Principal component analysis (PCA)

A statistical technique mainly used for reducing the dimensionality of datasets. PCA forms the basis of **multivariate data analysis** based on projection methods.

Intro

- principal component analysis (PCA) is one of the **oldest and most widely** used to reduce dimensionality. Its idea is simple: reduce the dimensionality of a dataset, while preserving as much variability[1] (i.e. statistical information) as possible.
- Earliest literature on PCA: **Pearson (1901)** and **Hotelling (1933)**
  - Pearson K. 1901; On lines and planes of closest fit to systems of points in space. *Phil. Mag.* 2, 559–572. (doi:10.1080/14786440109462720)
  - Hotelling H. 1933; Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* 24, 417–441, 498–520. (doi:10.1037/h0071325)

1. Also called spread or dispersion refers to how spread out a set of data is.

More about PCA

- PCA is predominantly used as a **dimensionality reduction technique** in domains like facial recognition, computer vision and image compression. It is also used for finding patterns in datasets of high dimension in the field of finance, data mining, bioinformatics, psychology, etc.

- Why it was not until decades later when computers became widely available to use PCA on datasets that were not significantly small?

[A sneak peek of future examples]

A simple method

Suppose a Dataset V of size n x p. This Dataset can be represented with p points in a n dimensional space.

$$V = [v_1, v_2, v_3, \ldots, v_p]_{n \times p}$$

A linear combination of v_i is sought with its maximum variance.

The linear combination is as follows:

$$[Va]_{n\times 1} = \sum_{i=1}^{p} a_i v_i$$

$$\text{where } a = [a_1, a_2, a_3, \ldots, a_p]^t_{p \times 1}$$

The variance of this linear combination is given by the equation:

$$\text{Variance}(Va) = a^t S a,$$

Where S is the covariance of columns of matrix V.

Our maximization problem to maximize the variance of the linear combination of columns in Dataset is:

$$\text{Max } a^tSa$$

For this problem to have a well-defined solution, the following restriction must be imposed:

$$a^ta=1$$

Using Lagrange multiplier strategy, the constraint can be implemented into the maximization problem:

$$\text{Max } [\ a^tSa - \lambda\,(a^ta - 1)\ ]$$

λ is Lagrange multiplier

By differentiating our modified maximization problem and equating it to zero, the following equation is produced:

$$Sa = \lambda a$$

This equation restricts acceptable vectors in a and λ to be the eigen vectors and eigen values of S, respectively.

- We can freely ignore small eigen values and their corresponding eigen vectors, as they do not carry essential information about the dataset in them.

Principal component of V:

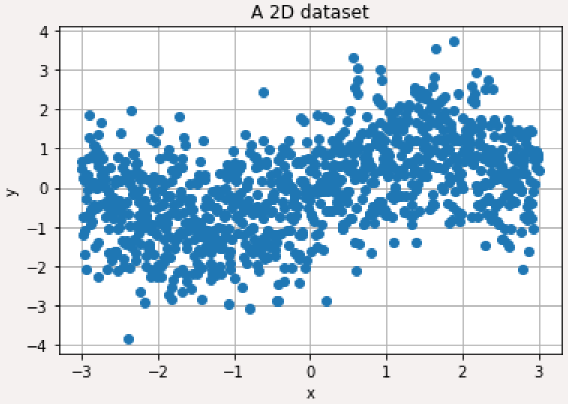
- **Va**, with the calculated eigen vectors plugged in, is the principal components of the dataset, and it has all important information compressed in itself.

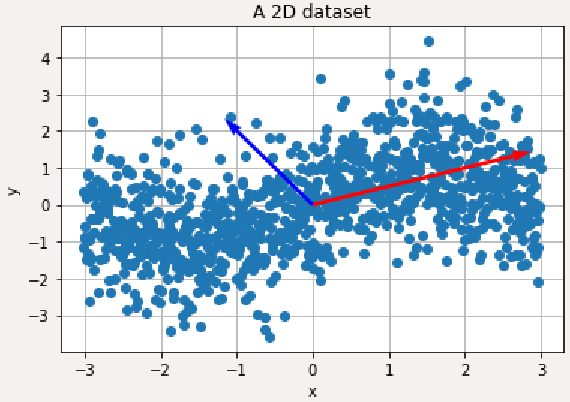
Short review of eigen vectors-values

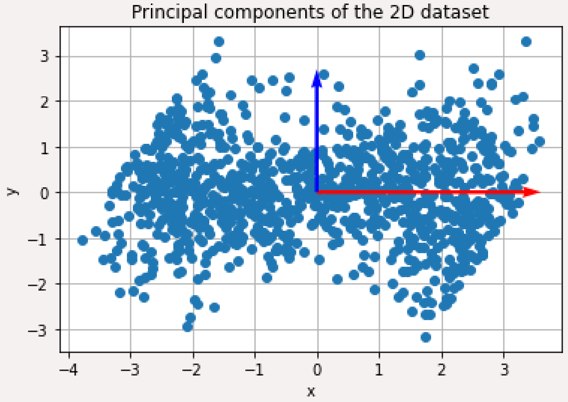
In essence, an eigenvector **v** of a linear transformation **T** is a nonzero vector that, when **T** is applied to it, does not change direction. Applying **T** to the eigenvector only scales the eigenvector by the scalar value λ, called an eigenvalue. This condition can be written as the equation:

$$T v = \lambda v$$

Suppose we have a 2D dataset. Eigen vectors corresponding to the dataset show in which direction data changes the most. Eigen value of each eigen vectors represents how much dataset is spread out in that direction.







Reconstruction of original dataset out of principal components:

Since covariance of columns of the original dataset, S, is symmetric, then its eigen vectors are **orthogonal** meaning $a_i \cdot a_j = 0$, if $i \neq j$.

Matrix a:

$$[a]_{p \times p} = [a_1, a_2, \ldots, a_p]^t_{p \times p}$$

To reconstruct an approximation of the original data:

- multiply Va (Principal Components of the original dataset) by a^t as follows:

$$V \times a \times a^t = V \times [a_1, a_2, \ldots, a_p]^t \times [a_1, a_2, \ldots, a_p]$$

$$= \begin{bmatrix} v_1 & v_2 & \ldots & v_p \end{bmatrix}_{1\times p} \begin{bmatrix} a_1^t a_1 & 0 & 0 & 0 \\ 0 & a_2^t a_2 & 0 & 0 \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & a_p^t a_p \end{bmatrix}_{p\times p}$$

$$= V$$

Different approaches to PCA

Following methods are commonly used to perform PCA on incomplete data as well as for accurate missing value estimation:

- Bayesian PCA
- Probabilistic PCA
- Nipals PCA

Other methods:

- Inverse Non-Linear PCA
- Conventional SVD PCA