Principal Component Analysis (PCA)

Thapar Institute of Engineering & Technology, Patiala

Introduction to PCA

- PCA is used to reduce the dimensions of a given problem to encode the basic features of a multivariate data (eg. an image) by using a few appropriate variables (dimensions).
- Suitable linear transformation (change of basis).
- PCA de-correlates the original data by finding the directions (dimensions) in which the variance is maximized.

Constructing the data matrix

Data matrix: $$X = \begin{pmatrix} | & | & & | \\ \mathbf{x_1} & \mathbf{x_2} & \cdots & \mathbf{x_n} \\ | & | & & | \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{x}}_1 \\ \tilde{\mathbf{x}}_2 \\ . \\ . \\ . \\ \tilde{\mathbf{x}}_m \end{pmatrix}$$

represents the mean subtracted data where each $\begin{matrix} | \\ \mathbf{x_k} \\ | \end{matrix}$ represents an m dimensional data point (vector) and there are n of them corresponding to n samples, and each of $\tilde{\mathbf{x}}_k = \left(\tilde{x}_{k,1} \;\; \tilde{x}_{k,2} \;\; \cdots \; \tilde{x}_{k,n}\right)$ represents the n sample observations (data points) of the k^{th} dimension.

Mean subtraction: $\tilde{\mathbf{x}}_k \longleftarrow \left(\tilde{\mathbf{x}}_k - \sum_{i=1}^{n} \frac{\tilde{x}_{k,i}}{n}\right)$.

Covariance matrix of the dataset

The covariance matrix of the mean subtracted data is

$$C_X = \frac{1}{n-1} XX^T = \frac{1}{n-1} \begin{pmatrix} \tilde{\mathbf{x}}_1\tilde{\mathbf{x}}_1^T & \tilde{\mathbf{x}}_1\tilde{\mathbf{x}}_2^T & \cdots & \tilde{\mathbf{x}}_1\tilde{\mathbf{x}}_m^T \\ \tilde{\mathbf{x}}_2\tilde{\mathbf{x}}_1^T & \tilde{\mathbf{x}}_2\tilde{\mathbf{x}}_2^T & \cdots & \tilde{\mathbf{x}}_2\tilde{\mathbf{x}}_m^T \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \tilde{\mathbf{x}}_m\tilde{\mathbf{x}}_1^T & \tilde{\mathbf{x}}_m\tilde{\mathbf{x}}_2^T & \cdots & \tilde{\mathbf{x}}_m\tilde{\mathbf{x}}_m^T \end{pmatrix}. \qquad (1)$$

- Optimal representation of the data should be s.t. off-diagonal terms of the covariance matrix must be zero.
- Retain those dimensions that lend themselves to exhibit the greatest *variability* within the data (maximal *variance* presents richer information). The diagonal entries of the covariance matrix in the transformed space should be as large as possible.

Linear transformation and principal components

X is transformed into a new representation Y (also $m \times n$ matrix) by a change of basis matrix P of dimensions $m \times m$ as follows:

$$Y = PX = \begin{pmatrix} \mathbf{p_1} \cdot \mathbf{x_1} & \mathbf{p_1} \cdot \mathbf{x_2} & \cdots & \mathbf{p_1} \cdot \mathbf{x_n} \\ \mathbf{p_2} \cdot \mathbf{x_1} & \mathbf{p_2} \cdot \mathbf{x_2} & \cdots & \mathbf{p_2} \cdot \mathbf{x_n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \mathbf{p_m} \cdot \mathbf{x_1} & \mathbf{p_m} \cdot \mathbf{x_2} & \cdots & \mathbf{p_m} \cdot \mathbf{x_n} \end{pmatrix}. \tag{2}$$

Here $P = \begin{pmatrix} ---\mathbf{p_1}--- \\ ---\mathbf{p_2}--- \\ \cdot \\ \cdot \\ \cdot \\ ---\mathbf{p_m}--- \end{pmatrix}$ where $---\mathbf{p_k}---$ form the new basis and are called principal components.

How to find the change of basis matrix P?

Covariance matrix of the transformed data C_Y by

1. maximizing the diagonal entries of C_Y (maximizing variance of every dimension), and
2. minimizing the off-diagonal entries of C_Y (minimizing covariance between dimensions).

What does the above mean?

- Find P that makes C_Y diagonal.
- Since the columns of P must form the new basis, we must have that P is an orthonormal matrix ($PP^T = I$).

Mathematically, what is P?

$C_Y = \frac{1}{n-1} YY^T = \frac{1}{n-1}(PX)(PX)^T = \frac{1}{n-1}(PX)(X^T P^T)$
$= \frac{1}{n-1} P(XX^T)P^T = PSP^T$
where $S = \frac{1}{n-1} XX^T$ is an $m \times m$ symmetric positive definite matrix and hence diagonizable: $S = EDE^T$ where E is an $m \times m$ orthonormal matrix whose columns are the eigenvectors of S and D is a diagonal matrix with the respective eigenvalues along the diagonal.

$P = E^T$ is the desired transformation matrix!

$C_Y = PSP^T = E^T(EDE^T)E = D$ because $E^T E = PP^T = I$.

How to find eigenvalues of S?

- By hand!
- By Singular Value Decomposition (SVD $\longleftarrow$ good idea!

How to recover the original data?

$$Y = PX = V^T X, \boxed{X = VY}.$$