University Echahid Hamma Lakhdar El Oued
Faculty of Exact Sciences
Department of Computer Science
Level: M1 IoT/AI Computer Science.

January 2025
Hourly: 01:30

Semester EXAM – Machine Learning CORRECTION

Exercise 01: Rearrange the following steps into the correct order for a standard Machine Learning project workflow (2pts):

a. Split the data into training and testing sets. **04**

b. Regularize the model and fine-tune its parameters. **08**

c. Collect or obtain the data. **01**

d. Evaluate the model using performance metrics (e.g., accuracy,RMSE).**07**

e. Train the model on the training set. **06**

f. Discover and visualize the data to understand its structure. **02**

g. Launch, monitor, and maintain the model in production. **09**

h. Select an appropriate Machine Learning algorithm. **05**

i. Prepare the data for use with Machine Learning algorithms. **03**

Exercise 02: Choose the correct answer for the following MCQ (16 pts).

1. **What is Machine Learning (ML)?**

A. A method of software development focused on building traditional, rule-based programs.
B. The study of how to make computers generate random results without using algorithms.
C. Subset of artificial intelligence focused on building systems that learn from data and improve performance over time without being explicitly programmed.
D. A process of writing optimized algorithms for high-speed data computation.

2. **What are the essential components required to develop a simple machine-learning model?**

A. A large dataset, a database management system, and a GPU.
B. Data set, an algorithm, and a framework or programming environment.
C. A programming language, cloud storage, and network bandwidth.
D. An IDE, a compiler, and test cases.

3. **How does a machine learning model identify a malignant cell from benign ones?**

A. By manually coding rules to differentiate malignant and benign cells.
B. By learning patterns and features from labelled datasets of malignant and benign cells during training.
C. By using random guesses and refining its accuracy over time without any prior data.
D. By analyzing the DNA of each cell and generating predictions without prior knowledge.

4. **What is a regression model in Machine Learning?**

A. A model used to classify data into predefined categories.
B. A model designed to predict numerical values based on input features.
C. A type of unsupervised learning model that clusters data points.
D. A deep learning model used for image recognition tasks.

5. **What is the key difference between Machine Learning and Deep Learning?**

A. Machine Learning is a subset of Deep Learning that requires less data for training models.
B. Deep Learning automates feature extraction and uses neural networks to learn complex patterns, while Machine Learning often relies on manual feature engineering.
C. Machine Learning can handle larger datasets than Deep Learning.
D. Deep Learning is used only for image-related tasks, while Machine Learning applies to all other domains.

6. **Which of the following statements correctly describes the differences between supervised, unsupervised, and reinforcement learning?**

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.

7. **Which machine learning task is described as predicting a target numeric value, such as the price of a car, based on features like mileage, age, and brand?**

A. Classification
B. Regression
C. Clustering
D. Reinforcement Learning

8. **Which type of Machine Learning model is most suitable for a robot learning how to walk through trial and error?**

A. Supervised Learning, as it uses labeled data to train the robot.
B. Unsupervised Learning, as it clusters walking patterns without labeled data.
C. Reinforcement Learning, as the robot learns by receiving rewards or penalties for its movements.
D. Regression Model, as it predicts the robot's walking speed based on input data.

9. **What does Model-Based Learning mean in Machine Learning?**

A. Using a trial-and-error approach to optimize decisions without a model.
B. Discovering the mathematical laws or relationships that the data follows and using them for predictions.
C. Learning directly from data without trying to understand its underlying structure.
D. Grouping data points based on similarities without any prior knowledge.

10. What does Instance-Based Learning mean in Machine Learning?

A. Building a mathematical model of the data to predict outcomes.

B. Comparing new cases with similar cases stored in memory to make predictions.

C. Using predefined rules to classify data without storing previous examples.

D. Grouping similar data points without relying on any past cases.

11. What does Overfitting the training data mean in Machine Learning?

A. The model performs poorly on both the training and testing datasets.

B. The model performs well on the training data but fails to generalize to new, unseen data.

C. The model uses all available data without creating a validation set.

D. The model is too simple to capture the underlying patterns in the training data.

12. Why using a Test Set in Machine Learning?

A. To train the model on new data.

B. To evaluate how well the model generalizes to unseen data.

C. To identify irrelevant features in the dataset.

D. To increase the size of the training data for better performance.

13. What does Hyperparameter Tuning mean in Machine Learning?

A. Automatically adjusting the model's weights during the training process.

B. Selecting the best set of predefined parameters (like learning rate or number of layers) to optimize a model's performance.

C. Fine-tuning the model on the test data for better generalization.

D. Replacing missing values in the dataset before training.

14. Which of the following is the best performance measure for regression projects?

A. Accuracy.

B. Mean Squared Error (MSE) or Root Mean Squared Error (RMSE).

C. Precision and Recall.

D. Confusion Matrix.

15. What is the best approach to select a machine learning algorithm (model) for training on a dataset?

A. Analyze and visualize the data to understand its structure, then choose a model suited to the data and problem type.

B. Use a model that performed well in a different project.

C. Select the most popular model, as it generally works best.

D. Pick the model with the most parameters, as it ensures better performance.

16. What does Cross-Validation evaluation mean in Machine Learning?

A. Splitting the data into a training set and a test set to evaluate model performance.

B. Testing the model on unseen data from a completely different dataset.

C. Dividing the dataset into multiple folds and using each fold as a test set while training on the remaining folds, then averaging the results to assess model performance.

D. Running the model multiple times on the training data to ensure consistency.

Exercise 03 (02pts)

A. Given the following true and predicted values:

- Y-actual = [3 , 5 , 2.5 , 7]
- Y-predicted = [2.8 , 5.3 , 2 , 6.8]

Calculate the **Root Mean Squared Error (RMSE)** for this data.

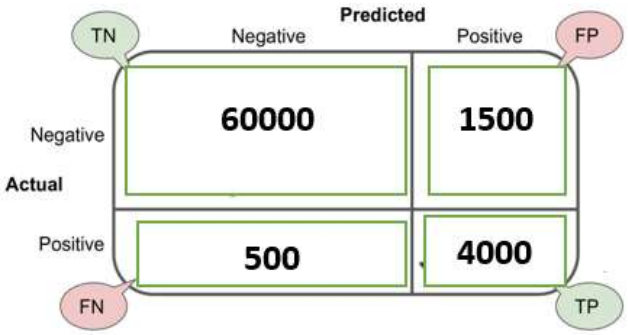
Answer:

$$\text{RMSE} = \sqrt{\frac{1}{4}\left[(3-2.8)^2 + (5-5.3)^2 + (2.5-2)^2 + (7-6.8)^2\right]}$$

B. Given the following confusion matrix:

Confision Matrix: ([[60000 (TN), 1500(FP)],
[500(FN), 4000 (TP)]])

Calculates the accuracy score.



Answer:

$$\text{Accuracy} = \frac{4000 + 60000}{4000 + 60000 + 1500 + 500}$$