AY: 2022-2023
L3-S5: Dept. of Electrical Engineering
Midterm Exam | Machine Learning
15/11/22 (12:30→13:30)
Teacher: A. Mhamdi

Full Name: .....
ID: .....
Class: .....
Room: .....
Time Limit: 1h

✂----------

This document contains 4 pages numbered from 1/4 to 4/4. As soon as it is handed over to you, make sure that it is complete. The 3 tasks are independent and can be treated in the order that suits you.

The following rules apply:

❶ **A handwritten double-sided A4** sheet is permitted.

❷ **The use of any electronic material**, except basic calculator, is prohibited.

❸ **Mysterious or unsupported answers** will not receive full credit.

❹ **If the provided space** is not sufficient, feel free to attach an additional sheet.

Do not write anything in this table.

Task	Points	Score
1	10	
2	5	
3	5	
Total	20	

Task N°1

⌛ 20mn | (10 points)

(a) (1 point) Machine Learning (ML) algorithms are meant to simulate:
○ intelligent machines. √ the human brain. ○ apes.

(b) (1 point) Why should everyone learn the basics of ML?
○ learning ML is fun and accessible to everyone.
○ ML strategies are being used in many industries and domains.
√ both of the above.

(c) (1 point) How is ML related to Artificial Intelligence (AI)?
○ AI is form of unsupervised ML.
√ ML is a type of AI that relies on learning through data.
○ AI focuses on classification, while ML is about clustering data.
○ ML and AI are the same thing.

(d) (1 point) In traditional computer programming, you input commands. What do you input with ML?
○ Patterns. ○ Rules. ○ Programs. √ Data.

(e) (1 point) Which of the following groups are not ML techniques?

- ○ Anomaly Detection.
- ○ Recommendation Systems.
- √ Numpy, Scipy & Scikit-Learn.
- ○ Classification.
- ○ Clustering.

(f) (1 point) You are running a company, and you want to develop learning algorithms to address each of two problems.

Problem #1: You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.

Problem #2: You would like software to examine individual customer accounts, and for each account decide if it has been hacked/compromised.

Should you treat these as classification or as regression problems?

- ○ Treat both as classification problems.
- ○ Treat both as regression problems.
- ○ Treat problem #1 as a classification problem, problem #2 as a regression problem.
- √ Treat problem #1 as a regression problem, problem #2 as a classification problem.

(g) (1 point) What is one reason not to use the same data for both your training set and your testing set?

- ○ You will almost certainly underfit the model.
- ○ You will pick the wrong algorithm.
- ○ You might not have enough data for both.
- √ You will almost certainly overfit the model.

(h) (1 point) To predict a quantity value. use

√ regression. ○ clustering. ○ classification.

(i) (1 point) You can use ML to:

- √ predict the likelihood of disease from a patient's medical history or reports.
- √ leverage weather data to predict weather events.
- √ understand the sentiment of a text.
- √ detect fake news to stop the spread of propaganda.

DO NOT WRITE ANYTHING HERE

✂----------

(j) (1 point) You work for a music streaming service and want to use supervised ML to classify music into different genres. Your service has collected thousands of songs in each genre, and you used this as your training data. Now you pull out a small random subset of all the songs in your service. What is this subset called?

○ Big Data. ○ Supervised Set. √ Test Data. ○ Data Cluster.

Task N°2 ⏳ 15mn | (5 points)

For some classification problem, we run the code given hereafter in order to display the confusion matrix, using the `crosstab` function.

[11]:
```
pd.crosstab(y_test, y_pred, \
        rownames=['Expected'], colnames=['Predicted'],\
        margins=True)
```

[11]:

Predicted	0	1	All
Expected			
0	52	15	67
1	30	38	68
All	82	53	135

For each case, calculate the precision, recall and f1-score metrics.

(a) (2 points) 1ST CASE: 0 is positive and 1 is negative.

TP	FP	TN	FN
52	30	38	15

Accuracy	Precision	Recall	F1-Score
0.667	0.634	0.776	0.69

(b) (2 points) 2ND CASE: 0 is negative and 1 is positive.

TP	FP	TN	FN
38	15	52	30

Accuracy	Precision	Recall	F1-Score
0.667	0.717	0.559	0.628

(c) (1 point) What does `pd` stand for?

PANDAS

DO NOT WRITE ANYTHING HERE

✂- -

Task Nº3 ⌛ 25mn | (5 points)

y is a random variable whose value is considered to be a linear function of some variable x:

x	-2	1	2	3
y	-5.5	0.5	2.5	4.5

The linear regression model $y = \theta_0 + \theta_1 x$ is used. Determine the values of θ_0, θ_1 using normal equation.

$$y = \begin{bmatrix} 5.5 \\ 0.5 \\ 2.5 \\ 4.5 \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} 1 & -2 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \Longrightarrow X^T X = \begin{bmatrix} 4 & 4 \\ 4 & 18 \end{bmatrix}$$

The estimated vector θ is

$$\hat{\theta} = \begin{bmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \end{bmatrix} = \begin{bmatrix} -1.5 \\ 2 \end{bmatrix}$$