

Assignment #1

머신러닝 이론과 실전

Due: 2021 년 9 월 6 일

1. Write a program to implement the multiple linear regression analysis. Do not use commands in packages such as 'sklearn' or 'statmodels'.
 - a. Make a python script program to accept file names for data and output from the user. Give the user a prompt to type-in the data file and output file names.

<Python example>

```
data_name=input("Enter the name of data file [(ex) harris.dat] : ") # data name
coding_fm=int(input("Select the data coding format(1 = 'a b c' or 2 = 'a,b,c'): ")) # data separator
separator_fm={coding_fm==1 : " "}.get(True, ",")
res_pos=int(input("Enter the column position of the response variable : [from 1 to p] : "))
header=input("Does the data have column header? (y/n) : ")
if(header=="y"): trdata=pd.read_csv(data_name, sep=separator_fm) # loading data
else : trdata=pd.read_csv(data_name, sep=separator_fm, header=None) # loading data
out_name=input("Enter the output file name to export [(ex) result.txt] : ")
```

- b. The program must print out the coefficient for each X variable
- c. The program must assume that the data file is located in the same folder with the program file.
- d. The program can be a naïve one, thus you don't have to worry about many issues such as missing values, collinearity, ANOVA table, etc.
- e. Turn in the program file at the course website. It must be executable without any modification on the program. After the run, it must generate one output file.
- f. Use a data file named "harris.dat" for testing. Assume that the first column is the response variable for this data.
- g. The output file generated by your program must look like the below (the sample output is fictitious).

Coefficients

Constant: 5.312

Beta1: 1.345

Beta2: .236

Beta3: -.439

Beta4: .457

ID, Actual values, Fitted values

1, 9.5, 9.8

2, 4.6, 4.8

3, -2.3, -3.2

(continue)

Model Summary

R-square = .5689

MSE = .234

과제로 제출할 보고서의 내용에는 다음과 같은 것이 포함되어야 합니다.

1. 파이썬 코드 내용 (코드의 내용을 과제의 보고서에 그대로 포함시키시기 바랍니다. 또한 파이썬코드는 파일로도 별도로 제출합니다. 제가 코드를 돌려보기 위해 코드파일도 필요합니다.)
2. output 파일의 내용 (파이썬코드를 실행시키면 생성되는 파일의 내용을 그대로 보고서에 포함시키시기 바랍니다. (제가 코드를 돌려보면 output 파일이 생성될것이기 때문에 여러분은 output 파일은 별도로 제출하지 않습니다.)