Assignment 01

To be solved INDIVIDUALLY

Submit by October 12, 2018, 23h59 by email to jaime.cardoso@fe.up.pt

1. Write a Python function to compute the predictions according to the mean Euclidean distance to the sample points of each class.
The function should have the following interface function `[prediction] = meanPrediction(dataClass1, dataClass2, dataUnknownClass)` where `dataClass1` is an array N_1xd; `dataClass2` is an array N_2xd; `dataUnknownClass` is an array $N_t$xd; and `prediction` is an array N_tx1. *d* is the dimension of the features.

Consider the data in the table from two different classes:

	C_1			C_2		
sample	x_1	x_2	x_3	x_1	x_2	x_3
1	−5.01	−8.12	−3.68	−0.91	−0.18	−0.05
2	−5.43	−3.48	−3.54	1.30	−2.06	−3.53
3	1.08	−5.52	1.66	−7.75	−4.54	−0.95
4	0.86	−3.78	−4.11	−5.47	0.50	3.92
5	−2.67	0.63	7.39	6.14	5.72	−4.85
6	4.94	3.29	2.08	3.60	1.26	4.36
7	−2.51	2.09	−2.59	5.37	−4.63	−3.65
8	−2.25	−2.13	−6.94	7.18	1.46	−6.66
9	5.56	2.86	−2.26	−7.39	1.17	6.30
10	1.03	−3.33	4.33	−7.50	−6.32	−0.31

a) Determine the training error on your samples using only the x_1 feature value. Make use of the function `meanPrediction` you wrote.
b) Repeat but now use two feature values, x_1 and x_2.
c) Repeat but use all three feature values.
d) Discuss your results. Is it ever possible for a finite set of data that the training error be larger for more data dimensions?

2. Peter is a very predictable man. When he uses his tablet, all he does is watch movies. He always watches until his battery dies. He is also a very meticulous man. He has kept logs of every time he has charged his tablet, which includes how long he charged his tablet for and how long he was able to watch movies for afterwards. Now, Peter wants to use this log to predict how long he will be able to watch movies for when he starts so that he can plan his activities after watching his movies accordingly.

You will be able to access Peter's tablet charging log by reading from the file "TabletTrainingdata.txt". The training data file consists of 100 lines, each with 2 comma-separated numbers. The first number denotes the amount of time the tablet was charged and the second denotes the amount of time the battery lasted.

Read an input (test case) from the console (stdin) representing the amount of time the tablet was charged and output to the console the amount of time you predict his battery will last.

```
#example to read test case
timeCharged = float(input().strip())

#example to output
print(prediction)
```

3. Prove that $\mathrm{E}[\mathrm{L}] \leq \mathrm{E}[\mathrm{L}|\mathrm{L} > \mathrm{n}]$ for any discrete random variable L, where the inequality is strict if $\mathrm{P}(\mathrm{L} \leq \mathrm{n}) > 0$. E[X] represents the expected value of the random variable X.