ML
LECTURE-27

BY

Dr. Ramesh Kumar Thakur
Assistant Professor (II)
School Of Computer Engineering

Previous Year Question (PCA)

- 1. Explain the Principal Component Analysis (PCA) and reduce the following dataset step-by-step from 2 dimensions to 1. [4]

Feature	Example 1	Example 2	Example 3	Example 4
x	4	8	13	7
y	11	4	5	14

- **Ans: Step-01:**
- **Get data.**
- **The given feature vectors are-**
- x1 = (4, 11)
- x2 = (8, 4)
- x3 = (13, 5)
- x4 = (7, 14)
- **Step-02:**
- **Calculate the mean vector (μ).**
- Mean vector (μ) = ((4 + 8 + 13 + 7) / 4, (11 + 4 + 5 + 14) / 4) = (8, 8.5)

Previous Year Question (PCA)

- **Step-03:**
- **Subtract mean vector (μ) from the given feature vectors.**
- $x1 - \mu = (4-8, 11-8.5) = (-4, 2.5)$
- $x2 - \mu = (8-8, 4-8.5) = (0, -4.5)$
- $x3 - \mu = (13-8, 5-8.5) = (5, -3.5)$
- $x4 - \mu = (7-8, 14-8.5) = (-1, 5.5)$

- **Step-04:**
- **Calculate the covariance matrix.**
- $m1 = (x1 - \mu)(x1 - \mu)^T = (-4, 2.5)(-4, 2.5)^T = \begin{bmatrix} 16 & -10 \\ -10 & 6.25 \end{bmatrix}$
- $m2 = (x2 - \mu)(x2 - \mu)^T = (0, -4.5)(0, -4.5)^T = \begin{bmatrix} 0 & 0 \\ 0 & 20.25 \end{bmatrix}$
- $m3 = (x3 - \mu)(x3 - \mu)^T = (5, -3.5)(5, -3.5)^T = \begin{bmatrix} 25 & -17.5 \\ -17.5 & 12.25 \end{bmatrix}$
- $m4 = (x4 - \mu)(x4 - \mu)^T = (-1, 5.5)(-1, 5.5)^T = \begin{bmatrix} 1 & -5.5 \\ -5.5 & 30.25 \end{bmatrix}$
- Covariance matrix $= (m1 + m2 + m3 + m4) / 4 = \begin{bmatrix} 10.5 & -8.25 \\ -8.25 & 17.25 \end{bmatrix}$

Previous Year Question (PCA)

- **Step-05:**
- **Calculate the eigen values and eigen vectors of the covariance matrix.**
- **λ is an eigen value for a matrix M if it is a solution of the characteristic equation $|M - \lambda I| = 0$.**

$$\left|\begin{bmatrix} 10.5 & -8.25 \\ -8.25 & 17.25 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right| = 0$$

- $(10.5 - \lambda)(17.25 - \lambda) - (-8.25)(-8.25) = 0$
- $113.0625 - 27.75\lambda + \lambda^2 = 0$
- Solving this quadratic equation, we get $\lambda = 22.7886, 4.96135$
- Thus, two eigen values are $\lambda 1 = 22.7886$ and $\lambda 2 = 4.96135$.
- Clearly, the second eigen value is very small compared to the first eigen value.
- So, the second eigen vector can be left out.
- Eigen vector corresponds to the greatest eigen value is principal component for the given data set.
- So. we find the eigen vector corresponding to eigen value $\lambda 1$.

Previous Year Question (PCA)

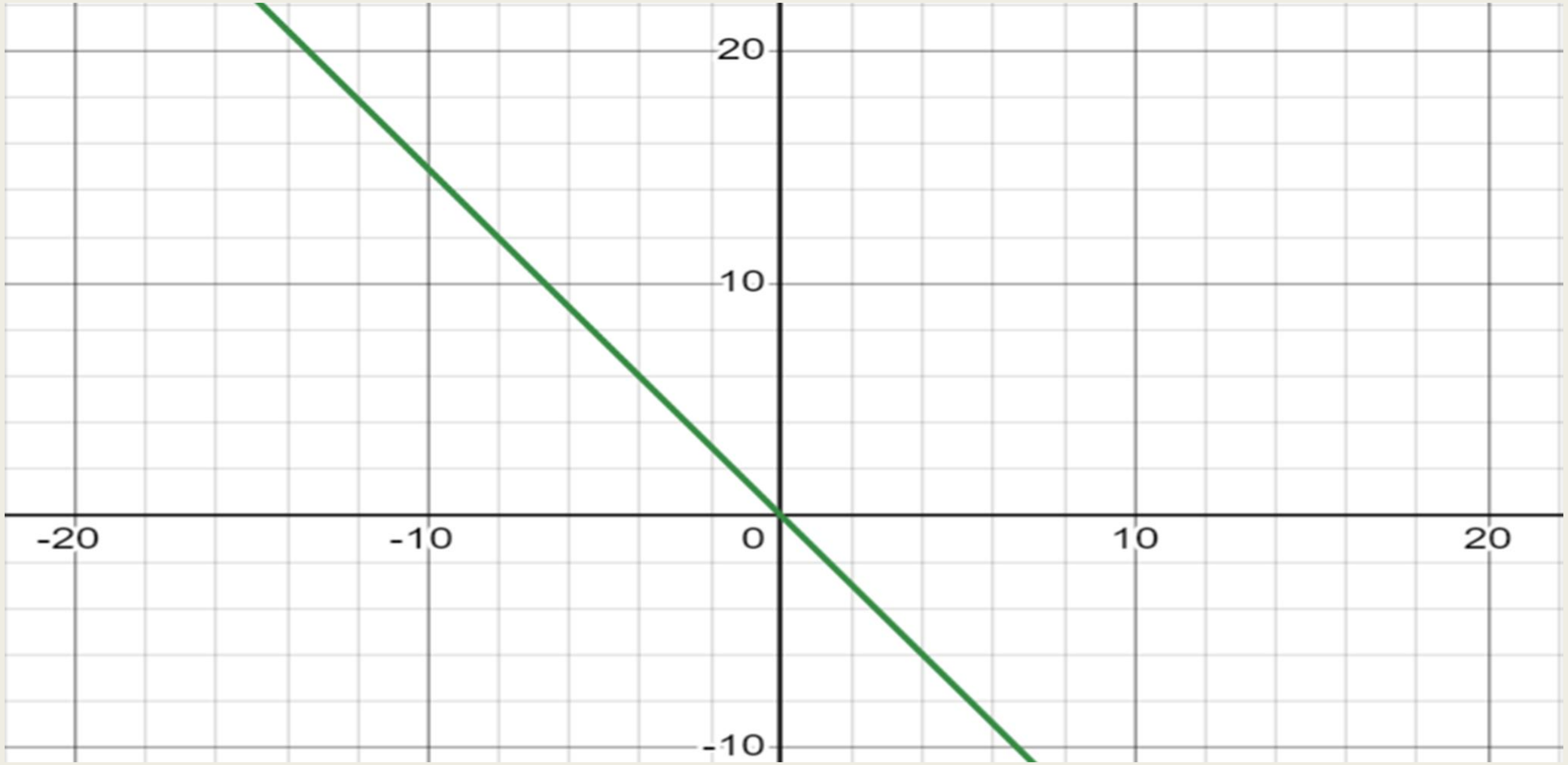
- We use the following equation to find the eigen vector-
- $MX = \lambda X$
- where- M = Covariance Matrix
-     X = Eigen vector
-     λ = Eigen value
- Substituting the values in the above equation, we get-

$$\begin{bmatrix} 10.5 & -8.25 \\ -8.25 & 17.25 \end{bmatrix} \begin{bmatrix} X1 \\ X2 \end{bmatrix} = 22.79 \begin{bmatrix} X1 \\ X2 \end{bmatrix}$$

- Solving these, we get-
- 10.5X1 - 8.25X2 = 22.79X1
- -8.25X1 + 17.25X2 = 22.79X2
- On simplification, we get-
- 12.29X1 = -8.25X2   ………(1)
- -8.25X1 = 5.54X2   ………(2)
- X1/5.54=X2/-8.25
- From (1) and (2), X1 = -0.67X2

Previous Year Question (PCA)

- From (2), the eigen vector is- $\begin{bmatrix} 5.54 \\ -8.25 \end{bmatrix}$
- Thus, principal component for the given data set is- $\begin{bmatrix} 5.54 \\ -8.25 \end{bmatrix}$
- Lastly, we project the data points onto the new subspace as-



Previous Year Question (PCA)

- 2. Explain the Principal Component Analysis (PCA) and reduce the following dataset step-by-step from 2 dimensions to 1.

Feature	Example 1	Example 2	Example 3	Example 4
x	2	1	0	-1
y	4	3	1	0.5

- **Ans: Homework.**

- 3. Create the reduced dimension of the given data from 4 to 2 using Principal Component Analysis (PCA).

Sampales	A	B	C	D
1	14	18	13	7
2	40	4	5	14
3	87	11	13	23
4	8	15	8	45

Previous Year Question (PCA)

- **Ans:-**
- **Step-01:**
- **Get data.**
- The given feature vectors are-
- x1 = (14, 18, 13, 7)
- x2 = (40, 4, 5, 14)
- x3 = (87, 11, 13, 23)
- x4 = (8, 15, 8, 45)

- **Step-02:**
- **Calculate the mean vector (μ).**
- Mean vector (μ) = ((14 + 40 + 87 + 8) / 4, (18 + 4 + 11 + 15) / 4, (13 + 5 + 13 + 8) / 4, (7 + 14 + 23 + 45) / 4) = (37.25, 12, 9.75, 22.25)

Previous Year Question (PCA)

- **Step-03:**
- **Subtract mean vector (μ) from the given feature vectors.**
- x1 – μ = (14-37.25, 18-12, 13-9.75, 7-22.25) = (-23.25, 6, 3.25, 15.25)
- x2 – μ = (40-37.25, 4-12, 5-9.75, 14-22.25) = (2.75, -8, -4.75, -8.25)
- x3 – μ = (87-37.25, 11-12, 13-9.75, 23-22.25) = (49.75, -1, 3.25, 0.75)
- x4 – μ = (8-37.25, 15-12, 8-9.75, 45-22.25) = (-29.25, 3, -1.75, 22.75)

- **Step-04:**
- **Calculate the covariance matrix.**
- m1 = (x1 – μ)(x1 – μ)T = (-23.25, 6, 3.25, 15.25)(-23.25, 6, 3.25, 15.25)T =

540.5625	-139.5	-75.5625	-354.5625
-139.5	36	19.5	91.5
-75.5625	19.5	10.5625	49.5625
-354.5625	91.5	49.5625	232.5625

- m2 = (x2 – μ)(x2 – μ)T = (2.75, -8, -4.75, -8.25)(2.75, -8, -4.75, -8.25)T =

7.5625	-22	-13.0625	-22.6875
-22	64	38	66
-13.0625	38	22.5625	39.1875
-22.6875	66	39.1875	68.0625

Previous Year Question (PCA)

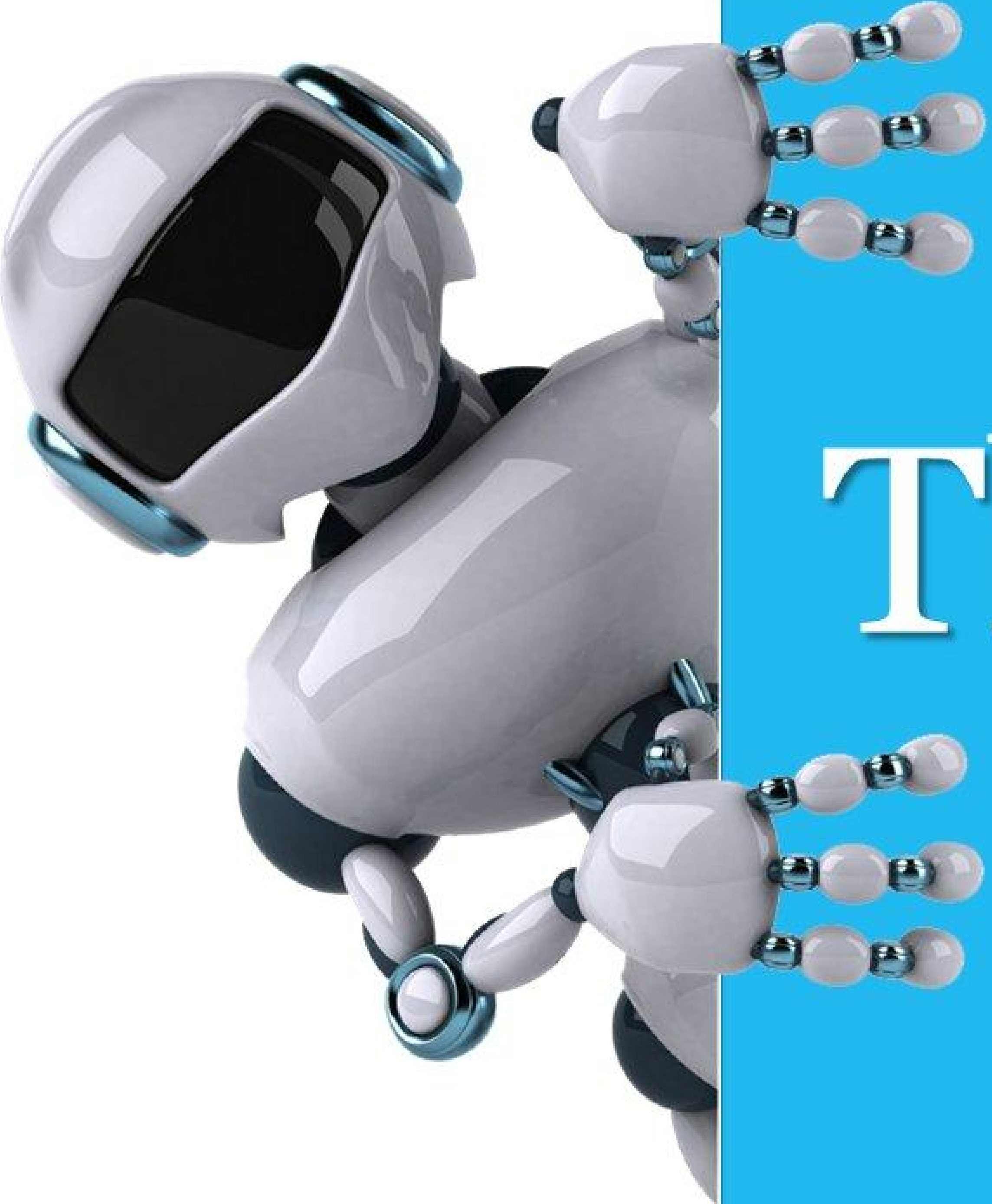
- $m3 = (x3 - \mu)(x3 - \mu)^T = (49.75, -1, 3.25, 0.75)(49.75, -1, 3.25, 0.75)^T =$

2475.0625	-49.75	161.6875	37.3125
-49.75	1	-3.25	-0.75
161.6875	-3.25	10.5625	2.4375
37.3125	-0.75	2.4375	0.5625

- $m4 = (x4 - \mu)(x4 - \mu)^T = (-29.25, 3, -1.75, 22.75)(-29.25, 3, -1.75, 22.75)^T =$

855.5625	-87.75	51.1875	-665.4375
-87.75	9	-5.25	68.25
51.1875	-5.25	3.0625	-39.8125
-665.4375	68.25	-39.8125	517.5625

- Covariance matrix = $(m1 + m2 + m3 + m4) / 4$
- **From here rest of the solution is Homework.**



Thank you