

Roopkishor Singh

May 18, 2017 · 4 min read

Singular Value Decomposition. Elucidated.

Mathematics is building block of Machine learning. I know math is hard to understand but it is much needed as well. Singular value decomposition (SVD) is one mathematical method used in various applications.



Not this SVD :P

Singular Value Decomposition is a matrix factorization method which is used in various domains of science and technology. Furthermore, due to recent great developments of machine learning, data mining and theoretical computer science, SVD has been found to be more and more important. It is not only a powerful tool and theory but also an art.

Singular value decomposition makes matrices become a “Language” of data science.

Matrix Factorization: It is a representation of a matrix into a product of matrices. There are many different matrix factorization and each used for different class of problems.

Mathematics behind SVD

For a $m \times n$ matrix(M) there exists a singular value decomposition of M , of the form

$$M=U\Sigma V^*$$

Singular value decomposition

where

- U is a $m \times m$ unitary matrix. (left singular vector)
- Σ is a $m \times n$ diagonal matrix with non-negative real numbers.
- V is a $n \times n$ unitary matrix . (right singular vector)
- V^* is the conjugate transpose of the $n \times n$ unitary matrix.
- The diagonal values of Σ are known as Singular values of M .

Conjugate Transpose: The conjugate transpose of a matrix interchanges the row and column index for each element.

Identity matrix: It is a square matrix in which all the elements of the principal diagonal are ones and all other elements are zeros.

Diagonal Matrix: It is a matrix in which the entries outside the main diagonal are all zero.

Unitary matrix: Matrices whose conjugate transpose is also its inverse, that is $UU^*=I$.

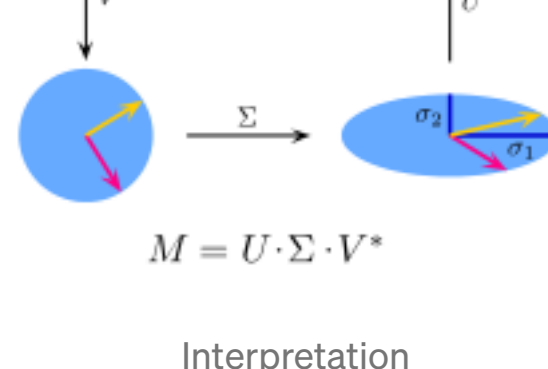
Singular Values: Basically it denotes the square root of the eigenvalues of XX^* where X is a matrix.

Intuition

1. Consider any 2 independent vectors lying in a plane both starting at the origin. Rotate the vectors by an arbitrary angle and importantly such that the vector with the larger x value has a lower absolute y value than the other. The 2 points will lie on the boundary of an ellipse centered on the origin.
2. Scale the ellipse(2 vectors) to a unit circle and inspect the angle between the two unit vectors.
3. Vary the angle of the rotation and repeat the scaling until the scaled unit vectors are orthogonal.
4. Finally rotate the unit vectors by angle β until they correspond to x-axis and y-axis.

The above process describes is the inverse of SVD.

The expression $U\Sigma V^*$ can be interpreted as a composition of three geometrical transformations: a rotation, a scaling, and another rotation. SVD is generally used for data compression in various fields. Other than data compression the resultant matrices has lots of wonderful properties.

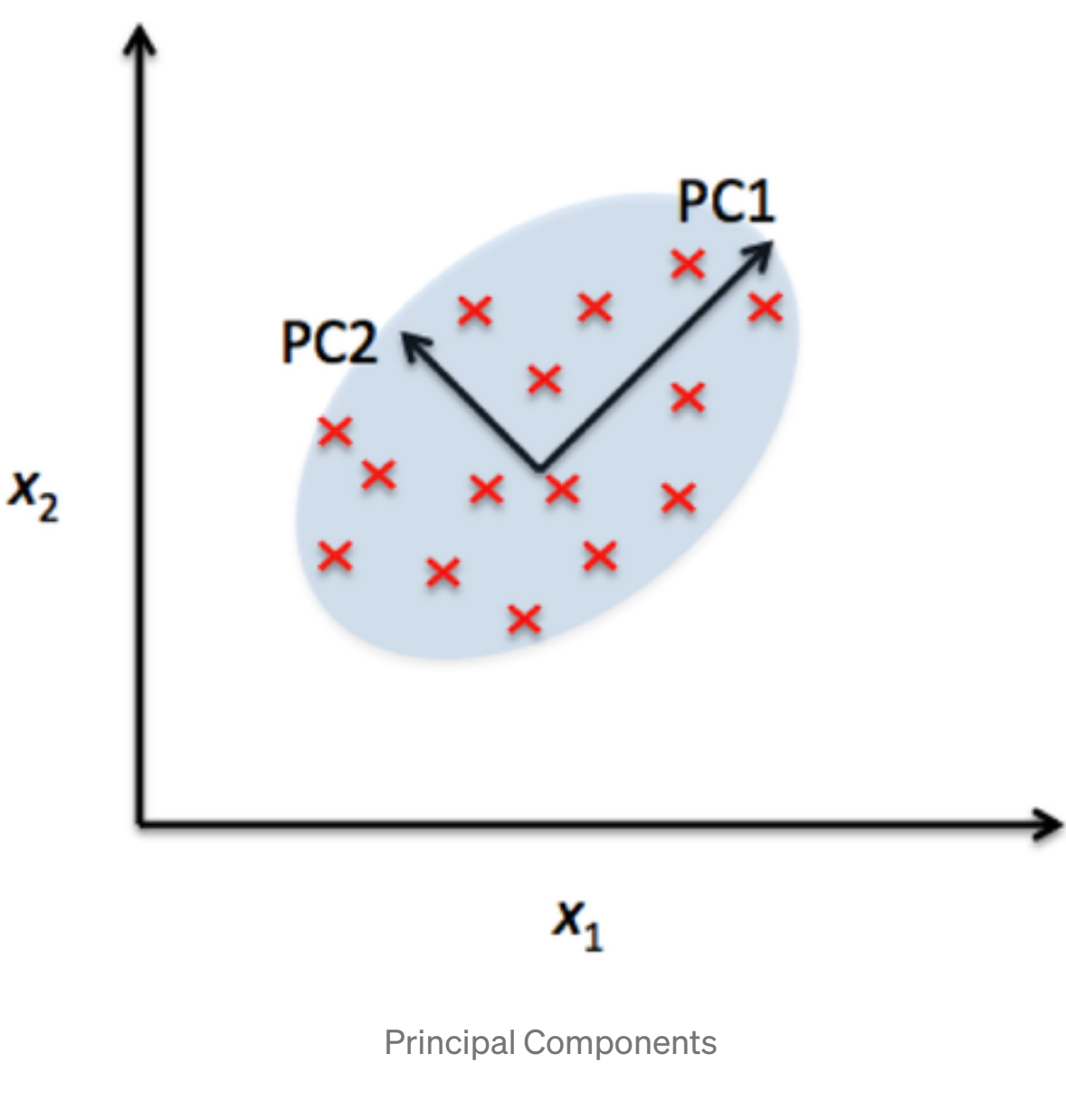


if M is symmetric positive definite its eigenvectors are orthogonal and we can write $M = Q\Sigma Q^*$. This is a special case of a SVD, with $U = V = Q$

For mathematical example, refer to this [link](#).

Principal Component Analysis

Principal components analysis is a procedure for identifying a smaller number of uncorrelated variables, called “principal components”, from a large set of data. The goal of principal components analysis is to explain the maximum amount of variance with the fewest number of principal components. For more details, go through this [blog](#).



Principal Components

Relation between PCA and SVD-

Let M be $m \times n$ matrix where m is the number of samples and n is the number of features

$$M=U\Sigma V^*$$

The covariance matrix of M is C

$$C=M^*M/(m-1)$$

$$C=V\Sigma U^*U\Sigma V^*/(m-1)$$

Note : The above is correct only if M is centered(scaled). Only then is covariance matrix equal to $M^*M/(m-1)$. That is why we need to center the initial dataset M for PCA.

The right singular vectors V are principal directions and Principal components are given by $MV=U\Sigma V^*V=U\Sigma$. To reduce the dimensionality of the data from m to k where $k < m$, select the first k columns of $U\Sigma$.

Other Applications

1. [Recommender System](#)
2. [Image Processing](#)
3. [Natural language Processing](#)

References

- [mit.edu](#)
- [nyu.edu](#)
- [yale.edu](#)

Footnotes

This blog tries to aware you about the basic mathematics of singular value decomposition and different application of singular value decomposition. Go through the references to understand it thoroughly.

♥ if this was good read.

P.S. — Suggestions would be highly appreciated.

👍 341



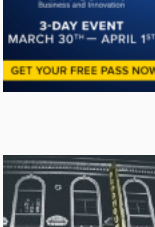
More from Data Science Group, IITR

Where Data Meets Intelligence . Code | Train | Predict

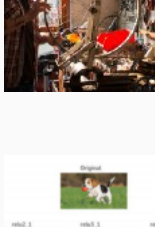
Follow

Read more from Data Science Group, IITR

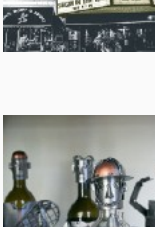
More from Medium



The ODSC Guide to Machine Learning Download, Making XAI Work, and Jobs



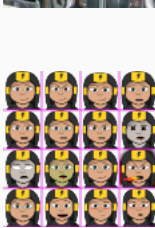
Deep Learning — from Prototype to Production



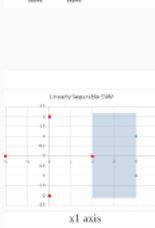
Transfer Learning using MXNet (Part 2/7)—sklearn Data Transformations



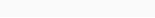
Image to Image Translation
It is an extremely arduous image processing ...



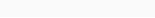
Machine Learning for Grandmas.
As Carlos García Moreno says : Artificial intel...



Overview of the YOLO Object Detection Algorithm



Different Ways of Improving Training Accuracy



Support Vector Machine (SVM)
[1] Introduction and Methodology

Get started

Sign In

Q Search



Roopkishor Singh

63 Followers

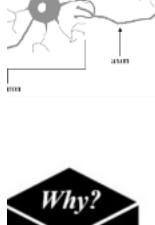
Data Scientist @ Goldman Sachs | IIT Roorkee'18

Follow

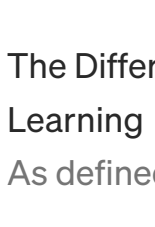


Related

Overview of ML Evaluation Metrics
Model evaluation is done by measur...



Main Principle Behind Training Neural Networks
— Empirical Risk...



LIME: Explaining Black-Box Machine Learning Models

The Different Types of Machine Learning
As defined by Arthur Samuel, machi...

Help Status Writers Blog Careers
Privacy Terms About Knowable