Assignment - 4 221501031.

1) Principal component are linear combinations of the original features in a dataset. they are derived through the process called principal component analysis (PCA), which Transforms the data into a new coordinates system where the axes (principal components) capture the maximum variance in the data.

The principal components are ordered such that

- First principal component (~~F~~PC1)
- Second " " (PC2)

Importance in Dimensionality Reduction.

1. Variance Maximization.
2. Reduction in Redundancy.
3. Feature compression.
4. ~~R~~ Noise Reduction.
5. Visualization.

2. Steps Involved in Performing PCA are :-

1. Data standardization :-

This steps ensures that all feature contribute equally especially when they are on different scales.

2. Compute the covariance matrix :-

The covariance matrix quantifies the variance b/w the pair of features.

3. Calculate the eigen values + eigen vectors.

calculates the amount of variance & direction of the principal component

221501031

4. Select the Top k principal components:-

- Rank the principal components by the magnitude of the eigen values choose the top k components.

5. Transform the Data:-

- This step results in a reduced datased with fewer dimensions while preserving the most critical information.

3) find the eigenvalues of the correlation matrix:-

$$R = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}$$

we solve the characteristic equation:-

$$\det(R - \lambda I) = 0$$

$$\begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 0$$

$$\begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = 0$$

$$\begin{bmatrix} 1-\lambda & 0.6 \\ 0.6 & 1-\lambda \end{bmatrix} = 0$$

$$\Rightarrow (1-\lambda)(1-\lambda) - (0.6)(0.6) = 0$$

$$\Rightarrow (1-\lambda)^2 - (0.36) = 0$$

$$\Rightarrow 1 - 2\lambda + \lambda^2 - 0.36$$

$$\Rightarrow \lambda^2 - 2\lambda + 0.64 \quad \left[\frac{\pm b \pm \sqrt{b^2 - 4ac}}{2a} \right]$$

$$\lambda_1 = 1.6, \ \lambda_2 = 0.4$$

Part-B

1) Eigen values & eigen vectors calculation

221501031

The eigen value (λ) & eigen vector (v)

$$\Sigma v - \lambda v$$

eigen values :-

$\lambda_1 = 6.53$. $\lambda_2 = 3.83$, $\lambda_3 = 1.64$

eigen vector :-

$$v_1 = \begin{bmatrix} -0.823 \\ -0.325 \\ -0.466 \end{bmatrix}, v_2 = \begin{bmatrix} 0.154 \\ -0.917 \\ 0.369 \end{bmatrix}, v_3 = \begin{bmatrix} -0.547 \\ 0.232 \\ 0.804 \end{bmatrix}$$

b) The principal components are determined by the eigenvectors. Each principal component compounds to a direction of maximum variance in the data :-

- First principal component : Along the direction of v_1
- second principal component . Along the direction of v_2
- Third principal component . Along the direction of v_3

c) The Num of principal Component to Retain :- . .

(i) Total variance = sum of the eigen values

$$\lambda_1 + \lambda_2 + \lambda_3 = 6.53 + 3.83 + 1.64 = 12$$

(ii) variance by λ_1 :-

$$\frac{6.53}{12} = 54.4\%$$

cumulative variance after λ_1 & λ_2

$$\frac{6.53 + 3.83}{12} =$$

Variance ratio (PC2) $= \frac{1.64}{12} = (13.7\%)$

Variance Ratio (PC3) $= \frac{3.83}{12} = 0.319\ (31.9\%)$

Cummulative explained variance :-

After PC1 : 54.4%.
After PC2 : 68.1%.
After PC3 : 100%.

To capture at least 85% we need two components
that is $(PC_1 + PC_2) = 68.1\%$.

(2) $\Sigma = \begin{pmatrix} 6 & 2 & 1 \\ 2 & 5 & 0 \\ 1 & 0 & 4 \end{pmatrix}$

$\det(\Sigma - \lambda I) = 0$

$\det\left(\begin{pmatrix} 6-\lambda & 2 & 1 \\ 2 & 5-\lambda & 0 \\ 1 & 0 & 4-\lambda \end{pmatrix}\right) = 0$

$\Rightarrow (6-\lambda)\begin{bmatrix} 5-\lambda & 0 \\ 0 & 4-\lambda \end{bmatrix} - 2\begin{bmatrix} 2 & 0 \\ 1 & 4-\lambda \end{bmatrix} + 1\begin{bmatrix} 2 & 5-\lambda \\ 1 & 0 \end{bmatrix}$

$\Rightarrow (6-\lambda)((5-\lambda)(4-\lambda)) - 2 \times 2\,(4-\lambda) + 1 \times -(5-\lambda)$

$\lambda^3 - 15\lambda^2 + 74\lambda - 120 = 0$

$\Rightarrow \lambda_1 = 8,\ \lambda_2 = 4,\ \lambda_3 = 3$

By solving eigen vectors we get -

$V_1 = \begin{pmatrix} 0.707 \\ 0.707 \\ 0 \end{pmatrix},\ V_2 = \begin{pmatrix} -0.577 \\ 0.577 \\ 0.577 \end{pmatrix},\ V_3 = \begin{pmatrix} 0.408 \\ -0.408 \\ 0.816 \end{pmatrix}$

Contribution of each Variable 221501031

variable 1 = 16.67%.

variable 2 = 16.67%.

variable 3 = 16.67%

For second component (V_2)

variable 1 : 25%.

variable 2 = 100%.

variable 3 = 100%.

3) $\Sigma = \begin{bmatrix} 7 & 3 & 2 \\ 3 & 6 & 1 \\ 2 & 1 & 5 \end{bmatrix}$

$\det(\Sigma - \lambda I) = 0$

eigen values

$\lambda_1 = 10.42$

$\lambda_2 = 4.39$

$\lambda_3 = 3.19$

eigen vectors:-

$V_1 = \begin{bmatrix} 0.729 \\ 0.577 \\ 0.374 \end{bmatrix}, V_2 = \begin{bmatrix} 0.039 \\ -0.577 \\ 0.816 \end{bmatrix}, V_3 = \begin{bmatrix} 0.687 \\ -0.577 \\ -0.471 \end{bmatrix}$

(b) $PC_i = V_i^T x$

where v_i is the i^{th} eigen vector,

PC_1 is the component along V_1

PC_2 is the compont along V_2

PC_3 is the component along V_3

b) How many principal component to Retain for 90% variance

221501031

Total variance = $\lambda_1 + \lambda_2 + \lambda_3$

$= 8 + 4 + 3 = 15$

The 90% thrashfold is

$0.9 \times 15 = 13.5$

variance of λi

$\lambda_1 = \frac{8}{15}$

$\lambda_2 = \frac{4}{15}$

$\lambda_3 = \frac{3}{15}$

The cummulation variance : $\frac{\lambda_1}{T.V} + \frac{\lambda_2}{T.V}$

First component $\frac{8}{15} = \cancel{0.6667}\ 0.5333$

First two componend $\cdot \frac{8+4}{15} = \frac{12}{15} = 0.8$

Since we need to retain 90% first <u>two</u> components Is sufficient

(c) calculate the contribution of each variable In a retained principal component.

For the first compont $(v_1) = [0.707\ 0.707\ 0]^T$

$(0.707)^2 = 0.4998$

$(0.707)^2 = 0.4998$

$(0)^2 = 0$

for the second Principal conuponent

$v_2 = [-0.577\ \ 0.577\ \ 0.577]^T$

$= [-0.3329\ \ 0.3329\ \ 0.3329]$

b) How many principal component to Retain for 90% variance

221501031

Total variance = $\lambda_1 + \lambda_2 + \lambda_3$

$= 8 + 4 + 3 = 15$

The 90% threshold is

$0.9 \times 15 = 13.5$

variance of λi

$\lambda_1 = \frac{8}{15}$

$\lambda_2 = \frac{4}{15}$

$\lambda_3 = \frac{3}{15}$

The cumulation variance : $\frac{\lambda_1}{T.V} + \frac{\lambda_2}{T.V}$

First component : $\frac{8}{15} = \cancel{0.6667}\ 0.5333$

first two component : $\frac{8+4}{15} = \frac{12}{15} = 0.8$

since we need to retain 90% first two components Is sufficient

(c) calculate the contribution of each variable in a retained principal component.

for the first compont $(v_1) = [\lambda \lambda \lambda]^T$ $[0.707\ 0.707\ 0]^T$

$(0.707)^2 = 0.4998$

$(0.707)^2 = 0.4998$

$(0)^2 = 0$

for the second principal component.

$v_2 = [-0.577\ \ 0.577\ \ 0.577]^T$

$= [-0.3329\ \ 0.3329\ \ 0.3329]$

c) Variance contribution.

Total variance $= \lambda_1 + \lambda_2 + \lambda_3 = 10.42 + 4.39 + 3.19 = 18$

For PC_1

$$\frac{\lambda_1}{T.V} \times 100 = \frac{10.42}{18} \times 100 = 57.88\%$$

$$\frac{\lambda_2}{T.V} \times 100 = \frac{4.39}{18} \times 100 \approx 24.37\%$$

$$\frac{\lambda_3}{T.V} \times 100 = \frac{3.19}{18} \times 100 \approx 17.75\%$$

Cumulative variance: To capture 95% of the variance

After PC_1: 57.88%

After $PC_1 + PC_2$: $57.88 + 24.37 = 82.25$

$PC_1 + PC_2 + PC_3 = 100\%$

all three components are required to capture 95% of the variance in this case.