University Echahid Hamma Lakhdar El Oued

Faculty of Exact Sciences
Department of Computer Science
Level: M1 IoT/AI Computer Science.

January 2024
Hourly: 01:30

Correction of Semester EXAM – Machine Learning

Exercise 01: Give the correct order of standard ML project steps:

- **a.** Discover and visualize the data. **2**
- **b.** Select an algorithm. **5**
- **c.** Launch, monitor, and maintain the model. **9**
- **d.** Collect or get the data. **1**
- **e.** Perform metrics (accuracy, RMSE, …) to evaluate the model. **7**
- **f.** Prepare the data for Machine Learning algorithms. **4**
- **g.** Split the data into train and test sets. **3**
- **h.** Train the model. **6**
- **i.** Regularize the obtained model and select the best parameters. **8**

Exercise 02: Choose the correct answer for the following MCQ

01-What is Machine Learning?

- a. A type of computer programming language.
- b. A method for teaching computers to learn from data. √
- c. A type of hardware used for processing data.
- d. A software for designing algorithms.

02-Which of the following is an example of supervised learning?

- a. Clustering.
- b. Regression. √
- c. Association rule learning.
- d. Reinforcement learning.

03-Which evaluation metric is commonly used for regression problems?

- a. Accuracy (F1 score).
- b. R-squared.
- c. Mean Squared Error (MSE). √
- d. Area Under the Receiver Operating Characteristic (ROC-AUC).

04-What is the primary goal of unsupervised learning?

- a. Predicting an output variable based on input data.
- b. Classifying input data into predefined categories.
- c. Discovering patterns and relationships in data. √
- d. Training a model to make sequential decisions.

05-What is the purpose of the "training set" in machine learning?

- a. To test the performance of the model.
- b. To validate the model's predictions.
- c. To provide data for learning the model. √
- d. To fine-tune hyperparameters.

06-What does the term "overfitting" refer to in machine learning?

a. When the model performs well on the training data but poorly on new data. √
b. When the model generalizes well to new data.
c. When the model's complexity is insufficient to capture patterns in the data.
d. When the model is perfectly fitted to the training data

07-Which evaluation metric is commonly used for binary classification problems?

a. Root Mean Squared Error (RMSE).
b. Accuracy (F1 score). √
c. R-squared.
d. Area Under the Receiver Operating Characteristic (ROC-AUC)

08-Which of the following statements regarding binary classification and multiple classification in machine learning is true?

a. Binary classification and multiple classification are two terms that refer to the same concept in machine learning.
b. Binary classification involves categorizing data into two classes, while multiple classification involves categorizing data into more than two classes. √
c. Binary classification is suitable for handling problems with more than two classes, while multiple classification is specifically designed for two-class problems.
d. Binary classification is only applicable to numerical data, whereas multiple classification is used exclusively for categorical data.

9-Which from the following statements reflect Regression Models?

a. Regression models are used for classification tasks, aiming to predict the probability of an instance belonging to a particular class.
b. The goal of regression models is to predict discrete labels for given input data.
c. Regression models predict continuous numerical values and are suitable for tasks such as predicting house prices or stock prices. √
d. Regression models are exclusively designed for handling categorical variables in a dataset.

10-Data cleaning is a crucial step in preparing a dataset for a standard regression model. Which of the following statements best reflects the data cleaning in this context?

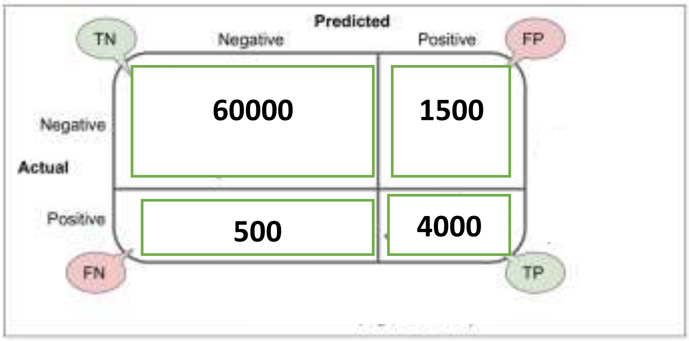
a. Data cleaning is only necessary if the dataset is too small; otherwise, larger datasets can compensate for any inconsistencies.
b. In regression modelling, data cleaning is essential to handle missing values, text values and feature scaling to ensure the reliability of the model's predictions. √
c. Data cleaning is primarily concerned with reshaping the dataset to fit the regression model's requirements and does not impact the model's accuracy.
d. The impact of data cleaning on regression models is minimal, as regression algorithms are inherently robust to noise and outliers.

Exercise 03 (06pts)

Given the following confusion matrix:

Confision Matrix: ([[60000, 1500],
[500, 4000]])

Calculates the standard **Accuracy**, **Precision**, **Recall** and **F1** scores.



		Predicted	
		Negative	Positive
Actual	Negative	60000	1500
	Positive	500	4000

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

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

$$F_1 = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = \frac{TP}{TP + \frac{FN + FP}{2}}$$