Midterm Take Home Exam

AI+X: Deep Learning

제출 마감일: 13:00, April 29 *Submit LMS online*

NO Questions on Kakao Talk.

- 시험이기 때문에 문제 내용에 대한 카톡 질문을 금지합니다. 단, 문서에 문제가 있다거나, 오류에 관한 경우에는 카톡으로 문의 남기셔도 좋습니다. 신중히 이용하세요.
- 답은 당연히 한글로 작성하셔도 됩니다.

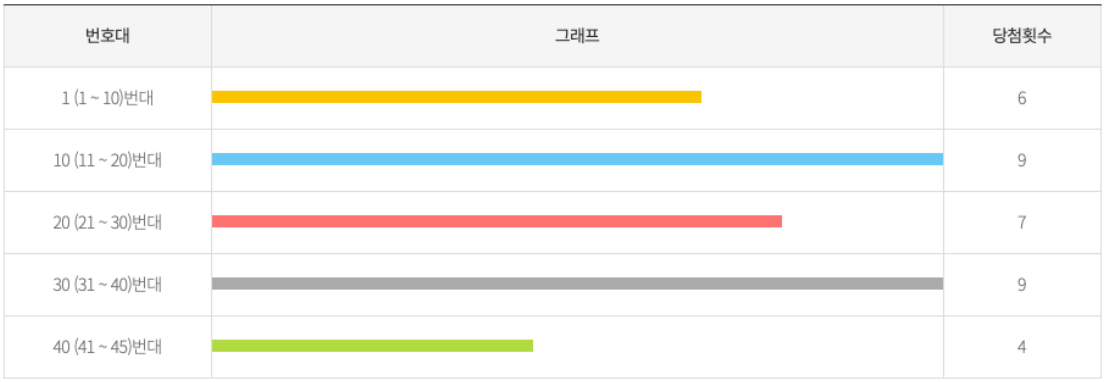
(For Task 1 to 2) You are asked to write a small data analysis script for weekly lottery numbers (6 digits including the bonus number; a single digit ranges from 1 to 45). The previous winning numbers are provided in 'lottery.csv' file, containing around 900+ weekly rounds from 2002/12/07 till 2021/11/13. You will use lottery.csv file to complete the following analysis tasks.

lottery.csv data format:

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

(For Task 3) You are asked to run a small program and show the results.

**Task-1 (10 pts): Write a statistical analysis script (python, R, Excel or whatever you like) to show the winning number distribution in ranges of 1s, 10s, 20s, 30s, and 40s. For example, the following figures shows the winning number distribution of the last five weeks:



Your job is to do this for the entire winning numbers (lottery.csv). If you provide the distribution in numbers (당첨횟수), it's 5 pts. If you provide the graph like the above, then it's 10 pts. If using python, pandas should be a good choice (http://pandas.pydata.org/). R is also fine. Please show your code and outputs. An alternative way to do this is using Excel. If you are using Excel, show the step by step making your output.

**Task-2 (10 pts): Feature engineering. Create a new feature and add it to the column list (to the lottery.csv). For example, you can compute the average value of all 7 numbers (including the bonus) and maybe use it as an extra feature to consider. Do not use this average example as your feature. Come up with your own and EXPLAIN your feature first. If you have a source code to compute your feature, please include them for submission (This is a bonus, it's not mandatory). If you can, show the first 10 lines of your new data, the most recent dates. Python, R, Excel or whatever you like to try is fine by me.

**Task-3 (10 pts): Download 'handwriting_recognition.py' from the LMS. Do the following. These files should be under the same directory to run correctly.

Download 'IAM_Words.zip' from LMS. // 파일크기가 700 MB 가 넘어서 조금 걸립니다.

```
$ unzip -qq IAM_Words.zip
$ mkdir data
$ mkdir data/words
$ tar -xf IAM_Words/words.tgz -C data/words
$ mv IAM_Words/words.txt data
```

You will need to install packages like keras (e.g., pip install). Show the sequence of your commands to install the required packages and run the python file. Also show the output images and texts of running 'handwriting_recognition.py' (e.g., screen capture or save image). You can certainly use your previous conda environment or WSL for Windows (if it works for you).

**Task-4: Write one paragraph explaining your tasks and any difficulties you had. (several sentences should be enough.) Even if you can't do the whole assignment, submit as much as you can (with explanation why you can't do this).

Submission:

Do the above tasks. Put the outputs of all tasks to a single PDF file: including source code (if any), steps, graphs, and your paragraph.

Suggestions:

Take a very similar step like the examples we saw in class. You will show me how to investigate the dataset and code. It's completely up to you for your choice of tools.