PRIOR KNOWLEDGE

Supervised Learning

Supervised learning is the type of machine learning in which machines are trained using well "labelled" training data, and on the basis of that data, machines predict the output. The labelled data means some input data is already tagged with the correct output.

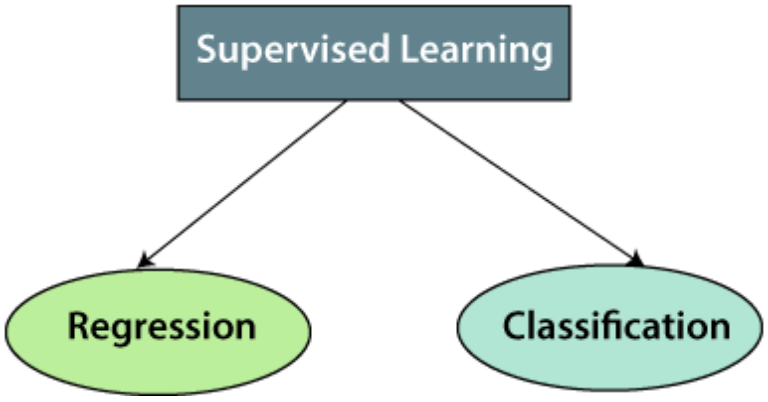
In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly. It applies the same concept as a student learns in the supervision of the teacher.

Supervised learning is a process of providing input data as well as correct output data to the machine learning model. The aim of a supervised learning algorithm is to find a mapping function to map the input variable(x) with the output variable(y).

In the real-world, supervised learning can be used for Risk Assessment, Image classification, Fraud Detection, spam filtering, etc.

Types of supervised Machine learning Algorithms:

Supervised learning can be further divided into two types of problems:



1. Regression

Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc. Below are some popular Regression algorithms which come under supervised learning

- Linear Regression
- Regression Trees
- Non-Linear Regression
- Bayesian Linear Regression
- Polynomial Regression

2. Classification

Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc.

Spam Filtering,

- Random Forest
- Decision Trees
- Logistic Regression
- Support vector Machines

RISK ASSESSMENT

Risk assessment has a primary role in safety-critical industries. However, it faces a series of overall challenges, partially related to technology advancements and increasing needs. There is currently a call for continuous risk assessment, improvement in learning past lessons and definition of techniques to process relevant data, which are to be coupled with adequate capability to deal with unexpected events and provide the right support to enable risk management. Through this work, we suggest a risk assessment approach based on machine learning. In particular, a deep neural network (DNN) model is developed and tested for a drive-off scenario involving an Oil & Gas drilling rig. Results show reasonable accuracy for DNN predictions and

general suitability to (partially) overcome risk assessment challenges. Nevertheless, intrinsic model limitations should be taken into account and appropriate model selection and customization should be carefully carried out to deliver appropriate support for safety-related decision-making.

IMAGE CLASSIFICATION

Image classification is a supervised learning problem: define a set of target classes (objects to identify in images), and train a model to recognize them using labeled example photos. Early computer vision models relied on raw pixel data as the input to the model. However, as shown in Figure 2, raw pixel data alone doesn't provide a sufficiently stable representation to encompass the myriad variations of an object as captured in an image. The position of the object, background behind the object, ambient lighting, camera angle, and camera focus all can produce fluctuation in raw pixel data; these differences are significant enough that they cannot be corrected for by taking weighted averages of pixel RGB values

CLUSTERING

It is basically a type of unsupervised learning method. An unsupervised learning method is a method in which we draw references from datasets consisting of input data without labeled responses. Generally, it is used as a process to find meaningful structure, explanatory underlying processes, generative features, and groupings inherent in a set of examples.

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them.

Unsupervised learning

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.

Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore the machine is restricted to finding the hidden structure in unlabeled data by itself.

For instance, suppose it is given an image having both dogs and cats which it has never seen.

MARKET BASKET ANALYSIS

Nowadays Machine Learning is helping the Retail Industry in many different ways. You can imagine that from forecasting the performance of sales to identify the buyers, there are many applications of machine learning(ML) in the retail industry. “Market Basket Analysis” is one of the best applications of machine learning in the retail industry. By analyzing the past buying behavior of customers, we can find out which are the products that are bought frequently together by the customers.

SEMANTIC CLUSTERING

We describe a semantic clustering method designed to address shortcomings in the common bag-of-words document representation for functional semantic classification tasks. The method uses WordNet-based distance metrics to construct a similarity matrix, and expectation maximization to find and represent clusters of semantically-related terms. Using these clusters as features for machine learning helps maintain performance across distinct, domain-specific vocabulary while reducing the size of the document representation. We present promising results

FLASK library

A Flask extension typically has flask in its name as a prefix or suffix. If it wraps another library, it should include the library name as well. This makes it easy to search for extensions, and makes their purpose clearer.

A general Python packaging recommendation is that the install name from the package index and the name used in import statements should be related. The import name is lowercase, with words separated by underscores (_). The install name is either lower case or title case, with words separated by dashes (-). If it wraps another library, prefer using the same case as that library’s name.

Here are some example install and import names:

- `Flask-Name` imported as `flask_name`
- `flask-name-lower` imported as `flask_name_lower`
- `Flask-ComboName` imported as `flask_comboname`
- `Name-Flask` imported as `name_flask`

Configuration Techniques

- Configuration per application instance, through `app.config` values. This is configuration that could reasonably change for each deployment of an application. A common example is a URL to an external resource, such as a database. Configuration keys should start with the extension's name so that they don't interfere with other extensions.

- Configuration per extension instance, through `init` arguments. This configuration usually affects how the extension is used, such that it wouldn't make sense to change it per deployment.