University Echahid Hamma Lakhdar - El Oued

Faculty of Exact Sciences

Department of Computer Science **JAN 2026**

Level: M1 IoT/AI Computer Science. **Hourly: 01:30**

SEMESTER EXAM – Machine Learning

Exercise 01 (6 pts):

A - Cite three cases where machine learning should be used instead of standard programming

1- Problems with **no explicit or stable rules**
2- Tasks involving **unstructured data**
3- Problems requiring **probabilistic or approximate answers**

B- Give three real-world applications that cannot be effectively solved using traditional rule-based programming, but can be successfully addressed using machine learning.

1- Image recognition
2- Speech recognition
3- Natural language understanding

C- Cite three essential topics in Machine Learning

1- Classification
2- Regression
3- clustering

Exercise 02: (14pts) Answer the following questions:

1. What does Machine Learning (ML) refer to, choose all correct answers?

A. A computational paradigm in which models infer patterns and decision rules from data rather than relying solely on explicitly programmed logic.
B. A subset of artificial intelligence that enables systems to improve their performance on a task through experience, without being explicitly programmed for each scenario.
C. A collection of deterministic algorithms whose behavior and outputs are fully specified at design time.
D. A set of statistical and optimization techniques used exclusively for descriptive data analysis, without predictive capabilities.
E. An approach to problem solving in which model parameters are automatically adjusted based on data to optimize a defined objective function.
F. A methodology focused on simulating human emotions and cognitive states directly within software systems.

2. **Which of the following are essential elements for building a basic Machine Learning model? Choose all correct answers.**

A. A large dataset.
B. A learning algorithm.
C. A suitable programming framework or environment.
D. Cloud storage infrastructure.
E. High network bandwidth.
F. A graphical processing unit (GPU).

3. **Which statement best describes a regression model in Machine Learning? One correct answer.**
   A. A model that classifies input data into a fixed set of discrete categories.
   B. A model that learns to predict continuous numerical outcomes based on input features.
   C. An unsupervised learning approach that groups similar data points into clusters.
   D. A deep learning model mainly designed for image and visual pattern recognition.
   E. A reinforcement learning approach that learns optimal actions through rewards and penalties.
   F. A statistical model used exclusively for text classification tasks.

4. **Which of the following statements correctly describes the differences between supervised, unsupervised, and reinforcement learning? One correct answer.**

A. Supervised learning uses labeled data, unsupervised learning uses unlabeled data, and reinforcement learning involves learning through trial and error with rewards and penalties.
B. Supervised learning uses unlabeled data, unsupervised learning uses labeled data, and reinforcement learning uses predefined rules.
C. Supervised learning is used for classification tasks, unsupervised learning is only for clustering, and reinforcement learning is used only in robotics.
D. Supervised learning uses rules to make decisions, unsupervised learning learns directly from outcomes, and reinforcement learning learns from historical data.

5. **Dataset Analysis: Energy Consumption Prediction (4 points)**

The following dataset represents daily energy consumption for different buildings.

Building_ID	Outdoor_ Temperature (°C)	Occupancy	Operating_ Hours	Building_ Size (m²)	Energy_ Consumption (kWh)
B001	15	120	10	1500	320
B002	30	200	12	2200	610
B003	10	80	8	1200	250
B004	25	160	11	1800	520
B005	18	100	9	1400	300

The objective is to **predict the Energy_Consumption** using the remaining variables.

5.1 Learning Paradigm: Which type of learning paradigm best describes this problem?

A. Supervised Learning
B. Unsupervised Learning
C. Reinforcement Learning

5.2 Machine Learning Task: Which Machine Learning task is most appropriate for this problem?

A. Classification
B. Regression
C. Clustering
D. Dimensionality Reduction
E. Anomaly Detection

5.3 Target Variable: Identify the target (output) variable **in the dataset.**

..... **Energy_Consumption**.....

6. Scenario: Smart Traffic Light Optimization

A city is developing a **smart traffic light system** called **SmartFlow**. The system adjusts the duration of traffic lights based on traffic conditions. After each adjustment, the system receives feedback in the form of a reward.

Attempt_ID	Green_Light_Duration (s)	Traffic_Density (cars/min)	Average_Wait_Time (s)	Reward
1	30	45	80	-6
2	40	50	55	+3
3	50	60	35	+9
4	25	40	90	-8
5	45	55	40	+7

- **Positive rewards** indicate reduced congestion and smoother traffic flow.
- **Negative rewards** indicate increased waiting time and traffic congestion.

Which type of Machine Learning model is most suitable for optimising the traffic light system through repeated interaction with its environment?

……Reinforcement Learning model……

What is the main objective of the learning process in this scenario?

A. To classify traffic density into predefined categories
B. To minimise average vehicle waiting time over time
C. To cluster similar traffic patterns
D. To predict future traffic density without interaction

7. **Advanced Scenario: Predicting production in Modern Agriculture**

An agricultural scientist is developing a system to **estimate production per hectare** based on environmental and soil conditions.

Field_ID	Soil_Moisture (%)	Nitrogen_Level (ppm)	Rainfall (mm)	Production (tons/ha)
F001	35	42	520	6.2
F002	28	30	410	4.8
F003	40	50	600	7.1

Two different learning strategies are considered for predicting the yield of a new field:

- **Strategy A:**
  The system assumes that crop yield can be represented by a fixed functional relationship between the input variables and the output. Training consists of estimating a limited set of parameters, which are then used to make predictions.
- **Strategy B:**
  The system does not build an explicit global model. Instead, it stores historical field data and predicts yield for a new field by identifying and averaging the yields of the most similar fields.

Which option correctly characterizes **Strategy A** and **Strategy B**?
A. Strategy A → Instance-based learning, Strategy B → Model-based learning
B. Strategy A → Model-based learning, Strategy B → Instance-based learning
C. Strategy A → Unsupervised learning, Strategy B → Supervised learning
D. Strategy A → Clustering-based approach, Strategy B → Classification-based approach

Suppose now the estimated model is:

$$\hat{y} = 0.5267\,SM + 0.0967\,N - 0.03133\,R$$

(Where SM=Soil_Moisture, N=Nitrogen_Level, R=Rainfall.)

Using the model above, evaluate its performance using **RMSE (the Root Mean Square Error)**.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2}$$

$$RMSE \approx 0.0043 \text{ tons per hectare}$$