INSTITUTE OF TECHNOLOGICAL STUDIES OF BIZERTE

AY: 2025-2026
MIDTERM | **Machine Learning**
Oct. 2025

L3-S5: Dept. of Electrical Engineering
Teacher: A. Mhamdi
Time Limit: 1h

This document contains **6** pages numbered from **1** to **6**. Upon receiving it, verify completeness. The **2** tasks are independent and can be solved in any order you prefer. The following rules apply:

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

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

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

❹ **Express results** in scientific notation with 3 sig figs *(e.g.,* $0.01234 \rightarrow 1.23 \times 10^{-2}$*)*.

❺ **Task N°2:** Correct answers earn points as indicated. There is no negative scoring.

Task N°1

⌛ 25mn | (8 points)

We consider the following dataset ($x_0 = 1$):

Table 1: Housing energy efficiency prediction.

House Area (x_1)	Insulation Thickness (x_2)	Window Quality (x_3)	Energy Consumption (y)
100	5	3	200
120	6	4	180
150	4	2	220
200	7	5	150
180	5	4	170

The output y can be expressed as a linear function of house characteristics *(i.e.,* $y = x^T \cdot \theta$*)*. An estimate of the parameter vector θ is given by:

$$\hat{\theta} \approx [266.667,\ -0.076,\ 3.806,\ -25.520]^T$$

(a) (2 points) Predict energy consumption for a house with: Area=160, Window Quality=6, and Insulation Thickness=4.

$$\hat{y} = \underbrace{[1,\ 160,\ 4,\ 6]}_{x^T} \cdot \underbrace{[266.667,\ -0.076,\ 3.806,\ -25.520]^T}_{\hat{\theta}}$$
$$\approx 1.17 \times 10^2$$

(b) (6 points) Compute each of the following metrics:

MAE **RMSE** **MAPE**

Given the value of $\hat{\theta}$, we can compute the predicted output:

$$\hat{y} = X \cdot \hat{\theta} \approx [201.537,\ 178.303,\ 219.451,\ 150.509,\ 169.937]^T$$

The error vector ε is:

$$\varepsilon = y - \hat{y} \approx [-1.537,\ 1.697,\ 0.549,\ -0.509,\ 0.063]^T$$

Mean Absolute Error (MAE)

$$\begin{aligned} \text{MAE} &= \frac{1}{5}\sum_{i=1}^{5}|\varepsilon_i| \\ &\approx 8.71 \times 10^{-1} \end{aligned}$$

Root Mean Squared Error (RMSE)

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{1}{5}\sum_{i=1}^{5}\varepsilon_i^2} \\ &\approx 1.08 \end{aligned}$$

Mean Absolute Percentage Error (MAPE)

$$\begin{aligned} \text{MAPE} &= \frac{1}{5}\sum_{i=1}^{5}\left|\frac{\varepsilon_i}{y_i}\right| \cdot 100\% \\ &\approx 4.67 \times 10^{-1}\ \% \end{aligned}$$

AY: 2025-2026 | Full Name: ....
L3-S5: Dept. of Electrical Engineering | ID: ....
Midterm | Machine Learning | Class: AII3..................
Oct. 2025 | Room: ....
Teacher: A. Mhamdi | Time Limit: 1h

Answer Sheet

Task N°2 — 35mn | (12 points)

(a) (½ point) What is supervised learning?
- ◯ Learning without labeled data
- √ Learning with labeled data
- ◯ Learning through reinforcement
- ◯ Learning through clustering

(b) (½ point) Which of the following is an example of supervised learning?
- ◯ K-means clustering
- √ Linear regression
- ◯ Principal Component Analysis (PCA)
- ◯ Hierarchical clustering

(c) (½ point) What is the primary goal of regression analysis?
- √ To predict a continuous output variable
- ◯ To categorize data into classes
- ◯ To reduce dimensionality
- ◯ To group similar data points

(d) (½ point) In the simple linear regression equation $y = \theta x + b$, what does θ represent?
◯ Input feature value ◯ Prediction √ Slope ◯ y-Intercept

(e) (½ point) Linear Regression is primarily used for what type of problems?
√ Regression ◯ Classification ◯ Clustering ◯ Dimensionality Reduction

(f) (½ point) In a machine learning project, the dataset used to train the model is called:
√ Training Set ◯ Validation Set ◯ Test Set ◯ Evaluation Set

(g) (½ point) Which are metrics used to evaluate regression models?
☐ Accuracy ☐ Precision √ MSE √ MAE

DO NOT WRITE ANYTHING HERE

(h) (½ point) What is the main difference between classification and regression?

- √ Classification predicts discrete labels, while regression predicts continuous values
- ◯ Classification is always unsupervised, while regression is supervised
- ◯ There is no difference
- ◯ Regression uses only linear models

(i) (½ point) What is the primary goal of machine learning?

- ◯ To write explicit programs for every task.
- √ To enable computers to learn and improve from experience without being explicitly programmed.
- ◯ To only process and store large amounts of data.
- ◯ To replace human intelligence entirely.

(j) (½ point) The process of dividing a dataset into a training set and a test set is crucial to:

- √ Evaluate the model's performance on unseen data and check for overfitting.
- ◯ Increase the model's speed.
- ◯ Make the model more complex.
- ◯ Reduce the size of the dataset.

(k) (½ point) Multiple Linear Regression differs from Simple Linear Regression because it:

- √ uses more than one feature (independent variable) to predict the target.
- ◯ predicts multiple target variables simultaneously.
- ◯ is a classification algorithm.
- ◯ uses multiple cost functions.

(l) (½ point) If a linear regression model has a high error on both the training and test data, the model is likely:

◯ A classification model ◯ Perfectly fit √ Underfit ◯ Overfit

(m) (½ point) What does a residual represent in linear regression?

- ◯ The slope of the regression line.
- ◯ The intercept of the regression line.
- ◯ The total error of the model.
- √ The difference between the predicted value and the actual value.

DO NOT WRITE ANYTHING HERE

✂----------

(n) (½ point) How do you start a comment in Python?

○ // √ # ○ /* ○ <!--

(o) (½ point) What will be the output of the following code?

```
print("Hello"[1])
```

○ H √ e ○ l ○ o

(p) (½ point) Which of the following methods can be used to add an element to a list?

□ add() □ push() √ append() √ insert()

(q) (½ point) What will be the output of this code?

```
print(bool(0))
```

√ False ○ True ○ None ○ Error

(r) (½ point) Which operator is used to check if two variables are equal?

○ = √ == ○ != ○ <

(s) (½ point) What is the output of the following code?

```
print("5" + "5")
```

○ 10 √ 55 ○ Error ○ 5 5

(t) (½ point) What will be the output of the following code snippet?

```
x = "abc"
print(x * 2)
```

√ abcabc ○ abab ab ○ aabbcc ○ abc 2

(u) (½ point) What is NumPy primarily used for in Python?

○ Creating web applications and APIs.

○ Advanced statistical modeling and machine learning algorithms.

√ Efficient handling and operations on multi-dimensional arrays and matrices.

○ Data visualization and plotting.

(v) (½ point) What is a key performance advantage of using NumPy arrays over native Python lists for numerical data?

○ Python lists are actually faster for mathematical operations.

○ NumPy arrays automatically visualize the data they contain.

Do not write anything here

✂- -

√ NumPy arrays are stored in one continuous memory location and operations are implemented in pre-compiled C code, making them much faster.

◯ NumPy arrays can only store numeric data, making them simpler.

(w) (½ point) Which of the following is the correct way to import the NumPy library with the conventional alias?

◯ `import np as numpy`

√ `import numpy as np`

◯ `from numpy import all`

◯ `use numpy`

(x) (½ point) You have a 1D NumPy array `arr = np.array([1, 2, 3, 4, 5])`. What does the operation `arr[1:4]` return?

◯ `array([1, 2, 3])`

◯ `array([2, 3, 4, 5])`

√ `array(2, 3, 4])`

◯ `array([1, 2, 3, 4])`