WHAT IS MACHINE LEARNING?

MACHINE LEARNING

Machine Learning is a process of analyzing data so that future predictions can be made.

Spam or no spam

Facial recognition

Recommender system

FORMAL DEFINITION

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 the experience E.

- Tom Mitchell

WHY IS THERE SO MUCH HYPE ABOUT DATA?

- Storage is cheap
- People are exchanging information freely
- Examples:
  - 300 hours of video are uploaded to YouTube every minute.
  - 350,000 tweets are sent per minute

DATA IS GOLDEN

From the existing data we want to discover patterns and meaning.

We use Machine Learning

TYPES

Supervised learning – we are given the input as well as the target value for each input.

Unsupervised learning – there is no target, we let the model determine the pattern.

Reinforcement learning – is a semi-supervised learning.

REGRESSION VS CLASSIFICATION

In Regression, the goal is to find an equation that better represents the data so that the equation can be used for future predictions.

Regression is about predicting a continuous quantity.

In classification, the goal is to label the data or assign class labels to the data.

Classification is about predicting discrete class labels.

WHAT TOOLS TO USE

Apart from understanding the techniques, one has to know how to clean-up and prepare data to run a specific algorithm.

Numpy – is used for numerical computing, i.e. matrix manipulation.

Pandas – is used to clean-up and prepare data.

Scikit learn – is used to run an algorithm on a specific data.

Additional fields – Statistics, Probability, Linear Algebra.