END3971 – AI & ES

Mehmet G. Güler - Ebru Geçici

First Due 30.12.2022@23.45 (Sharp)
Last Due 01.01.2023@23.45 (Sharp)

Assignment 2.a

- *You can work on the project with groups of 2-4 students (not more, not less).*
- *This is the second assignment of END3971 AI & ES lecture.*
- *You should submit it through online.yildiz.edu.tr by zipping your code, otherwise you will not be able to upload it. Every group must submit only one assignment and all of the students name should be written at the top of the jupiter file. Otherwise those who do not have their names cannot get any points.*
- *The assignment should be named as follows:*

  StudentNumber1_StudentNumber2_StudentNumber3.ipynb

  for example it can be 2005800303_20161543_1203413541.ipynb
- *Late assignments are accepted at the cost of 20 points per day. For example if you submit at 23.**5**5 on 30.12.2022 (10 minutes later than the first due date) or 14.35 on 31.12.2022 , you will get at max 80 points. However this is only true for two days. You cannot submit after 01.01.2023 23:45*
- *Please make sure that the code works in my computer, i.e., sometimes I receive codes that are reading some input files that are not provided hence the code does not work.*

- In this assignment you are asked to find a regression model to find the price of a second hand car using Python.
- You will find the data for 2900 cars in the supplement excel file and the data for 94 cars in a zip file that can be downloaded from here. The cars have some numeric features like age, km, etc and some categorical features like gear type, shift type, etc. Well, you should know how to deal with the categorical variables from your stats course. If not, you can check it from Prof. Google! You can also find it in our lecture notes.
- You can use any regression model, i.e., (linear, polynomial, ridge, or lasso) or KNN or SVM. Please don't use any other model/algorithm other than these.
- You will submit your project in two parts.
  - **Part 1-a:** (50%) The Python code that includes what you have done. It should go over all of the processes that you go through: Data manipulation, different regularization values, adding new features, etc. You won't write a report, hence this code should have all of the details, i.e., add comments to your code so that we can understand what you did. In other words, you should embed your report into your code.
  - **Part 1-b:** (30%)You have a data of 94 cars which have all of their features but no prices. You will use your model to make predictions for these cars. What I want you to do is to upload a text file which gives the price of these 95 cars. An example file is given here. I will take your predictions given in your Results.txt file and calculate the R^2 directly from there.
  - **Part 2:** (20%) You should provide the R^2 values of at least 3 models that you tried. Of course these 3 models should include the one that you have chosen. You should provide all of the details of the models. This part is only a summary of what you have tried in your project. You will only write your models and then give their R^2 values. Please indicate which model you have chosen. The output should be a table like the following:

END3971 – AI & ES

Mehmet G. Güler - Ebru Geçici

First Due 30.12.2022@23.45 (Sharp)
Last Due 01.01.2023@23.45 (Sharp)

Model	R^2 (train, CV, etc)	…. (add any columns when needed)
Ridge, alpha=0.2		
Lasso, alpha =0.3		
SVM		

- So you will submit three parts for the project:
    - Jupyter file that includes your full code and report like explanations
    - Results.txt file that includes the price of 95 cars given in the test file.
    - Table.docx that includes the table above
- Note that it is impossible to come up with the same model by any two groups since picking the train/test split is random and there are infinitely many regularization values to be chosen. If two groups have exactly the same result, then I will divide the grade between the groups.