}$

4) specificity = Negative recall = $\dfrac{\text{True (-)}}{\text{True (-) + False (+)}} = \dfrac{2}{5}$

5) False positive = $\dfrac{3}{3+2} = \dfrac{3}{5} = \boxed{0.6}$

6) F1-score = $2\left[\dfrac{\text{precision . recall}}{\text{precision + recall}}\right] = \dfrac{1}{\dfrac{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}}{2}}$

$= 2\left[\dfrac{\frac{4}{7} \times \frac{4}{5}}{4/7 + 4/5}\right] = 2\left[\dfrac{0.57 \times 0.80}{0.57 + 0.80}\right]$

$= \left(\dfrac{0.456}{1.37}\right) \times 2$

$= 0.333 \times 2 = 0.666$

FPR = $\dfrac{0}{5}$ =

7) ROC - plot

TPR = $\dfrac{2}{3}$ =

Threshold = [0.7, 0.5, 0.4, 0.2]

Threshold = 0.7

1	0.85	1
0	0.60	0
1	0.70	1
1	0.40	0
0	0.55	0
1	0.50	0
0	0.65	0
0	0.35	0
1	0.60	0
0	0.20	0

	original (+)	original (-)
Predicted (+)	2	0
Predicted (-)	3	5

Threshold = 0.4

Actual	Score	Predicted
1	0.85	1
0	0.60	1
1	0.70	1
1	0.40	1
0	0.55	1
1	0.50	1
0	0.65	1
0	0.35	0
1	0.60	1
0	0.20	0

	O(+)	O(-)
P(+)	5	3
P(-)	0	2

$FPR = \frac{3}{5} = \underline{\underline{0.6}}$ $\quad TPR = \frac{5}{5} = \underline{\underline{1}}$

Threshold = 0.2

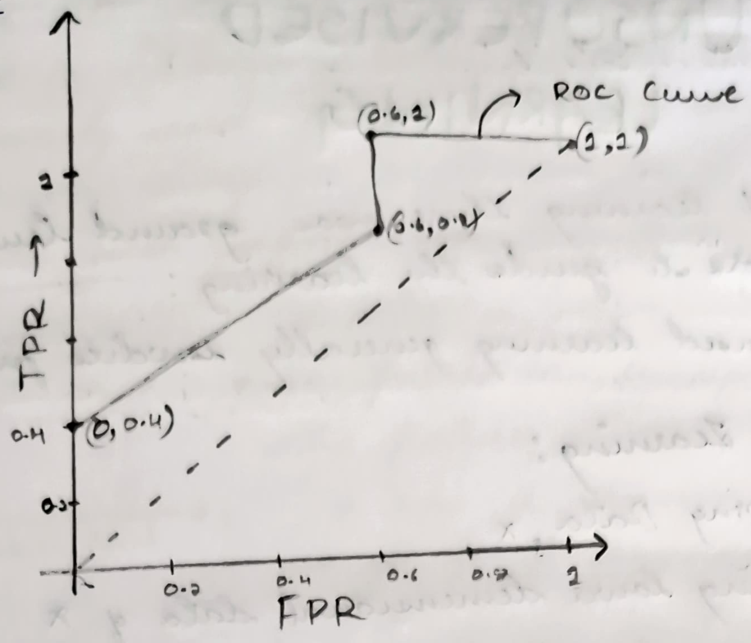
Actual	Score	Predicted
1	0.85	1
0	0.60	1
1	0.70	1
1	0.40	1
0	0.55	1
1	0.50	1
0	0.65	1
0	0.35	1
1	0.60	1
0	0.20	1

	O(+)	O(-)
P(+)	5	5
P(-)	0	0

$FPR = \frac{5}{5} = \underline{1}$ $\quad TPR = \frac{5}{5} = \underline{\underline{1}}$

Threshold	FPR	TPR
0.7	0	0.4
0.5	0.6	0.8
0.4	0.6	1
0.2	1	1

⇒ ROC Curve



⇒ AUC (AUC > 0.5)

$$AUC = \sum_{i=1}^{4} \frac{(FPR_i - FPR_{i-1}) * (TPR_i + TPR_{i-1})}{2}$$

$$= \cancel{\frac{(0.6 - 0) * (0.8 + 0.4)}{2}} + \cancel{(0.6 - 0.6)}$$

$$= \frac{(1 - 0.6) * (1 + 1)}{2} + \frac{(0.6 - 0.6) * (1 + 0.8)}{2}$$

$$+ \frac{(0.6 - 0) * (0.8 + 0.4)}{2}$$

$$= \frac{(0.4) * \cancel{2}}{\cancel{2}} + 0 + \frac{0.6 * 1.2}{2}$$

$$= 0.4 + 0 + 0.36$$

$$= 0.76$$

Eg:-

$$b = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \qquad a = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$a^T = \begin{bmatrix} 1 & 2 \end{bmatrix}$$

$$a \times a^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{2\times1} \times \begin{bmatrix} 1 & 2 \end{bmatrix}_{1\times2} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

$$p = \frac{a a^T b}{a^T a}$$

$$a^T \times a = \begin{bmatrix} 1 & 2 \end{bmatrix}_{1\times2} \times \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{2\times1} = 1+4 = 5$$

$$= \frac{\underset{2\times2}{\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}} \underset{2\times1}{\begin{bmatrix} 3 \\ 4 \end{bmatrix}}}{5}$$

$2\times2 \qquad 2\times1$

$$= \frac{\begin{bmatrix} 11 \\ 22 \end{bmatrix}}{5}$$

2×1

$$= \begin{bmatrix} 11 \\ 22 \end{bmatrix} * \frac{1}{5}$$

$$\begin{bmatrix} 5 \\ 10 \end{bmatrix} \times 0.5$$

$$= \begin{bmatrix} 2.2 \\ 4.4 \end{bmatrix}$$

$$\begin{bmatrix} 5 \times .5 \\ 10 \times .5 \end{bmatrix}$$

Example:

$$X = \begin{bmatrix} 2 & 100 \\ 4 & 200 \\ 6 & 300 \end{bmatrix}_{3\times 2} \xrightarrow{\text{apply PCA}} Z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}_{3\times 1}$$

$$\mathbb{R}^2 \longrightarrow \mathbb{R}^1$$

① Standardize each columns

Column - 1

$\mu_1 = 4$

$$\sigma_1 = \sqrt{\frac{\Sigma (x_i - \mu)^2}{N-1}}$$

$$= \sqrt{\frac{(2-4)^2 + (4-4)^2 + (6-4)^2}{3}}$$

$$= \sqrt{\frac{2^2 + 0 + 2^2}{3-1}}$$

$$= \sqrt{\frac{4+4}{3-1}}$$

$$= \sqrt{\frac{8}{3-1}}$$

$$= \sqrt{4}$$

$$= 2$$

$\mu_2 = 200$

$$\sigma_2 = \sqrt{\frac{(100-200)^2 + (200-200)^2 + (300-200)^2}{2}}$$

$$= \sqrt{\frac{(100)^2 + (100)^2}{2}}$$

$$= \sqrt{\frac{20000}{2}}$$

$$= \sqrt{10,000}$$

$\sigma_2 = 100$

$$X\_std = \begin{bmatrix} -1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}$$

Column - 1

$\frac{2-4}{2} = -1$, $\frac{4-4}{2} = 0$

$\frac{6-4}{2} = 1$

Colum - 2

$\frac{100-200}{100} = -1$ $\frac{200-200}{100} = 0$

$\frac{300-200}{100} = 1$

② Compute co-variance matrix :-

$$\frac{1}{n-1} X_{std}^T X_{std} = [\quad]_{2\times 2}$$

$$\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} = \frac{\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}}{2} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

③ Compute eigen values and eigen vectors of the co-variance matrix

④ Decide the needed % of variance explained and choose 'k' principle components.

⑤ Transform X into new space of principle components

⑥ Use this to train a model.

3) $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ $\lambda I = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$

$$|A - \lambda I| = \begin{vmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{vmatrix} = 0$$

$$0 = (1-\lambda)(1-\lambda) - (1\times 1)$$

$$= \cancel{1} - \lambda - \lambda + \lambda^2 - \cancel{1}$$

$$= \lambda^2 - 2\lambda$$

$$2\cancel{\lambda} = \lambda^{\cancel{2}}$$

$\lambda = 2$, $\lambda = 0$

$\lambda = 2$

eigen vector

$$\begin{bmatrix} 1-2 & 1 \\ 1 & 1-2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0$$

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0$$

$-x + y = 0$

$v_1 = \begin{bmatrix} t \\ t \end{bmatrix}$

$x = y$

$v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$v_1 = t\hat{i} + t\hat{j}$

$|v_1| = \sqrt{t^2 + t^2}$

$1 = \sqrt{2t^2}$

$1^2 = 2t^2$

$\frac{1}{2} = t^2$

$t = \frac{1}{\sqrt{2}}$

$|v_1| = 1$

$\sqrt{t^2 + t^2} = 1$

$\lambda = 0$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$x + y = 0$

$x = -y$

$v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \Rightarrow \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

(4) variance $\Rightarrow$ $\geq 95\%$

(5) $Z_{pca} = X_{std} \times V_2$

$$= \begin{bmatrix} -1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}_{3\times 2} \times \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}_{2\times 1}$$

eigen vector with

$$= \begin{bmatrix} -1/\sqrt{2} - \frac{1}{\sqrt{2}} \\ 0 \\ 1/\sqrt{2} + 1/\sqrt{2} \end{bmatrix}_{3\times 1}$$

$$= \begin{bmatrix} -2/\sqrt{2} \\ 0 \\ 2/\sqrt{2} \end{bmatrix}_{3\times 1}$$

Eg:-

	x_1	x_2
1	1	1
2	1	2
3	2	2
4	8	8
5	8	9
6)	9	8

$K = 2$

1) $C_1 = \{2, 3, 4, 6\}$

$C_2 = \{1, 5\}$

① First iteration

C_1 :- $[1,2]$, $[2,2]$, $[8,8]$, $[9,8]$

C_2 :- $[1,1]$, $[8,9]$

a) Compute Centroid

Centroid for $C_1 = [5,5]$

Centroid for $C_2 = [4.5, 5]$

$w(k)$ for sample ②

$$①\ C_1 = \sqrt{(5-1)^2 + (5-2)^2} = \sqrt{25} = 5$$

$$C_2 = \sqrt{(1-4.5)^2 + (5-2)^2} = \sqrt{21.25} = \approx 4.5$$

∴ sample ② belongs to cluster 2 or C_2

$w(k)$ for sample ①

$$C_1 = \sqrt{(5-1)^2 + (5-1)^2} = \sqrt{32}$$

$$C_2 = \sqrt{(4.5-1)^2 + (5-1)^2} = \sqrt{28.25}$$

∴ sample ① belongs to cluster 2 or C_2

w(k) for sample ③

$C_1 = \sqrt{(5-2)^2 + (5-2)^2} = \sqrt{18}$

$C_2 = \sqrt{(4.5-2)^2 + 3^2} = \sqrt{15.25}$

∴ sample ③ belongs to C_2

w(k) for sample ④

$C_1 = \sqrt{(8-5)^2 + (8-5)^2} = \sqrt{18}$

$C_2 = \sqrt{(8-4.5)^2 + (8-5)^2} = \sqrt{21.25}$

sample ④ belongs to C_1

w(k) sample ⑤

$C_1 = \sqrt{(8-5)^2 + (9-5)^2} = \sqrt{25}$

$C_2 = \sqrt{(8-4.5)^2 + (9-5)^2} = \sqrt{28.25}$

sample ⑤ belongs to cluster 1 or C_1

w(k) for sample ⑥

$C_1 = \sqrt{(9-5)^2 + (8-5)^2} = \sqrt{25}$

$C_2 = \sqrt{(9-4.5)^2 + (8-5)^2} = \sqrt{29.25}$

sample ⑥ belongs to C_1

$C_1 = \{4, 5, 6\}$

$C_2 = \{1, 2, 3\}$

Compute / Do the same steps again & again until centroid remains same