Principal Component Analysis

Prerequisite-

- Machine learning fundamentals and feature space
- Linear algebra and statistics

Objectives-

- Motivation for principal component analysis(PCA).
- Understanding what is PCA.
- Steps to compute PCA and Example.

Motivation and Meaning-

Machine learning model efficiency largely depends on the amount of data. The more and better the data we will have a better chance to train an efficient predictive model. However that large data also comes with certain limitations like the curse of dimensionality. With a high dimensional data there is a high possibility to have inconsistency and redundancy in data values. This will pose complexity over computation and analysis processing of data. To mitigate this limitation, a technique is proposed to reduce the dimensions while selecting most significant features of data, here principal component analysis comes into play.

Principal component Analysis is a method to identify the patterns in data using their similarities and dissimilarities between the sample points. The patterns within data are hard to find especially when we cannot visualize it graphically(the data is in high dimension). The principal component analysis is a powerful tool to explore data with its hidden patterns and reduce the dimensions. The role of PCA can be understood by a toy example shown in figure 1. Let us be interested in studying the motion of the ideal spring. To do that we have attached a ball of mass m to a frictionless spring and release it with a small distance from the equilibrium. As a result the spring starts oscillating and because it is frictionless so it will oscillate indefinitely along the axis at a set frequency. Because initially we don't know how many dimensions are important so we assume to record the two dimensional motion of the ball from three dimensions by placing three cameras around the system and project them. This will generate the three different distributions of oscillation and the big question is how do we get a simple equation of x from this data. [*Figure Source: A tutorial on principal component analysis- Jonathon Shlens*]

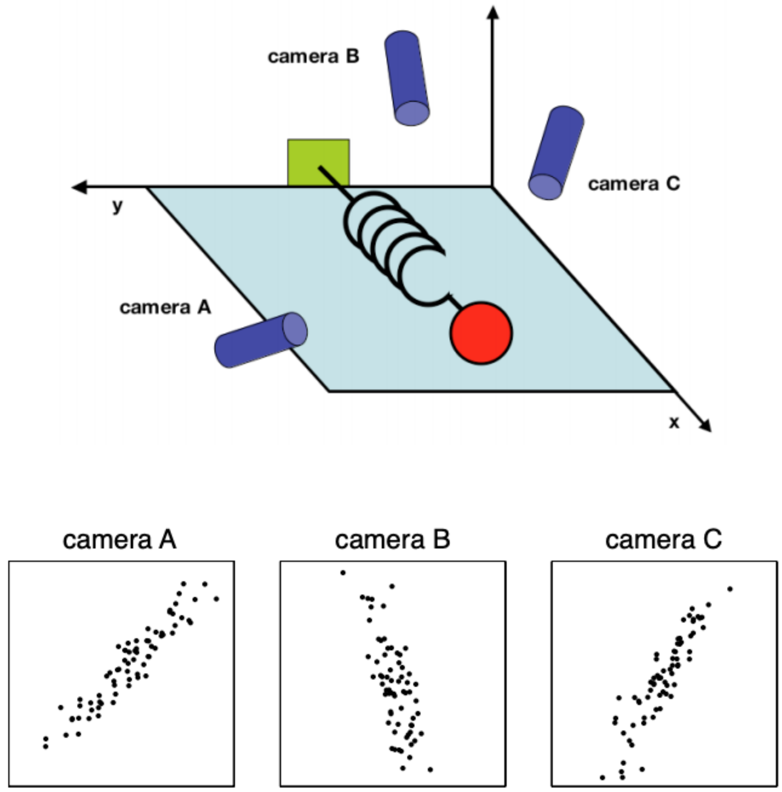

Figure 1 A toy example to understand PCA

The main goal of principal component analysis is to explore the meaningful basis to project a dataset. The hope is that the new basis will remove the noise and explore hidden patterns within data. The intuition to re-express the data to a new basis is understood through figure 2 which shows the data from camera A point of view. It is clear from the distribution that the largest direction of variation is not along the original basis of recording (x, y) but along the new line called the best fit line. The figure represents the noise and signal variance by two lines on data.

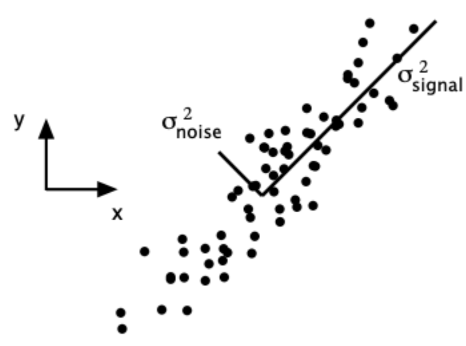

Figure 2 Change of basis (Camera A projection)

So, we can understand principal component analysis (PCA) in many ways like:

- Principal Component Analysis is a mathematical approach, which is used for better interpretation of your data.
- The main purpose of Principal Component Analysis is to reduce high dimensional data into low dimensional space.
- Principal Component Analysis is an unsupervised machine learning technique which finds insights of data without having prior knowledge. It reduces data by projecting geometrically onto a lower dimensional basis known as principal components.

When to use PCA

- When you have to reduce the number of features or dimensions in the data.
- When you have to check whether the features are independent of each other or not.

Steps of principal Component Analysis

The steps of principal component analysis are summarized as shown in following chart-

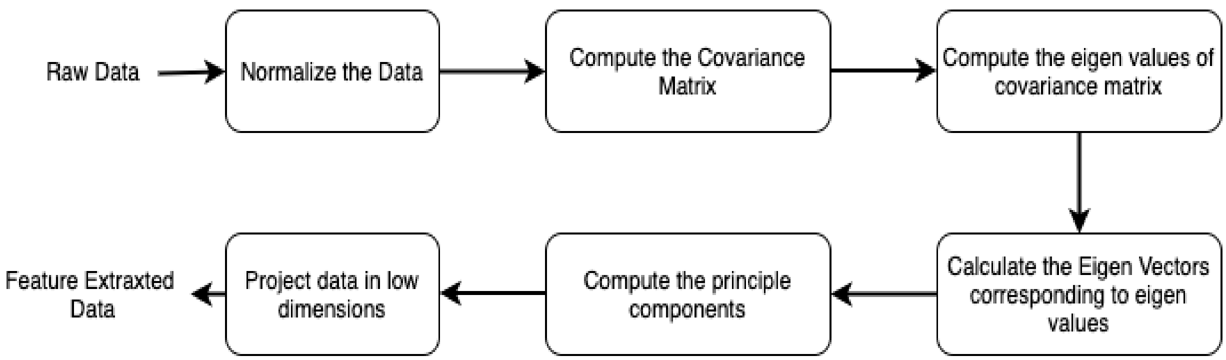

Figure 3 PCA steps

Step 1- Normalize is one of the fundamental steps of data processing, and aims to get an unbiased result of the model. It is the process to bring all the data variables within the same range of values (-1, 1). The normal value of variable is calculated using:

$$z = \frac{x - \bar{x}}{\sigma}$$

where $\bar{x}$ *is the mean and* σ *is standard deviation of distribution.*

Step 2- Second step to compute PCA is to calculate covariance matrix. A covariance matrix describes the correlation between variables within a dataset. It helps to identify heavily dependent variables. Mathematically a covariance matrix for three features data set may be defined as:

$$C = [cov(x,x)\ cov(x,y)\ cov(x,z)\ cov(y,x)\ cov(y,y)\ cov(y,z)\ cov(z,x)\ cov(z,y)\ cov(z,z)\ ]$$

Where,

$$cov(x,\ x) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})}{n-1}$$

and

$$cov(x,\ y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

- Cov(x,x) is actually the variance of variable x.
- Cov(x, y) is the variance of variable x w. r. t. variable y. cov(x, y) = cov(y, x)

The covariance value within the matrix denotes the dependency of two variables with each other. A negative value denotes the two variables are indirectly proportional to each other and a positive value is for directly proportional relationship.

Step 3- The step 3 is related to finding the mathematical constructs of covariance matrix to identify principal components. These mathematical constructs are eigenvalues and corresponding eigenvectors. The eigenvectors are nothing but the principal components of our covariance matrix and represent the axis of new feature space whereas eigenvalues are the magnitude of those vectors. In other words we can say that eigenvalues actually describe the contribution of each vector in terms of variance.(A high magnitude for eigen value denotes high variance along its eigen vector in feature space)

Eigenvectors and Eigenvalues- A vector whose direction remains the same ever after applying linear transformation is called an eigenvector. This can be expressed as:

$$A.x\ =\ \lambda.x$$

Where A is covariance square matrix with x its eigenvector and λ is a constant. The eigenvalues of the matrix A is obtained by solving the equation-

$$|A\ -\ \lambda I| = 0$$

Step 4- principal components can be computed by arranging eigenvalues with corresponding eigenvectors in descending order. The higher value eigen vectors have more significance over the data and form principal components whereas the lower value eigen vectors can be removed in order to reduce the dimensions.

Step 5- Finally the original data can be projected using principal components in reduced dimensions. This can be done by multiplying the transpose of the original dataset with the transpose of computed vectors.

Example- Let we have been given the following data value in two dimensions.

X_1 : 2.5, 0.5, 2.2, 1.9, 3.1, 2.3, 2.0, 1.0, 1.5, 1.2

X_2 : 2.4, 0.7, 2.9, 2.2, 3.0, 2.7, 1.6, 1.1, 1.6, 0.9

[[2.5, 0.5, 2.2, 1.9, 3.1, 2.3, 2. , 1. , 1.5, 1.2],
[2.4, 0.7, 2.9, 2.2, 3. , 2.7, 1.6, 1.1, 1.6, 0.9]]

Step 1- Standardize the values:

$$\overline{X_1} = \frac{\Sigma X_i}{N} = \frac{2.5+0.5+2.2+1.9+3.1+2.3+2.0+1.0+1.5+1.2}{10} = 1.82$$

$$\overline{X_2} = \frac{\Sigma X_i}{N} = \frac{2.4+0.7+2.9+2.2+3.0+2.7+1.6+1.1+1.6+0.9}{10} = 1.91$$

X'_1: 0.68, -1.32, 0.38, 0.08, 1.28, 0.48, 0.18, -0.82, -0.32, -0.62

X'_2: 0.49, -1.21, 0.99, 0.29, 1.09, 0.79, -0.31, -0.81, -0.31, -1.01

Step 2- Covariance Matrix-

$$C = [0.6018,\ 0.6042, 0.6042,\ 0.7165\ ]$$

Step 3- Calculate eigenvalue of covariance matrix C.

$$\lambda_1 \lambda_2 = [0.052,\ 1.27]$$

and the eigenvectors are:

$$V = [[-\ 0.7397818,\quad -\ 0.67284685]\ [\ 0.67284685,\quad -\ 0.7397818\ ]]$$

Step 4- Reduce the dimension and create a new feature vector (Highest eigenvalue vector is the principal component of the data). Here we choose the eigenvector with a bigger eigenvalue and leave the smaller one.

$$V = [0.6728,\quad -\ 0.7397\ ]$$

Step 5- Project the data on new feature space-

$$Y = X\,V^T = \left[X_{1,}\ X_2\right]\left[V_1,\ V_2\right]^T$$

Y =[[1.75, 0.35, 1.54, 1.33, 2.17, 1.61, 1.4 , 0.7 , 1.05, 0.84],

[-1.68, -0.49, -2.03, -1.54, -2.1 , -1.89, -1.12, -0.77, -1.12, -0.63]]

Advantages of PCA

- Removes the correlated attributes.
- Help to reduce overfitting.
- Improves the data visualization.
- It also help to improve the performance of Algorithm,

Disadvantages of PCA

- Data normalization must be needed before applying PCA
- Some level of information loss.
- Independent variables are becoming less interpretable.

****