SGD:-

Given: $\theta_0 = \theta_1 = \theta_2 = 0$, $\alpha = 0.1$, $x_0 = 1$.

For 1st iteration:- $h_\theta(x)^{(1)} = \theta_0 x_0^{(1)} + \theta_1 x_1^{(1)} + \theta_2 x_2^{(1)}$

x_1	x_2	y
1	2	3
2	1	4
3	3	5

$h_\theta(x) = 0\times1 + 0\times1 + 0\times2 = 0$.

We know,

$$\theta_j = \theta_j + \alpha \sum_{i=1}^{m} \left[y^{(i)} - h_\theta(x)^{(i)}\right] x_j^{(i)}$$

$$\theta_j = \theta_j - \alpha \sum_{i=1}^{m} \left[h_\theta(x)^{(i)} - y^{(i)}\right] x_j^{(i)}$$

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{3}\left[h_\theta(x)^{(i)} - y^{(i)}\right]^2 = \frac{1}{2}\times(0-3)^2 = 4.5$$

$\theta_0 = 0 - 0.1[0-3]\times 1$

$= -0.1(-3) = 0.3$

$\theta_1 = 0 - 0.1(-3)\times 1$

$= 0.3$

$\theta_2 = 0 - 0.1(-3)\times 2$

$= 0.6$

$\therefore$ For 1st iteration:

$\theta_0 = 0.3$, $\theta_1 = 0.3$, $\theta_2 = 0.6$

$J = 4.5$

For 2nd iteration:

$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2$

$= 0.3\times1 + 0.3\times2 + 0.6\times1$

$= 0.3 + 0.6 + 0.6 = 1.5$

$\theta_0 = \theta_0 - 0.1[1.5-4]\times x_0$

$= 0.3 - (0.1\times -2.5\times1)$

$= 0.3 + 0.25 = 0.55$

$\theta_1 = \theta_1 - 0.1[1.5-4]\times x_1$

$= 0.3 - (0.1\times -0.25\times 2)$

$= 0.3 + 0.05 = 0.35$

$\theta_2 = \theta_2 - 0.1[1.5-4]\times 1$

$= 0.6 - (0.1\times -0.25)\times 1$

$= 0.6 + 0.025 = 0.625$

$J(\theta) = \frac{1}{2}\times[1.5-4]^2$

$= 3.125$

$\therefore$ For 2nd iteration:

$\theta_0 = 0.55$, $\theta_1 = 0.35$, $\theta_2 = 0.625$

$J = 3.125$

For 3rd iteration:

$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2$

$= 0.55\times1 + 0.35\times3 + 0.625\times3$

$= 3.475$

$\theta_0 = 0.55 - 0.1(3.475-5)\times1$

$= 0.55 - 0.1\times(-1.525)$

$= 0.7025$

$\theta_1 = 0.35 - 0.1(3.475-5)\times3$

$= 0.35 + 0.4575$

$= 0.8075$

$\theta_2 = 0.625 - 0.1(-1.525)\times3$

$= 0.625 + 0.4575$

$= 1.08$

$J = \frac{1}{2}\times(3.475-5)^2$

$= 1.1628$

$\therefore$ For 3rd iteration:

$\theta_0 = 0.7025$, $\theta_1 = 0.8075$,

$\theta_2 = 1.08$, $J = 1.1628$

GD :-

apsara

Date: 16.01.25

Steps :-

① compute predictions.

$h_\theta(x^{(1)}) = 0$

$h_\theta(x^{(2)}) = 0$

$h_\theta(x^{(3)}) = 0$

② compute gradients.

ⓐ compute errors/residuals

$E^{(1)} = h_\theta(x^{(1)}) - y^{(1)}$

$E^{(2)} = h_\theta(x^{(2)}) - y^{(2)}$

$E^{(3)} = h_\theta(x^{(3)}) - y^{(2)}$

ⓑ calculate gradient.

$\frac{\delta J}{\delta \theta_0} = \sum_{i=1}^{m} (h_\theta(x)^{(i)} - y^{(i)}) x_0^{(i)}$

$= -3\times1 + (-4)\times1 + (-5)\times1$

$= -12.$

$\frac{\delta J}{\delta \theta_1} = (-3)\times1 + (-4)\times2 + (-5)\times3 = -26.$

$\frac{\delta J}{\delta \theta_2} = (-3)\times2 + (-4)\times1 + (-5)\times3$

$= -25.$

③ update parameters:

$\theta_0 = \theta_0 - \alpha . \frac{dJ}{\delta \theta_0} = 0 - 0.1\times(-12) = 1.2$

$\theta_1 = \theta_1 - \alpha \frac{\delta J}{\delta \theta_1} = 0 - 0.1\times(-26) = 2.6$

$\theta_2 = \theta_2 - \alpha \frac{\delta J}{\delta \theta_2} = 0 - 0.1\times(-25) = 2.5$

$J(\theta) = \frac{1}{2}\times(-12)^2 =$

Initially, $\theta = \begin{bmatrix}\theta_0\\\theta_1\\\theta_2\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$

$\boxed{J = 94.95}$

After 1st iteration.

$\theta = \begin{bmatrix}1.2\\2.6\\2.5\end{bmatrix}$

After 2nd iteration.

$\theta = \begin{bmatrix}-1.02\\-2.41\\-2.6\end{bmatrix}$

$J = 25678.56$

As J has become very large, so now reduce $\alpha = 0.01$

when $\alpha = 0.01$.

After 1st iteration :

$\theta_0 = 0.19$, $\theta_1 = 0.43$, $\theta_2 = 0.41$

$J = 5.4$.

After 2nd iteration :

$\theta_0 = 0.25$, $\theta_1 = 0.55$, $\theta_2 = 0.53$.

$J = 3$.

Eg:- Compute a 3-fold cross validation for the given dataset along with accuracy & standard deviation.

	x_1	x_2	y
#1 ←	2	-1	1
#2 ←	0.5	1.2	0
#3 ←	1	2	1
#4 ←	-3	-2	1
#5 ←	4	0.1	0

$F_1 : \{\theta_1, \theta_2\} = \{-1.8, 2.8\}$

$F_2 : \{\theta_1, \theta_2\} = \{2.1, 3.1\}$

$F_3 : \{\theta_1, \theta_2\} = \{1.9, 4\}$

Fold 1: Train set = { }
Test set = { }
Accuracy:

Fold 2: Train set = { }
Test " = { }
accuracy.

Fold 3: Train = { }
Test = { }
accuracy.

overall:
ACC = Avg (acc1, acc2, acc3)
SD = SD (acc1, acc2, acc3)

Page No.:

Fold 1:

Train set: {#1, #2, #3, #4}

Test set: {#5}

~~Fold 2:~~

~~Train set: {#1, #3, #5}~~ ~~Train {~~

~~Test set: {#2, #4}~~

Fold 2:

Train set: {#3, #4, #5}

Test set: {#1, #2}

Fold 3:

Train set: {#1, #2, #5}

Test set: {#3, #4}

As θ values are given so only do the calculation for test sets.

Fold 1

For #5 : $x = [4 \quad 0.1]$

$\theta^T x = [-1.8 \quad 2.8][4 \quad 0.1]$

$= (-1.8 \times 4) + (2.8 \times 0.1)$

$= -7.2 + 0.28$

$= -6.92$

$g(z) = \frac{1}{1+e^{-6.92}}$

$= \frac{1}{1013.3}$

$= 0.00098$

$\vec{y} = \cancel{0 - 0.00098}$

$\cancel{= -0.00098}$

As $y_{pred} = 0.00098$ & $y_{given} = 0$ which are almost same so give val = 1.

$\therefore$ Accuracy $= \frac{1}{1} \times 100 = 100\%$.

Fold 2

For #1 : $\theta^T = [2.1 \quad 3.1]$

$x = [2 \quad -1]$

$\theta^T x = [2.1 \quad 3.1][2 \quad -1]$

$= 2.1 \times 2 + 3.1(-1)$

$= 4.2 - 3.1$

$= 1.1$

$g(z) = \frac{1}{1+e^{-1.1}}$

$= \frac{1}{1+0.332} = \frac{1}{1.332}$

$= 0.750$.

$g(z) \approx y_{given}$ $\therefore$ val = 1

For #2 : $x = [0.5 \quad 0.2]$

$\theta^T x = [2.1 \quad 3.1][0.5 \quad 0.2]$

$= 2.1 \times 0.5 + 3.1 \times 0.2$

$= 1.05 + 0.62$

$= 1.67$

$g(z) = \frac{1}{1+e^{-1.67}} = \frac{1}{1+0.188}$

$= \frac{1}{1.188} = 0.841$

$g(z) \neq y_{given} = 0$, $\therefore$ val = 0.

Accuracy $= \frac{1+0}{2} \times 100 = 50\%$.

Fold 3

For #3 : $\theta^T = [1.9 \quad 4]$

$x = [1 \quad 2]$

$\theta^T x = [1.9 \quad 4][1 \quad 2]$

$= 1.9 \times 1 + 4 \times 2$

$= 1.9 + 8$

$= 9.9$

$g(z) = \frac{1}{1+e^{-9.9}} = \frac{1}{1+0.00005}$

$= \frac{1}{1.00005} = 0.9999$

$g(z) \approx y_{given} = 1$ $\therefore$ val = 1.

For #4 : $\theta^T = [1.9 \quad 4]$

$x = [-3 \quad -2]$

$\theta^T x = [1.9 \quad 4][-3 \quad -2]$

$= -3 \times 1.9 + (-2) \times 4$

$= -5.7 - 8 = -13.7$.

$g(z) = \frac{1}{1+e^{13.7}} = \frac{1}{1+890911.1}$

$= 0.000001$

$g(z) \neq y_{given} = 1$ $\therefore$ val = 0.

Q. Dataset :-

	x_1	x_2	y
#1	1	2	1
#2	3	6	5
#3	5	1	2

$|I| = 3$ & $k = 2$., for x_1, $S = \{2, 4\}$, x_2, $S = \{1.5, 4\}$

$\boxed{\text{For } x_1 \text{ \& } 2}$

$I^+_{x_1,2} = \{x_1 \geq 2\} = \{2,3\}$

$I^-_{x_1,2} = \{x_1 < 2\} = \{1\}$

$\hat{y}^+ = \frac{5+2}{2} = \frac{7}{2} = 3.5$

$\hat{y}^- = 1.$

$E^+_{x_1,2} = (5-3.5)^2 + (2-3.5)^2 = 2.25 + 2.25 = 4.5$

$E^-_{x_1,2} = (1-1)^2 = 0.$

$E_{x_1,2} = 4.5 + 0 = \boxed{4.5}$

$\boxed{\text{For } x_1 \text{ \& } 4}$

$I^+_{x_1,4} = \{x_1 \geq 4\} = \{3\}$

$I^-_{x_1,4} = \{x_1 < 4\} = \{1,2\}.$

$\hat{y}^+ = 2.$

$\hat{y}^- = \frac{5+1}{2} = 3.$

$E^+_{x_1,4} = (2-2)^2 = 0$

$E^-_{x_1,4} = (1-3)^2 + (3-5)^2 = 4+4 = 8.$

$E_{x_1,4} = 8+0 = \boxed{8}$

$\boxed{\text{For } x_2, 1.5}$

$I^+_{x_2,1.5} = \{x_2 \geq 1.5\} = \{1,2,3\}$

$I^-_{(x_2,1.5)} = \{x_2 < 1.5\} = \{3\}.$

$\hat{y}^+ = \frac{1+5}{3} = \frac{6}{3} = 3$

$\hat{y}^- = 2.$

$E^+_{x_2,1.5} = (3-1)^2 + (3-5)^2 + \cancel{(2.66-2)^2}$

$= \cancel{2.75 + 5.47 + 0.43}\ 4+4$

$= \cancel{8.65} = 8$

$E^-_{x_2,1.5} = (2-2)^2 = 0.$

$E_{x_2,1.5} = \cancel{8.65 + 0 = 8.65}\ 8+0 = \boxed{8}$

$\boxed{\text{For } x_2, 4}$

$I^+_{x_2,4} = \{x_2 \geq 4\} = \{2\}.$

$I^-_{(x_2,4)} = \{x_2 < 4\} = \{1,3\}.$

$\hat{y}^+ = 5$

$\hat{y}^- = \frac{1+2}{2} = \frac{3}{2} = 1.5$

$E^+_{x_2,4} = (5-5)^2 = 0.$

$E^-_{(x_2,4)} = (1.5-1)^2 + (1.5-2)^2$

$= 0.25 + 0.25$

$= 0.5$

$E_{x_2,4} = 0.5 + 0 = \boxed{0.5}.$

Set $j = x_2$ & $s = 4$. = min $E_{x_2,4}$

Node = $\boxed{x_2 \geq 4}$

$\boxed{1.5}$ $\boxed{5} \rightarrow \hat{y}^+$

$\hat{y}^-$

Date: __.__.__

Page No.: _____

① Entropy → To measure homogenity.

$$-\sum_{i=1}^{K} P_i \log_2 (P_i)$$

⊛ IG = Information Gain.

eg:- p(+ve) = 6.

n(-ve) = 6.

$$H\left(\frac{p}{p+n}, \frac{n}{p+n}\right) = -\frac{p}{p+n}\log_2\frac{p}{p+n} - \frac{n}{p+n}\log_2\frac{p}{p+n}.$$

Generalized to $\sum_{i=1}^{K} - P_i \log_2 P_i$.

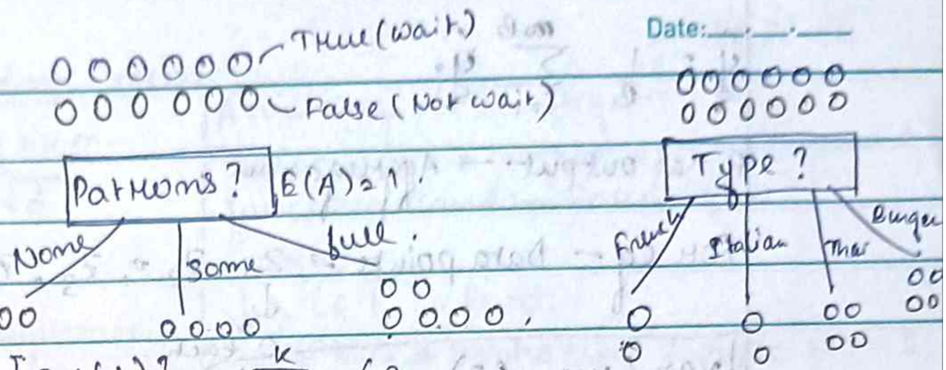
$$H\left(\frac{6}{12}, \frac{6}{12}\right) = -\frac{6}{12}\log_2\frac{6}{12} - \frac{6}{12}\log_2\frac{6}{12}.$$

$$= -\frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right).$$

$$= \cancel{-\frac{1}{2}\times(-0.301)} - \cancel{\frac{1}{2}\times(-0.301)}.$$

$$= -\frac{1}{2}\times(-1) - \frac{1}{2}\log(-1)$$

$$= \frac{1}{2}+\frac{1}{2} = 1$$



Min.

Expected entropy $[EH(A)] = \sum_{i=1}^{k} \left(\frac{P_i + n_i}{P+n}\right) \cdot H\left(\frac{P_i}{P_i+n_i}, \frac{n_i}{P_i+n_i}\right)$

$IG(A) = H\left(\frac{P}{P+n}, \frac{n}{P+n}\right) - EH(A)$

Max.

11/02/25

$H(\text{patrons}) = H\left(\frac{P}{P+n}, \frac{n}{P+n}\right) = H\left(\frac{6}{12}, \frac{6}{12}\right) = 1$.

$EH(\text{patrons}) = \sum_{i=1}^{3} -\frac{P_i + n_i}{P+n} \cdot H\left(\frac{P_i}{n_i+P_i}, \frac{n_i}{n_i+P_i}\right)$.

$= -\left[\frac{P_1 + n_1}{12} \cdot H\left(\frac{0}{2}, \frac{2}{2}\right) + \frac{P_2 + n_2}{12} H\left(\frac{4}{4}, \frac{0}{4}\right) + \frac{P_3+n_3}{12} H\left(\frac{2}{6}, \frac{4}{6}\right)\right.$

$= -\left[\frac{2}{12} H(0,1) + \frac{4}{12} H(1,0) + \frac{6}{12} H\left(\frac{1}{3}, \frac{2}{3}\right)\right]$

$= \left[\frac{2}{12}(-0\log_2 0 - 1\log_2 1) + \frac{4}{12}(-1\log_2 1 - 0\log_2 0) + \frac{6}{12}\left(-\frac{2}{6}\log_2 \frac{2}{6} - \frac{4}{6}\log_2 \frac{4}{6}\right)\right]$

$1.585 \times \frac{1}{3} + \frac{2}{3} \times 0.585$

$0.5283 + 0.39$

$= 0.455$.

Gini Index. same as IG.

$IG(\text{Patrons}) = H(\text{Patrons}) - EH(\text{patrons})$

$= 1 - 0.455 = 0.541$

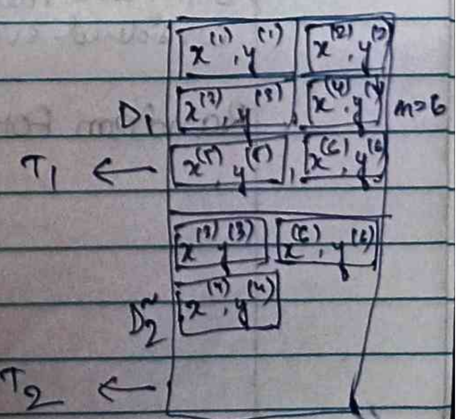
Ensemble Methods

Bagging / Boosting

Bagging / Bootstrap aggregation

creating a new dataset from existing dataset → Bootstrapping.

B datasets with replacements in B bags. containing $T_1, T_2, T_3 \ldots T_B$.



Inference :- $x \rightarrow \hat{y}$

$x \rightarrow T_1 \rightarrow \hat{y}_1$ $\quad x \rightarrow T_2 \rightarrow \hat{y}_2$ $\quad x \rightarrow T_3 \rightarrow \hat{y}_3$ $\quad x \rightarrow T_B \rightarrow \hat{y}_B$.

Gradient Boosting / Gradient Boosted Machine Decision. (GBM)

Date: 18.02.25

Gradient Boosting.

Boosted Regression Tree [AdaBoost Regression]

eg:2

x_1	x_2	y
1	2	4
2	3	5
3	4	6
4	5	7

$F(x) = 0 \; ; \; f_0(x) = 0$

For Iteration-1

$r = y$

x_1	x_2	r
1	2	4
2	3	5
3	4	6
4	5	7

a) Fit a model $f_1(x) = r$.

$\boxed{x_1 \geq 2.5}$

$\boxed{4.5}$ $\boxed{6.5}$

b) Update the model, $\lambda = 0.1$.

$F(x) = f_0(x) + \lambda f_1(x)$

c) Update the residuals.

$r^{(1)} = r^{(1)} - \lambda f_1(x^{(1)}) \rightarrow 4 - [0.1 \times 4.5] = 3.55$

$r^{(2)} = r^{(2)} - \lambda f_1(x^{(2)}) \rightarrow 5 - [0.1 \times 4.5] = 4.55$

$r^{(3)} = r^{(3)} - \lambda f_1(x^{(3)}) \rightarrow 6 - [0.1 \times 6.5] = 5.35$

$r^{(4)} = r^{(4)} - \lambda f_1(x^{(4)}) \rightarrow 7 - [0.1 \times 6.5] = 6.35$

For Iteration-2.

$$r = \begin{bmatrix} 3.55 \\ 4.55 \\ 5.35 \\ 6.35 \end{bmatrix}$$

x_1	x_2	r
1	2	3.55
2	3	4.55
3	4	5.35
4	5	6.35

a) Fit a model $\rightarrow$

$\boxed{x_1 \geq 2.5}$

Average (3.55, 4.55) $\leftarrow \boxed{4.05}$ $\boxed{5.85} \rightarrow$ Average (5.35, 6.35)

b) update the model,

$F(x) = f_0(x) + \lambda f_1(x) + \lambda f_2(x)$

c) update Residuals

$r^{(1)} = r^{(1)} - \lambda f_2(x^{(1)}) \rightarrow 3.55 - 0.1 \times (4.05) = 3.145$

$r^{(2)} = r^{(2)} - \lambda f_2(x^{(2)}) \rightarrow 4.55 - 0.1(4.05) = 4.145$

$r^{(3)} = r^{(3)} - \lambda f_2(x^{(2)}) \rightarrow 5.35 - 0.1(5.85) = 4.765$

$r^{(4)} = r^{(4)} - \lambda f_2(x^{(2)}) \rightarrow 6.35 - 0.1(5.85) = 5.765$

Date:

Misclassification error - Takes weight of all misclassified samples ~~Takes weights of all~~ & then we need to minimize it.

step b) E.

$$\epsilon = \frac{\sum_{i=1}^{N} w^{(i)} \, I\{y^{(i)} \neq f_b(x^{(i)}; \theta_b)\}}{\sum_{i=1}^{N} w^{(i)}}.$$

Note. $\epsilon = 0$, then all classification is right.

$\lambda_b = \frac{1}{2}\log\left(\frac{1-\epsilon}{\epsilon}\right)$. if $\epsilon = 0$; $w \downarrow \rightarrow$ minimizes mis calculations.

$\lambda_b \uparrow \rightarrow \downarrow$ inverse roles.

↳ gives high value to correctly predicted features.

step c) $w^{(m)} = w^{(m)} * e^{-\lambda_b} \rightarrow [0,1] < 1.$

$w^{(m)} = w^{(m)} * e^{\lambda_b} \rightarrow [\infty] > 1$

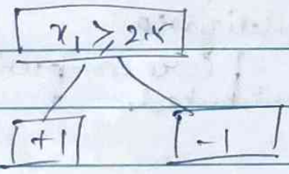
Example :-

x_1	x_2	y	w	$\hat{y}$
1	2	+1	1	
2	3	+1	1	
3	3	+1	1	
4	5	-1	1	
5	5	-1	1	
6	6	-1	1	

No. of weak learners = 5.

① Iteration 1:-

construct tree $f_1(x_1; 2.5)$.



$\epsilon = \frac{|X|}{6} = 0.1667.$

$\lambda_1 = \frac{1}{2}\log\left(\frac{1-0.1667}{0.1667}\right) = 0.8045.$

$w^3 = 1 \times e^{-0.8045}$ / $w^{\{1,2,4,5,6\}} = 1 \times e^{+0.8045}.$

$w^3 = 0.447$ / $w^{\{1,2,4,5,6\}} = 2.235.$

x_1	x_2	y	w	$\hat{y}_1$	w
1	2	+1	1	+1	2.235
2	3	+1	1	+1	2.235
3	3	+1	1	-1	0.447
4	5	-1	1	-1	2.235
5	5	-1	1	-1	2.235
6	6	-1	1	-1	2.235

Page No.:

samples along with test samples to compute y-pred.

11/03/25

Dataset :-

x_1	x_2	y
1	2	0
3	4	0

Here $n = 2$ & $\phi(x) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_1 x_2 \end{bmatrix}$ } $\phi(x) \in \mathbb{R}^3$

$\theta \in \mathbb{R}^3$

$\downarrow$

$[\theta_1 \ \theta_2 \ \theta_3]^T$

$x = [x_1 \ x_2]^T$

$x \in \mathbb{R}^2$

$\theta = \sum_{i=1}^{n} \beta_i \phi(x^{(i)})$ Find $\theta = \begin{bmatrix} \ \\ \ \end{bmatrix}_{3 \times 1}$

(Dimension of $\beta = 2$ / 2 no. of samples)

Assume :- β_i as β_1, β_2.

Step 1 :- Compute $\phi(x)^{(i)}$:-

$\phi(x^{(1)}) = \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix}$

$\phi(x^{(2)}) = \begin{bmatrix} 9 \\ 16 \\ 12 \end{bmatrix}$

Now, $\theta = \beta_1 \phi(x^{(1)}) + \beta_2 \phi(x^{(2)})$.

$= \beta_1 \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix} + \beta_2 \begin{bmatrix} 9 \\ 16 \\ 12 \end{bmatrix}$

$\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} \beta_1 + 9\beta_2 \\ 4\beta_1 + 16\beta_2 \\ 2\beta_1 + 12\beta_2 \end{bmatrix}$

similar to L_2 norm.

apsara

Date: 13.03.25

Subject to $\sum_{j=1}^{p} \beta_j^2 = 1$ — (2)

$y^{(i)}(\beta_0 + \beta_1 x_1^{(i)} + \beta_2 x_2^{(i)} + \dots + \beta_p x_p) \geq M$ — (3). for all $i = 1, 2, \dots n$

Constrained optimization problem.

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

3 lines :- (hyperplanes)

$2x_1 + 3x_2 - 5 = 0$ — (1)

$-x_1 + 4x_2 + 7 = 0$ — (2)

$5x_1 - 12x_2 + 10 = 0$ — (3)

Find optimal hyperplane for the given dataset.

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

→ For line (1),

$$d = \frac{|(-5) + 2x_1 + 3x_2|}{\sqrt{2^2 + 3^2}} = \frac{|(-5) + 2x_1 + 3x_2|}{\sqrt{13}}$$

For sample (1),

$$d_1 = \frac{|(-5) + 2\times 3 + 4\times 3|}{\sqrt{13}} = \frac{13}{\sqrt{13}} = \sqrt{13} = 3.605$$

For sample (2),

$$d_2 = \frac{|(-5) + 2\times 2 + 3\times 3|}{\sqrt{13}} = \frac{8}{\sqrt{13}} = 2.22$$

For sample (3),

$$d_3 = \frac{|(-5) + 2\times 1 + (-1)\times 3|}{\sqrt{13}} = \frac{6}{\sqrt{13}} = 1.664$$

For sample (4),

$$d_4 = \frac{|(-5) + 2\times(-2) + 3\times 1|}{\sqrt{13}} = \frac{6}{\sqrt{13}} = 1.664$$

Margin $(M_1) = \min(d_1, d_2, d_3, d_4) = d_3 \,\&\, d_4 = 1.664$

Page No.:

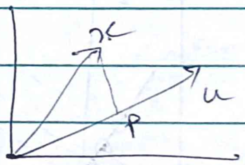
eg:- $b = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ $a = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ $a^T = [1 \ \ 2]$

$P = \frac{aa^T}{a^Ta} b$ $a^Ta = [1 \ \ 2] \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 4+1 = 5$

$= \frac{\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix}}{5}$ $aa^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{2\times1} [1 \ \ 2]_{1\times2} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$

$= \frac{\begin{bmatrix} 3+8 \\ 6+16 \end{bmatrix}}{5} = \frac{1}{5}\begin{bmatrix} 11 \\ 22 \end{bmatrix} = \begin{bmatrix} 2.2 \\ 4.4 \end{bmatrix}$

PCA



u to be subspace of unit length.

$\vec{x}$ projected point of $\vec{x}$ = $proj(u) \cdot \vec{x}$

$= \frac{uu^T}{|u^Tu|} \vec{x}$ → unit vector.

$= (\vec{x} \ u) u^T$.

Find u such that :-

$\underset{u}{\arg\max} \frac{1}{m} \sum_{i=1}^{m} \| proj(u) \cdot x^{(i)} \|^2$

$\hookrightarrow (\vec{x} \ u) \ u^T$ → unit vector.

$= \boxed{\underset{u}{\arg\max} \frac{1}{m} \sum_{i=1}^{m} (x^{(i)} \cdot u)^2}$ → The norm of a scaler multiplied with a unit vector is equal to square of the scaler.

$= \underset{u}{\arg\max} \ u^T \underbrace{\left(\frac{1}{m} \sum_{i=1}^{m} x^{(i)} \cdot x^{(i)T} \right)}_{\text{covariance matrix of } X} u$.

↳ sol^n: Eigen vector of X.

Page No.:

apsara

Date: __.__.____

Confusion matrix :-

	True +ve	True -ve
PP	4	3
PN	1	2

① Accuracy $= \frac{TP+TN}{Total} = \frac{4+2}{10} = \frac{6}{10}$

$= 0.6$

② Precision $= \frac{TP}{TP+FP} = \frac{4}{4+3} = \frac{4}{7}$

$= 0.57$

③ Recall/Sensitivity TPR $= \frac{TP}{TP+FN} = \frac{4}{4+1} = \frac{4}{5} = 0.80$

④ Specificity $= \frac{TN}{TN+FP} = \frac{2}{2+3} = \frac{2}{5} = 0.40$

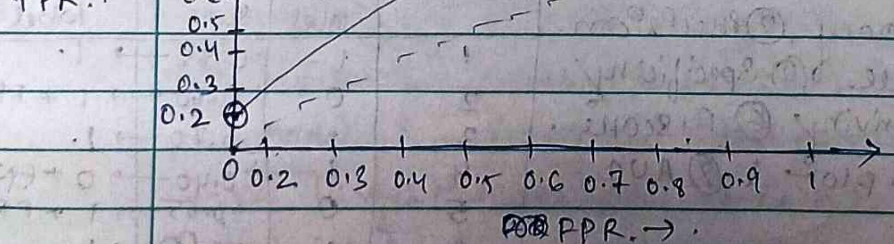
⑤ ~~Sensitivity~~ FPR $= \cancel{\frac{TP}{TP+FN} = \frac{4}{4+1} = \frac{4}{5} = 0.80} = 1-0.40 = 0.60$

⑥ F1 Score $= 2 \cdot \frac{Precision \times Recall}{Precision + Recall} = 2 \cdot \frac{0.57 \times 0.80}{0.57+0.80} = \frac{0.912}{1.37}$

$= 0.665$

⑦ ROC plot :- TPR | FPR

Thresholds	TP	FP	TN	FN	TPR	FPR (1-specificity)
0.2	5	4	1	0	1	1-0.2 → 0.8
0.4	5	3	2	0	1	1-0.4 → 0.6
0.5	4	3	2	1	0.8	1-0.4 = 0.6
0.7	1	0	5	4	0.2	1-1 = 0



⑧ AUC $= \sum_{i=1}^{4} (FPR_i - FPR_{i-1}) * \frac{(TPR_i + TPR_{i-1})}{2}$

$= \sum_{i=1}^{4} \frac{(0.6-0)*(0.8+0.2)}{2} + \frac{(0.8-0.6)*(1+1)}{2}$

$= 0.3 + 0.2 = 0.5$

Page No.: ____

For line ② :-

$$d = \frac{|7 - x_1 + 4x_2|}{\sqrt{(-1)^2 + 4^2}} = \frac{|7 - x_1 + 4x_2|}{\sqrt{17}}$$

For sample ①,

$$d_1 = \frac{|7 - 3 + 4 \times 4|}{\sqrt{17}} = 4.850.$$

For sample ②,

$$d_2 = \frac{|7 - 2 + 3 \times 4|}{\sqrt{17}} = \frac{17}{\sqrt{17}} = 4.123$$

For sample ③,

$$d_3 = \frac{|7 - 1 + (-1) \times 4|}{\sqrt{17}} = \frac{2}{\sqrt{17}} = 0.485$$

For sample ④

$$d_4 = \frac{|7 - (-2) + 4 \times 1|}{\sqrt{17}} = 3.152$$

~~For for sample ⑤~~, $M_2 = \min(d_1, d_2, d_3, d_4) = d_3 = 0.485$

For line ③ :-

$$d = \frac{|10 + 5x_1 - 12x_2|}{13}.$$

For sample ①,

$$d_1 = \frac{|10 + 5 \times 3 - 12 \times 4|}{13} = 1.769.$$

For sample ②,

$$d_2 = \frac{|10 + 5 \times 2 - 12 \times 3|}{13} = 1.230.$$

For sample ③,

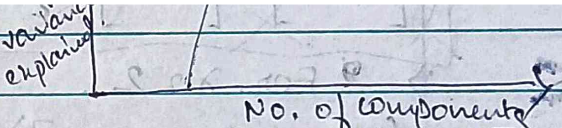
$$d_3 = \frac{|10 + 5 \times 1 + 12 \times 1|}{13} = 2.076$$

For sample ④,

$$d_4 = \frac{|10 - 2 \times 5 - 12 \times 1|}{13} = 0.923$$

$M_3 = \min(d_1, d_2, d_3, d_4) = d_4 = 0.923.$

∴ The max of M_1, M_2 & M_3 is M_1 therefore line ① is optimal hyperplane.



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

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

Step 1: Standardize each column:

$C_1 = \frac{|2-4|}{2} = -1, \ \frac{|4-4|}{2} = 0, \ \frac{|6-4|}{2} = 1.$

Mean $= \frac{2+4+6}{3} = \frac{12}{3} = 4.$

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

For C_2, Mean $= \frac{100+200+300}{3} = \frac{\cancel{600}^{200}}{\cancel{3}} = 200$

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

$\therefore \frac{100-200}{100} = -1, \ \frac{200-200}{100} = 0, \ \frac{100-300}{100} = 1.$

$X_{std} = \begin{bmatrix} -1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}$

Step 2: $\frac{1}{n-1} X_{std}^T X_{std} = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

Step 3: Compute eigen values & eigen vectors of the covariance matrix.

$\det(A - \lambda I) = 0.$

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

$\det\left(\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right) = \det \begin{bmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{bmatrix}$

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

$\Rightarrow 1 - 2\lambda + \lambda^2 = 1.$

$\Rightarrow \lambda(2-\lambda) = 0.$

$\Rightarrow \lambda(2-\lambda) = 0.$

$\therefore \lambda_1 = 0$ or $\lambda_2 = +2.$

Step 3. $(A - \lambda I)V = 0.$

~~$\left(\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right) V = 0$ For $\lambda = 0$, $\left(\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \right) V = 0$~~

$\Rightarrow$ For $\lambda = 2.$

$\left(\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \right) V_2 = 0 \Rightarrow \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} V = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\Rightarrow \cancel{V_2 = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}^{-1}} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ $\quad y - x = 0 \Rightarrow x = y$; $\quad x - y = 0 \Rightarrow x = y.$

~~Step 4 Zspace = $X_{std} \times V_2$~~ $= \begin{bmatrix} -1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} \cancel{\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}}$

$\therefore v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Normalizing, $v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$

Step 5. $Z_{space} = X_{std} \times V_2$

$= \begin{bmatrix} -1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} -2/\sqrt{2} \\ 0 \\ 2\sqrt{2} \end{bmatrix}_{3\times1}$

eg: Step 1. $c_1 = \{2, 3, 4, 6\}$, $c_2 = \{1, 5\}$.

	x_1	x_2
#1	1.	1.
#2	1	2
#3	2.	2
#4	8.	8
#5	8.	9
#6	9	8.

$K = 2$.

1st iteration:-

centroid of c_1 = $\begin{bmatrix} 5 \\ 5 \end{bmatrix}$

centroid of c_2 = $\begin{bmatrix} 4.5 \\ 5 \end{bmatrix}$

Step 2: $d_{11} = \sqrt{(1-5)^2 + (1-5)^2} = \sqrt{4^2 + 4^2} = 5.65$

$d_{12} = \sqrt{(4.5-1)^2 + (5-1)^2} = \sqrt{12.25 + 16} = 5.31$

Page No.:

$c_1 = \{4, 5, 6.\}$

$c_2 = \{1, 2, 3.\}$

$\min(d_{11}, d_{12}) = d_{12}$, so Sample #1 in c_2

* ELBOW Method

For sample 2:-

$d_{21} = \sqrt{(1-5)^2 + (2-5)^2} = \sqrt{16 + 9} = 5.$

$d_{22} = \sqrt{(1-4.5)^2 + (2-5)^2} = \sqrt{12.25 + 9} = 4.6$ ✓

$\min(d_{21}, d_{22}) = d_{22}$, sample #2 in c_2

For sample 3:-

$d_{31} = \sqrt{(2-5)^2 + (2-5)^2} = \sqrt{(-3)^2 + (-3)^2} = \sqrt{18} = 4.24.$

$d_{32} = \sqrt{(2-4.5)^2 + (2-5)^2} = \sqrt{(-2.5)^2 + (-3)^2} = \sqrt{15.25} = 3.90$ ✓

$\min(d_{31}, d_{32}) = d_{32}$, Sample # 3 in c_2.

For sample 4:-

$d_{41} = \sqrt{(8-5)^2 + (8-5)^2} = \sqrt{18} = 4.24.$

$d_{42} = \sqrt{(8-4.5)^2 + (8-5)^2} = \sqrt{21.25} = 4.60.$

$\min(d_{41}, d_{42}) = d_{41}$ = sample #4 in c_1

For sample 5:-

$d_{51} = \sqrt{(8-5)^2 + (9-5)^2} = 5$

$d_{52} = \sqrt{(8-4.5)^2 + (9-5)^2} = 5.41,$

$\min(d_{51}, d_{52}) = d_{51}$ = sample # 5 in c_1.

For sample 6:-

$d_{61} = \sqrt{(9-5)^2 + (8-5)^2} = 5.$

$d_{62} = \sqrt{(9-4.5)^2 + (8-5)^2} = 5.41$

$\min(d_{61}, d_{62}) = d_{61}$ = Sample # 6 in c_1.

Example :- Perform single linkage with Euclidean dist as measure.

	x_1	x_2
A	1	1
B	2	1
C	4	3
D	5	4
E	6	5

(A) (B) (C) (D) (E)

$^nC_2$ pairwise dist $= {}^5C_2 = \frac{5\times 4}{2} = 10$.

$d(AB) = \begin{bmatrix}1 & 1\end{bmatrix} \begin{bmatrix}2 & 1\end{bmatrix} = \sqrt{(2-1)^2+(1-1)^2} = 1$.

$d(AC) = \begin{bmatrix}1 & 1\end{bmatrix} \begin{bmatrix}4 & 3\end{bmatrix} = \sqrt{(4-1)^2+(3-1)^2} = 3.6$.

$d(AD) = \begin{bmatrix}1 & 1\end{bmatrix} \begin{bmatrix}5 & 4\end{bmatrix} = \sqrt{(5-1)^2+(4-1)^2} = 5$.

$d(AE) = \begin{bmatrix}1 & 1\end{bmatrix} \begin{bmatrix}6 & 5\end{bmatrix} = \sqrt{(6-1)^2+(5-1)^2} = 6.4$.

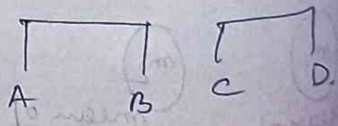
$d(BC) = 2.82$, $d(BD) = 4.2$, $d(BE) = 5.6$

$d(CE) = 2.82$, $d(CD) = 1.4$, $d(DE) = 1.4$.

	A	B	C	D	E
A	0	1	3.6	5	6.4
B	1	0	2.82	4.2	5.6
C	3.6	2.82	0	1.4	2.82
D	5	4.2	1.4	0	1.42
E	6.4	5.6	2.82	1.42	0

Merge (A & B) as distance value = 1 (least of all).

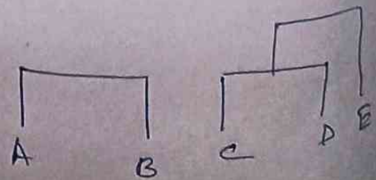
	AB	C	D	E
AB	0	2.8	4.2	5.6
C	2.8	0	1.4	2.8
D	4.2	1.4	0	1.4
E	5.6	2.8	1.4	0



min (CA, CB) = min (3.6, 2.8).

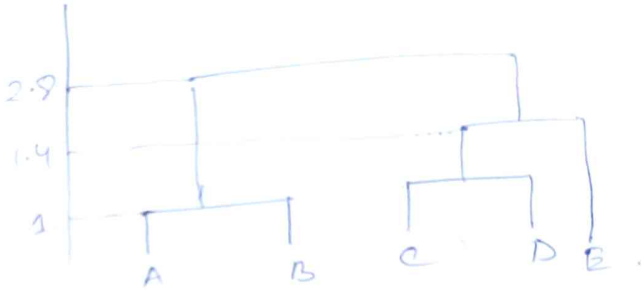
Merge C & D.

	AB	CD	E
AB	0	2.8	4.2
CD	2.8	0	1.4
E	4.2	1.4	0



Merge CDE.

	AB	CDE
AB	0	2.8
CDE	2.8	0



"To buy this product".

a
abu
allow
buy
⋮
more
⋮
play
product

$50{,}000 \times 1$ → $[0, 0, 0, 1, 0, 0, 0, 1, 0, 0]^T$

eg	x_1	x_2	x_3	class	Prob $y=1$ (Spam)	Prob $y=0$ (NoSpam)
#1	1	0	1	Spam	0	1/4
#2	1	1	0	Spam	0	1/4
#3	1	1	0	Spam	2/7	0
#4	0	0	1	No spam	2/7	0
#5	0	0	0	No spam	1/7	0
#6	1	1	1	No spam		
#7	0	1	0	Spam		
#8	1	0	0	Spam		
#9	0	1	0	Spam		
#10	1	0	0	Spam		
#11	1	1	0	No spam		

x_1	x_2	x_3	Prob ($y=1$ spam)	Prob $y=0$ (NoSpam)
0	0	0	0	1/4
0	0	1	0	1/4
0	1	0	2/7	0
0	1	1	2/7	0
1	0	0	1/7	0
1	0	1	2/7	1/4
1	1	0	0	1/4
1	1	1		

Naive Baye's Assumption → x's are independent given y
Conditional Independence assumption.

$P(x_{2087} | y) = P(x_{2087} | y, x_{38193})$.

eg- $X =$ 1) Free win now — $Y =$ Spam
2) win a prize — Spam
3) Hello how are you. — Not spam
4) Let's win it. — Not "
5) Free lunch today. — Not "

→ Training set.

word "free"	word "win"	label.
#1 yes	yes	spam (1)
#2 NO	yes	spam (1)
#3 NO	NO	Not spam (0)
#4 NO	yes	Not spam (0)
#5 yes.	NO	Not spam (0)

~~Step 1: Compute Prior~~.

Test set

Text message = "Free win".

Goal :- Prob. of $y = 1$ given text message (x).

→ $P(\text{spam}) = 2/5 = 0.4$
$P(\text{Not spam}) = 3/5 = 0.6$.
→ Prior computation.

Step 1 : compute Prior :-

~~P(free | label = spam) = P(~~.

Step 2 : compute likelihood :-

For Spam :- Free appears = 1/2 , $P(\text{Free} = \text{yes} \mid \text{spam} = 1) = \frac{1}{2}$.

~~For Not Spam~~ $P(\text{win} = \text{yes} \mid \text{spam} = 1) = 2/2$.

For Not spam class :- $P(\text{Free} = \text{yes} \mid \text{Not spam}) = 1/3$.

$P(\text{win} = \text{yes} \mid \text{Not spam}) = 1/3$.

Inference phase :-
Free = yes
win = yes

$P(\text{spam} \mid X) = P(X \mid \text{spam}) \cdot P(\text{spam})$.
$P(X \mid \text{spam}) = P(X_1, X_2 \mid \text{spam})$
$= P(X_1 \mid \text{spam}) \cdot P(X_2 \mid \text{spam})$

Page No.:

apsara

Date:___.___.___

$$P(\text{spam} \mid x) = P(x_1 \mid \text{spam}) \cdot P(x_2 \mid \text{spam}) \cdot P(\text{Spam})$$

$$= P(\text{Free} = \text{yes} \mid \text{spam}) \cdot P(\text{win} = \text{yes} \mid \text{spam}) \times P(\text{spam})$$

$$= \frac{1}{\cancel{2}} \times \frac{\cancel{2}}{\cancel{2}} \times \frac{\cancel{2}}{5} = \frac{1}{5} = 0.2$$

So, $$P(\text{Not spam} \mid x) = P(\text{Free} = \text{yes} \mid \text{Not spam}) \cdot P(\text{win} = \text{yes} \mid \text{Not spam}) \times P(\text{Not spam})$$

$$= \frac{1}{3} \times \frac{1}{\cancel{3}} \times \frac{\cancel{3}}{5} = \frac{1}{15} = 0.066$$

As $0.2 > 0.066$ ∴ text msg is $\boxed{\text{Spam}}$