Assignment 2: 숙제 2

AI+X: Deep Learning

DUE: Nov. 5 *Submit in Class*

(Learning)

This is a group assignment. Report your group members.

First, you will try a titanic survival prediction script. The script is provided. Refer to Task-2.

Next, you are asked to write a small data analysis program for predicting next lottery number (6 digits including the bonus number; a single digit ranges from 1 to 45). The previous winning numbers are provided in 'lottery.csv' file, containing all 761 weekly rounds from 2002/12/07 till 2016/08/20. You will use this file to complete the following analysis tasks.

lottery.csv data format:

round, date, first, second, third, fourth, fifth, sixth, bonus

**Task-1: Report your group members (4-5 people).

Member 1: Name, Major, Student ID
Member 2: Name, Major, Student ID
Member 3: Name, Major, Student ID
Member 4: Name, Major, Student ID

**Task-2: Download 'assign-titanic-lottery.zip' having 'titanic-assign.R', 'train.csv', and 'test.csv'. Install R and run it.

A few ways to install R:

1. For Linux or mac, you can use homebrew to install R or just download R studio - https://rstudio.com/.
2. For Windows and others, download and install R studio. There are many nice blogs that you can refer to. (e.g., https://www.guru99.com/download-install-r-rstudio.html)
3. Or you can install R in conda environment.

After R installation, install the required packages to run the given script (study it!). Please create the graphs in the script and print them out (black and white printing is just fine) for submission. For example, I used homebrew to install R and run in the following order.

>> brew install r
>> sudo r
>> install.packages("ggplot2")
>> …
>> …
>> Rscript titanic-assign.R
your graphs are created in the same directory.

**Task-3: Write a statistical analysis script to display the most frequently appeared number to the least. Use pandas (http://pandas.pydata.org/) for this task. Please print out your script for submission. An alternative way to do this is using Excel, but I strongly suggest that you try python and pandas for fun.

Example: $> ./your_script.py lottery.csv
Sample output: 1 -> 134 times
2 -> 117 times
3 -> 112 times …

**(BONUS) Task-4: Create a modified lottery data format by adding at new columns. For example, you can add "win" column indicating '0'-lose and '1'-win. Then your lottery data set would look like the following:

round, date, first, second, third, fourth, fifth, sixth, bonus, win

Likewise, please add at least two columns (including 'win') that you like, such as 'weather', 'daysofweek', or whatever. What you want to add and how many choices for a single column are completely up to you. Please print the first 20 lines of your modified data set (including csv header) for submission. If you have a source code for doing this, please print them out for submission.

SUBMISSION:

Do the above four tasks. Print them out: Task-1, graphs from Task-2, source code and outputs from Task-3 and 4.

DO NOT WORRY. We will discuss more in class.