Machine Learning - An Introduction

# Introduction

Machine Learning is a subfield of computer science that gives ‘computers the ability to learn without being explicitly programmed.’ It is mainly used to implement tasks that can’t be explicitly programmed with a set of rules. For example, consider a task of categorising an email as spam or not spam. There are no clear set of rules and boundaries to say whether an email is spam and also they dynamically change all the time. So, it is difficult to track the exact definition of spam mail. The problem can be solved with the help of machine learning. The program tries to learn from the data already existed i.e., emails that are already classified as spam or not and it evolves itself every time you mark an email as spam. This is only one example of application of machine learning. The implications and the applications of machine learning are vast and enormous and can be revolutionary.

Machine Learning overlaps with many other subfields of computer science like data mining, artificial intelligence, artificial neural networks, natural language processing. Data Scientist, ‘the sexiest job of 21st century’ as quoted by Harvard Business Review has many in common with Machine Learning and can be treated as a prerequisite for Data Science domain.

# Categories and subcategories

Machine learning can be broadly divided into 3 categories. They are:

1. Supervised Learning: The computer is given a set of inputs and desired outputs. Using the data supplied, it needs to predict the output. So, the system needs to identify the general rule that can give a certain output to a certain input.
2. Unsupervised Learning: The systems is given some data and it needs to find any hidden patterns in the data and needs to categorize the given data using the patterns. The main challenge is to find the structure in the input by itself without any assistance from the user.
3. Reinforcement Learning: The system needs to perform a goal by learning to interact in the environment. For example, A system learning to drive a car. It needs to interact with the different controls in the car and know what does each of the control do. It then needs to perform a goal like driving to a destination keeping the set of rules in mind and updating itself when needed.

Each of the categories can still be sub categorised or can be categorised differently depending on the desired output.

Supervised Learning can still be classified into:

1. Classification
2. Regression

Classification is predicting the output into discrete categories and regression is predicting within continuous categories. In other words, classification is classifying output into discrete categories and regression is output as continuous function. Our previous example i.e., classifying an email whether spam or not is a classification problem. Predicting the age of a person based on their picture is a regression problem.

Unsupervised Learning can still be classified into clustering and non-clustering.

Clustering is grouping the data into different groups based on similarities of data or related by different variables. Grouping the samples of DNA based on their similarities or related by variables like lifespan, location is a clustering problem. Non-clustering is finding a structure in a chaotic environment. Separating the sounds and voices of different individuals in an audio recording is a non-clustering problem.

# Implementation of algorithms in different languages

These machine learning concepts can be implemented into any language depending on the support and availability of machine learning specific packages. Some of the popular ones include weka in java, scikit, TensorFlow in python etc.. Python, R, Octave/Matlab, Java are popular and preferred languages for machine learning.

# Approach for a machine learning problem

Solving a machine learning problem includes different steps. Some of them are:

1. Identify the problem and its related assumptions.
2. Prepare the data.
3. Implement different algorithms on the data and observe the results.
4. Take steps to improve the results.
5. Move into production.

Each of the steps still has many sub-steps and many ways to achieve desired results depending on the problem.

# Conclusion

There is no thing such as perfect machine learning algorithm for a problem. Each of the categories whether classification or clustering or others, has different algorithms that can be implemented to solve the problem. The challenge is to select the algorithm based on their merits and demerits and to tune the algorithm to its specific needs.

# References & Acknowledgements

* Wikipedia [page](https://en.wikipedia.org/wiki/Machine_learning) for machine learning
* Machine Learning [course](https://www.coursera.org/learn/machine-learning/) by Andrew N.G on coursera
* [machinelearningmastery.com](http://machinelearningmastery.com/start-here/)