Roll no
2019111019

Date:
Page:

Statistical Methods in AI

Q1 a) Singular value decomposition is more generalizable.
SVD can be applied to any matrix of shape $m \times n$ ($m, n \in \mathbb{N}$)
However to apply Eigen value decomposition on a matrix X,
X must be diagonizable. That is, X should be a square matrix, and there must exist an invertible matrix Y of dimensions same as X, such that $Y \cdot X \cdot Y^{-1}$ is a diagonal matrix.

b) Let $M = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix}$

We want to find SVD for M.

$\Rightarrow M = U \Sigma V^T$

where $UU^T = VV^T = I = V^T V = U^T U$

and Σ is a diagonal matrix with non-negative entries.

We know $M^T M = V \Sigma^T \Sigma V^T$ since $U^T U = I$

$\Rightarrow M^T M \cdot V = V \Sigma^T \Sigma$

$\Rightarrow M^T M \cdot \underbrace{v_i}_{\text{column vector of } V} = v_i \sigma_i^2$

v_i is eigenvector of $M^T M$.

$$M^T M = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix} = \begin{bmatrix} 333 & 81 \\ 81 & 117 \end{bmatrix}$$

To find eigen vector and eigen value -

$\det(M^T M - \lambda I) = 0$

Date:
Page:

$$\det\begin{pmatrix} 333-\lambda & 81 \\ 81 & 117-\lambda \end{pmatrix} = 0$$

$$\Rightarrow \quad (333-\lambda)(117-\lambda) - 81^2 = 0$$

$$\Rightarrow \quad \lambda^2 - 450\lambda + 32400 = 0$$

$$\Rightarrow \quad (\lambda-90)(\lambda-360) = 0$$

So the eigenvalues are 90 and 360.
To find eigenvectors, we compute vectors in nullspace

$$M^TM - 90I = \begin{bmatrix} 243 & 81 \\ 81 & 27 \end{bmatrix} \qquad V = \begin{bmatrix} -1/\sqrt{10} \\ 3/\sqrt{10} \end{bmatrix}$$

$$M^TM - 360I = \begin{bmatrix} -27 & 81 \\ 81 & -243 \end{bmatrix} \qquad V = \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix}$$

So we got eigen values 360 and 90.
Using this we get Σ as:

$$\Sigma = \begin{bmatrix} \sqrt{360} & 0 \\ 0 & \sqrt{90} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \\ 0 & 0 \end{bmatrix}$$

We form V using eigenvectors corresponding to these eigen values.

$$V = \begin{bmatrix} 3/\sqrt{10} & -1/\sqrt{10} \\ 1/\sqrt{10} & 3/\sqrt{10} \end{bmatrix}$$

Date:
Page:

To find U, we use $MV = U\Sigma$

$$Mv_i = u_i \Sigma_i \Rightarrow u_i = \frac{Mv_i}{\Sigma_i}$$

Using this:

$$u_1 = \begin{pmatrix} 1/3 \\ 2/3 \\ 2/3 \end{pmatrix} \qquad u_2 = \begin{pmatrix} 2/3 \\ 1/3 \\ -2/3 \end{pmatrix}$$

u_3 = eigenvector for MM^T for eigenvalue 0.

$$\det(MM^T - \lambda I) = 0$$

$$\begin{vmatrix} 80-\lambda & 100 & 40 \\ 100 & 170-\lambda & 140 \\ 40 & 140 & 200-\lambda \end{vmatrix} = 0$$

$$-\lambda(\lambda-90)(\lambda-360) = 0$$

Needed vector is nullspace of:

$$MM^T - \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix} \qquad v = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix}$$

$$\therefore \quad U = \begin{bmatrix} 1/3 & 2/3 & 2/3 \\ 2/3 & 1/3 & -2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix}$$

So, $M = U\Sigma V^T$ is equal to:

$$\begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & 2 \end{bmatrix} = \begin{bmatrix} 1/3 & 2/3 & 2/3 \\ 2/3 & 1/3 & -2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 3/\sqrt{10} & -1/\sqrt{10} \\ 1/\sqrt{10} & 3/\sqrt{10} \end{bmatrix}^T$$

Q2 a) Correct options = BC

Explanation -

A - Diagonal elements of D are equal means all eigen values are equal. Hence the data would be equally spread in all eigen vector directions and PCA will not be useful.

B - If all elements of D are not equal, that means there exist unequal eigen values. Hence PCA will select eigenvectors based on rank of eigen values. So PCA will be useful.

C - While doing PCA, the line itself will become the principal eigen vector. Spread along rest perpendicular eigen vectors will be 0. So 0 eigen value will lead to ~~dete~~ D not being full rank.

D - V is always full rank.

E - This ~~is~~ isn't correct as we can take a set of points, do mean shifting and show that non-zero eigen values can be found. So D can be full rank.

Date:
Page:

b) False

PCA reduces dimensions but its main focus is reducing loss of information in general, and does not attempt to preserve the information useful in data classification.

Date:
Page:

Q3. a) Prior probability is the probability of an event occuring before any data is collected or some evidence is taken in account. Posterior probability is the updated probability after taking into consideration the new information.

eg Lets say probability of a group of people ~~person~~ having COVID is P. This is our prior probability. If we learn that the group of people have recieved both doses of their vaccines, then the probability of them having COVID won't be P anymore but will get updated because of the new evidence. This updated probability (P(COVID| 2 DOSES)) is the posterior probability.

b) P(Headache&sore throat) = 0.2

P (Flu) = 0.5

P (Headache&sore throat| Flu) = 0.9

P(Flu | Headache&sore throat)

$$= \frac{P(\text{Headaache\&sore throat} \mid \text{Flu}) \times P(\text{Flu})}{P(\text{Headache sore throat})}$$

$$= \frac{0.9 \times 0.5}{0.2}$$

$$= 0.225$$