Principal Component Analysis

Introduction : In order to analyze a complicated set of data, we need a tool capable of distinguishing the main trends in it. Principal Component Analysis (PCA) is especially well suited for such a task. First we will present two problems that could benefit from PCA. Then, we will dive into the linear algebra behind PCA and finally we will present the solution of the two first problems.

Problems : First, imagine you have a huge set of data on the European Union for a lot of economic indicators. Analyzing such a table is a complicated task and we would like to visualize it on a 2D graph. Choosing two indicators would not be a good solution. We would prefer having the two most significant directions in the data structure. Secondly, suppose you would like to send an image of 784 bits to your friend. Nevertheless, we want to reduce the number of bits to send by compressing the image. Again, what would be the best way to keep the most important information in the image ?

Building PCA : Considering our dataset is a sample of a system, the goal is to formulate different criterions to select important dynamics. Next, we will do a *change of basis* that optimizes these criterions.

1. **Signal to Noise Ratio :** This indicator attests the precision of the data. In a 2 dimensional case, the SNR would be the ratio between the variance in one direction and the variance in the *orthogonal* direction. Assuming the direction with largest variance contains the dynamic of interest, maximizing the variance corresponds to finding an optimized rotation of the dataset.
2. **Redundancy :** In order to keep only relevant informations, we do not want to record any information that could be expressed in terms of a single variable. The covariance between two quantities is an indicator that measures the degree of *linear relationship* and that we should minimize.

Generalizing these criterions to a m-dimensional case with a dataset X, the *covariance matrix* : Σ_X gathers the variance of each measurement in the diagonal terms and the covariance of each pair of measurements in the off-diagonal terms.

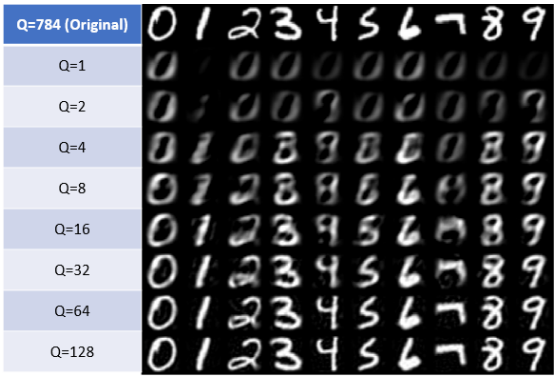
Linear algebra : The optimal form we seek is therefore a diagonal matrix of covariance. We want to find $P : Y = PX$ and Σ_Y is diagonal. To do that, we use the SVD of X : $X = USV^T$

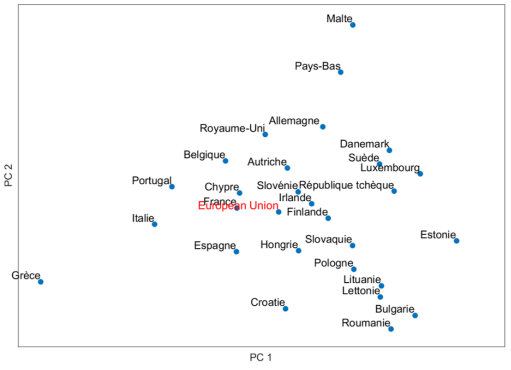
$$\Sigma_Y = \frac{1}{N-1}YY^T = \frac{1}{N-1}PXX^TP^T = \frac{1}{N-1}PUSV^TVS^TU^TP^T = \frac{1}{N-1}PUS^2U^TP^T$$

We see that $P = U^T$ will make Σ_Y diagonal. The columns of P are the *principal components*, each dimension is *uncorrelated* from each other and the diagonal terms of Σ_Y indicates how important each dimension is.

If we want to reduce the number of dimensions in our data, we can take only the columns of U that gather the most variance. By doing so, U_Q is no longer orthogonal : $U_Q^TU_Q = I \neq U_QU_Q^T$. Selecting $P = U_Q^T$, we project the data on a lower dimensional space. We can back-project the data in the original space and end up with : $\hat{X} = U_QU_Q^TX$.

Solution of the problems : We can present the solution of the original problems. We see that sending only 32 bits already allows to have a good representation of the image, thanks to PCA. We can also project the European Union dataset on two principal component axes to analyse its structure.





Limitations : PCA is a linear technique that projects data on an hyperplane. This is not efficient when the data is spread non-linearly in the space. Also, this technique is non-parametric and cannot take into account known properties of the data. To solve these issues, *Kernel PCA* first apply a non-linear transformation on the dataset before doing regular PCA. Also, PCA relies only on mean and variance computations, which assumes the samples are gaussian distributed. PCA may not lead to interesting directions when it is not the case. For that reason, ICA will rely on a stronger criterion than uncorrelation : *statistical independence*.