Principal Component Analysis (PCA)

Oliver Zhao

1 Description

Principal component analysis (PCA) uses orthogonal transformations to create a set of linearly uncorrelated values called principal components. Each principal component is a linear combination of the different features of the original data set. Consequently, PCA may not be sufficient for data sets with complex nonlinear relationships. The first principal component captures the maximal amount of variance in the data set, with subsequent principal components capturing less and less of the variance. This method is mainly used for dimension reduction, so that computations in models can be performed more efficiently. A less common usage is for visualizing high-dimension data sets in 2D or 3D.

For example, consider a set of features $X_1, \dots, X_p$ in a p-dimensional data set. If we have n samples, then our data set can be expressed in a matrix $\mathbf{X}$ where its dimensions are $n \times p$. It is assumed that the variables in $\mathbf{X} = (X_1, \dots, X_p)$ are centered around zero. The first principal component is the normalized linear combination of the features

$$Z_1 = w_{11}X_1 + w_{21}X_2 + \cdots + w_{p1}X_p, \tag{1.1}$$

That has the largest variance. The elements $w_{11}, \dots, w_{p1}$ are known as the weights of the first principal component, where $\mathbf{w_1} = (w_{11}, \dots, w_{[1}^T)$ is the weight vector of the first principal component. The first principle component weight vector solves the following optimization problem

$$\underset{||\mathbf{w_1}||=1}{\arg\max} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{p} w_{j1} x_{ji} \right)^2 \right\}. \tag{1.2}$$

The optimization problem can be solved through eigen decomposition of the covariance matrix $\mathbf{A}$ of the data set. The covariance of two features X and Y is defined as

$$cov(X,Y) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{x})(Y_i - \bar{y}), \tag{1.3}$$

Where $\bar{x}$ and $\bar{y}$ are the averages of the variables X and Y, and X_i and Y_i are the ith elements of the variables X and Y. Note that we use the notation X and Y for our variables instead of X_1 and X_2 for the equation above for clarity. The covariance matrix is a square matrix of $p \times p$ dimensions, where element a_{ij} is defined as

$$a_{ij} = cov(X_i, X_j). \tag{1.4}$$

Notice that the covariance matrix is a symmetric matrix since $a_{ij} = a_{ji}$. The eigenvalues of the covariance matrix $\mathbf{A}$ are the roots of the characteristic equation

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \mathbf{0}. \tag{1.5}$$

Where I is the identity matrix. We shall denote the eigenvalues in decreasing order as $\lambda_1, \dots, \lambda_p$, with the corresponding eigenvectors as $\mathbf{v}_1, \dots, \mathbf{v}_p$. The number of principal components we wish to isolate is equal to the number of eigenvectors we shall use to construct our eigenvector matrix $\mathbf{W}$. It we want k principal components, our eigenvector matrix will be a $p \times k$ matrix defined as

$$\mathbf{W} = [\mathbf{v}_1 \dots \mathbf{v}_k]. \tag{1.6}$$

Note that the direction of the eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ are in the same direction of the corresponding principal components. However, we must normalize the magnitude so that the vectors are unit vectors, hence the weight vector $\mathbf{w}_i$ of the i-th principal component is computed as

$$\mathbf{w_i} = \mathbf{v_i} / ||\mathbf{v_i}||. \tag{1.7}$$

Additionally, the fraction of the original data set variance that the i-th principal component captures is calculated as

$$\text{Fraction of Variance Captured} = \frac{\lambda_i}{\sum_{j=1}^{k} \lambda_j}. \tag{1.8}$$

In order to transform the original data set $\mathbf{X}$ into a new data set $\mathbf{X_{new}}$, we apply the following transformation

$$\mathbf{X_{new}} = (\mathbf{W}^T \mathbf{X}^T)^T, \tag{1.9}$$

Resulting in a $n \times k$ transformed data set with k features and n samples, with each row containing a different sample and each column containing a different feature.

2 Algorithm

Description: Consider a matrix $\mathbf{X}$ that characterizes a data set, where the mean of each column is centered at zero and there are n rows for the number of samples and p columns for the number of features. The weights associated with the i-th principal component are stored in the vector $\mathbf{w}_i = (w_{i1}, \dots, w_{ip})$.

---

1. Find the covariance matrix $\mathbf{A}$, where its elements a_{ij} is defined as

$$a_{ij} = cov(X_i, X_j). \tag{2.1}$$

2. Perform eigenvalue decomposition on the covariance matrix $\mathbf{A}$ and solve for the eigenvalues $\lambda_1, \dots, \lambda_p$, where

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \mathbf{0}. \tag{2.2}$$

3. Solve for the corresponding eigenvectors $\mathbf{v_1}, \dots, \mathbf{v_p}$, where

$$\mathbf{A}\mathbf{v_i} = \lambda_i \mathbf{v_i}. \tag{2.3}$$

4. Determine the number of principal components desired for the data set transformation and then transform the data set, where

$$\mathbf{X_{new}} = (\mathbf{W}^T \mathbf{X}^T)^T \tag{2.4}$$

5. Implement dimension-reduced data set in machine learning model or plot for data visualization.