Module

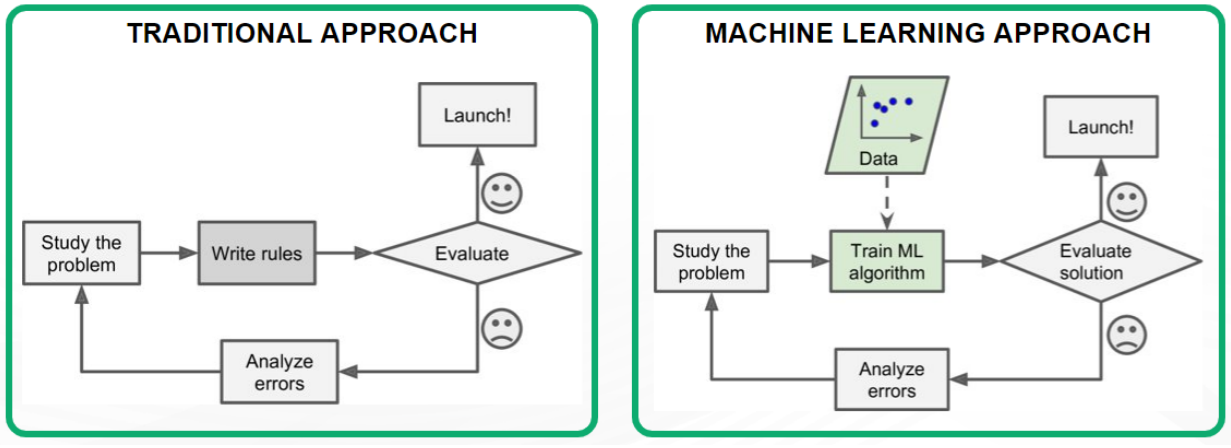
AI and Machine Learning Introduction

The process of machine learning

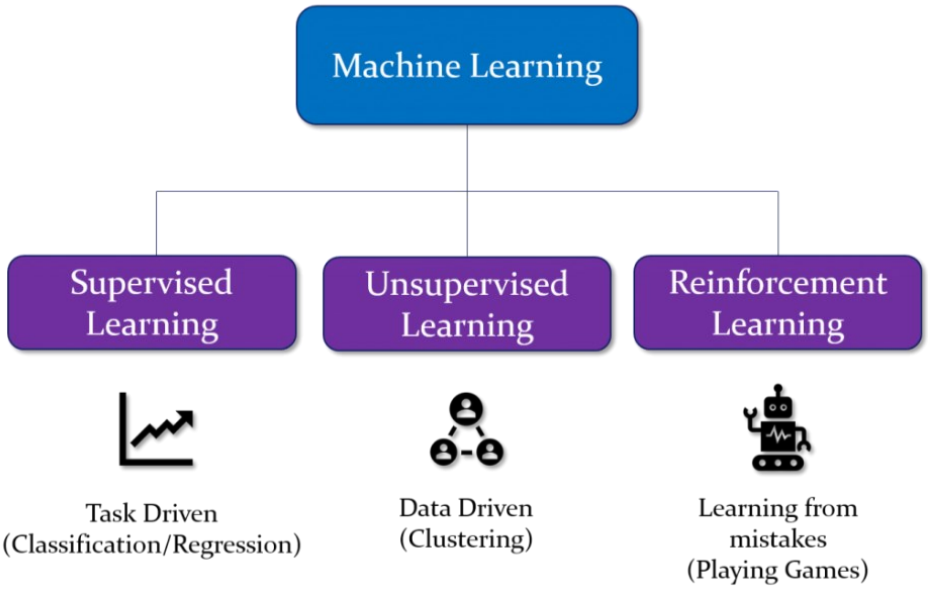
What is Machine Learning?

Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.

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 M. Mitchell

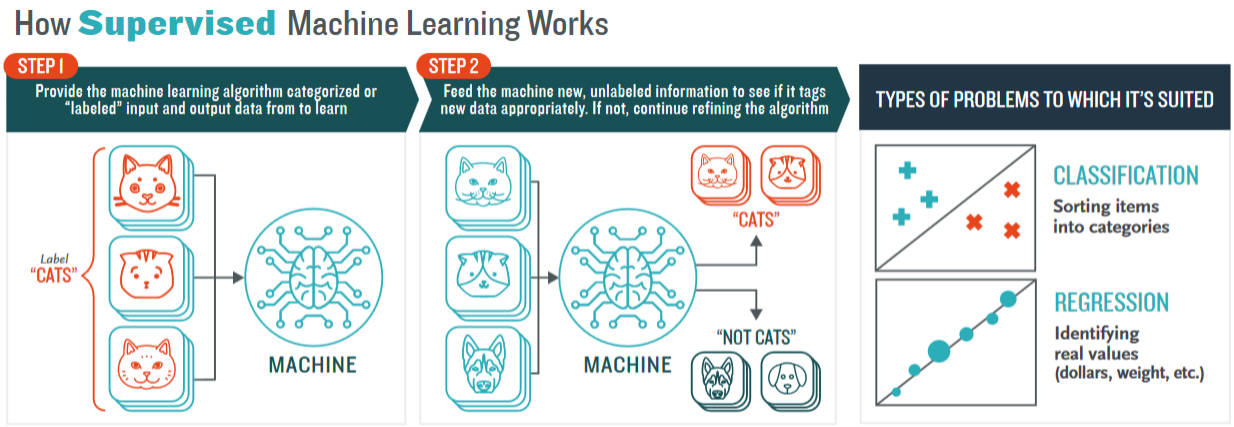


Supervised and unsupervised learning



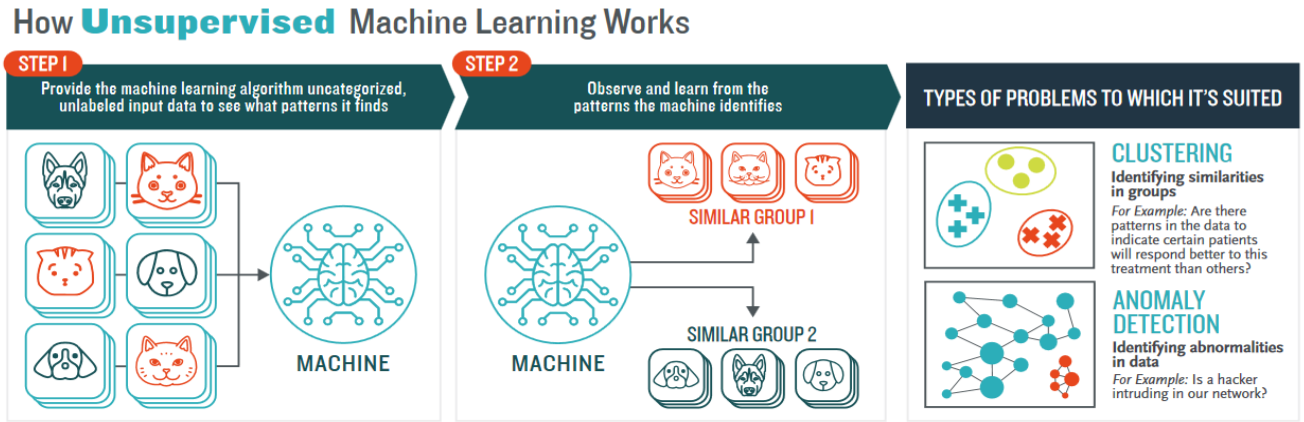
1.) Supervised Learning

Learning with labeled training data. It is called supervised learning because the process of learning from the training data by a machine can be related to a teacher supervising the learning process of a student who is new to the subject. Hence teacher is the training data.



2.) Unsupervised Learning

Unsupervised learning is a machine learning concept where unlabeled and unclassified information is analyzed to discover hidden knowledge.



Regression and classification problems

Whereas when we are trying to predict a real values variable, the problem falls under the category of regression problem.

When we are trying to predict a categorical or nominal variable, the problem is known as classification problem.

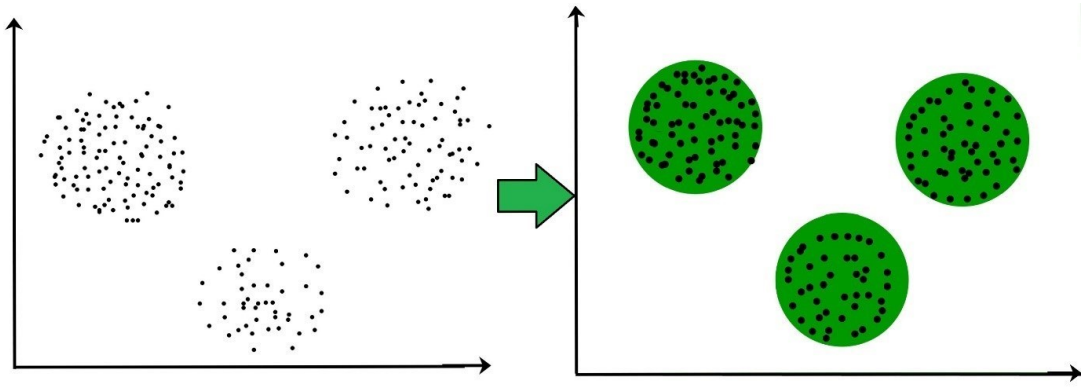
Examples -

- Prediction of results of a game based on past analysis of results (Classification)
- Predicting whether a tumor is malignant or benign (Classification)
- Price prediction in domains such as real estate, stocks etc. (Regression)

Clustering and Anomaly Detection

Clustering refers to a broad set of techniques for finding subgroups, or clusters, in a data set on the basis of the characteristics of the objects within that data set in such a manner that the objects within the group are similar but are different from the objects from the other groups.

Application – Recommender Systems, Customer Segmentation.



Anomaly Detection is the technique of identifying rare events or observations which can raise suspicions by being statistically different from the rest of the observations.

Based on the above assumptions, the data is then clustered using a similarity measure and the data points which are far off from the cluster are considered to be anomalies.

Application – Credit Card Fraud Detection

Recommendation System

Recommendation systems are an important class of machine learning algorithms that offer "relevant" suggestions to users.

Example:

- Netflix movie Suggestion.
- Product Suggestion on Amazon, Ebay etc.

Dimensionality Reduction

Dimensionality reduction refers to techniques for reducing the number of input variables in training data. The curse of dimensionality refers to various phenomena that arise when analyzing and organizing data in high-dimensional spaces.

Some of the techniques used to reduce dimensionality are

- Principle Component Analysis
- Wavelet Transform
- Backward Feature Selection
- Forward Feature Selection

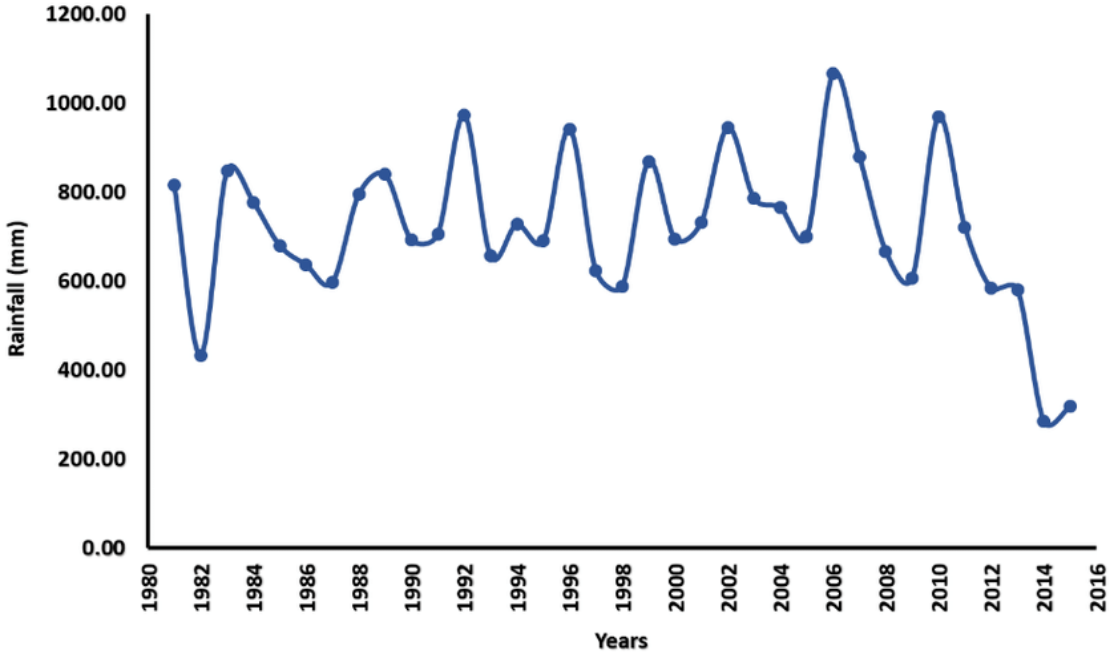
Timeseries Forecasting

What is time series?

Time series is a sequence of observations of categorical or numeric variables indexed by a date, or timestamp. It can be univariant or multivariant.

Examples for Datasets of timeseries

- Stock Price data
- Whether data
- Cryptocurrency price data
- Equipment’s Vibration data



The idea of time series forecasting is to Analyze historical data from the time perspective, identify the patterns, and yield short or long-term predictions by considering the captured patterns and identifying how target variables will change in the future.

What Makes a Machine Learning Expert?

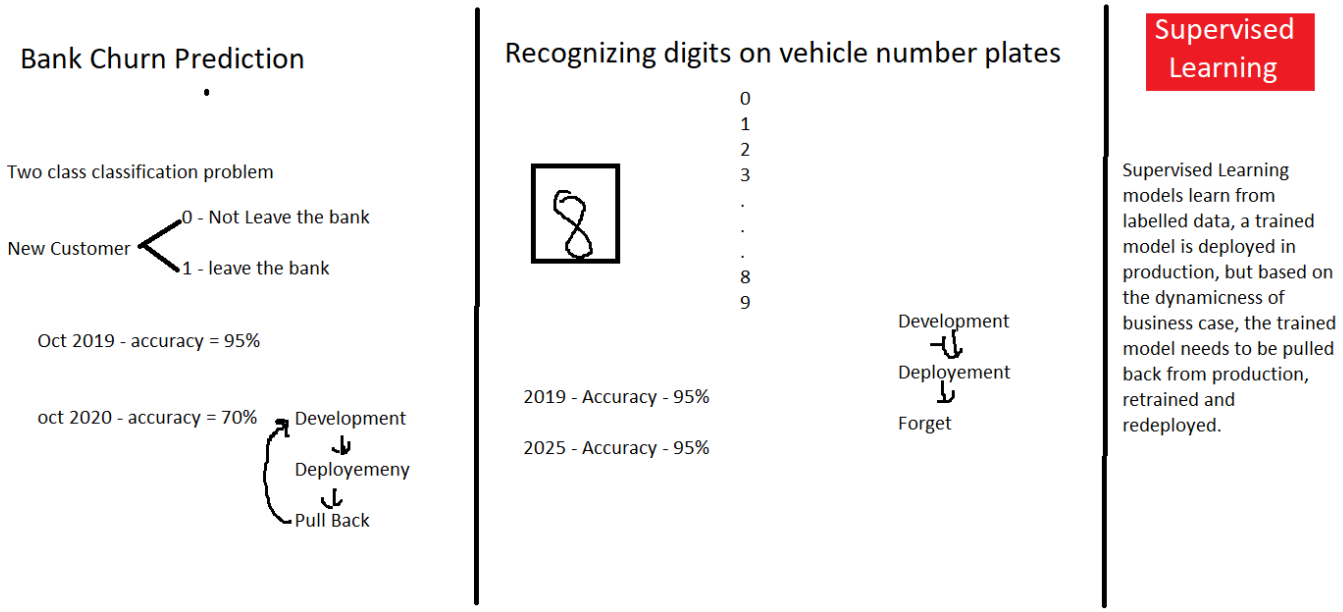
The technical skills needed for a machine learning engineer. Here is a list of technical skills a machine learning engineer is expected to possess.

- Applied Mathematics
- Neural Network Architectures
- Physics
- Data Modelling and Evaluation
- Advances Signal Processing Techniques
- Natural Language Processing
- Audio and video Processing
- Reinforcement Learning

What to Learn to Become a Machine Learning Developer?

- Languages-R/Python for Machine Learning
- Statistics for Data Science
- Math behind Machine Learning Algorithms
- Packages for Machine Learning – Tensor Flow, keras etc.
- Optimization, boosting Techniques
- Big data Technologies-Hadoop, Apache Spark,

Understanding In House Development V/S Outsourcing Strategy



Majority of the data science solutions need periodic retraining of the model depending on the sensitivity of the business case.

Majority of Computer vision and NLP solutions are one time training solutions and do not need periodic retraining of the model.

The ML solutions which often need periodic retraining of the model should be developed in house whereas the ML solutions where the ML models need to be trained only one time should be outsourced. The decision to outsource or develop in house also depends on data sensitivity. If data to be used for ML model is more sensitive, then the project should be developed internally.

Machine Learning Development Platform by Cloud Service Providers

Cloud Computing is essential part of Machine Learning. This is the solution for many smaller and mid-level companies that don't want to build, test, and implement their own machine learning algorithms from scratch.

These companies can focus on their core business and obtain value addition from Machine Learning without needing to become experts. So, they get increasing profits while decreasing their risk of investment which means it's a win-win situation for all.

Below Are some of the cloud computing service providers for Machine Learning-

1.Amazon Web Services

Amazon Web Services is a cloud computing platform that is a subsidiary of Amazon. It was launched in 2006 is currently one of the most popular cloud computing platforms for machine learning. AWS provides various products for Machine Learning like

- Amazon Sage Maker – This is used to create and train machine learning models
- Amazon Augmented AI – This is used to implement a human review of the machine learning models
- Amazon Forecast – This uses machine learning to increase the forecast accuracy
- Amazon Translate – This uses machine learning and natural language processing for language translation
- Amazon Personalize – This creates personal recommendations in machine learning systems
- AWS Deep Learning AMI's – This is used for Deep Learning solutions
- Amazon Polly – This is used to convert text into life-like speech

2. Microsoft Azure

Microsoft Azure is a cloud computing platform created by Microsoft. It was initially released in 2010 and is a popular cloud computing platform for machine learning and data analytics. Some of the Microsoft Azure products for machine learning are:

- Microsoft Azure Cognitive Service – This provides smart cognitive services for applications.

- Microsoft Azure Databricks – This provides Apache Spark-based analytics
- Microsoft Azure Bot Service – This provides smart and intelligent bot services that can be scaled
- Microsoft Azure Cognitive Search – This is a Machine Learning based service for mobile and web applications
- Microsoft Azure Machine Learning – This is used to create and deploy machine learning models on the cloud

3. Google Cloud

The Google Cloud Platform is a cloud computing platform that is provided by Google. It was launched in 2008 and it provides the same infrastructure for companies that Google also uses in its internal products. Google Cloud provides various products for machine learning such as:

- Google Cloud Auto ML – This is used for training an Auto ML machine learning model and its development
- Google Cloud AI Platform – This is used for creating, training, and managing ML models
- Google Cloud Speech-to-Text – This is a speech recognition system for transmitting from speech to text and it supports 120 languages.
- Google Cloud Vision AI – This is used to create machine learning models for cloud vision that detect text, etc.
- Google Cloud Text-to-Speech – This is a speech creation system for transmitting from text to speech
- Google Cloud Natural Language – This is for natural language processing for analysing and classifying text

4. IBM Cloud

The IBM Cloud Platform is a cloud computing platform offered by IBM. It provides various cloud delivery models that are public, private, and hybrid models. IBM Cloud provides various products for machine learning such as:

- IBM Watson Studio – This is used to build machine learning and artificial intelligence models as well as preparing and analysing data
- IBM Watson Speech-to-Text – This is a speech recognition system for converting speech and audio into written text
- IBM Watson Text-to-Speech – This is a speech creation system for converting text into natural-sounding audio
- IBM Watson Natural Language Understanding – This is for natural language processing for analysing and classifying text
- IBM Watson Visual Recognition – This uses machine learning to search visual images and classify them
- IBM Watson Assistant – This is used for creating and managing virtual assistants

Summary

- Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.
- Machine Learning algorithms can be categorized into supervised, unsupervised and reinforcement learning.
- The ML solutions which often need periodic retraining of the model should be developed in house whereas the ML solutions where the ML models need to be trained only one time should be outsourced.
- Popular cloud service providers such as Microsoft Azure, AWS, Google Cloud Platform, and IBM Cloud offer multiple services for ML project development to deployment.