Higher Institute of Technological Studies of Bizerte

AY: 2022-2023
Exam | AI-ECUE221
05/06/23 (09:00→10:30)

M1-S2: Dept. of Electrical Engineering
Teacher: A. Mhamdi
Time Limit: $1\frac{1}{2}$ h

This document contains **6** pages numbered from **1/6** to **6/6**. As soon as it is handed over to you, make sure that it is complete. The **2** 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°2:** No penalty for wrong answers.

Task N°1

⌛ 30mn | (4 points)

Given the `Julia` code below:

[1]:
```
y_true = [1,1,0,1,0,1,0,0,1,0,1]
y_pred = [0,0,0,1,1,1,0,0,1,0,1];
```

`y_true` denotes the true or actual outcome values in a dataset, while `y_pred` refers to the predicted outcome values generated by a machine learning model or other predictive algorithm.

[2]:
```
using EvalMetrics
```

[3]:
```
cm = ConfusionMatrix(y_true, y_pred)
```

[3]:
```
ConfusionMatrix{Int64}(6, 5, 4, 4, 1, 2)
```

[4]:
```
using NamedArrays
```

[5]:
```
NamedArray(getproperty.(Ref(cm), [:tn :fn ; :fp :tp]),
                (["False", "True"], ["False", "True"]),
                ("pred", "true"))
```

(a) (2 points) Reproduce and fill in the confusion matrix, output of cell #5.

[5]:
```
2×2 Named Matrix{Int64}
pred   true   False   True

False           4       2
True            1       4
```

(b) (2 points) Compute the following metrics:

- ✔ *Accuracy*
- ✔ *Precision*
- ✔ *Recall*
- ✔ *F1-score*

By applying the formulas, we get:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = 0.727$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = 0.8$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = 0.667$$

$$\text{f1 - score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = 0.727$$

Higher Institute of Technological Studies of Bizerte

AY: 2022-2023	**Full Name:** .....
M1-S2: Dept. of Electrical Engineering	**ID:** .....
**Exam	AI-ECUE221**	**Class:** .....
05/06/23 (09:00→10:30)	**Room:** .....
Teacher: A. Mhamdi	**Time Limit:** 1½ h

✂----------

Answer Sheet

Task N°2 ⌛ 20mn | (16 points)

(a) (1 point) What is a machine learning model?

- ◯ A linear function
- ◯ A mechanism for improving models
- √ Generic program made specific by data.
- ◯ A set of high-level statistics and visualizations that help you understand the shape of your data.

(b) (1 point) In k-nearest neighbors, the closer you are to a neighbor, the more likely you are to

- √ share common characteristics
- ◯ be part of the root node
- ◯ have an **Euclidean** connection
- ◯ be part of the same cluster

(c) (1 point) You work for a large pharmaceutical company whose data science team wants to use unsupervised learning machine algorithms to help discover new drugs. What is an advantage to this approach?

- ◯ You will be able to prioritize different classes of drugs, such as antibiotics.
- ◯ You can create a training set of drugs you would like to discover.
- √ The algorithms will cluster together drugs that have similar traits.
- ◯ Human experts can create classes of drugs to help guide discovery.

(d) (1 point) We should use multiple linear regression to predict a dependent variable that is growing exponentially with time.

- ◯ Yes
- √ No

Do not write anything here

(e) (1 point) Your data science team is working on a machine learning product that can act as an artificial opponent in video games. The team is using a machine learning algorithm that focuses on rewards: *If the machine does some things well, then it improves the quality of the outcome.*
How would you describe this type of machine learning algorithm?

- ◯ Semi-supervised machine learning
- ◯ Supervised machine learning
- ◯ Unsupervised machine learning
- √ Reinforcement learning

(f) (1 point) What is one of the most effective way to correct for underfitting your model to the data?

- ◯ Create training clusters
- ◯ Remove predictors
- ◯ Use reinforcement learning
- √ Add more predictors

(g) (1 point) In supervised machine learning, data scientist often have the challenge of balancing between underfitting or overfitting their data model. They often have to adjust the training set to make better predictions. What is this balance called?

- ◯ The under/over challenge
- ◯ Balance between clustering classification
- √ Bias-variance trade-off
- ◯ The multiclass training set challenge

(h) (1 point) Which of the following is NOT supervised learning?

- ◯ Decision Tree
- ◯ Linear Regression
- √ PCA
- ◯ Naive Bayesian

(i) (1 point) What does it mean to underfit your data model?

- ◯ There is too little data in your training set.

Do not write anything here

- ○ There is too much data in your training set.
- √ There is not a lot of variance but there is a high bias.
- ○ Your model has low bias but high variance.

(j) (1 point) Which choice is the best example of labeled data?

- √ a spreadsheet
- ○ 20000 recorded voicemail messages
- ○ 100000 images of automobiles
- ○ hundreds of gigabytes of audio files

(k) (1 point) K-means clustering is what type of machine learning algorithm?

- ○ Reinforcement
- ○ Supervised
- √ Unsupervised
- ○ Classification

(l) (1 point) Which is the best explanation for how a model training algorithm works?

- √ Model training algorithms incrementally adjust model parameters to minimize loss function.
- ○ Model training algorithms slowly add or remove model parameters until the model fits the dataset it was trained on.
- ○ Model training algorithms train a separate model for each datapoint and then pick the one with the lowest loss function.
- ○ Model training algorithms are used on a prepared machine learning model to make predictions on new unseen data.

(m) (1 point) You're working on a binary classification task, to classify if an image contains a cat ('1') or doesn't contain a cat ('0'). What loss $\mathcal{L}$ would you choose to minimize in order to train a model?

- ○ $\mathcal{L} = y \ln \hat{y} + (1 - y) \ln (1 - \hat{y})$
- √ $\mathcal{L} = -y \ln \hat{y} - (1 - y) \ln (1 - \hat{y})$
- ○ $\mathcal{L} = \|y - \hat{y}\|_2^2$
- ○ $\mathcal{L} = \|y - \hat{y}\|_2^2 + \text{constant}$

Do not write anything here

(n) (1 point) Does Linear Regression have assumptions such as linearity or multivariate normality that need to be checked?

- √ True
- ◯ False

(o) (1 point) In the 1983 movie **WarGames**, the computer learns how to master the game of chess by playing against itself. What machine learning method was the computer using?

- ◯ Binary learning
- ◯ Supervised learning
- ◯ Unsupervised learning
- √ Reinforcement learning

(p) (1 point) Which project might be best suited for supervised machine learning?

- ◯ Data scrubbing
- √ Predicting a risk score
- ◯ Tax filing software
- ◯ Spreadsheet consolidation