|9/22

4 - The clustering process is continued untill all the points are visited.

|9/22

Module 6

* EM Algorithm -

It is an approach for maximum likeli hood estimation (MLE) in the presence of latent variables The most discussed application of EM Algorithm is for clustering with mixture model. The EM Algorithm is an iterative approach that cycles between 2 steps i.e estimation step(E) and maximoization step(M)

E: Classify the data using current theory

M: Generate the best theory using the current classification of data.

Step E generates the expected classification for each example and Step M generates the most likely theory given the classified data.

Eg) Imagine there are two coins A and B. One is more likely to get heads, the other more likely to get tails. You pick one at random and toss it and then find out which one was it. Suppose we do it five times i.e

28/9/22

pick a coin randomly, toss it ten times and record the number of heads and tails. Then get the average number heads for each coin. This helps in finding out the maximum likelihood. Consider the following reasons.

Sets	Coin A H	Coin A T	Coin B H	Coin B T
1			5	5
2	9	1		
3	8	2		
4			4	6
5	7	3		
Total	24	6	9	11

$$\theta_A(h) = \frac{24}{30} = 0.8 \qquad \theta_B(h) = \frac{9}{20} = 0.45$$

(5 sets 10 tosses per set)

From Above Data Coin A yields head 80% of time and Coin B 40% of time

Eg) What if we are given only the results of our coins to ~~find~~ answer ① guess the % of heads that each coin yields ② guess which coin was

Scanned with CamScanner

8/9/22

picked for each set of 10 coin tosses

Steps to do -

Step1) Assign random averages to both the coins

Step2) For each of the five rounds of 10 coin tosses do

a) Check the % of heads (heads)

b) Find the probability of it coming from each coin

c) Compute the expected number of heads -

Using that probability as a weight multiply it number of heads

d) Record these numbers

e) Recompute new means for point A and point B

Step3) With these new mean values go back to Step 2)

Step4) Step 2) and Step3) are repeated until the mean values are stabled / settled.

Sets	H	T
1	5	5
2	9	1
3	8	2
4	4	6
5	7	3
Total	24	6

28/9/22.

(randomly assigned)

① Assume $\theta_A = 0.6$ and $\theta_B = 0.5$

θ_i is average number of heads for each coin

Consider ~~row num~~ turn 1 which has $\frac{5}{10}$ number of heads and $\frac{5}{10}$ number of tails

② Compute the likelihood that it was a coin A or coin B using binomial Distribution ~~P(X)~~ =

$$\boxed{P(k) = \theta^k (1-\theta)^{n-k}}$$

Likelihood of A = $P_A^{h} (1 - P_A)^{n-h}$

$= 0.6^5 (1-0.6)^{10-5}$

$= 0.0007962624$

Likelihood of B = $P_B^{h} (1 - P_B)^{n-h}$

$= 0.5^5 (1-0.5)^5$

$= 0.0009765625$

Convert the likelihood of A and B to a probability of getting heads by normalizing the likelihood using the formula

$$\theta = \frac{A}{A+B}$$

for A = $0.0007962624 / (0.0007962624 + 0.0009765625)$

$= 0.45$

3/9/22

for B = 0.55

(2) Estimate likely number of heads and tails for A and B

~~for A = 0.45~~ estimates of ~~A~~ heads for A = $\theta_A \times$ number of heads

$= 0.45 \times 5$
$= \underline{\underline{2.2}}$

for A~~B~~ estimates of tails = 0.45×5
$= \underline{\underline{2.2}}$

for B estimates of head = 0.55×5
$= \underline{\underline{2.75}}$

for B estimates of tail = $\theta_B \times$ number of tails
$= 0.55 \times 5$
$= \underline{\underline{2.75}}$

Do the same for all 5 runs

Likelihood of A = $P_A^h(1-P_A)^{n-h}$
$= 0.6^9(1-0.6)^{10-9}$
$= 0.6^9(1-0.6)^1$
$= 0.0040310784$

Likelihood of B = $P_B^h(1-P_B)^{n-h}$
$= 0.5^9(1-0.5)^{10-9}$
$= 0.5^9(1-0.5)^1$
$= 0.0009765625$

28/9/22

Convert the likelihood of A & B to a probability of getting heads by normalizing the likelihood using the formula

$$\theta = \frac{A}{A+B}$$

for A $= \frac{0.0040310784}{0.0040310784 + 0.0009765625}$

$= 0.8049855172 = 0.8$

for B $= 0.1950144628 = 0.2$

Estimate likely no of heads & tails for A & B

for A estimates of head $= \theta_A \times$ no of heads

$= 0.8 \times 9$

$= \underline{\underline{7.2}}$

for A estimates of tails $= 0.8 \times 1$

$= \underline{\underline{0.8}}$

for B estimates of heads $= \theta_B \times$ no of head

$= 0.2 \times 9$

$= \underline{\underline{1.8}}$

for B estimates of tails $= 0.2 \times 1$

$= \underline{\underline{0.2}}$

Scanned with CamScanner

8/9/22

	Coin A H	Coin A T	Coin B H	Coin B T
0.45 * A, 0.55 * B	2.2	2.2	2.8	2.8
0.8 * A, 0.2 * B	7.2	0.8	1.8	0.2
0.73 * A, 0.27 * B	5.9	1.5	2.1	0.5
0.35 * A, 0.65 * B	1.4	2.1	2.6	3.9
0.65 * A, 0.35 * B	4.5	1.9	2.5	1.1
Total	21.3	8.6	11.7	8.4

Compute the new probabilites for each coin using $\frac{H}{H+T}$ that gives you the new maximized parameter θ for each coin

$$\theta_A = \frac{21.3}{21.3+8.6} = 0.71$$

$$\theta_B = \frac{11.7}{11.7+8.4} = 0.58$$

With these new values of probabilities repeat both E Step & M Step no of times till your θ value gets settled i.e conversions to fixed probabilities

The conversion values for these example are $\theta_A = 0.80$ and $\theta_B = 0.52$

These values can be used to answer the question like (1) % of heads

28/9/22

that each coin yields
$\theta_A = 0.8$ & $\theta_B = 0.5$
(ii) which coin yields more no of heads
(iii) which was picked for each set of coin tosses

Set	H	T	P(h)	
1	5	5	0.5	B
2	9	1	0.9	A
3	8	2	0.8	A
4	4	6	0.4	B
5	7	3	0.7	A
Total	24	6		

6/10/22

★ Dimensionality Reduction –

1) Principal Component Analysis – It is a way of identifying pattern in a data & expressing a data in such a way that it highlights their ~~xx~~ similarities & dissimilarities. The Advantage of PCA is that once we have found pattern in the Data & Compress the Data, i.e reducing the no of Dimensions without much loss of information. This technique is very much useful in image compression

Steps in PCA –

① Get some Data we call it as matrix 'A'

② Calculate the Mean for each attribute

③ Substract the mean from the given matrix 'A'

A - Mean

This produce the dataset whose mean is zero. This matrix is called as Data Adjust Matrix 'B'

④ Calculate the co-variance matrix of matrix B

$$C = \begin{bmatrix} 6_{xx} & 6_{xy} \\ 6_{yx} & 6_{yy} \end{bmatrix} \qquad C = \frac{B^T * B}{(n-1)} \; (n \to \text{dimens}$$

Scanned with CamScanner

6/10/22

(5) Calculate the ideal value for the co-variance matrix where λ_1, λ_2 & so on are eigen values. Then calculate the eigen vector of co-variance matrix in such a way that 'n' eigen vector if co-variance matrix ~~has~~ is having 'n' columns.

$$|C - \lambda I| = 0$$

$\lambda_1, \lambda_2, \ldots$

(6) Vectors

$$U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} 1 \\ U_2 \end{bmatrix} \lambda_1$$

$$V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} V_1 \\ 1 \end{bmatrix} \lambda_2$$

(7) Normalize eigen vectors because the eigen vectors should be unique vectors

$$U = \frac{1}{\sqrt{U_1^2 + U_2^2}} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$$

$$V = \frac{1}{\sqrt{V_1^2 + V_2^2}} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$$

(8) For ~~the~~ ming feature vector in which arrange the features in the descending order of eigen values

$$Fr = \begin{bmatrix} U_1 & V_1 & \cdots \\ U_2 & V_2 & \cdots \end{bmatrix} \text{ where } \lambda_1 > \lambda_2$$

$\lambda_1 \lambda_2$

10/22

(9) Getting Old Data Back it is a row original data

(A)

$$\boxed{A' = F_V^T * B^T + \text{Original Mean}}$$

8] Perform the PCA Steps on following ~~step1~~ Dataset

(A)

x	y
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2.0	1.6
1.0	1.1
1.5	1.6
1.1	0.9

$[10 \times 2]$

$n = 10$

A] Mean $(\bar{y}) = \frac{19.1}{10} = 1.91$

Mean $(\bar{x}) = \frac{18.1}{10} = 1.81$

(B)

/10/22

(1) (x - Mean)

x	y
0.69	0.49
-1.31	-1.21
0.39	0.99
0.09	0.29
1.29	1.09
0.49	0.29
0.19	-0.31
-0.81	-0.81
-0.31	-0.31
-0.71	-1.01

$\Rightarrow C = \frac{B^T * B}{n-1}$

$C = \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix}$

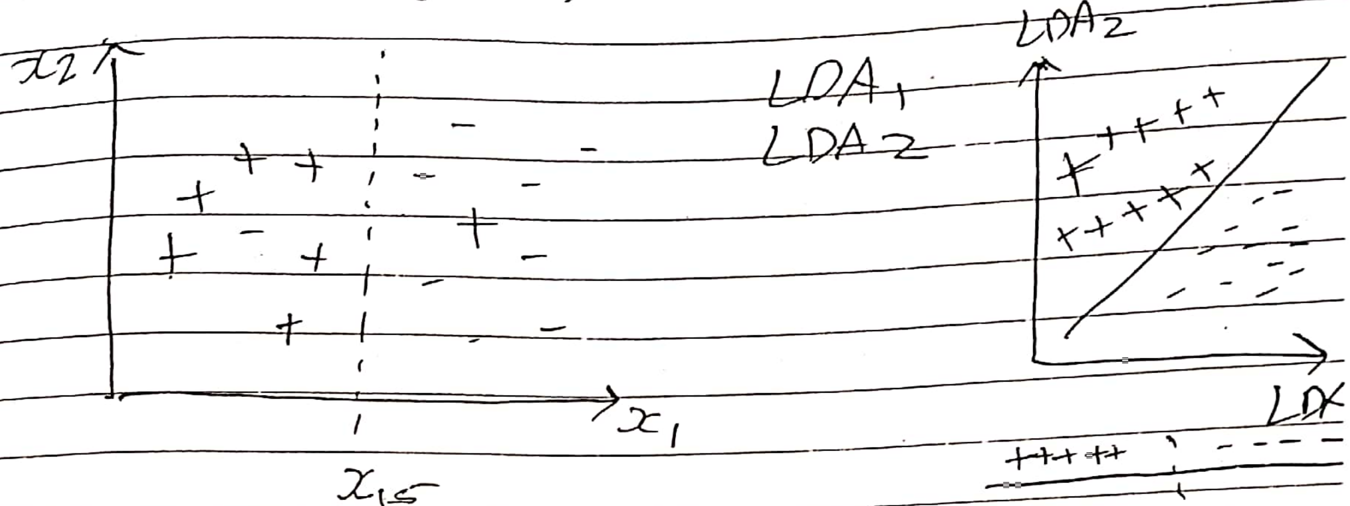
$C = \begin{bmatrix} 0.616555 & 0.615444 \\ 0.615444 & 0.2165556 \end{bmatrix}$

$\Rightarrow |C - \lambda I| = 0$

$\begin{bmatrix} 0.616555 - \lambda & 0.615444 \\ 0.615444 & 0.21655556 - \lambda \end{bmatrix} = 0$

/10/22

* Linear Discriminant Analysis (LDA) :-
(LDA)

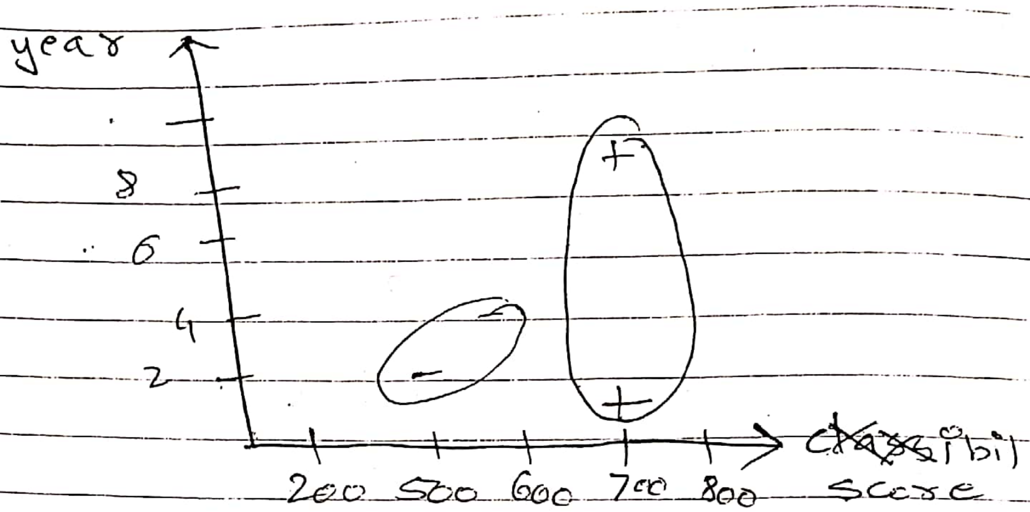


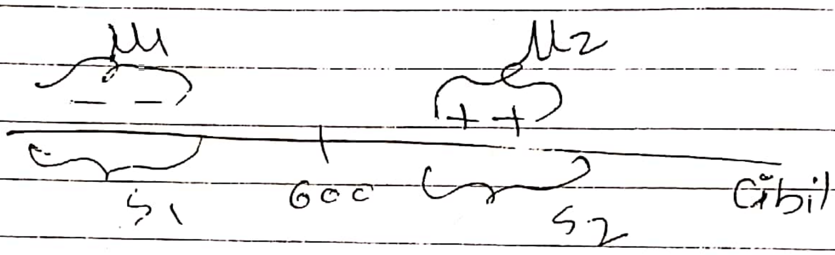
LDA (Linear Discriminant Analysis) is a Dimensionality Reduction technique commonly used for supervised classification problems. It is used for modelling difference in groups i.e separating 2 or more classes. It is used to project feature in higher dimension space into a lower dimension space

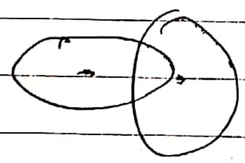
# years	Cibil score	Class
2	500	Bad
4	550	Bad
1	700	Good
8	750	Good

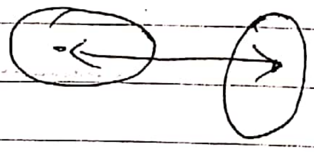
Scanned with CamScanner

12/10/22









LDA uses both the axes (cibil score & no of years) to create a new axis & projects data into a new axis in a way to maximize the separation of the two categories & hence reducing the 2-D graph to a 1D graph.
Two criterias are used by LDA to create a new axis.

12/10/22

1. Maximize the distance between the means of the 2 classes
2. Minimize the variation ~~of~~ within the class

$$\boxed{= \frac{(\mu_1 - \mu_2)^2 \quad \text{max}}{S_1^2 + S_2^2 \quad \text{min}}}$$

→ The major steps in LDA Analysis is

Step 1 → Commute within Class Scatter Matrix (SW)

Step 2 → Commute between Class Matrix (SB

Step 3 → Find the best LDA projection vector

The whole LDA process can be defined with the following steps -

1. Commuting the within class & between class scatter matrix
2. Commmuting the eigen vectors & eigen values for the scatter matrices
3. Sorting the eigen values & selecting top k (Definitely k<n)
4. Creating a new matrix that will contain the eigen vectors map to k eigen values
5. Obtaining the new features by

Scanned with CamScanner

2/10/22

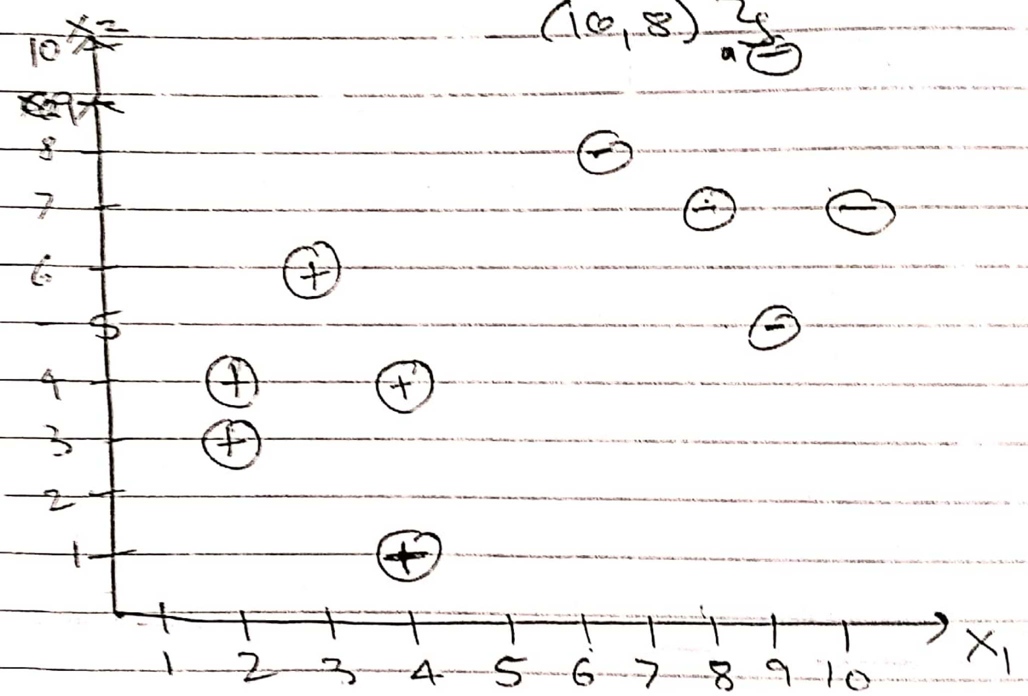
taking the dot product of the data & the matrix from step (4)

Q) 8)

$G_1 = X_1 = (x_1, x_2) = \{(4,1), (2,4), (2,3), (3,6), (4,4)\}$

$G_2 = X_2 = (x_1, x_2) = \{(9,10), (6,8), (9,5), (8,7), (10,8)\}$

A)



Step 1 – Commute within Class Scatter Matrix i.e S_w

$$\boxed{S_w = S_1 + S_2}$$

where S_1 is covariance matrix of class one & S_2 is covariance matrix of class two

$$\boxed{S_1 = \frac{1}{n} \sum_{x \in G_1} (x - \mu_1)(x - \mu_1)^T}$$

where $\mu_1 \rightarrow$ mean of class one

2/10/22

$$\boxed{S_2 = \frac{1}{n} \sum_{x \in C_2} (x - \mu_2)(x - \mu_2)^T}$$

$$\mu_1 = \left\{ \frac{4+2+2+3+4}{5}, \frac{1+4+3+6+4}{5} \right\}$$

$$= \{3, 3.60\}$$

$$\mu_2 = \left\{ \frac{9+6+9+8+10}{5}, \frac{10+8+5+7+8}{5} \right\}$$

$$= \{8.4, 7.6\}$$

$$(x - \mu_1)_{C_1} = \begin{bmatrix} 1 & -1 & -1 & 0 & 1 \\ -2.6 & 0.4 & -0.6 & 2.4 & 0.4 \end{bmatrix}$$

~~Step~~ Calculate $(x - \mu_1) \cdot (x - \mu_1)^T$ for each x value so we will have 5 such matrices

$$\begin{bmatrix} 1 \\ -2.6 \end{bmatrix} \begin{bmatrix} 1 & -2.6 \end{bmatrix}$$

$$\begin{bmatrix} 1 & (-2.6) \\ -2.6 & 6.76 \end{bmatrix} \text{ --- } (1)$$

$$\begin{bmatrix} -1 \\ 0.4 \end{bmatrix} \begin{bmatrix} -1 & 0.4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & -0.4 \\ -0.4 & 0.16 \end{bmatrix} \text{ --- } (2)$$

$$\begin{bmatrix} -1 \\ -0.6 \end{bmatrix} \begin{bmatrix} -1 & -0.6 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0.6 \\ 0.6 & 0.36 \end{bmatrix} \text{ --- } (3)$$

$$\begin{bmatrix} 0 \\ 2.4 \end{bmatrix} \begin{bmatrix} 0 & 2.4 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 \\ 0 & 5.76 \end{bmatrix} \text{ --- } (4)$$

$$\begin{bmatrix} 1 \\ 0.4 \end{bmatrix} \begin{bmatrix} 1 & 0.4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0.4 \\ 0.4 & 0.16 \end{bmatrix} \text{ --- } (5)$$

12/10/22

Adding matrix ①, ② ③, ④, ⑤ , we get

$$\begin{bmatrix} 4 & -2 \\ -2 & 13.2 \end{bmatrix}$$

Divide by 5

$$S_1 = \begin{bmatrix} 0.8 & -0.4 \\ -0.4 & 2.64 \end{bmatrix}$$

Similarly, covariance matrix for class 2 is

$$S_2 = \begin{bmatrix} 1.84 & -0.04 \\ -0.04 & 2.64 \end{bmatrix}$$

$$S_W = S_1 + S_2 = \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{bmatrix}$$

Step 2 - Compute between class scatter matrix S_B

$$S_B = (\mu_1 - \mu_2) \cdot (\mu_1 - \mu_2)^T$$

$$\begin{bmatrix} -5.4 & -4 \end{bmatrix} \begin{bmatrix} -5.4 \\ -4 \end{bmatrix}$$

$$= \begin{bmatrix} 29.16 & 21.6 \\ 21.6 & 16 \end{bmatrix}$$

2/10/22

Step 3 Find the Best LDA Projection Vector (V)

Note → We find this using eigen vector having largest eigen value

$\boxed{S_W^{-1} S_B V = \lambda V}$ —— ⓐ (V has 2 components)

$|S_W^{-1} S_B V - \lambda V| = 0$

$$\begin{bmatrix} 11.89-\lambda & 8.81 \\ 5.08 & 3.76\lambda \end{bmatrix} = 0$$

$(11.89-\lambda)\cancel{8}(3.76\lambda) - (8.81 \times 5.08) = 0$

So, $\boxed{\lambda = 15.65}$

Substituting λ in equation ⓐ

$$\begin{bmatrix} 11.89 & 8.81 \\ 5.08 & 3.76 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = 15.65 \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$$

$$\begin{bmatrix} 11.89V_1 & V_2 8.81 \\ 5.08V_1 & 3.76V_2 \end{bmatrix} = \begin{bmatrix} 15.65V_1 \\ 15.65V_2 \end{bmatrix}$$

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} 0.91 \\ 0.39 \end{bmatrix}$$

#6 12/10/22

★ Comparison Between PCA And LDA

PCA	LDA
① PCA is unsupervised Dimensionality reduction	① LDA is supervised Dimensionality reduction
② Focuses on finding a feature subspace maximizes the separability between groups	② Focuses on capturing the direction of maximum variation in the dataset
③ PCA less superior to LDA	③ LDA is superior to PCA
④ PCA requires fewer computations	④ LDA requires significantly more computation than PCA for large Datasets
⑤ Deals with data in its entirely for PCA without paying any particular attention to the underlying class structure	⑤ Deals Directly with Discrimination between classes
⑥ Applications - Prominent field of criminal investigation is beneficial	⑥ Applications - Classification problem in speech recognition

19/10/22

Singular Value Decomposition (SVD) -

SVD is ~~the~~ one of the several techniques that can be used to reduce the dimensionality i.e the no. of columns of the dataset. More columns normally means more time required to built modules & score data. If some columns have no predicte values, this means wasted time or hardly those columns contribute in prediction rather they contribut noise to the model.

Singular Value -

Let A is the matrix of $m \times n$ dimension so $A^T \cdot A$ is $n \times n$ symetric matrix. The eigen values of $A^T \cdot A$ are $\lambda_1, \lambda_2, \lambda_3$, so on. Assume $\lambda_1 \geq \lambda_2 \geq \lambda_3 \ldots$ & so on ≥ 0

Let $\sigma_i = \sqrt{\lambda_i}$ $\sigma_1 = \sqrt{\lambda_1}$, $\sigma_2 = \sqrt{\lambda_2}$, $\sigma_3 = \sqrt{\lambda_3}$ .. & so on so $\sigma_1 > \sigma_2 > \sigma_3 >$ .. & so on

Here, σ_i is the singular value of Matrix A

In SVD, we decompose the original matrix in 3 submatrices in such a way that $\boxed{A = U \Sigma V^T}$

19/10/22

A is m×n given matrix i.e given dataset
$A_{m\times n}$ [Dataset]
~~$U_{m\times m}$~~
U is m×m - Orthogonal matrix
V is n×n - Orthogonal matrix
& Σ is m×n - Diagonal matrix

$$A \cdot A^T = (U\Sigma V^T) \cdot (U \Sigma V^T)^T$$
$$= U\Sigma \underbrace{V^T V}_{} \Sigma^T U^T$$
$$= U \underline{\underline{\Sigma \Sigma^T}} U^T$$
$$= U \Sigma^2 U^T$$

U is the matrix of eigen vectors of $A \cdot A^T$

$$A^T \cdot A = (U\Sigma V^T)^T (U \Sigma V^T)$$
$$= V\Sigma^T \underbrace{U^T U}_{} \Sigma V^T$$
$$= V \underline{\underline{\Sigma^T \Sigma}} V^T$$
$$= V\Sigma^2 V^T$$

Vectors in U are called as left singular value of A & Vectors in V are called as right singular values of A.

19/10/22

The main intuation behind SVD is that the matrix A transforms a set of orthogonal vectors (V) to another set of orthogonal vectors (U) with scaling factor σ. So σ is called the singular value corresponding to the respective singular vectors U & V.

Q) Find SVD for the following matrix

$$A = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$

A) To find U ~~i.e.~~ $\cancel{U = A A}$ find $A \cdot A^T$

$$\begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 9+1+1 & \cancel{-3}+\cancel{3}+1 \\ \cancel{-3}+\cancel{3}+1 & 1+9+1 \end{bmatrix}$$

$$= \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}$$

(2 columns. so eigen values = 2)

$$\lambda = \frac{T \pm \sqrt{T^2 - 4D}}{2}$$

$\lambda_1 = 10$

$\lambda_2 = 12$

$|B - \lambda I| = 0$

$$\begin{vmatrix} 11-\lambda & 1 \\ 1 & 11-\lambda \end{vmatrix} = 0$$

$(11-\lambda)(11-\lambda) - 1 = 0$

$121 - 11\lambda - 11\lambda + \lambda^2 - 1 = 0$

$\lambda^2 - 22\lambda + 120 = 0$

19/10/22

$(\lambda-10)(\lambda-12)=0$

λ_1 Put
$$\begin{vmatrix} 11-\lambda_1 & 1 \\ 1 & 11-\lambda_1 \end{vmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = 0$$

$$\lambda_{10} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

λ_2 Put

$$\begin{vmatrix} 11-\lambda_2 & 1 \\ 1 & 11-\lambda_2 \end{vmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = 0$$

$$\lambda_{12} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= \begin{array}{cc} \lambda_2 & \lambda_1 \end{array} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

By the rule corresponding highest value in the first column. Thus we have above matrix. Normalize the vectors in this matrix

(divide each element by $\sqrt{\ }$ of square & summation of all vectors in matrix)

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

$$U = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

19/10/22

$$A = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$

To find V, find $A^T A$

$$A^T A = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 9+1 & 3-3 & 3-1 \\ 3-3 & 1+9 & 1+3 \\ 3-1 & 1+3 & 1+1 \end{bmatrix}$$

$$= \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix}$$ (3 columns so 3 eigen values)

$$|B - \lambda I| = 0$$

$$\begin{vmatrix} 10-\lambda & 0 & 2-\lambda \\ 0 & 10-\lambda & 4-\lambda \\ 2-\lambda & 4-\lambda & 2-\lambda \end{vmatrix} = 0$$

$$(10-\lambda)\left[(10-\lambda)(2-\lambda) - (4-\lambda)^2\right] + (2-\lambda)\left[(10-\lambda)(2-\lambda)\right] = 0$$

$$\cancel{(10-\lambda)\,[20 - 2\lambda - 10\lambda\ \lambda^2}$$

$$\lambda^3 - 22\lambda^2 + 132\lambda = 0$$

$\lambda_1 = 0$

$\lambda_2 = 10$

$\lambda_3 = 12$

16 19/10/22

eigen vector

for $\lambda_1 = 0$ $\quad \lambda_2 = 10$ $\quad \lambda_3 = 12$

$$\begin{bmatrix} 1 \\ 2 \\ -5 \end{bmatrix} \quad \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} , \quad \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 2 & 1 \\ 2 & -1 & 2 \\ 1 & 0 & -5 \end{bmatrix}$$

Normalize

$$V = \begin{bmatrix} 1/\sqrt{6} & 2/\sqrt{5} & 1/\sqrt{30} \\ 2/\sqrt{6} & -1/\sqrt{5} & 2/\sqrt{30} \\ 1/\sqrt{6} & 0 & -5/\sqrt{30} \end{bmatrix}$$

$$V^T = \begin{bmatrix} 1/\sqrt{6} & 2\sqrt{6} & 1/\sqrt{6} \\ 2\sqrt{5} & -1/\sqrt{5} & 0 \\ 1/\sqrt{30} & 2/\sqrt{30} & -5/\sqrt{30} \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}$$

$$= \left[\begin{array}{cc|c} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{array}\right]$$

Matrix summation can be found by taking square

19/10/22

root of the non zero eigen values in descending order.

$\therefore$ we get above matrix

The third column $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is taken so as to ~~make~~ take multiplication possible.

$$A = U \Sigma V^T$$

$$A = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \\ 2/\sqrt{5} & -1/\sqrt{5} & 0 \\ 1/\sqrt{30} & 2/\sqrt{30} & -5/\sqrt{30} \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{12}/\sqrt{2} - \sqrt{10}/\sqrt{2} & 0 - \sqrt{10}/\sqrt{2} & 0 - 0 \\ \sqrt{12}/\sqrt{2} - 0 & 0 + \sqrt{10}/\sqrt{2} & 0 - 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \\ 2/\sqrt{5} & -1/\sqrt{5} & 0 \\ 1/\sqrt{30} & 2/\sqrt{30} & -5/\sqrt{30} \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{1} & -\sqrt{5} & 0 \\ \sqrt{6} & \sqrt{5} & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \\ 2/\sqrt{5} & -1/\sqrt{5} & 0 \\ 1/\sqrt{30} & 2/\sqrt{30} & -5/\sqrt{30} \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{1}/\sqrt{6} - 2 - 0 & 2\sqrt{1}/\sqrt{6} - 1 - 0 & \sqrt{1}/\sqrt{6} - 0 - 0 \\ 1 - 2 - 0 & 2 - 1 - 0 & 1 - 0 - 0 \end{bmatrix}$$

~~$= [\;-1 \quad 1 \quad 1\;]$~~