Artificial Intelligence with Python
Assignment 5

Problem 1: Diabetes

Investigate the model for predicting Diabetes disease progression by adding more explanatory variables to it in addition to bmi and s5.

a) Which variable would you add next? Why?

b) How does adding it affect the model's performance? Compute metrics and compare to having just bmi and s5.

d) Does it help if you add even more variables?

Include your own findings and explanations in code comments or inside triple quotes """...""".

Each step above is worth a point. You need 2 points in order to complete this problem.

Problem 2: Profit prediction

Consider the dataset 50_Startups.csv which contains data for companies' profit etc.

0) Read the dataset into pandas dataframe paying attention to file delimeter.

1) Identify the variables inside the dataset

2) Investigate the correlation between the variables

3) Choose appropriate variables to predict company profit. Justify your choice.

4) Plot explanatory variables against profit in order to confirm (close to) linear dependence

5) Form training and testing data (80/20 split)

6) Train linear regression model with training data

7) Compute RMSE and R2 values for training and testing data separately

Include your own findings and explanations in code comments or inside triple quotes """...""".

Problem 3: Car mpg

Consider car performance data from the file Auto.csv.

1) Read the data into pandas dataframe

2) Setup multiple regression X and y to predict 'mpg' of cars using all the variables except 'mpg', 'name' and 'origin'

3) Split data into training and testing sets (80/20 split)

4) Implement both ridge regression and LASSO regression using several values for alpha

5) Search optimal value for alpha (in terms of R2 score) by fitting the models with training data and computing the score using testing data

6) Plot the R2 scores for both regressors as functions of alpha

7) Identify, as accurately as you can, the value for alpha which gives the best score

Include your own findings and explanations in code comments or inside triple quotes """...""".