principal component analysis (PCA)

Authors

Alexander Jung*, Department of Computer Science, Aalto University, Espoo, Finland
*Corresponding author: alex.jung@aalto.fi

Abstract

Principal component analysis (PCA) is a dimensionality reduction method that maps the feature vectors $\mathbf{x} \in \mathbb{R}^d$ of a dataset to shorter vectors $\mathbf{z} = \mathbf{W}\mathbf{x} \in \mathbb{R}^{d'}$ by a linear feature transformation. The matrix $\mathbf{W}$ is chosen so that the original feature vectors can be reconstructed from $\mathbf{z}$ by a linear map with the minimum sum of squared errors over the dataset. The rows of the optimal $\mathbf{W}$ are the eigenvectors of the sample covariance matrix of the centered feature vectors that belong to its d' largest eigenvalues, so PCA is computed by an eigenvalue decomposition (EVD). PCA is a form of empirical risk minimization (ERM) with the squared reconstruction error as loss, and an autoencoder with linear encoder and decoder learns the same subspace. A typical application is the visualization of a dataset with thousands of features in a scatterplot of its first two principal components.

Definition

Consider a dataset

$$\mathcal{D} = \big\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}\big\}$$

consisting of data points characterized by feature vectors $\mathbf{x}^{(r)} \in \mathbb{R}^d$ for $r = 1, \ldots, m$. PCA determines, for a given number $d' < d$, a linear feature transformation

$$\boldsymbol{\Phi}^{(\mathbf{W})} : \mathbb{R}^d \to \mathbb{R}^{d'} : \mathbf{x} \mapsto \mathbf{W}\mathbf{x}$$

such that the new feature vectors $\mathbf{z}^{(r)} = \mathbf{W}\mathbf{x}^{(r)}$ allow reconstructing the original features with minimum linear reconstruction error (Bishop, 2006; Hastie et al., 2009; Jung, 2022):

$$\min_{\mathbf{R} \in \mathbb{R}^{d \times d'}} \sum_{r=1}^{m} \Big\| \mathbf{x}^{(r)} - \mathbf{R}\mathbf{W}\mathbf{x}^{(r)} \Big\|_2^2.$$

PCA can be viewed as a form of empirical risk minimization (ERM) using the loss function $L\left(\mathbf{x}, \mathbf{W}\right) = \Big\| \mathbf{x} - \widehat{\mathbf{R}}\mathbf{W}\mathbf{x} \Big\|_2^2$ with a reconstruction matrix $\widehat{\mathbf{R}}$ that achieves the above minimum reconstruction error. It turns out that this ERM problem can be solved by a matrix $\mathbf{W} = \big(\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(d')}\big)^T$ whose rows are given by d' eigenvectors corresponding to the d' largest eigenvalues of the matrix:

$$\widehat{\mathbf{Q}} = \frac{1}{m} \sum_{r=1}^{m} \mathbf{x}^{(r)} \big(\mathbf{x}^{(r)}\big)^T = \mathbf{X}^T\mathbf{X}.$$

Note that $\widehat{\mathbf{Q}}$ coincides with the sample covariance matrix of $\mathcal{D}$ if its sample mean vanishes. The positive semi-definite (psd) matrix $\widehat{\mathbf{Q}}$ allows for an eigenvalue decomposition (EVD) of the following form (Horn and Johnson, 2013; Meyer, 2000):

$$\widehat{\mathbf{Q}} = \sum_{j=1}^{d} \lambda_j \mathbf{u}^{(d)} \big(\mathbf{u}^{(d)}\big)^T = \begin{pmatrix} \mathbf{u}^{(1)} & \cdots & \mathbf{u}^{(d)} \end{pmatrix} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix} \begin{pmatrix} \big(\mathbf{u}^{(1)}\big)^T \\ \vdots \\ \big(\mathbf{u}^{(d)}\big)^T \end{pmatrix}.$$

This decomposition consists of decreasing nonnegative eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d \geq 0$ and corresponding eigenvectors $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(d)}$ that form an orthonormal basis of $\mathbb{R}^d$. A common application is visualizing high-dimensional gene expression data by projecting thousands of measured gene activities onto two or three principal components that capture the dominant sources of variation.

Cross References

feature transformation, feature learning, dimensionality reduction, EVD

References

1. Bishop (2006). Pattern Recognition and Machine Learning. Springer Science+Business Media, New York, NY, USA.

2. Horn and Johnson (2013). Matrix Analysis. 2nd ed., Cambridge Univ. Press, New York, NY, USA.

3. Jung (2022). Machine Learning: The Basics. Springer Nature, Singapore, Singapore.

4. Meyer (2000). Matrix Analysis and Applied Linear Algebra. SIAM, Philadelphia, PA, USA.

5. Hastie et al. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed., Springer Science+Business Media, New York, NY, USA.