

Introduction to Machine Learning

Cong Cuong PHAM

November 29, 2016

Abstract

This document introduces some fundamental notions of Machine Learning.

1 Introduction

1.1 What is Machine Learning?

- ML is a method of data analysis that automates analytical model building.
- Using algorithms that iteratively learn from data, ML allows computers to find hidden insights without being explicitly programmed where to look.

1.2 What is it used for?

- | | |
|--|----------------------------------|
| • Fraud detection. | • Recommendation Engines. |
| • Web search results. | • Customer Segmentation. |
| • Real-time ads on web pages. | • Text Sentiment Analysis. |
| • Credit scoring and next-best offers. | • Predicting Customer Churn. |
| • Prediction of equipment models. | • Pattern and image recognition. |
| • New pricing models. | • Email Spam filtering. |
| • Network intrusion detection. | • Financial Modeling. |

1.3 ML Process

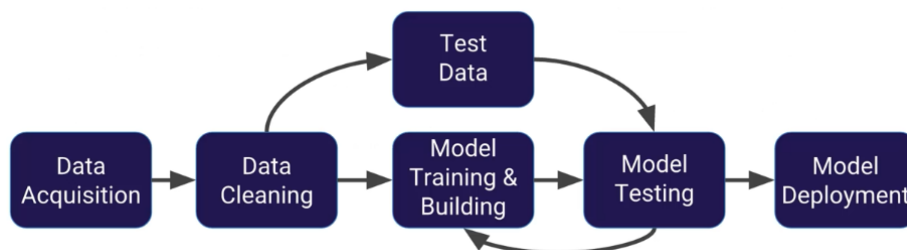


Figure 1: Machine Learning Process

1.4 Types of Machine Learning Algorithms

- Supervised Learning.
- Unsupervised Learning.
- Reinforcement Learning.

1.4.1 Supervised Learning

Is provided with a set of labeled data and is trying to predict a label based of known features.

The learning algorithm receives a set of inputs along with the corresponding correct outputs, and the algorithm learns by comparing its actual output with correct outputs to find errors. It then modifies the model accordingly.

Though methods like classification, regression, prediction and gradient boosting, supervised learning uses patterns to predict the values of the label on additional unlabeled data.

Supervised learning is commonly used in applications where historical data predicts likely future events.

Example: Anticipate when credit card transactions are likely to be fraudulent or which insurance customer is likely to file a claim; Predict the price of house based on different features for house for which we have historical price data.

1.4.2 Unsupervised Learning

Is used against data that has no historical labels. The system is not told the “right answer”. The algorithm must figure out what is being shown.

The goal is to explore the data and find some structure within. It can find the main attributes that separate customer segments from each other.

Popular techniques include self-organizing maps, nearest-neighbor mapping, k-means clustering and singular value decomposition.

Example: Segment text topics, recommend items and identify data outliers.

1.4.3 Reinforcement Learning

Is often used for robotics, gaming and navigation. With reinforcement learning, the algorithm discovers through trial and error which actions yield the greatest rewards.

This type of learning has three primary components: the agent (the learner or decision maker), the environment (everything the agent interacts with) and actions (what the agent can do). The objective is for the agent to choose actions that maximize the expected reward over a given amount of time. The agent will reach the goal much faster by following a good policy.

The goal in reinforcement learning is to learn the best policy.