Portfolio evidence

Fundamental Concepts of Machine Learning

Rafael Marí Reyna

Robotics 9B

Due Date: September 12th, 2023

Teacher Victor Alejandro Ortiz

Machine Learning

Supervised learning

Supervised learning is a machine learning approach that's defined by its use of labeled datasets. These datasets are designed to train or "supervise" algorithms into classifying data or predicting outcomes accurately. Using labeled inputs and outputs, the model can measure its accuracy and learn over time.

Supervised learning, as the name indicates, has the presence of a supervisor as a teacher. Basically supervised learning is when we teach or train the machine using data that is well-labelled. Which means some data is already tagged with the correct answer. After that, the machine is provided with a new set of examples(data) so that the supervised learning algorithm analyses the training data(set of training examples) and produces a correct outcome from labeled data.

Supervised learning can be separated into two types of problems when data mining: classification and regression, which I will explain later.

Unsupervised learning

Unsupervised learning uses machine learning algorithms to analyze and cluster unlabeled data sets. These algorithms discover hidden patterns in data without the need for human intervention (hence, they are "unsupervised").

Unsupervised learning is the training of a machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here the task of the machine is to group unsorted information according to similarities, patterns, and differences without any prior training of data.

Unsupervised learning models are used for three main tasks: clustering, association and dimensionality reduction.

Differences between supervised and unsupervised learning

	Supervised learning	Unsupervised learning
Data	Uses labeled data with known answers or outputs	Processes unlabeled data. There are no predefined answers (i.e., no desired output is given)
Goals	To make a prediction (e.g., the future value of a house) or a classification (e.g., correctly identify spam emails)	To explore and discover patterns, structures, or relationships in large volumes of data
General tasks	Classification, regression	Clustering, dimensionality reduction, association learning
Can be applied to	Sentiment analysis, stock market prediction, house price estimation	Medical image analysis, product recommendations, fraud detection
Human supervision	Requires human intervention to provide labeled data for training	Does not require human intervention/explicit guidance
Accuracy	Tends to have higher accuracy because it learns from labeled examples with known answers	Accuracy evaluation is harder and more subjective because there are no correct answers

Probabilistic Model

A Probabilistic model in machine learning is a mathematical representation of a real-world process that incorporates uncertain or random variables. The goal of probabilistic modeling

is to estimate the probabilities of the possible outcomes of a system based on data or prior knowledge.

Probabilistic models are an essential component of machine learning, which aims to learn patterns from data and make predictions on new, unseen data. They are statistical models that capture the inherent uncertainty in data and incorporate it into their predictions. Probabilistic models are used in various applications such as image and speech recognition, natural language processing, and recommendation systems. In recent years, significant progress has been made in developing probabilistic models that can handle large datasets efficiently.

Classification

Classification is used to categorize input data into predefined classes or categories. By training with labeled data, the computer learns to recognize and differentiate various features or characteristics associated with each class. For example, in image classification, the goal may be to identify objects in an image. Similarly, classification can be used to predict discrete outcomes, like determining whether it will rain on a given day.

Regression

Regression is a type of classification where we forecast a number instead of a category. With regression, the predicted outcomes are real values, such as the expected price of a house (based on information like square footage or location). Regression can assist companies with sales predictions by considering variables such as weather, social media presence, or inbound tourists.

Differences between classification and regression.

Classification	Regression
The target variables are discrete.	The target variables are continuous.
In this algorithm, we try to find the best possible decision boundary which can separate the two classes with the maximum possible separation.	In this algorithm, we try to find the best-fit line which can represent the overall trend in the data.
Evaluation metrics like Precision, Recall, and F1-Score are used here to evaluate the performance of the classification algorithms.	Evaluation metrics like Mean Squared Error, R2-Score, and MAPE are used here to evaluate the performance of the regression algorithms.
Here we face the problems like binary Classification or Multi-Class Classification problems.	Here we face the problems like Linear Regression models as well as non-linear models.
Input Data are Independent variables and categorical dependent variable.	Input Data are Independent variables and continuous dependent variable.
Output is Categorical labels.	Output is Continuous numerical values.

References

J. Delua, “Supervised vs. Unsupervised Learning: What’s the Difference?,” IBM Blog, Mar. 12, 2021. https://www.ibm.com/blog/supervised-vs-unsupervised-learning/

K. Nikolopoulou, “Supervised vs. Unsupervised Learning: Key Differences,” Scribbr, Jul. 06, 2023. https://www.scribbr.com/ai-tools/supervised-vs-unsupervised-learning/

“Probabilistic Model in Machine Learning - Javatpoint,” www.javatpoint.com. https://www.javatpoint.com/probabilistic-model-in-machine-learning

“Probabilistic Models in Machine Learning,” GeeksforGeeks, May 09, 2023. https://www.geeksforgeeks.org/probabilistic-models-in-machine-learning/

“ML | Classification vs Regression - GeeksforGeeks,” GeeksforGeeks, Jan. 08, 2019. https://www.geeksforgeeks.org/ml-classification-vs-regression/

“Regression vs. Classification: What’s the Difference?,” Statology, Oct. 25, 2020. https://www.statology.org/regression-vs-classification/