Principal component analysis (PCA)

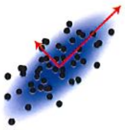
Seminar in Matrix Computations

John DE WASSEIGE   Marie VALENDUC

Problem statement

Let $X \in \mathbb{R}^{I \times J}$ be a data table containing I observations and J variables. Our problem is :

> *How to extract the most important information from* X *and express it using new variables?*



This can be solved using **principal component analysis** (PCA), which is one of the most popular multivariate techniques in **statistics** and **machine learning**.

Goals of PCA

- **extract** the most important information from a data table,
- **compress** the size of a data set by keeping only this important information.

Application 1 : Image compression

Use lower rank approximation matrix to compress data. On the example below, the original image can be represented by a rank 500 matrix and the 50 highest principal components values contain more than 70% information.

Application 2 : Eigenfaces

Apply PCA on N pictures: find N eigenvectors, or **eigenfaces**, keep M of these. You can represent an image with M coefficients $\longrightarrow$ **data compression**.

Used for face recognition, age simulation, etc.

Method

1. Preprocessing step : The columns of X are centered s.t.

$$\frac{1}{I}\sum_{i=1}^{I} X_{i,j} = 0 \quad 1 \leq j \leq J \tag{1}$$

2. Decorrelation step : Find the SVD of X, *i.e.*, $X = U\Sigma V^T$, and compute the *factor score matrix* F as

$$F = XV \tag{2}$$

3. Optional step : Project supplementary observations x_{sup} onto the principal components

$$f_{\text{sup}}^T = x_{\text{sup}}^T V \tag{3}$$

Why does the SVD work?

We want new variables, called **factor scores**, that are written as a linear combination of the original variables s.t.,

1. the factor scores explain as much as possible the **variance** of X,
2. the factor scores are pairwise **orthogonal**,
3. the coefficients of the linear combination are **finite**.

We define the factor score matrix $F := XV$ with $F \in \mathbb{R}^{I \times L}$ and $V \in \mathbb{R}^{J \times L}$ ($\text{rank}(X) = L$). We must solve :

$$\begin{array}{lll} \underset{V \in \mathbb{R}^{J\times L}}{\text{maximize}} & \left\|V^T X^T X V\right\| & \text{(Cond. 1)} \\ \text{subject to} & V^T V = I. & \text{(Cond. 3)} \end{array}$$

Using a diagonal matrix of Lagrange multipliers Λ,

$$\mathcal{L} = \text{trace}\left\{V^T X^T X V - \Lambda\left(V^T V - I\right)\right\}, \tag{4}$$

$$\frac{\partial \mathcal{L}}{\partial V} = 2X^T X V - 2V\Lambda = 0 \tag{5}$$

This leads to

$$X^T X = V \Lambda V^T \tag{6}$$

V is the matrix of eigenvectors of $X^T X$ associated to Λ (see Remark 2.2 in course notes).

Geometrical interpretation

- **Rotation** of the original axes,
- **Projection** of the data onto the principal components.