[CSC 5825 Fall 2017]

Due. Before the Class of Oct, 18, 2017

Full credit: 100 points with extra 10 points bonus

October 3, 2017

Question 1. (30 points in total) Write the log likelihood function for multinomial samples (10 points) and derive equation 4.6:

$$\hat{p}_i = \frac{\sum_t X_i^t}{N}$$

on textbook page 68 (20 points).

Note that on page 68,"$x_1, x_2, ..., x_K$ are the indicator variables where x_i is 1 if the outcome is state i and 0 otherwise."

$$x_i^t = \begin{cases} 1 & \text{if experiment t chooses state i} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Question 2. Dimension Reduction via Principal component analysis (PCA) (80 points). The data is scraped from the website `https://sofifa.com` by extracting the Player personal data and Player Ids and then the playing and style statistics. You can download it from `https://www.kaggle.com/thec03u5/fifa-18-demo-player-dataset/data`. The complete dataset includes: Player personal attributes (Nationality, Club, Photo, Age, Value etc.) Player performance attributes (Overall, Potential, Aggression, Agility etc.) Player preferred position and ratings at all positions.

Implement a PCA algorithm and apply it to the FIFA data set (you can NOT use singular value decomposition (SVD) function directly for solving the problem but you can use it for validation) (50 points). Visualize your results by plotting the first two principle components (20 points). Note only the continuous variables can be used for PCA analysis and you can use categorical variables such as Nationality, club etc to validate your PCA results. Draw conclusions based on your analysis (bonus 10 points). Note your conclusions do not have to the same and the correct answers are not unique.

Submission Instructions

To earn the full credit, your must type your solutions with sufficient details either using LaTeX or Microsoft Word **embedded** with source code and plots. Please submit it through the blackboard website. You will get points off for not following these requirements.

Late homework (w/o acceptable documents) will be accepted with penalty. 20% off penalty if late for 24 hours or less. 40% off penalty if late between 24 hours and 48 hours. 60% off penalty if late between 48 hours and 72 hours. 80% off penalty if late between 72 hours and 96 hours. No homework late more than 96 hours will be accepted.