Principal Component Analysis

- Dimensionality reduction technique -→ to alleviate curse of dimensionality.

- Concepts from projection in linear algebra

Consider a data set of observations $\{\mathbf{x}_n\}$ where $n = 1, \ldots, N$, and $\mathbf{x}_n$ is a Euclidean variable with dimensionality D. Our goal is to project the data onto a space having dimensionality $M < D$ while maximizing the variance of the projected data.

Principal component analysis, or PCA, is a technique that is widely used for applications such as dimensionality reduction, lossy data compression, feature extraction, and data visualization (Jolliffe, 2002). It is also known as the *Karhunen-Loève* transform.

PCA can be defined as the orthogonal projection of the data onto a lower dimensional linear space, known as the *principal subspace*, such that the variance of the projected data is maximized

$$z = \mathbf{w}^T \mathbf{x}$$

We assume that the training samples have zero mean

$$z_1 = \mathbf{w}_1^T \mathbf{x}$$

$$\mathrm{Var}(z_1) = \mathbf{w}_1^T \Sigma \mathbf{w}_1$$

We seek $\mathbf{w}_1$ such that $\mathrm{Var}(z_1)$ is maximized subject to the constraint that $\mathbf{w}_1^T \mathbf{w}_1 = 1$. Writing this as a Lagrange problem, we have

$$\max_{\mathbf{w}_1} \mathbf{w}_1^T \Sigma \mathbf{w}_1 - \alpha(\mathbf{w}_1^T \mathbf{w}_1 - 1)$$

Taking the derivative with respect to $\mathbf{w}_1$ and setting it equal to 0, we have

$$2\Sigma\mathbf{w}_1 - 2\alpha\mathbf{w}_1 = 0, \text{ and therefore } \Sigma\mathbf{w}_1 = \alpha\mathbf{w}_1$$

we choose the eigenvector with the largest eigenvalue for the variance to be maximum. Therefore the principal component is the eigenvector of the covariance matrix of the input sample with the largest eigenvalue, $\lambda_1 = \alpha$.

$$\mathbf{w}_1^T\Sigma\mathbf{w}_1 = \alpha\mathbf{w}_1^T\mathbf{w}_1 = \alpha$$

The second principal component, $\boldsymbol{w}_2$, should also maximize variance, be of unit length, and be orthogonal to $\boldsymbol{w}_1$. This latter requirement is so that after projection $z_2 = \boldsymbol{w}_2^T \boldsymbol{x}$ is uncorrelated with z_1. For the second principal component, we have

$$\max_{\boldsymbol{w}_2} \boldsymbol{w}_2^T \boldsymbol{\Sigma} \boldsymbol{w}_2 - \alpha(\boldsymbol{w}_2^T \boldsymbol{w}_2 - 1) - \beta(\boldsymbol{w}_2^T \boldsymbol{w}_1 - 0)$$

Taking the derivative with respect to $\boldsymbol{w}_2$ and setting it equal to 0, we have

$$2\boldsymbol{\Sigma}\boldsymbol{w}_2 - 2\alpha\boldsymbol{w}_2 - \beta\boldsymbol{w}_1 = 0$$

$$2\boldsymbol{w}_1^T\boldsymbol{\Sigma}\boldsymbol{w}_2 - 2\alpha\boldsymbol{w}_1^T\boldsymbol{w}_2 - \beta\boldsymbol{w}_1^T\boldsymbol{w}_1 = 0$$

We have,

$$\mathbf{w}_1^T \mathbf{w}_2 = 0.$$

$$\mathbf{w}_1^T \Sigma \mathbf{w}_2 = \mathbf{w}_2^T \Sigma \mathbf{w}_1 = \lambda_1 \mathbf{w}_2^T \mathbf{w}_1 = 0$$

Hence

$$\beta = 0$$

$$\Sigma \mathbf{w}_2 = \alpha \mathbf{w}_2$$

which implies that $\mathbf{w}_2$ should be the eigenvector of $\boldsymbol{\Sigma}$ with the second largest eigenvalue, $\lambda_2 = \alpha$. Similarly, we can show that the other dimensions are given by the eigenvectors with decreasing eigenvalues.

Because $\boldsymbol{\Sigma}$ is symmetric, for two different eigenvalues, the eigenvectors are orthogonal. If $\boldsymbol{\Sigma}$ is positive definite ($\boldsymbol{x}^T\boldsymbol{\Sigma}\boldsymbol{x} > 0$, for all nonnull $\boldsymbol{x}$), then all its eigenvalues are positive. If $\boldsymbol{\Sigma}$ is singular, then its rank, the effective dimensionality, is k with $k < d$ and $\lambda_i, i = k+1, \ldots, d$ are 0 (λ_i are sorted in descending order). The k eigenvectors with nonzero eigenvalues are the dimensions of the reduced space. The first eigenvector (the one with the largest eigenvalue), $\boldsymbol{w}_1$, namely, the principal component, explains the largest part of the variance; the second explains the second largest; and so on.

$$\boldsymbol{z} = \mathbf{W}^T(\boldsymbol{x} - \boldsymbol{m})$$

where the k columns of $\mathbf{W}$ are the k leading eigenvectors of $\mathbf{S}$, the estimator to $\boldsymbol{\Sigma}$. We subtract the sample mean $\boldsymbol{m}$ from $\boldsymbol{x}$ before projection to center the data on the origin. After this linear transformation, we get to a k-dimensional space whose dimensions are the eigenvectors, and the variances over these new dimensions are equal to the eigenvalues

- Define additional Principal Components in an incremental fashion

- Choose a new direction that maximize the projected variance and orthogonal to those already considered.