

Introduction to Machine Learning

 ddas.tech/introduction-to-machine-learning/

September 26, 2022

Created By: Debasis Das (Sep 2022)

Table of Contents

- [What is Machine Learning](#)
- [Feature Extraction & Data Representation](#)
- [Preprocessing required for Feature Extraction](#)
- [Why statistical methods are used in machine learning?](#)
- [Applications of Machine Learning](#)
- [Types of Machines Learning](#)
- [Supervised Learning](#)
- [Unsupervised Learning](#)
- [Reference](#)

What is Machine Learning

- Simple classification
- Predictions
- Mathematical modelling
- Pattern recognition etc

All of the above requires learning on the part of the machines

Enabling a machine to learn rather than explicitly programming it

Machine learning is not explicit programming to do something rather the machine should learn from experience or historical data to predict future data.

One kind of machine learning can be to learn from a set of images (eg Cats and dogs) and when a new image is fed into the model it should be able to predict if the image is that of a cat or a dog.

The initial set of data is called training data set using which the machine learning model is built/created.

The task here is to enable a machine to understand the concept of cats and dogs from a set of images.

The images that are fed into the model as a training data set can be called as raw data (Although models can be built on raw data itself, it is desirable to get some more information from the raw data). This step is called **Feature Extraction**

Feature Extraction & Data Representation

- Numerical representation for the raw data.
- The feature representation can be
 - Numerical
 - Categorical (eg: Colors, Shapes etc)
 - Ordinal (eg: Size (being Large, Medium, Small), Temperature – The values does carry some meaning)

Preprocessing required for Feature Extraction

- **Segmentation** (separating the foreground image from the background image)
- **Filtering** (Removing noise samples)
- **Transformation** (eg: Geometrical transformation to compensate for some anomalies, or transformation for color space etc)

Why statistical methods are used in machine learning?

- When the data is noisy(measurement noise) – features are represented using random variables/ vectors
- Inaccuracy of the assumed model
- Inherent ambiguity in the real world problems

Applications of Machine Learning

- Computer Vision
- Speech recognition
- Natural Language Processing
- Medical Informatics
- Robotics
- Computational Biology
- Information Technology
 - Spam Detection
 - Web Image Search
 - Recommendation
 - Information Filtering
 - Community Detection
 - Adaptive advertisement
 - Sentiment Analysis

- Finance
 - Credit Risk Assessment
 - Fraud Detection
 - Stock Market Prediction
 - Algorithmic trading
 - Return forecasting

Types of Machines Learning

- **Supervised Learning** (Training samples have labels/class)
- **Unsupervised Learning** (the training data does not have labels, here the training data can be divided into groups based on the feature set/structure of data)
- **Reinforcement Learning** – Learning to take actions to maximize the notion of reward (Goals, Reward functions are used). Used in robotics.

Supervised Learning

Task of learning a function that maps an input to an output based on example input-output pairs.

- Support-vector machines
- Linear regression
- Logistic regression
- Naive Bayes
- Linear discriminant analysis
- Decision trees
- K-nearest neighbor algorithm
- Neural networks (Multilayer perceptron)
- Similarity learning

Unsupervised Learning

Unsupervised learning, also known as unsupervised machine learning, uses machine learning algorithms to analyze and cluster unlabeled datasets.

These algorithms discover hidden patterns or data groupings without the need for human intervention. Its ability to discover similarities and differences in information make it the ideal solution for

- exploratory data analysis
- cross-selling strategies,
- customer segmentation,
- image recognition.

Types of Unsupervised Learnings are as follows

- Clustering
 - K-Means clustering
 - Hierarchial Clustering
 - DBScan Clustering
- Neural Networks

Unsupervised Learning is used in

- News Section
- Computer Vision (object recognition)
- Medical Imaging
- Anomaly detection
- Customer persona definitions
- Recommendation engines

Reference

- Book – Introduction to Data Mining (PANG-NING TAN, MICHAEL STEINBACH, ANUJ KARPATNE, VIPIN KUMAR)
- Book – Pattern Recognition and Machine Learning (By Christopher M Bishop)