

Machine Learning

A short writeup by Harine Vidyasekaran

What is Machine Learning?

Machine Learning (ML) is a sub-field of Artificial intelligence (AI), which is broadly defined as the capability of a machine to imitate “intelligence”, which means intelligent human behaviour. More specifically, it gives computers the capability to learn without being explicitly programmed and it is an approach to data analysis that involves building and adapting models, which allow programs to "learn" through experience.

How does it work?

In ML, algorithms are constructed and trained (with the use of data called training data and testing data) so that they adapt their models to improve their ability to make predictions, and to find patterns and correlations in large data sets to make the best decisions and predictions based on that analysis. ML applications improve with use and become more accurate the more data they have access to.

Types

The basic layers of AI, to understand the whole picture:

AI > ML > Neural networks > Deep learning.

AI processes data to make decisions and predictions. AI is the parent of all the machine learning subsets beneath it. Within the first subset is machine learning; within that is deep learning, and then neural networks within that.

There are 7 steps of Machine Learning:

1. Gathering the required data
2. Preparing that data
3. Choosing a ML model
4. Training using the data
5. Evaluation or testing
6. Hyperparameter tuning
7. Prediction

Specifically, types of ML:

Machine learning is comprised of different types of machine learning models, using various algorithmic techniques. Depending upon the nature of the data and the desired outcome, one of four learning models (shown next) can be used. Within each of those models, one or more algorithmic techniques may be applied – relative to the data sets in use and the intended results.

1. Supervised machine learning models are trained with labelled data sets, which allow the models to learn and grow more accurate over time. Eg: an algorithm would be trained with pictures of dogs and other things, all labelled by humans, and the machine would learn ways to identify pictures of dogs on its own. Supervised machine learning is the most common type used today.

2. Semi-supervised learning: This approach to machine learning involves a mix of two ML types above and below. Data scientists may feed an algorithm mostly labelled training data, but the model is free to explore the data on its own and develop its own understanding of the data set.

3. Unsupervised machine learning: The program looks for patterns in unlabelled data. Unsupervised machine learning can find patterns or trends that people aren't explicitly looking for. Eg: an unsupervised machine learning program could look through online sales data and identify different types of clients making purchases.

4. Reinforcement machine learning: trains machines through trial and error to take the best action by establishing a reward system. Reinforcement learning can train models to play games or train autonomous vehicles to drive by telling the machine when it made the right decisions, which helps it learn over time what actions it should take.

Applications

Applications of machine learning are all around us – in our homes, our shopping carts, our entertainment media, and our healthcare. ML is also behind chatbots and predictive text, language translation apps, the shows Netflix suggests to you, and how your social media feeds are presented. It can include giving enterprises a view of trends in customer behaviour and business operational patterns which help companies understand their audience better. More importantly, we happen to be in the “Data is the new oil” digital era. Just like oil, which is worth more when transformed into more constructive objects (gasoline, cosmetics, and various other products), when data is processed, analysed, and utilized efficiently, it will have much greater value. But the traditional methods of data analysis does not work as we progress time, for the sheer amount of data that humans and machines produce is ginormous. For this, we turn to automation and ML.

For applications of ML, Python is the best programming language due to the various benefits it has. Other languages for ML applications are: C++, JavaScript, Java, C# etc.

Important and common applications include:

1. Product/Video recommendation – recommending products and videos for the user based on their past likes and searches
2. Image recognition – identifying parts of an image, face detection, pattern recognition
3. Wildlife preservation – recording and analysing behaviour of animals so as to provide help regulate and monitor their populations
4. Predict diseases – Identifying patterns and symptoms in a patient's history and making an analysis
5. Algorithmic trading - important data is extracted to automate or support imperative investment activities like buying or selling stocks.