Fundamentals of Data Science

Coding assignment (30 points in total)

Data

You are given a set of observations linking the amount of rain per year and field productivity in a dry area somewhere in Central America.

The data is stored in the input file `inputdata⊗.csv`. Symbol ⊗ denotes a single-digit number from 0 to 9.

IMPORTANT: You must use the input file where number ⊗ is the same as the last digit in your student number

Your input file is a CSV file (an ASCII file which can be inspected by opening it using any basic text editor, such as Notepad). The file contains two columns: amount of precipitations (in mm per year) and the productivity coefficient. It has up to 35 rows; the first row contains the names of the columns, all other rows correspond to each individual observation.

Coding task

Write a Python code, which will

- read the data into Python numpy array(s), (**0-4 points**)
- plot the data as a two-dimensional scatter plot, (**0-5 points**)
- create a linear regression model based on the data, and (**0-7 points**)
- plot the corresponding line over the original data. (**0-5 points**)

The resulting graphs should have adequate axis labels (**0-2 points**).

The code should have comments briefly explaining the significance of each line (**0-2 points**).

Question (5 points)

Based on the linear regression model, evaluate the productivity coefficient of the field if the amount of precipitations is X mm. The resulting value (predicted productivity coefficient) needs to be printed on the plot.

IMPORTANT: You must predict the productivity coefficient for value X given in the table below.

Last digit of your ID	File name	Value X
0	inputdata0.csv	360.0
1	inputdata1.csv	300.0
2	inputdata2.csv	280.0
3	inputdata3.csv	350.0
4	inputdata4.csv	380.0
5	inputdata5.csv	260.0
6	inputdata6.csv	245.0
7	inputdata7.csv	310.0
8	inputdata8.csv	290.0
9	inputdata9.csv	275.0