Introduction to Machine Learning

Every data has patterns, machines can mine these patterns from the data and can become intelligent. Artificial Intelligence is all about making machine intelligent so that it can mimic human intelligence.

Machine learning is the science of programming computers so they can learn from data. It is the subset of artificial intelligence. Humans learn from experience, this is the basic key idea of machine learning. Machine Learning algorithms learns patterns from the past data and gains experience to predict the new data.

But how they learn and identifies the pattern? Using mathematics. Machine learning algorithms uses mathematical concepts like Probability, Statistics, Linear Algebra and Calculus to learn from data.

But how? Every data will have a correlation with same set of data, for example if a set of input data is producing output data as 1 means other set of data which produces same output 1 will have some correlation with that input data. These correlation are identified using mathematical formula.

How Machine learning programs differs from the traditional programming methods?

In traditional programming we will have input data, we need to program the logic to get the specified output data. But in machine learning the program learns its own logic from the given input and output data.

Data from which the model learns is known as training data. Each input data is referred to as feature and the output data is referred to as label. Once the training is over we use a set of data to test the model accuracy, these data are referred to as test data.

Types of Machine Learning

* Supervised Learning: Machine learning models which are trained according to the output data are known as supervised learning model. These models uses both input and output data as their training data. Types of Supervised Learning,

Classification

Regression

* Unsupervised Learning: Machine Learning models which are trained on input data to group output data is known as unsupervised learning model. These models uses input data as its training data and groups the same set of output data. Types of Unsupervised Learning,

Clustering

Association

* Reinforcement Learning: Machine Learning models which interact with environment and discovers which action yields highest reward by trail and error are known as reinforcement learning models. These type of models are used in robotics, gaming and navigation.

Data Preprocessing

Data preprocessing is the initial and the most important step in developing machine learning algorithms. Dataset may contain unprocessed raw data, these data cannot be directly fed into the algorithm.

How it’s done? Dataset contains both numerical and text values, computer programs never understands the text values hence they must be converted to numerical values. Dataset will also have some missing values so they need to be updated with average value. Data range may also vary, for example one input data will be 0.1 and another will be 10000 so these values must be scaled.

All the above mentioned process are done by using Python packages like Numpy, Pandas and Scikit-learn preprocessing libraries.

Data visualization

It is not important to include all the input feature, some feature may be extremely related to the output and some may not. We can even drop the feature which has no impact on the output data.

But how to find the relationship between the data? Data Visualization is the process of visualizing the data which helps to find the relationship between the data. Python libraries like Matplotlib, Seaborn are used to do data visualization.

About this repository

This repository consists of demonstration projects of all machine learning algorithms with brief explanation.