Stat 897 Spring 2017 Data Analysis Assignment 6

Penn State

Due October 8, 2017

In this assignment we will again use the stock return data (`spreturns.Rda`) from DAA 5 part two. For the purposes of this exercise you may assume the mean returns are zero if it helps.

a) Restrict to the first 100 columns (first 100 stocks) in the data. Compute the covariance matrix of these stocks. Using the `corrplot` package, produce a visual of the *correlation* matrix. Do not print the covariance matrix. (I will take points off if you do because printing out a 100 x 100 matrix really isn't helpful, visuals are much better!)

b) Perform PCA on the stocks, and plot the variance explained. Do you see any natural "knee"?

c) How would you interpret the factor or first principal component? How would you describe the covariance matrix that results from keeping only the projection onto the first or first and second components?

d) Relate the covariance matrix method for PCA (eigendecomposition) to the direct SVD on the given **252** by **100** data matrix X.