Data Analytics

110-2 Homework #06
Due at 23h59, April 3, 2022; files uploaded to NTU-COOL

1. Implement a FA function in R/Python based on the PCA function you implemented in HW05. User can pass the dataset and specify the desired number of factors.
   a. (15%) Necessary outputs are:
      - the loading matrix **A**;
      - the factor matrix **F**;
      - the communality vector of h_i^2;
      - the uniqueness vector of ψ_i; and
      - the vector of the proportions of total variance contributed by the ith factor.
   b. (10%) Apply your FA function to the AutoMPG dataset and generate the necessary outputs given 2 factors are selected. Compare with the PCA results in HW05-EX3-b and discuss.

   Note: Directly applying the existed PCA library/package in your function loses all the 20 points in this exercise.

2. Transpose the ORL face dataset to let **X** be a 2576×400 data matrix. Perform the factor analysis on **X** with the FA function of your implementation in EX1.
   a. (10%) How many factors are needed to explain 50%, 60%, 70%, 80%, and 90% of the total variance?
   b. (10%) On condition of explaining 80% of the total variance, rescale the first factor into the range $[0, 255]$. Reshape the 2576×1 first factor into a 46×56 matrix. Plot an image from the 46×56 matrix using the rescaled factor.

3. Find a package/library to perform PLSR on the AutoMPG dataset. Take 300 cars "randomly" to build the model and the rest of 92 cars to test.
   a. (5%) Start with a single y (=mpg). What do you observe with the testing results?
   b. (10%) Use $y = [\text{mpg} \quad \text{model year}]$. What do you observe with the testing results? In particular, compare and discuss the testing results of the mpg between (a) and (b).