MTH 4600 Homework 5, Due 4/5

Principal Component Analysis

• In this exercise we subject the 30 components of the Dow Jones Industial Average (DJIA) to a principal component analysis. The file `DJI.txt` contains a time series of the daily percent price change for each of the 30 stocks for all of year 2020 (what a year). There are 253 items of data for each stock. I obtained this data from `www.investing.com/equities/united-states`. I have also included starter code `PCA-DJI.cpp`. This code opens `DJI.txt`, reads in the data, puts it into a 253×30 matrix X and prints to the screen a bit of information. For the following, you will want to steal from `TreasuryPCAnalysis.cpp`.

1. Compute the sample covariance and correlation matrices for the data. You don't need to report the 30×30 matrices, but report any pairs (identified by ticker) that have a correlation in excess of 0.8. Are there any negative correlations? Please comment on the results.

2. Compute the eigenvalues and eigenvectors of the covariance matrix. Report the first eigenvector (the one with the largest eigenvalue) with components identified by ticker. What is true about the first eigenvector that is not true of the remaining 29 eigenvectors? Problems 3 – 5 focus on an interpretation of the first eigenvector.

3. The Capital Asset Pricing Model says that $ER = r + \beta(EM - r)$, where R is a particular stock's return on any particular day, M is the overall market's return that same day, β is the stock's "beta", and r is the risk-free rate of interest. The stock's beta is a scaling factor measuring the stock's expected response to changes in the overall market. It is defined as

$$\beta \;=\; \frac{\mathrm{Cov}\,(R, M)}{\mathrm{Var}\,M}.$$

Take the daily market returns to be the daily returns of the total Dow average. These returns can be found in the file `TotalDow.txt` over the same 2020 period and are also read in by `PCA-DJI.cpp`. Use the data in `DJI.txt` and `TotalDow.txt` to estimate the betas of the 30 individual components. Produce a scatter plot of points of the form (β_i, q_i) where β_i is stock i's beta and q_i is the i^{th} component in the first eigenvector. Here $1 \le i \le 30$. The horizontal scale for beta should run from 0 to 2, and the vertical scale from 0 to .35. Comment on what you see.

4. For each i and t with $1 \le i \le 30$ and $1 \le t \le 253$, compute two numbers. The first is $x = \beta_i M_t$, where M_t is the total Dow's return on day t. The second is $y = Y_{t1} q_i$, where Y_{t1} is the first principal component of the t^{th} row of X and q_i is the i^{th} component of the first eigenvector. Produce a scatter plot of the 253×30 points where both the horizontal and vertical axes are scaled from -25 to 25.

5. Based on problems 3 and 4, give an interpretation of the first eigenvector.

6. Report the eigenvector corresponding to the second largest eigenvalue with components identified by ticker. Which components have a presence in excess of 0.15 in absolute value? Is any particular sector heavily represented in this eigenvector? How do you interpret this eigenvector? You can find sector information about the stocks at

`https://en.wikipedia.org/wiki/Dow_Jones_Industrial_Average.`