PAGE NO :
DATE :

4. Update centroids:
- For each cluster, compute the mean of the points assigned to it.
- Update the centroid with this mean

5. Convergence check:
- If cluster assignments do not change or centroids remain the same, stop

6. Output:
- Final centroids of all clusters
- Cluster labels for each data point.

Principal Component Analysis (PCA)

1. Standardize the Data.
- Given a dataset x of size $n \times d$,
$n \rightarrow$ no of samples.
$d \rightarrow$ number of features.

$$X_{centered} = X - \mu.$$

2. Compute the covariance Matrix:

- Calculate the covariance matrix of the centered data. This matrix captures the relationship between different features

$$C = \frac{1}{n-1} X^T_{centered} X_{centered}.$$

PAGE NO :
DATE :

3. Compute the Eigenvalues & Eigenvectors
- Find the eigenvalues and eigenvectors
Eigenvectors (directions of the new feature space)

Eigenvalues (variance)

4. Sort Eigenvalues and Eigenvectors:
- Sort the eigenvalues in descending order

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d.$$

5. Select the Top K Eigenvectors.
- Choose the top K eigenvectors corresponding to the K largest eigenvalues. These eigenvectors form a new basis for the reduced feature space.

6. Construction the Projection Matrix:
- Form a matrix W

$$W = [v_1, v_2, v_3 \ldots v_k]$$

7. Project the Data:
- Multiply the centered data $X_{centered}$ by the projection matrix W to obtain reduced dataset in the new K-dimensional space

21/4/25

$$X_{reduced} = X_{centered} W$$