Univerisdad Politecnica de Yucatan

Victor Ortiz

Machine Learning

Robotics .-9a

Axel Esdras De Flores Islas

Investigation task

Fundamental Concept.

Define the concept and characteristics of and characteristic of supervised and unsupervised learning

Supervised Learning:

Supervised learning is considered as a machine learning paradigm with which the algorithm obtains learning from data pre-established with the training function, labeled to have the ability to execute predictions or make decisions without human intervention. Its key features include:

Labeled Data:

In supervised learning, the training data set consists of input-output pairs, in which each input is linked to a predetermined output to generate the training (label).

Objective:

The general objective of supervised learning is to generate a mapping function (training) that has the capacity to predict the output data of new data of previously unestablished entries.

Feedback:

The algorithm enters an execution and learning loop from the labeled input and output data, where it compares its predictions with the already labeled data in order to reduce the probability of error.

Examples:

Common applications include image classification, speech recognition, and regression tasks.

Unsupervised Learning:

On the other hand, unsupervised learning is a machine learning approach in which the algorithm explores patterns and structures in unlabeled data. Its key features include:

Unlabeled Data:

In unsupervised learning, the training set consists only of input data without corresponding labels or categories.

Goal:

The primary goal of unsupervised learning is to discover inherent patterns, groupings, or representations within data without predefined categories.

No Feedback:

Unlike supervised learning, there is no explicit feedback in unsupervised learning. The algorithm independently explores data patterns.

Examples: Common applications include clustering (grouping similar data points), dimensionality reduction, and generative modeling.

Define the concept of a probabilistic model

"Probabilistic Models are one of the most important segments in Machine Learning, which is based on the application of statistical codes to data analysis. This dates back to one of the first approaches of machine learning and continues to be widely used today. Unobserved variables are seen as stochastic in probabilistic models, and interdependence between variables is recorded in a joint probability distribution. It provides a foundation for embracing learning for what it is. The probabilistic framework outlines the approach for representing and deploying model reservations. In scientific data analysis, predictions play a dominating role. Their contribution is also critical in machine learning, cognitive computing, automation, and artificial intelligence.

These probabilistic models have many admirable characteristics and are quite useful in statistical analysis. They make it quite simple to reason about the inconsistencies present across most data. In fact, they may be built hierarchically to create complicated models from basic elements. One of the main reasons why probabilistic modeling is so popular nowadays is that it provides natural protection against overfitting and allows for completely coherent inferences over complex forms from data." (1)

Explain the diferrences between supervised and unsupervised learning

1. Training data:

In this type of learning, the algorithm is trained with a set of labeled data. This means that for a set of input data, there is a labeled output, which will feed the understanding and training pattern of what the code should arrogate for the following inputs without established outputs. In this way we train our algorithm to predict inputs to outputs based on its previous training.

Example of understanding:

A young man (algorithm) who wants to learn to play basketball approaches a coach (developer), who teaches him the necessary rules (input data): such as fouls, fouls, violations. And the coach will indicate which actions generate the fouls and fouls (labeled output data). Thus, as time passes, the player generates a criterion on what actions will be violations, faults, fouls.

Unsupervised learning:

In unsupervised learning, the algorithm is trained on an unlabeled data set. Corresponding output labels are not provided. The goal of the algorithm is to find patterns, structures, or relationships within the data without predefined categories.

Example of understanding:

Continuing with the case of the player, he arrives at a court without prior knowledge (input data), and begins to play with some people who are there (the algorithm is running) and over time realizes which actions are fouls, fouls according to what he sees in the course of the game (he discovers patterns of behavior)

2. Objective:

Supervised Learning:

The main goal of supervised learning is to learn a mapping function from inputs to outputs. This allows the algorithm to make predictions or classifications of new, invisible data.

Unsupervised Learning

– The primary goal of unsupervised learning is to discover hidden patterns, cluster similar data points, or reduce the dimensionality of the data. It is not intended to predict specific output labels.

3. Comments:

Supervised Learning:

Supervised learning algorithms receive feedback during training. They compare their predictions with the true labels and adjust their internal parameters to minimize prediction errors. This feedback loop is crucial to improve the accuracy of the model.

Example of understanding:

continuing with the case of the player, throughout his training the coach makes comments and suggestions in order to generate a correct pattern of understanding parameters within the player. In addition, these comments teach the player that plays (input data) are performing poorly (unwanted output data) being crucial for better game performance (output data).

Unsupervised learning:

Unsupervised learning algorithms do not receive explicit feedback during training because there are no labels. They explore data independently, making discoveries based on statistical properties and relationships within the data set.

Example of understanding:

In the case of the player who is not trained, he does not receive specific comments from someone with experience (equiped output data) during his learning process. So he is exploring and creating learning from the creation of behavior patterns based on his learning.

4. Examples:

Supervised Learning :

Common applications of supervised learning include image classification, speech recognition, email spam detection, and regression (number value prediction) tasks.

Unsupervised learning:

Common applications of unsupervised learning include clustering (grouping similar data points), dimensionality reduction (reducing the complexity of the data), and generative modeling (creating new data samples that resemble the data of training).

5. Evaluation:

Supervised Learning :

Supervised learning models can be evaluated using metrics such as accuracy, precision, recall, F1 score, mean square error (MSE), etc. depending on the specific task.

Unsupervised learning:

Evaluating unsupervised learning models can be more challenging because there are no explicit labels to compare against. Evaluation is often based on the quality of the discovered patterns or the effectiveness of the model output in subsequent tasks.

Identify the difference between the concepts of regression and classification

"The most significant difference between regression vs classification is that while regression helps predict a continuous quantity, classification predicts discrete class labels. There are also some overlaps between the two types of machine learning algorithms.

- A regression algorithm can predict a discrete value which is in the form of an integer quantity
- A classification algorithm can predict a continuous value if it is in the form of a class label probability

Let's consider a dataset that contains student information of a particular university. A regression algorithm can be used in this case to predict the height of any student based on their weight, gender, diet, or subject major. We use regression in this case because height is a continuous quantity. There is an infinite number of possible values for a person's height.

On the contrary, classification can be used to analyse whether an email is a spam or not spam. The algorithm checks the keywords in an email and the sender's address is to find out the probability of the email being spam. Similarly, while a regression model can be used to predict temperature for the next day, we can use a classification algorithm to determine whether it will be cold or hot according to the given temperature values." (2)

In a simpler way, the classification process refers more to the interpretation of "1" and "0". Examples: Tall (1) / Short (0), Child (1) / Adult (0), Teacher (1) / Student (0). In a way in which it allows to generate an algorithm with multiple classifications but that in a primitive way falls on the operation of "1" and "0". Unlike the regressions manage to quantitatively predict the output data, unlike in the classification the output data is qualitative.

References:

https://www.simplilearn.com/tutorials/machine-learning-tutorial/what-are-probabilistic-models (1)

https://www.springboard.com/blog/data-science/regression-vs-classification/#:~:text=The%20most%20significant%20difference%20between,types%20of%20machine%20learning%20algorithms. (2)