AY: 2022-2023
MIDTERM EXAM | **AI-ECUE221**
13/04/23 (09:00→10:00)

M1-S2: Dept. of Electrical Engineering
Teacher: A. Mhamdi
Time Limit: 1h

This document contains **5** pages numbered from **1/5** to **5/5**. 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 **No document** is allowed in the examination room.

❷ **Any electronic material**, except basic calculator, is prohibited.

❸ **Round results** to the nearest thousandth *(i.e., third digit after the decimal point)*.

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

❺ **Task N°3:** Each correct answer will grant a mark with no negative scoring.

Task N°1

⌛ 25mn | (7 points)

Consider the following matrix of features X and the corresponding target vector y:

$$X = \begin{bmatrix} 0.25 & 1 \\ 3 & -1 \\ 2 & -0.5 \\ 1 & 0.5 \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} 1.075 \\ -1.05 \\ -0.425 \\ 0.525 \end{bmatrix}$$

We suppose that y is linearly dependent on features in X to which we need to add a bias.

Using the **stochastic gradient descent** algorithm (**SGD**), determine the parameter vector θ which maps X to y. The initial value of θ is $\begin{bmatrix} 0.1 & 0 & -0.1 \end{bmatrix}^T$. The learning rate is set at 0.1. Reproduce and fill in the following table on your paper.

k	1	2	3	4
$h_\theta(x_k)$	0	0.281	−0.740	−0.141
$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$	$\begin{bmatrix} 0.208 \\ 0.027 \\ 0.007 \end{bmatrix}$	$\begin{bmatrix} 0.0744 \\ -0.372 \\ 0.140 \end{bmatrix}$	$\begin{bmatrix} 0.106 \\ -0.31 \\ 0.125 \end{bmatrix}$	$\begin{bmatrix} 0.172 \\ -0.243 \\ 0.158 \end{bmatrix}$

Task N°2

⌛ 15mn | (6 points)

What will be the output of cell #5 after code showing hereafter is being executed.

[1]:
```
using DataFrames, MLJ
```

[2]:
```
Xdf = DataFrame(A=rand(-5:.2:3, 4), B=-1:2, C=rand(100:7:1000, 4))
```

[2]:

	A	B	C
	Float64	Int64	Int64
1	2.2	-1	639
2	2.8	0	632
3	-2.0	1	345
4	2.2	2	639

[3]:
```
describe(Xdf, :min, :max, :mean, :std)
```

[3]:

	variable	min	max	mean	std
	Symbol	Real	Real	Float64	Float64
1	A	-2.0	2.8	1.3	2.21811
2	B	-1	2	0.5	1.29099
3	C	345	639	563.75	145.871

[4]:
```
schema(Xdf)
```

[4]:

names	scitypes	types
A	Continuous	Float64
B	Count	Int64
C	Count	Int64

[5]:
```
sc = Standardizer(count=true)
Xsc = machine(sc, Xdf) |> fit! |> MLJ.transform
```

```
[ Info: Training machine(Standardizer(features = Symbol[], …), …).
```

[5]:

	A	B	C
	Float64	Float64	Float64
1	0.405751	-1.1619	0.515868
2	0.676252	-0.387298	0.46788
3	-1.48775	0.387298	-1.49962
4	0.405751	1.1619	0.515868

AY: 2022-2023 | Full Name: ....
M1-S2: Dept. of Electrical Engineering | ID: ....
Midterm Exam | AI-ECUE221 | Class: ....
13/04/23 (09:00→10:00) | Room: ....
Teacher: A. Mhamdi | Time Limit: 1h

✂- -

Answer Sheet

Task N°3 ⌛ 20mn | (7 points)

(a) (½ point) How is Machine Learning (ML) related to Artificial Intelligence (AI)?
- ○ AI focuses on classification, while ML is about clustering data.
- ○ AI is form of unsupervised ML.
- √ ML is a type of AI that relies on learning through data.
- ○ ML and AI are the same thing.

(b) (½ point) Your ML system is using labeled examples to try to predict future data, compare that data to the predicted result, and then the model. What is the best description of this ML method?
- √ Supervised learning
- ○ Unsupervised learning
- ○ Semi-supervised learning
- ○ Semi-reinforcement learning

(c) (½ point) What is a well-designed/well-fitted model?
- ○ One that has been trained with labeled training data
- ○ One that has been trained with an exhaustive set of all conditions and permutations in the training data
- ○ One that has been trained and tested with the same data
- √ One that has a high degree of accuracy and is able to accurately predict results

(d) (½ point) Your data science team wants to use ML to better filter out spam messages. The team has gathered a database of 100000 messages that have been identified as spam or not spam. If you are using supervised ML, what would you call this data set?

○ ML algorithm √ Training set ○ Big data test set ○ Data cluster

DO NOT WRITE ANYTHING HERE

(e) (½ point) Asian user complains that your company's facial recognition model does not properly identify their facial expressions. What should you do?

- ○ Retrain your model with updated hyperparameter values.
- √ Include Asian faces in your training data and retrain your model.
- ○ Include Asian faces in your test data and retrain your model.
- ○ Retrain your model with smaller batch sizes.

(f) (½ point) When we discuss "STATE", it is seen a categorical variable. When facing these type of variables - what step is required?

- ○ Including all columns
- ○ Removing all columns.
- √ Using dummy variables.

(g) (½ point) Which of the following groups are not ML techniques?

- √ **Flux** and **MLJ**
- ○ Classification and clustering
- ○ Anomaly detection and recommendation systems

(h) (½ point) Why is it important for ML algorithms to have access to high-quality data?

- ○ It will take too long for programmers to scrub poor data.
- ○ If the data is high quality, the algorithms will be easier to develop.
- √ If the data is low quality, you will get inaccurate results.
- ○ Low-quality data requires much more processing power than high-quality data.

(i) (½ point) The "Regression" technique in ML is a group of algorithms that are used for:

- ○ Finding items/events that often co-occur *(e.g., grocery items that are usually bought together by a customer)*.
- ○ Predicting a class/category of a case *(e.g., a cell is benign or malignant, or a customer will churn or not)*.
- √ Predicting a continuous value *(e.g., the price of a house based on its characteristics)*.

DO NOT WRITE ANYTHING HERE

(j) (½ point) To predict a quantity value. use "_____".
◯ classification √ regression ◯ clustering ◯ dimensionality reduction

(k) (½ point) Supervised learning deals with labeled data, while unsupervised learning deals with unlabeled data. √ True ◯ False

(l) (½ point) In the context of calculus, what is $\frac{df}{dx}$?
◯ Equivalent to f divided by x
√ The derivative of f wrt x
◯ The prediction function
◯ The derivative of x

(m) (½ point) Which of the below is a popular method to handle missing data in a given column?
◯ Replace with the standard deviation of the column.
◯ Replace with the min or max of the column.
√ Replace with the mean of the column.

(n) (½ point) With traditional programming, the programmer typically inputs commands. With ML, the programmer inputs "_____"
◯ algorithms
√ data
◯ supervised learning
◯ unsupervised learning