Linear Algebra for Machine Learning
Lebanese University - Faculty of Sciences
Principal Component Analysis (PCA)

By Dr. Youssef SALMAN

Step 1. Centering the Data

Let

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

be the data matrix with n observations (rows) and p variables (columns).

The mean of each variable j is

$$\mu_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \qquad j = 1, \dots, p.$$

Define the vector of means

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}, \quad \text{and let } \mathbf{1} = [1, 1, \dots, 1]^\top \in \mathbb{R}^n.$$

The **centered data matrix** is obtained by removing the column means:

$$X_c = X - \mathbf{1}\boldsymbol{\mu}^\top.$$

Now every column of X_c has mean zero:

$$\frac{1}{n}\mathbf{1}^\top X_c = \mathbf{0}.$$

Step 2. Computing the Covariance Matrix

The sample covariance matrix of the centered data is

$$S = \frac{1}{n-1} X_c^\top X_c.$$

Each element s_{jk} of S is

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \mu_j)(x_{ik} - \mu_k),$$

and represents the covariance between variables j and k.

The matrix S is symmetric ($S = S^\top$) and positive semidefinite, which guarantees real non-negative eigenvalues.

Step 3. Eigenvalue Decomposition

We solve the eigenvalue problem:

$$S v_i = \lambda_i v_i, \qquad i = 1, \ldots, p,$$

where

- λ_i is the i^{th} eigenvalue of S,
- v_i is the corresponding eigenvector (of unit norm),
- all eigenvectors are orthogonal: $v_i^\top v_j = 0$ for $i \neq j$.

Arrange the eigenvalues in decreasing order:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0.$$

Collect the eigenvectors in the matrix

$$V = [v_1 \ v_2 \ \ldots \ v_p],$$

so that

$$S = V \Lambda V^\top, \quad \Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p).$$

The columns of V represent the directions (axes) of the **principal components**.

—

Step 4. Forming the Loading Matrix

The matrix V is orthonormal:

$$V^\top V = V V^\top = I_p.$$

Each column v_i defines a new axis in the transformed coordinate system. The matrix V is sometimes called the *loading matrix*, since it shows how each principal component is formed as a linear combination of the original variables.

If X_c is your centered data, the original variables can be expressed as combinations of the new axes:

$$X_c = ZV^\top,$$

where Z are the new coordinates (the principal component scores).

Step 5. Projection onto the New Axes (Principal Components)

To transform the data into the new coordinate system, we project each observation onto the eigenvector directions:

$$Z = X_c V.$$

- $Z \in \mathbb{R}^{n\times p}$ is the matrix of **scores**.
- Each column $z_i = X_c v_i$ is the i^{th} **principal component**.
- Each row of Z gives the coordinates of an observation in the PCA space.

Goal of this step: re-express the data using uncorrelated variables (principal components) that capture decreasing amounts of variance.

Geometrically, this corresponds to rotating the coordinate system so that the new axes align with directions of maximum variance.

Step 6. Variance of the Principal Components and Explained Variance

The covariance of the new variables Z is

$$\mathrm{Cov}(Z) = \frac{1}{n-1} Z^\top Z = \frac{1}{n-1}(V^\top X_c^\top X_c V) = V^\top S V.$$

Since $S = V\Lambda V^\top$, we get

$$\mathrm{Cov}(Z) = \Lambda.$$

Therefore:

- The variance of the i^{th} principal component is λ_i.

- The components are uncorrelated because $\mathrm{Cov}(Z)$ is diagonal.

The **total variance** of the data is

$$\mathrm{tr}(S) = \sum_{i=1}^{p} \lambda_i.$$

The proportion of variance explained by the i^{th} component is

$$\mathrm{EVR}_i = \frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j},$$

known as the **explained variance ratio**.

Interpretation:

- Large λ_i means the component captures more of the data's variability.
- The first few components (largest λ_i) usually explain most of the information.
- Dimensionality reduction: keep only $k < p$ components such that $\sum_{i=1}^{k} \mathrm{EVR}_i$ reaches a desired threshold (e.g. 90%).

Summary of All Steps

1. Compute column means and center the data: $X_c = X - \mathbf{1}\mu^\top$.
2. Compute covariance matrix: $S = \frac{1}{n-1} X_c^\top X_c$.
3. Solve eigenvalue problem $Sv_i = \lambda_i v_i$.
4. Form orthonormal loading matrix $V = [v_1, \ldots, v_p]$.
5. Compute scores (projection): $Z = X_c V$.
6. Verify variances and compute explained variance ratios: $\mathrm{Cov}(Z) = \Lambda$, $\mathrm{EVR}_i = \lambda_i / \sum_j \lambda_j$.

These steps constitute the mathematical foundation of Principal Component Analysis (PCA). The transformation $Z = X_c V$ decorrelates the variables and orders the new axes according to decreasing variance, enabling compression, visualization, and noise reduction while preserving as much information as possible.