Practical Session 3

Aitor Muguruza

Department of Mathematics, Imperial College London
E-mail address: AM5115@ic.ac.uk

Tuesday 7th February, 2017

For this session it is strongly recommended to use R and the *princomp* function in order to perform PCA. Moreover, the *summary* function gives the components weights.

Question 1: PCA on equity markets

Choose a set of stocks of your choice that aim to represent different sectors of the economy (use at least 8 stocks) and download their quotes from either Yahoo Finance or Google Finance. You will then need to transform stock prices into returns, for this we will use log-returns, i.e.:

$$\mu_{t_i} = \log\left(\frac{S_{t_{i+1}}}{S_{t_i}}\right), \quad i = 1, ..., N$$

where N is the number of data samples we have.

1. **Clean the data:** Replace the missing values by replacing them by the average between of previous and next value.
2. **PCA:** Run a PCA on the data.
3. **Conclusion:** Interpret the meaning of the first 3 components.
4. **The market and Component 1:** Usually, the firt component is interpreted as the market. Compute the correlation between a portfolio created by using the weights of the first component and an index that represents the economy e.g. S&P 500. Comment your results.

Question 2 : PCA on interest rate markets

Load the file *UStreasury.csv* and perform the following analysis:

1. **PCA:** Run a PCA on the data.
2. **Conclusion:** Interpret the meaning of the first 3 components.
3. Determine whether it is sensible to assume that the data is **Gaussian**.