

Supervised and Unsupervised Machine Learning Algorithms

What is supervised machine learning and how does it relate to unsupervised machine learning?

In this post you will discover supervised learning, unsupervised learning and semis-supervised learning. After reading this post you will know:

- About the classification and regression supervised learning problems.
- About the clustering and association unsupervised learning problems.
- Example algorithms used for supervised and unsupervised problems.
- A problem that sits in between supervised and unsupervised learning called semi-supervised learning.

Let's get started.



Supervised and Unsupervised Machine Learning Algorithms

Photo by [US Department of Education](#), some rights reserved.

Supervised Machine Learning

The majority of practical machine learning uses supervised learning.

Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output.

$$Y = f(X)$$

The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.

It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. We know the correct answers, the algorithm iteratively makes predictions on the training data and is corrected by the teacher. Learning stops when the algorithm achieves an acceptable level of performance.

Supervised learning problems can be further grouped into regression and classification problems.

- **Classification:** A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”.
- **Regression:** A regression problem is when the output variable is a real value, such as “dollars” or “weight”.

Some common types of problems built on top of classification and regression include recommendation and time series prediction respectively.

Some popular examples of supervised machine learning algorithms are:

- Linear regression for regression problems.
- Random forest for classification and regression problems.
- Support vector machines for classification problems.

Unsupervised Machine Learning

Unsupervised learning is where you only have input data (X) and no corresponding output variables.

The goal for unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data.

These are called unsupervised learning because unlike supervised learning above there is no correct answers and there is no teacher. Algorithms are left to their own devices to discover and present the interesting structure in the data.

Unsupervised learning problems can be further grouped into clustering and association problems.

- **Clustering:** A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.
- **Association:** An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

Some popular examples of unsupervised learning algorithms are:

- k-means for clustering problems.
- Apriori algorithm for association rule learning problems.

Semi-Supervised Machine Learning

Problems where you have a large amount of input data (X) and only some of the data is labeled (Y) are called semi-supervised learning problems.

These problems sit in between both supervised and unsupervised learning.

A good example is a photo archive where only some of the images are labeled, (e.g. dog, cat, person) and the majority are unlabeled.

Many real world machine learning problems fall into this area. This is because it can be expensive or time-consuming to label data as it may require access to domain experts. Whereas unlabeled data is cheap and easy to collect and store.

You can use unsupervised learning techniques to discover and learn the structure in the input variables.

You can also use supervised learning techniques to make best guess predictions for the unlabeled data, feed that data back into the supervised learning algorithm as training data and use the model to make predictions on new unseen data.

Summary

In this post you learned the difference between supervised, unsupervised and semi-supervised learning. You now know that:

- **Supervised:** All data is labeled and the algorithms learn to predict the output from the input data.
- **Unsupervised:** All data is unlabeled and the algorithms learn to inherent structure from the input data.
- **Semi-supervised:** Some data is labeled but most of it is unlabeled and a mixture of supervised and unsupervised techniques can be used.

Do you have any questions about supervised, unsupervised or semi-supervised learning? Leave a comment and ask your question and I will do my best to answer it.

Source: <https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/>