Mithilesh Singh

PCA

What is Principal Component Analysis?

Principal Component Analysis technique used for EDA exploratory data analysis, especially in dimensionality reduction.

How does Principal Component Analysis work?

PCA achieves dimensionality reduction by **finding the principal components** of the data points.

The principal components are the **vectors of highest variance**.

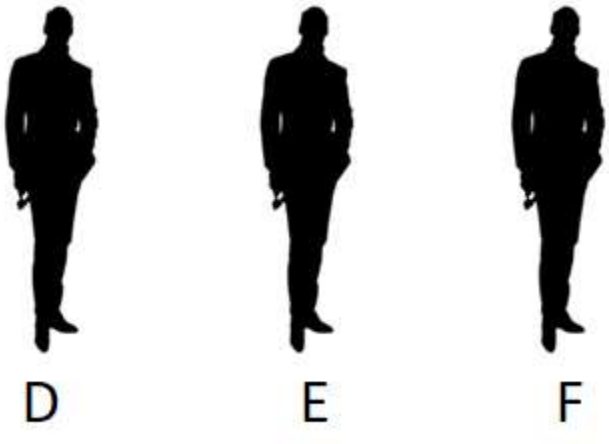
In other words, they are the directions where the data points vary the most. PCA finds these vectors that explain most of the variance observed in the data points.

Variance is Both Our Enemy and Our Friend

The greater the variance, the more the information. Vice versa.

The silhouette of three friends whom we need to identify based on their height differences.

Without a doubt, we are going to say that Who is Person A, Person B, and Person C.



The silhouette of three similarly-tall friends D,E,F whom we need to identify is very difficult.

In the same way, when our data has a higher variance, it holds more information.
PCA->maximum variance

PCA Intuition

If you have a large data sets with thousands/millions of observations (rows) and hundreds of different variables (columns)

While having more data is always great, sometimes they have so much information in them, we would have impossibly long model training time and the curse of dimensionality starts to become a problem.

Direct elimination of variables is the obvious way but it has clearly impact on the *information* content of your data set. Too much or a wrong eliminations and your data set becomes useless, too less and the data set remains large and difficult to analyze.

This is our **first assumption** to keep in mind:

higher variance => higher information content.

We can compare PCA with writing a book summary.

Finding the time to read a 1000-pages book is a luxury that few can afford. Wouldn't it be nice if we can summarize the most important points in just 100-200 pages so that the information is easily digestible?

We may lose some information in the process, but, at least we get the big picture.

How does PCA work?

It's a two-step process. We can't write a book summary if we haven't read or understood the content of the book.

PCA works the same way — understand, then summarize.

Understanding data, the PCA way

PCA is defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the ***greatest variance*** *by some scalar projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.*

In the eyes of PCA, variance is an objective and mathematical way to quantify the amount of information in our data.

Variance *is* information.

In addition to focusing on the high variance data concept

Three Assumptions are:

- variance is related to information content and should be maximized
- redundant variables and high correlations between variables are a form of noise that should be minimized
- correlations between variables are linear

another example:-

Guess who's who based on their height and weight.

Round two.

Person	Height (cm)	Weight (kg)
Alex	145	68
Ben	160	67
Chris	185	69

The same set of friends and their respective height and weight.

In the beginning, we only had height. Now, we have more amount of data (weight and height).

— how PCA summarizes our data, or more accurately, reduces dimensionality.

Summarizing data with PCA

Personally, the weight differences are so small (small variance), it doesn't help us to differentiate our friends at all. we still had to rely mostly on height to make our guesses.

Intuitively, we have just reduced our data from 2-dimensions to 1-dimension. The idea is that we can selectively keep the variables with higher variances and then forget about the variables with lower variance.

Can we keep *both*?

Perhaps, with a different perspective.

Up until now, we have only been looking at the variance of height and weight individually. Instead of limiting ourselves to choose just one or the other, why not combine them?

PCA combines both height and weight to create two brand new variables. It could be 30% height and 70% weight, or 87.2% height and 13.8% weight, or any other combinations depending on the data that we have.

These two new variables are called the **first principal component (PC1)** and the **second principal component (PC2)**. Rather than using height and weight on the two axes, we can use PC1 and PC2 respectively.

Note:- Performing a PCA will give us N number of principal components, where N is equal to the dimensionality of our original data.

From this list of principal components, we generally choose the least number of principal components that would explain the most amount of our original data. (means maximum amount of information)

Principal components

Using eigenvalues and eigenvectors, we can find the main axes of our data. The first main axis (also called “first principal component”) is the axis in which the data varies the most. The

second main axis (also called “second principal component”) is the axis with the second largest variation and so on.

We know that for two variables covariance formula is:-

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

Covariance represents a dispersion measure that include the concept of linear “**synchronicity**” between two variables in relation to their respective means.

From Covariance to Covariance Matrix

Considering that the covariance formula we can see the covariance matrix of all the covariance between all of the variables in our data set.

$$C_x = \begin{vmatrix} Var(X_1) & Cov(X_1, X_2) & \dots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & Var(X_2) & \dots & Cov(X_2, X_n) \\ \dots & \dots & \dots & \dots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \dots & Var(X_n) \end{vmatrix}$$

.

Thus, PCA is a method that brings together:

1. A measure of how each variable is associated with one another. (Covariance matrix.)

2. The directions in which our data are dispersed. (Eigenvectors.)

3. The relative importance of these different directions. (Eigenvalues.)

PCA combines our predictors and allows us to drop the Eigenvectors that are relatively unimportant.

Steps of PCA

The PCA algorithm consists of the following steps.

1. Standardizing data by subtracting the mean and dividing by the standard deviation
2. Calculate the Covariance matrix.
3. Calculate eigenvalues and eigenvectors
4. Merge the eigenvectors into a matrix and apply it to the data. This rotates and scales the data. The principal components are now aligned with the axes of our features.
5. Keep the new features which account for the most variation and discard the rest.