

Quiz 2

11:30am-12:15pm, Friday, September 29, 2017

Save the required plots as images, name them appropriately, and submit them together with your other documents.

Use the data stored in file `SCLC_study_output_filtered.csv` for all of the questions. Rows in the data table correspond to observations and columns correspond to variables. The first column contains sample names and the first row contains variable names. So be sure to exclude the first row from PCA analysis.

Perform PCA of the data without standardizing the variables and answer the following questions.

- (1) (20 points) What is the total variance of the original variables and what is the total variance of the PCs. Are they equal?
- (2) (15 points) What is the covariance between the first PC and the second PC?
- (3) (15 points) Plot the scores plot using the first and second PCs. In your scatter plot, use red dots for the first 20 samples (rows 1 to 20) in the raw data table and use blue dots for the remaining 20 samples (rows 21 to 40).
- (4) (15 points) How many PCs do you need to keep in order to keep 75% of the variance in the data?
- (5) (15 points) Plot the loadings plot.

Standardize each variable and perform PCA again and answer the following questions.

- (6) (20 points) What is the total variance of all of the PCs. Is it equal to the total variance of the original variables before standardization?