

Assignment #3

머신러닝 이론과 실전

Due: 2021 년 9 월 20 일

1. Write a Python code to implement the gradient descent algorithm for multiple linear regression analysis.
 - a. Focus on the estimation of the coefficients for each X variable.
 - b. Use a data file named "harris.dat" for checking. The first column is the response variable for this data.
 - c. Compare the estimated coefficients with those by the package.
 - d. The output file generated by your program must look like the below (the sample output is fictitious).

```
Coefficients by Gradient Descent Method
```

```
-----
```

```
Constant:  5.312  
Beta1:    1.345  
Beta2:    .236  
Beta3:   -.439  
Beta4:    .457
```

```
Coefficients by Statmodels
```

```
-----
```

```
Constant:  5.312  
Beta1:    1.345  
Beta2:    .236  
Beta3:   -.439  
Beta4:    .457
```

2. Modify your program in Assignment #2 to do followings.
 - a. Prompt the user to enter the training/test data file name.
 - b. Prompt the user whether to run regression or classification.
 - c. If regression is chosen, perform the linear regression as in Assignment #1.
 - d. If classification is chosen, make the program to implement (i) LDA that can handle more than two classes.
 - e. Perform (i) LDA. Use a data file named 'veh.dat' for the training and 'veh.test.dat' as the test data.
 - f. The output file for classification generated by the program must look like below. (The numbers are fictitious).

```
Confusion Matrix (Training)
```

```
-----  
                        Predicted Class  
                        1      2      3      4  
Actual   1      239     14      6      8  
Class    2       12    153      5     12  
          3        2      4     98      2  
          4        3      6      8    123
```

```
Model Summary (Training)
```

```
-----  
Overall accuracy = .793
```

```
Confusion Matrix (Test)
```

```
-----  
                        Predicted Class  
                        1      2      3      4  
Actual   1      239     14      6      8  
Class    2       12    153      5     12  
          3        2      4     98      2  
          4        3      6      8    123
```

```
Model Summary (Test)
```

```
-----  
Overall accuracy = .793
```