Management Consultants in Big Data

BigData4 Analytics

Machine Learning Demystified

By Jackie Down

Implementing machine learning into business operations can be daunting, but it is now necessary to move businesses forward who are wishing to drive growth and innovation. Of course, data is the key and undoubtedly a major hurdle!

Cutting-edge companies have realised that data is a strategic asset that needs to be harvested for rich insights that can help fuel their respective businesses. They have also recognised that there must be a culture conducive to turning data into something that's actionable from the top down. Until now, most of today's data creation has been from the ever-increasing powerful mobile devices and the various social media channels. The second wave of data growth, which will significantly increase the data growth is the Internet of Things (IoT). This data will continue to grow as more and more source systems than ever come online over the coming years. With all this data comes the challenge of integrating all these new and complex data sources across businesses.

Here is where the application of machine learning can help. In traditional computer programming, one writes specific instructions for the computer to process the input it is provided and produce an output. For example, the input can be an application for a credit card, the computer program is an instruction to process this application, extract the useful pieces of information, compare it with other data and produce an output, which in this case would be a recommendation to accept or reject the credit card application.

In contrast, a machine learning program does not have a specific instruction set on which credit card applications to accept or reject, but instead would learn from the input data it has been provided with and progressively improve its performance automatically through experience.

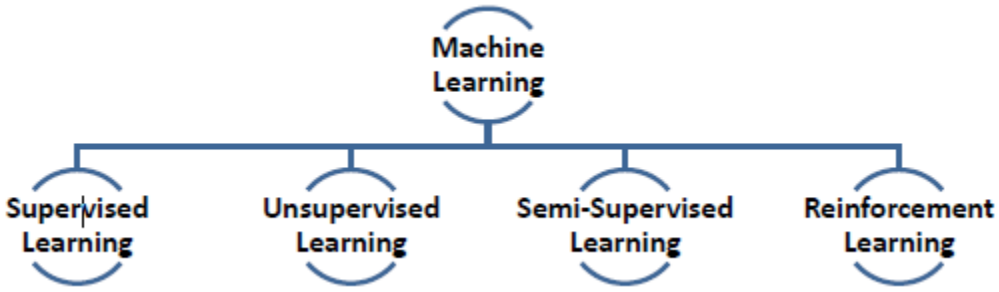
There are some tasks that are more easily 'programmed' through conventional programming, while some others are more amenable to machine learning. Humans do certain things involuntarily that we cannot articulate in words. For example, I can tell you how I button my coat, but I cannot explain why I recognise someone's face. It is hard to write a traditional computer program to identify a face. However, you can load

the computer with thousands of images of faces along with a picture of the face you want it to recognise until it can decipher the difference.

Machine learning isn't magic, but it has the capacity to help companies develop powerful revenue generating products and solutions. It opens yet another way in which we can understand data and make useful recommendations without necessarily knowing exactly how humans solve problems and make recommendations. Traditional analytics tools are limited by data volume and the need for human interaction to specify program execution, machine learning offers the scale, speed and accuracy needed to truly uncover the full value of big data. High-performance machine learning can analyse all a data set rather than extrapolate against a sample of it. This scalability not only allows predictive solutions to be more accurate, it also illustrates the importance of software speed to interpret billions of data points in real-time and to analyse live streaming data. Also, unlike traditional analysis, machine learning thrives on growing datasets. The more data fed into the system, the more it can learn and apply the results to higher quality insights.
There is tremendous opportunity in the application of machine learning to discover valuable insights that can lead to better and faster business decisions.

Types of Machine Learning terminology demystified

All Machine Learning tasks can be classified in several categories; the main ones are shown below:

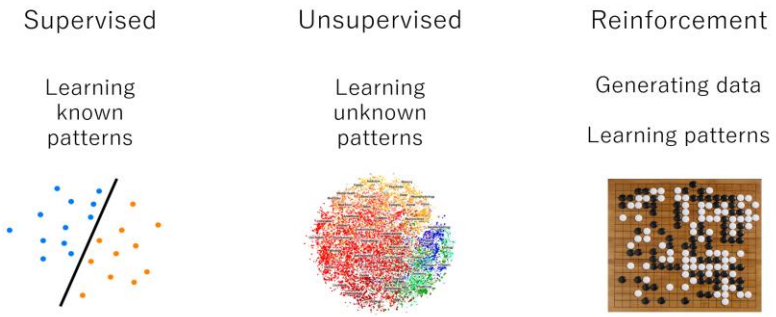


Supervised Learning - Learning a model from training data where the right answers (labels) are known up front. This is easier to explain using an example. Let us imagine that we want to teach a computer to distinguish pictures of cats and dogs. We can ask some of our friends to send us pictures of cats and dogs adding a tag 'cat' or 'dog'. Labelling is usually done by human annotators to ensure a high quality of data. So now we know the true labels of the pictures and can use this data to **"supervise"** our algorithm in learning the right way to classify images. Once our algorithm learns how to classify images we can use it on new data and predict labels ('cat' or 'dog' in our case) on previously unseen images.

Unsupervised machine learning: means that we derive a learning algorithm of the labels we used in supervised learning. We just provide machine learning with a large amount of data and characteristic of each observation (single piece of data). As an example, imagine your friends were not very helpful and forgot to label the images of cats and dogs that they have sent. However, you still want to split this data into 2 categories. You can employ unsupervised ML (in this case a technique called clustering) to separate your images in two groups based on some inherent features (characteristics) of the pictures.

Semi-Supervised Learning: learn from both labelled and unlabelled data. It has been discovered that the use of unlabeled data together with a small amount of labeled data can improve accuracy considerably. The collection of unlabeled data is inexpensive relative to labeled data. Often labeled data is scarce and unlabeled data is plentiful.

R**einforcement learning**. This can be easily illustrated by an example of learning to play chess. As input to this problem information is received about whether a game played was won or lost. So, ML does not have every move in the game labelled as successful or not, but only has the result of the whole game. Therefore, the algorithm can play a lot of games and each time gives bigger "weights" to those moves that resulted in a winning combination.



In this article, we tried to bridge the gap between the world of technology and statistics, and provide a short insight about the mysterious world of machine learning.

www.bigdata4analytics.com