Machine Learning

17CS73

SYLLABUS

Introduction:

Well posed learning problems, Designing a Learning system, Perspective and Issues in Machine Learning.

Concept Learning:

Concept learning task, Concept learning as search, Find-S algorithm, Version space, Candidate Elimination algorithm, Inductive Bias.

SYLLABUS

Decision Tree Learning:

Decision tree representation, Appropriate problems for decision tree learning, Basic decision tree learning algorithm, hypothesis space search in decision tree learning, Inductive bias in decision tree learning, Issues in decision tree learning.

SYLLABUS

Artificial Neural Networks:

Introduction, Neural Network representation, Appropriate problems, Perceptrons, Backpropagation algorithm.

SYLLABUS

Bayesian Learning:

Introduction, Bayes theorem, Bayes theorem and concept learning, ML and LS error hypothesis, ML for predicting probabilities, MDL principle, Naive Bayes classifier, Bayesian belief networks, EM algorithm

SYLLABUS

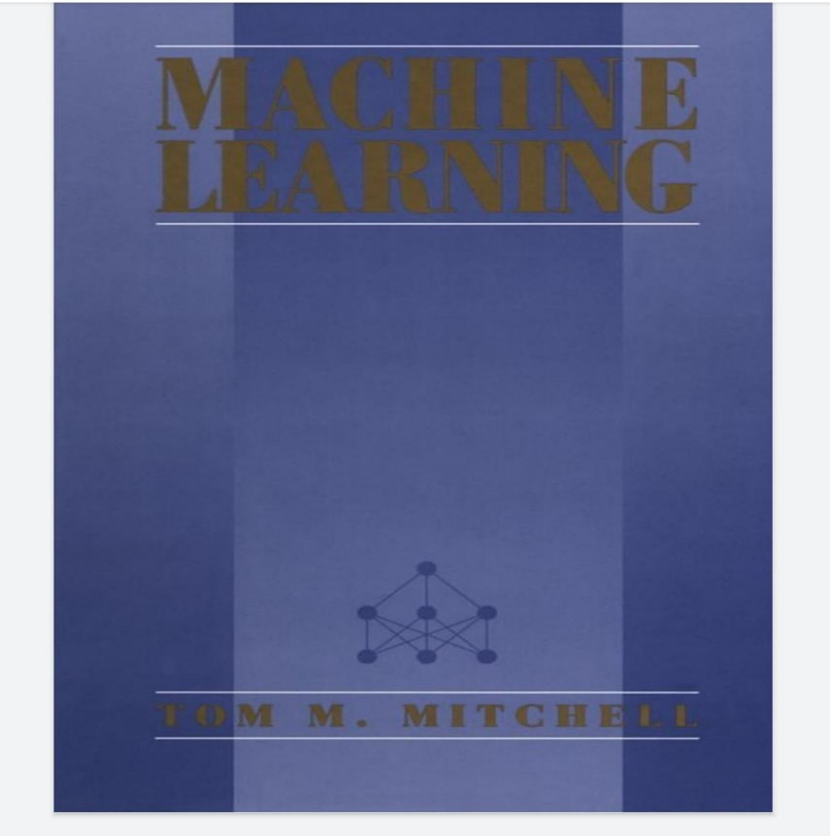
Evaluating Hypothesis:

Motivation, Estimating hypothesis accuracy, Basics of sampling theorem, General approach for deriving confidence intervals, Difference in error of two hypothesis, Comparing learning algorithms.

Instance Based Learning:

Introduction, k-nearest neighbor learning, locally weighted regression, radial basis function, cased-based reasoning,

Reinforcement Learning: Introduction, Learning Task, Q Learning



Machine Learning

Tom M. Mitchell

What Is Machine Learning?

Machine Learning is the science (and art) of programming computers so they can learn from data.

Here is a slightly more general definition:

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

—Arthur Samuel, 1959

What Is Machine Learning?

And a more engineering-oriented one:

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 Mitchell, 1997

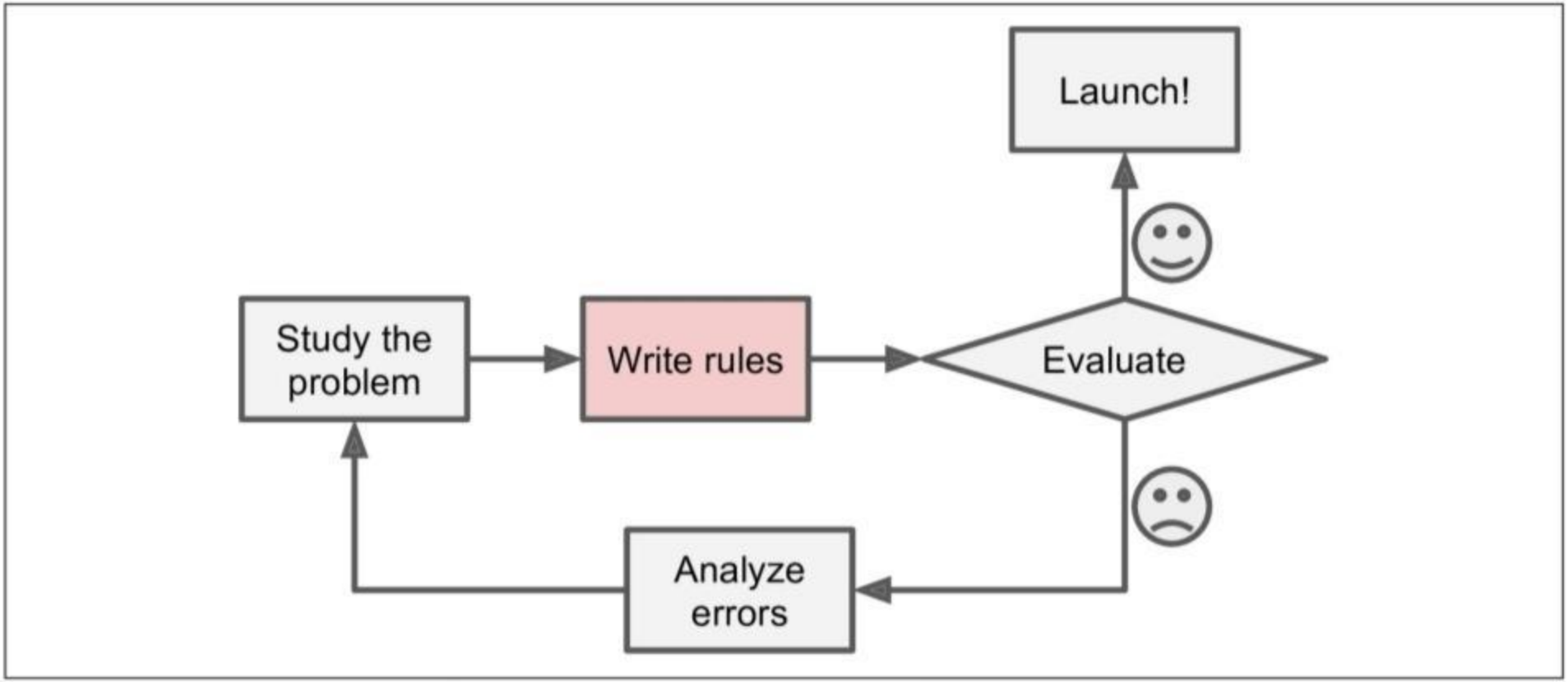


Figure 1-1. The traditional approach

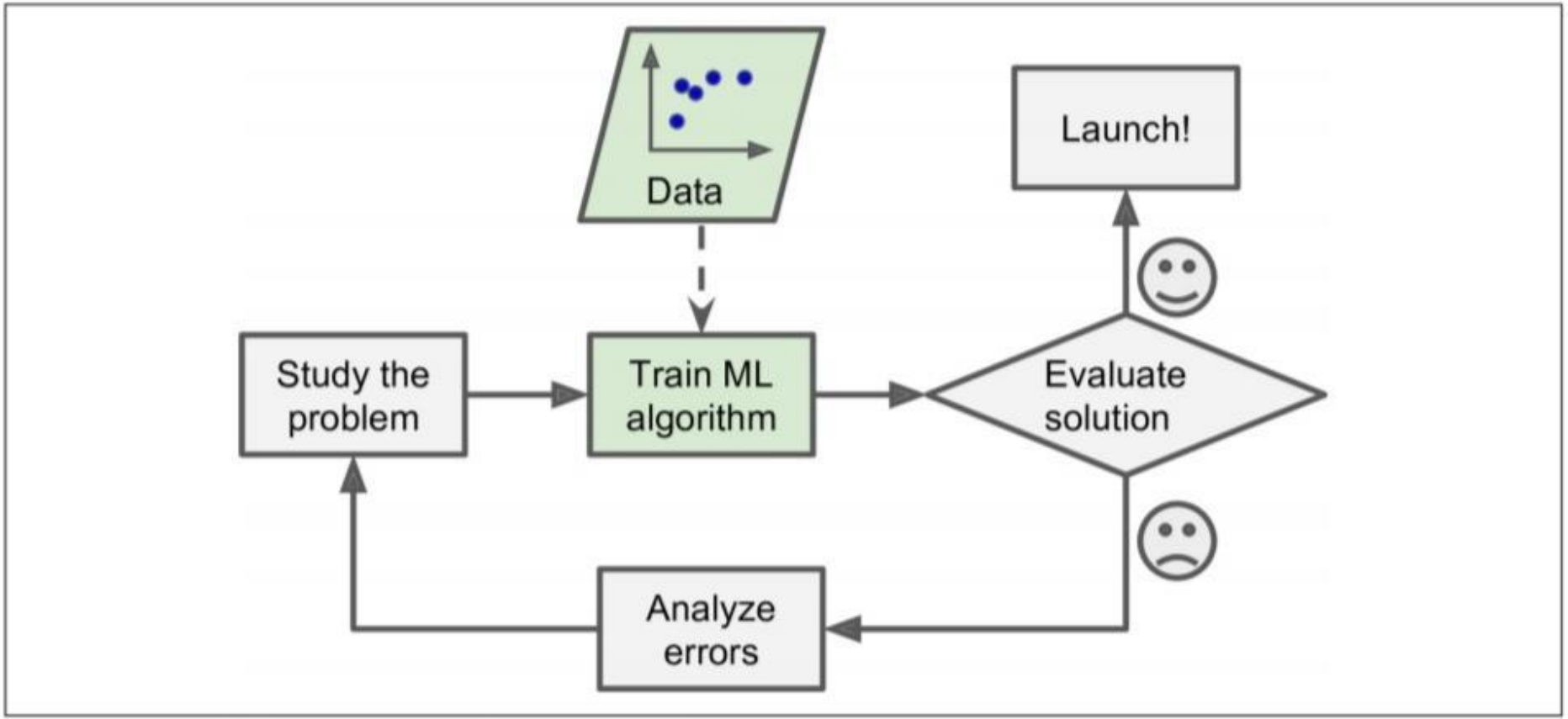


Figure 1-2. Machine Learning approach

To summarize, Machine Learning is great for:

- Problems for which existing solutions require a lot of hand-tuning or long lists of rules: one Machine Learning algorithm can often simplify code and perform better.
- Complex problems for which there is no good solution at all using a traditional approach: the best Machine Learning techniques can find a solution.
- Fluctuating environments: a Machine Learning system can adapt to new data.
- Getting insights about complex problems and large amounts of data.

Types of Machine Learning Systems

There are so many different types of Machine Learning systems that it is useful to classify them in broad categories based on:

• Whether or not they are trained with human supervision (supervised, unsupervised, semi supervised, and Reinforcement Learning)

• Whether or not they can learn incrementally on the fly (online versus batch learning)

• Whether they work by simply comparing new data points to known data points, or instead detect patterns in the training data and build a predictive model, much like scientists do (instance-based versus model-based learning)

Supervised/Unsupervised Learning

There are four major categories:

1. Supervised learning
2. unsupervised learning
3. semi supervised learning
4. Reinforcement Learning.

Supervised/Unsupervised Learning: Supervised learning

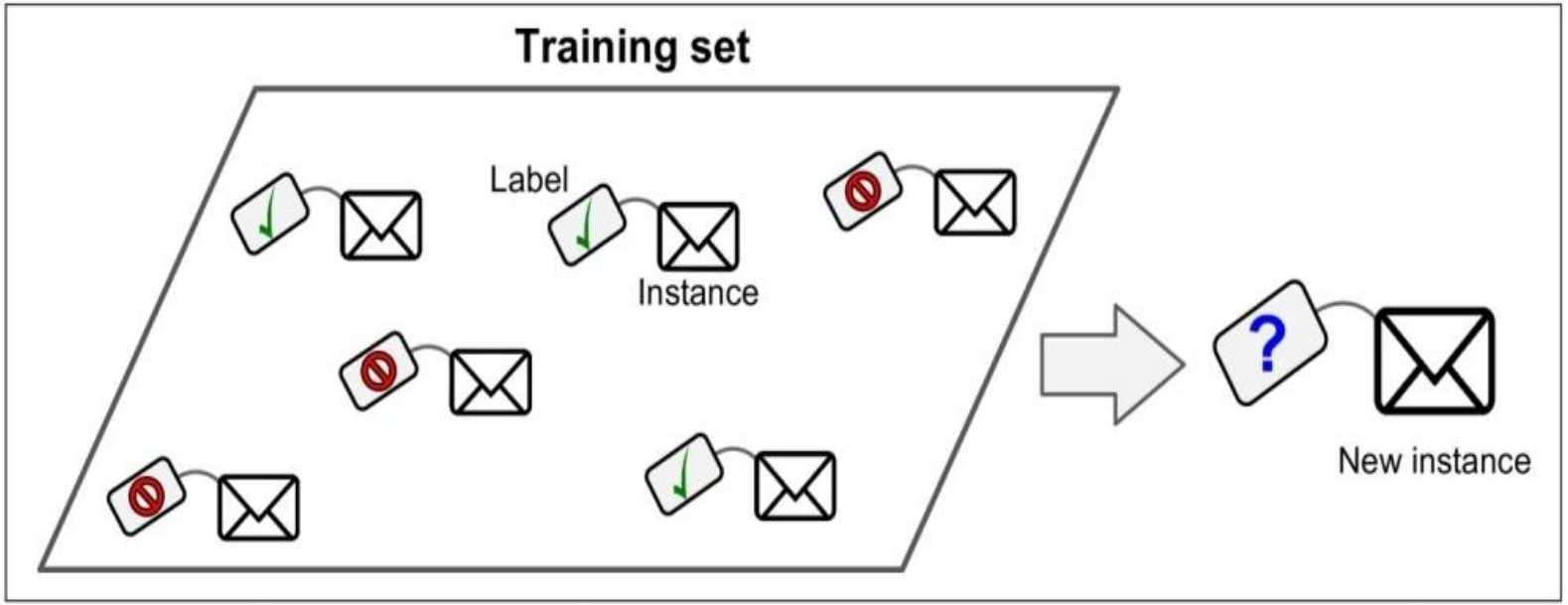


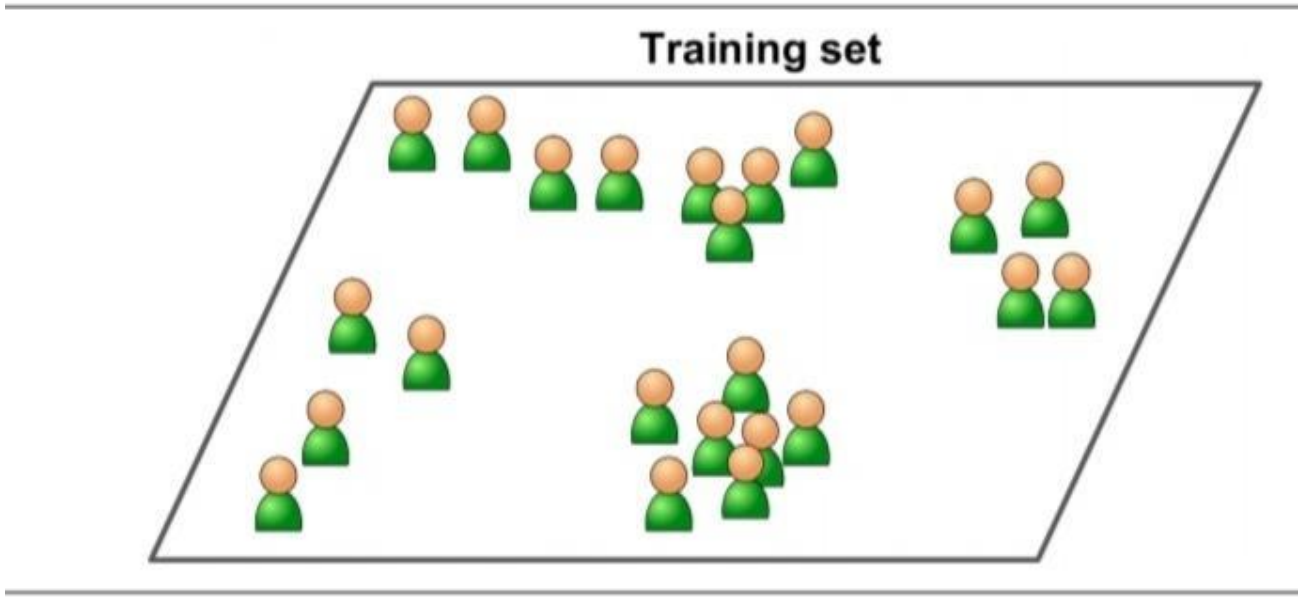
Figure 1-5. A labeled training set for supervised learning (e.g., spam classification)

Supervised/Unsupervised Learning: Supervised learning

Here are some of the most important supervised learning algorithms:

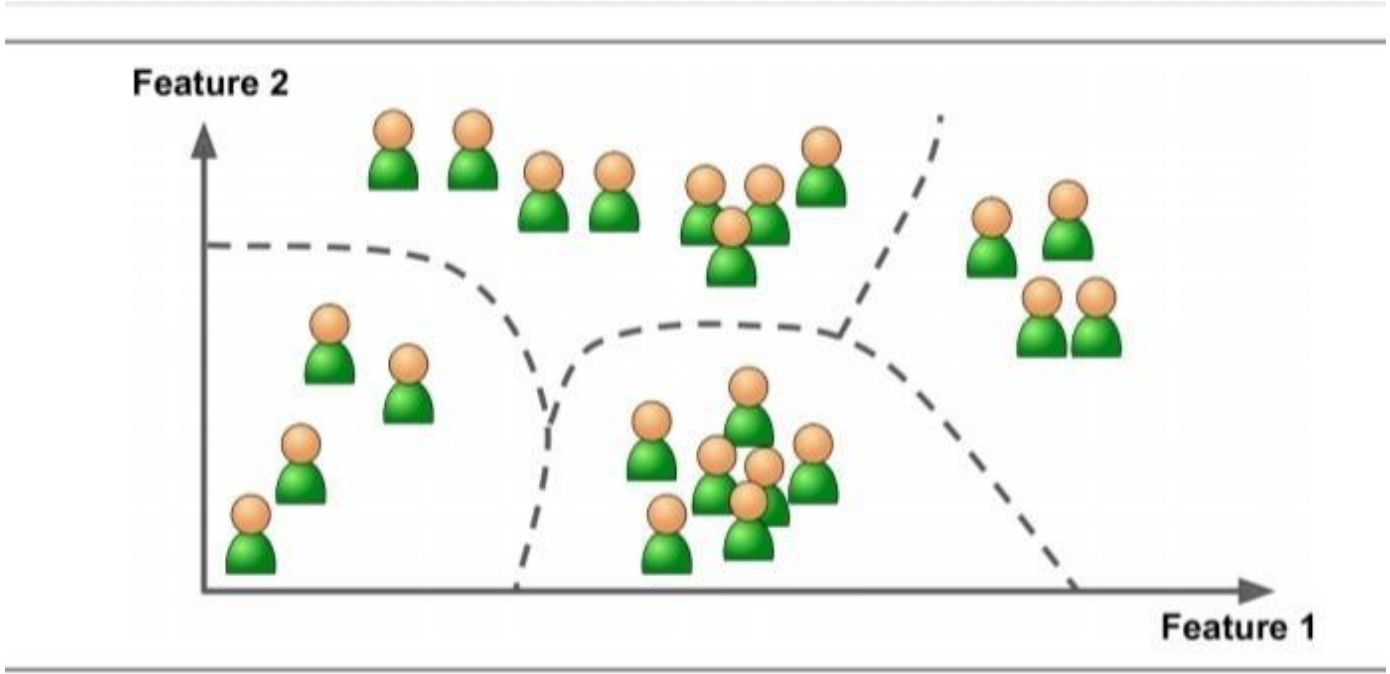
- k-Nearest Neighbors
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVMs)
- Decision Trees and Random Forests
- Neural networks

Supervised/Unsupervised Learning: Unsupervised learning



An unlabeled training set for unsupervised learning

Supervised/Unsupervised Learning: Unsupervised learning



Clustering

Supervised/Unsupervised Learning: Unsupervised learning

Here are some of the most important unsupervised learning algorithms :

- Clustering
  - — k-Means
  - — Hierarchical Cluster Analysis (HCA)
  - — Expectation Maximization
- Visualization and dimensionality reduction
  - — Principal Component Analysis (PCA)
  - — Kernel PCA
  - — Locally-Linear Embedding (LLE)
  - — t-distributed Stochastic Neighbor Embedding (t-SNE)
- Association rule learning
  - — Apriori
  - — Eclat

Supervised/Unsupervised Learning: Semi Supervised learning

Some algorithms can deal with partially labeled training data, usually a lot of unlabeled data and a little bit of labeled data. This is called semi supervised learning

Main Challenges of Machine Learning

- Insufficient Quantity of Training Data
- Poor-Quality Data
- Irrelevant Features
- Overfitting the Training Data
- Underfitting the Training Data