AY: 2025-2026 | **L3-S5: Dept. of Electrical Engineering**
EXAM | **Machine Learning** | **Teacher: A. Mhamdi**
Dec. 2025 | **Time Limit: 1½ h**

This document contains **7** pages numbered from **1** to **7**. Upon receiving it, verify completeness. The **4** 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 **Round results** to the nearest thousandth *(i.e., the third digit after the decimal point)*.

❺ **Task N°4:** Correct answers earn points as indicated. There is no negative scoring.

Task N°1

⌛ 20mn | (4 points)

A data scientist has built a logistic regression model to predict whether a student will pass an exam (1 = Pass, 0 = Fail) based on the number of hours they studied. The model has been fitted to data and the resulting coefficients are:

θ_0: -4.0 *(Intercept)*

θ_1: 1.0 *(Coefficient for hours)*

The decision rule is to predict **"Pass"** if the predicted probability is ≥ 0.5.

(a) (1 point) For every additional hour studied, how do the log-odds of passing change?

The coefficient θ_1 means for every additional hour studied, the log-odds of passing increase by 1.0.

$$\begin{aligned}
\text{log-odds}|_{\text{hrs}=1} &= -3.0 \\
\text{log-odds}|_{\text{hrs}=2} &= -2.0 \\
\text{log-odds}|_{\text{hrs}=3} &= -1.0 \\
\text{log-odds}|_{\text{hrs}=4} &= 0.0 \text{ (decision boundary)}
\end{aligned}$$

(b) (1 point) What are the log-odds of passing for a student who studied for 0 hours.

The equation for the predicted log-odds of passing is:

$$\begin{aligned}\text{log-odds}|_{\text{hrs}=0} &= \theta_0 + \theta_1 \times \text{hrs} = \text{-}4.0 + 1.0 \times \text{hrs} \\ &= \text{-}4.0 + 1.0 \times 0 \\ &= \text{-}4.0\end{aligned}$$

(c) (2 points) For someone who studied 5 hours, what would the model predict?

The equation for the predicted probability of passing is:

$$\begin{aligned}\text{P(Pass)} &= \frac{1}{1 + e^{-(-4.0+1.0\times\text{hrs})}} \\ &= \frac{1}{1 + e^{-(-4.0+1.0\times 5)}} \\ &= \frac{1}{1 + e^{-1}} \\ &= 0.731\end{aligned}$$

Since the calculated probability is greater than the threshold of 0.5, the model would predict **"Pass"**.

Task N°2 ⌛ 20mn | (4 points)

Imagine a neural network with the following structure:

Input Layer: 2 inputs (x1, x2)

Hidden Layer: 2 neurons (h1, h2) using the ReLU activation function[1].

Output Layer: 1 neuron (y) using the Sigmoid activation function[2].

Weights and Biases:

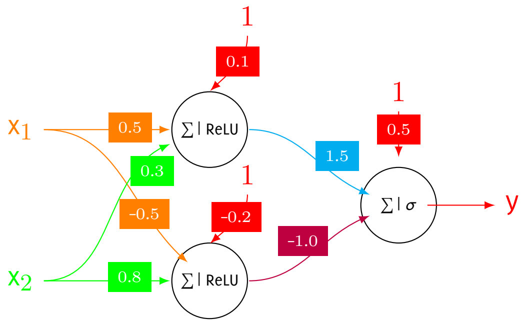
$$\mathcal{W}^{[0]} = \begin{bmatrix} 0.5 & 0.3 \\ \text{-}0.5 & 0.8 \end{bmatrix} \quad \text{and} \quad \mathbf{b}^{[0]} = \begin{bmatrix} 0.1 \\ \text{-}0.2 \end{bmatrix} \tag{1}$$

$$\mathbf{w}^{[1]} = \begin{bmatrix} 1.5 & \text{-}1.0 \end{bmatrix} \quad \text{and} \quad \mathbf{b}^{[1]} = 0.5 \tag{2}$$

(a) (2 points) Draw a sketch for the corresponding NN architecture.

[1] $\text{ReLU}(z) = \max(0, z)$

[2] $\sigma(z) = \frac{1}{1 + e^{-z}}$



(b) (2 points) Calculate the final output of the network, $\hat{y}$, for the input: $x_1 = 1.5$ and $x_2 = \text{-}1.0$.

$$
\begin{aligned}
\mathbf{z}^{[0]} &= \mathcal{W}^{[0]} \times \begin{bmatrix} 1.5 \\ \text{-}1.0 \end{bmatrix} + \mathbf{b}^{[0]} = \begin{bmatrix} 0.55 \\ \text{-}1.75 \end{bmatrix} \\
\mathbf{a}^{[0]} &= \mathrm{ReLU}(\mathbf{z}) = \begin{bmatrix} 0.55 \\ 0.0 \end{bmatrix} \\
z^{[1]} &= \mathbf{w}^{[1]} \times \mathbf{a}^{[0]} + \mathbf{b}^{[1]} = 1.325 \\
a^{[1]} &= \sigma\left(z^{[1]}\right) = 0.79 \\
\hat{y} &= a^{[1]} = 0.79
\end{aligned}
$$

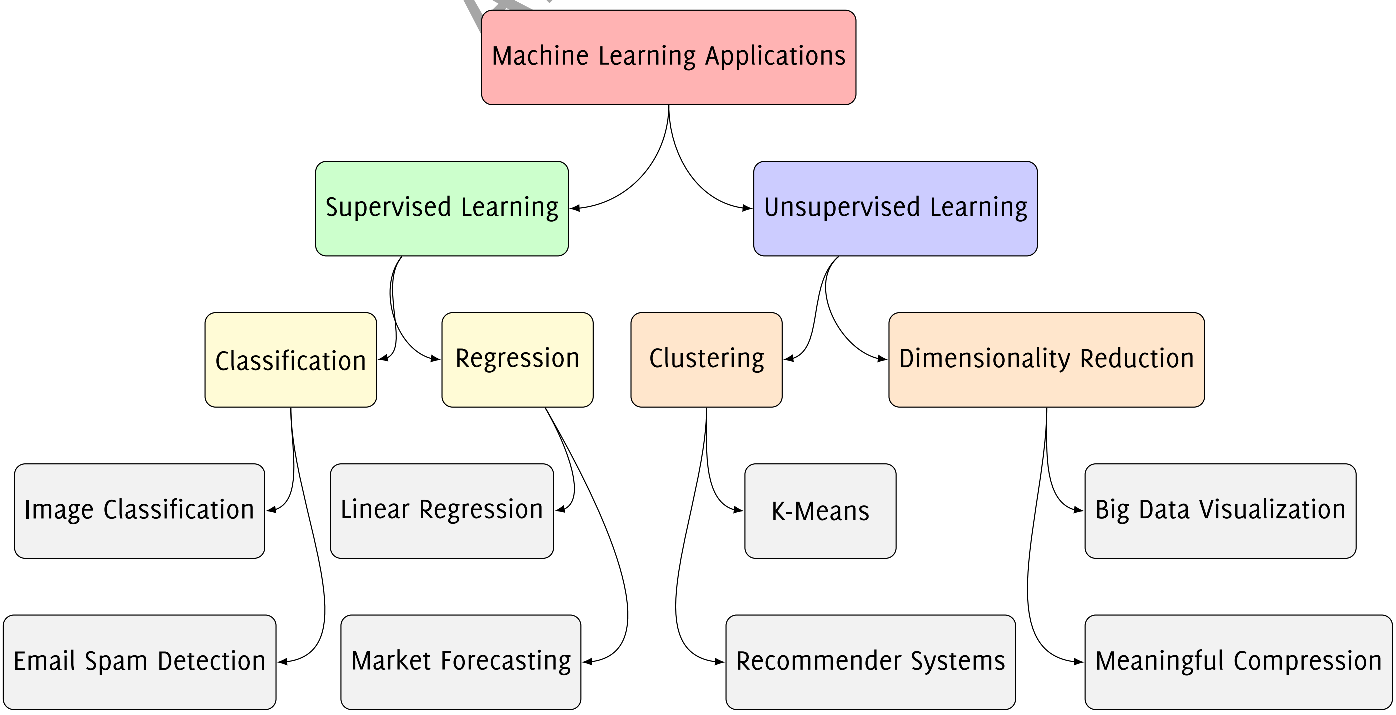
AY: 2025-2026	**Full Name:** ....
L3-S5: Dept. of Electrical Engineering	**ID:** ....
Exam | **Machine Learning**	**Class:** **AII3**..................
Dec. 2025	**Room:** ....
Teacher: A. Mhamdi	**Time Limit:** $1\frac{1}{2}$ **h**

Answer Sheet

Task N°3 ⌛ 15mn | ($3\frac{1}{2}$ points)

Categorize the following machine learning applications into their appropriate learning paradigms by placing each one in the correct branch:

Recommender Systems	Meaningful Compression	Big Data Visualization
Classification	Image Classification	Email Spam Detection
Unsupervised Learning	Linear Regression	Market Forecasting
Supervised Learning	Clustering	Dimensionality Reduction
K-Means	Regression	



DO NOT WRITE ANYTHING HERE

✂--

Task N°4 ⌛ 35mn | ($8\frac{1}{2}$ points)

(a) ($\frac{1}{2}$ point) What is the correct syntax to create a function in Python?

- ◯ `function myFunction():`
- ◯ `create myFunction():`
- √ `def myFunction():`
- ◯ `func myFunction():`

(b) ($\frac{1}{2}$ point) Which of the following data types is immutable?

◯ List ◯ Dictionary √ Tuple ◯ Set

(c) ($\frac{1}{2}$ point) Which of the following raises an exception?

√ `x = 1 / 0` ◯ `x = 10` ◯ `x = "Hello"` ◯ `x = [1, 2, 3]`

(d) ($\frac{1}{2}$ point) Which keyword is used to handle exceptions?

◯ `catch` √ `try` ◯ `handle` ◯ `except`

(e) ($\frac{1}{2}$ point) What does the `len()` function do?

- √ Returns the number of elements in an object
- ◯ Converts a string to lower case
- ◯ Joins two lists together
- ◯ Checks if an object is empty

(f) ($\frac{1}{2}$ point) Which of the following is a Python built-in function to convert a string to an integer?

√ `int()` ◯ `string()` ◯ `convert()` ◯ `parseInt()`

(g) ($\frac{1}{2}$ point) Which algorithm is commonly used for classification tasks?

◯ K-means √ k-NN ◯ PCA ◯ DBSCAN

(h) ($\frac{1}{2}$ point) What is clustering in the context of unsupervised learning?

- ◯ Assigning labels to data points
- √ Grouping similar data points together
- ◯ Predicting future values
- ◯ Analyzing time-series data

(i) ($\frac{1}{2}$ point) Which of the following is an example of unsupervised learning?

◯ SVM ◯ ANN √ K-means ◯ Logistic Regression

DO NOT WRITE ANYTHING HERE

✂- -

(j) (½ point) Which of the following is **NOT** a common clustering algorithm?

- ◯ K-means
- ◯ Hierarchical clustering
- ◯ DBSCAN
- √ Logistic Regression

(k) (½ point) You have split your data into training and testing sets using `train_test_split`. Before training your model, it is considered a best practice to:

- ◻ Fit the scaler on the entire dataset (both training and test sets) to ensure consistent scaling.
- √ Fit the scaler on the training data only, then transform both the training and test sets using that scaler.
- ◻ Transform the training data with the scaler, and then fit the scaler on the test data.
- √ It is not necessary to scale data when using regression models.

ANSWERS

(l) (½ point) Which approach would you use to identify customer segments in a dataset?

◯ Regression ◯ Classification √ Clustering ◯ Time-series analysis

(m) (½ point) The primary function of an activation function in an artificial neuron is to:

- ◯ Reduce the computational cost of the network.
- ◯ Increase the number of parameters in the model.
- √ Introduce non-linearity into the network, allowing it to learn complex patterns.
- ◯ Normalize the input data before processing.

(n) (½ point) During the training of a neural network, backpropagation is the process used to:

- ◯ Propagate the input data forward through the layers to generate an output.
- ◯ Randomly initialize the weights and biases before training begins.
- √ Calculate the error of the network's prediction and update the weights backwards from the output layer to the input layer.
- ◯ Select the best architecture (number of layers and neurons) for a given problem.

(o) (½ point) What does NLP stand for?

- ◯ Natural Language Programming

Do not write anything here

✂- -

- ◯ New Language Processing
- ◯ Neural Language Processing
- √ Natural Language Processing

(p) (½ point) The "large" in Large Language Models (LLMs) primarily refers to:

- ◯ The vast amount of disk space required to store a single prompt.
- ◯ The long time it takes to generate a single response for the user.
- √ The enormous number of parameters (weights and biases) in the neural network.
- ◯ The massive size of the vocabulary they can understand.

(q) (½ point) When a Large Language Model (LLM) like ChatGPT generates a response, it is best described as:

- ◯ Retrieving the most relevant pre-written paragraph from a massive database of text.
- ◯ Performing a logical deduction to find the single "correct" answer to the user's query.
- √ Calculating the statistical likelihood of the next word based on its training data and previous context.
- ◯ Randomly selecting words from its vocabulary to create novel sentences.