

Machine Learning

What is Machine Learning?

- Machine Learning is the science (and art) of programming computers so they can learn from data.
- Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed. —Arthur Samuel, 1959
- A computer program is said to learn from experience E with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with experience E . —Tom Mitchell, 1997

Machine Learning is great for:

- Problems for which existing solutions require a lot of hand-tuning or long lists of rules: one Machine Learning algorithm can often simplify code and perform better.
- Complex problems for which there is no good solution at all using a traditional approach: the best Machine Learning techniques can find a solution.
- Fluctuating environments: a Machine Learning system can adapt to new data.
- Getting insights about complex problems and large amounts of data.

Examples of Applications:

- Detecting tumours in brain scans
- Automatically classifying news articles
- Analysing images of products on a production, typically performed using convolutional neural networks
- Detecting credit card fraud
- Building an intelligent bot for a game

Types of Machine Learning Systems:

There are so many different types of Machine Learning systems that it is useful to classify them in broad categories based on:

- Whether or not they are trained with human supervision (supervised, unsupervised, semi supervised, and Reinforcement Learning)
- Whether or not they can learn incrementally on the fly (online versus batch learning)
- Whether they work by simply comparing new data points to known data points, or instead detect patterns in the training data and build a predictive model, much like scientists do (instance-based versus model-based learning) These criteria are not exclusive; you can combine them in any way you like. For example, a state-of-the-art spam filter may learn on the fly using a deep neural network model trained using examples of spam and ham; this makes it an online, model based, supervised learning system.

Supervised learning:

In the data there is always a predefined outcome.

The 2 groups of problems: Classification and Regression.

- Classification: predicts which class a given sample of data (sample of descriptive features) is part of (discrete value).
- Regression: predicts continuous value.

Unsupervised learning:

There are not predefined and known set of outcomes and look for hidden patterns and relations in the data

Here are some of the most important **supervised learning algorithms**:

- k-Nearest Neighbours
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVMs)
- Decision Trees and Random Forests
- Neural networks

Here are some of the most important **unsupervised learning algorithms**:

- Clustering
- Anomaly detection and novelty detection
- Visualization and dimensionality reduction
- Association rule learning

Machine Learning Steps:

1. Data Collection: gathering data from various sources.
Types of data: Qualitative (nominal, ordinal) and Quantitative (continuous, discrete)
2. Data Pre-processing/Preparation: cleaning data to have homogeneity.
3. Data Visualization and Feature Engineering: transforming results into visual graphs.
4. Choose a Model: selecting the right algorithm.
5. Train and Testing/Evaluating the Model: gaining insights from the model's results.
6. Deployment: integrate a model into an existing production environment.