Machine Learning Introduction and Learning Plan



In this tutorial we will see the brief introduction of Machine Learning and preferred learning plan for beginners.

## What Is Machine Learning

We are using machine learning every day without even noticing it. Probably you are reading this tutorial because search engine algorithm suggested it to you. Few examples of everyday machine learning uses are,

* Every time we do Google search, result are not sorted alphabetically or date wise, a machine learning algorithm sorts it based on relevance
* Gmail uses machine learning algorithm to block spam email. Imagine using mail service like Gmail without spam filter?
* Google photos use machine learning algorithm for sorting you and your loved one's photos.

In so many ways, machine learning algorithms are either making decisions for us or helping us to make more informed decisions.

Unlike traditional fixed rule based programming machine learning is a science of getting computers to learn without being explicitly programmed.

## Why Should We Learn It?

You must have heard the buzzword **"Data Is New Oil"**. So if data is going to be a new oil machine learning is the best tool to make use of exabytes of data generated every day

Any individual, organization or a country must use the machine learning tools to analyze the data so that they can make well-informed decision. If they failed to do it they will be left behind.

## Learning Plan

In my opinion the best way to learn anything is to understand the concepts, practice it and then use it to solve some real world problems.

I will follow the below four-step plan for every machine learning algorithm. \* Step1: Understand the concepts using easy to understand intuitions \* Step2: Write python code to implement the algorithm from scratch. \* Step3: Test the algorithm with some test datasets. \* Step4: Use open source library like scikit-learn to test same algorithm

Since scikit-learn library has all the machine learning algorithms implemented, it is very tempting to just start using it. But if you are beginner, I would recommond you to follow above plan so that you will learn inner workings of every algorithm and it will make huge difference in long run.

In all my tutorials I am going to use some mathematical models and python code to test the algorithms. You don't need to be professional programmer. If you have any kind of programming experience then you should be able to understand the examples.

During this course you will learn many new concept and use many new libraries likes panda, numpy etc. I will try to explain everything in detail in my tutorials.

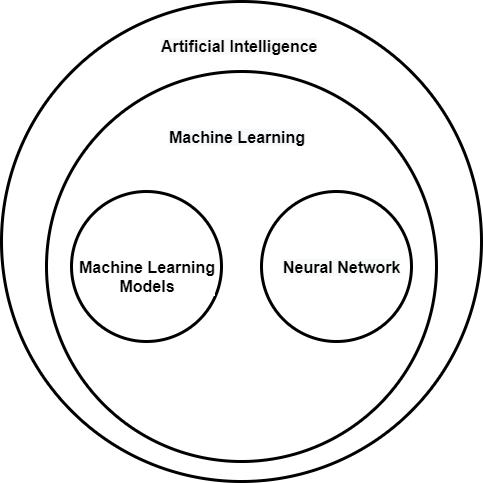
Without further ado lets start with it...



## Definition and Types of Machine Learning

"A program is said to learn from experience **'E'** with respect to some task **'T'** and some performance measure **'P'**, if its performance(P) at task(T) improves with experience(E)" By Tom Mitchell

Machine learning is subset of Artificial intelligence. To solve machine learning problems we either use mathematical models, or we can use neural network. Machine learning problems are of two type, \* Supervise learning \* Unsupervised learning



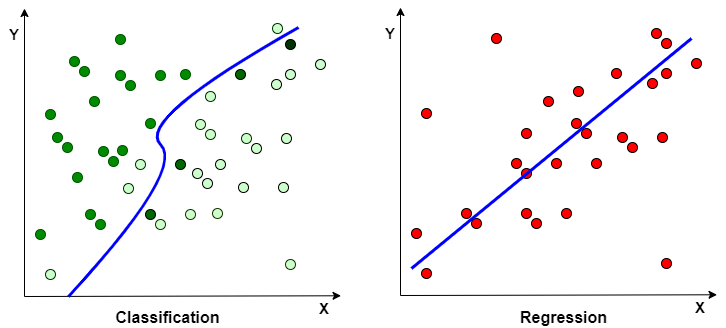
## Supervised Learning

In supervised learning our dataset contains the input values and expected output values. For example, consider below dataset of house size and price

|  |  |
| --- | --- |
| size(sqft) | price(K) |
| 1200 | 3000 K |
| 1300 | 4000 K |
| 1400 | 5000 K |

So in supervise learning we have to understand the relationship between input and output values so that we can predict the values based on input data. Regression and classification are two types of supervised learning problems \* **Regression problem**: If the prediction is some kind of continuous value like price of the house then it is called as regression problem

* **Classification problem**: If the prediction is some kind of classification like house will sale or not then it is called as classification problem



## Unsupervised Learning

In case of unsupervised learning our dataset is without any set of output values or labels. Unsupervised learning is a learning setting where we give the algorithm a ton of unlabeled data and ask it to find some kind of structure in it.

Since our data is unlabeled unsupervised learning allows us to approach a problem with little or no idea what our result should look like

Clustering and association are two type of unsupervised learning problems

* **Clustering**: Clustering problem is used to fin inherent grouping in data. e.g. In case of e-commerce finding a group of customers based on purchasing behavior
* **Association**: Association problem is used to find rules that describe the large portions of data. e.g. In e-commerce finding a rule, why customers that buy X items also tends to buy Y items



## 

## Machine Learning Models

In this series we are going to learn below machine learning models

|  |  |
| --- | --- |
| Linear Regression | Univariate linear regression |
| Multivariate linear regression | Logistic regression |
| Single class (binary) classification | Multi class classification |
| Decision tree | Random Forest |
| Support vector machine | K fold cross validation |
| K means clustering | Naive Bayes |
| Hyper parameter tuning |  |