# SVD

PCA is closely related to singular value decomposition (SVD). In fact, the two are so intimately related that the names are often used interchangeably. What we will see though is that SVD is a more general method of understanding change of basis.

SVD is based on simple, interpolatable Linear Algebra.

## How its calculated and interpolated

We start with , which is a matrix with rows and *m* columns, we have n samples, each with m features. We also assume that . *X* is a tall, skinny matrix.

It can be converted to an orthogonal matrix, a diagonal matrix, and another orthogonal matrix (or a rotation, a stretch and a second rotation):

is*n x n*, are called igen-faces, they give a base that X can be represented. U is the left singular vector, only the first columns are important;

is*n x m*, it’s diagonal,

is*m x m*, the right singular vector,

Each column of tell a mixture of to make up

## Matrix approximation

Economy version: , is U of the first columns, is of the first row.

Truncate version: , is U of the first columns,is of the first row.

## Dominant conditions

This is a matrix.

The covariance matrix of X is CovX = , it’s symmetry and semi-positive.

Let

Then

**CovX =**

We also know that semi-positive matrix can be decomposed to it’s igen-vectors:

**CovX =**

Similarly,

This is intuitive way of interpret SVD, yet it’s not an efficient way to calculate U,V. There are better ways, such as QR decomposition.

The projection of original matrix onto principal axes:

# PCA

In many practical applications, high-dimensional data have most of their variation in lower-dimensional space that can be found using dimension reduction techniques. One of the techniques is principal components analysis, PCA.

PCA finds structure in the covariance or correlation matrix and uses this structure to locate low-dimensional subspaces containing most of the variation in the data.

PCA starts with a sample,, (n>>m)

or in Matrix form

samples, each of which is of -dimension random vectors, with mean vector and covariance matrix , m\*m.

PCA can be applied to either the sample covariance matrix or the correlation matrix. The correlation matrix is the covariance matrix of the standardized variables.

isan orthogonal matrix whose columns are the eigen-vectors of **.**  are the corresponding eigenvalues. The columns are arranged so that the eigen values are ordered from large to small. We also assume no tie among the eigenvalues, which almost certainly will be true in actual applications.

A normal linear combination of is of the form **,** where . The first principal component is the normal linear combination with the greatest variance. The variation in the direction is: Var(

The first component principal maximizes this variance over **.** The maximizer is the **,** the eigen vector corresponding to the largest eigenvalue, is called the first principal axis.

The projections **,** onto this vectorare called the first principal component of principal component scores. Requiring that the norm of be fixed is essential, because otherwise **the variance is** unbounded and there is no maximizer.