

# Machine Learning Overview

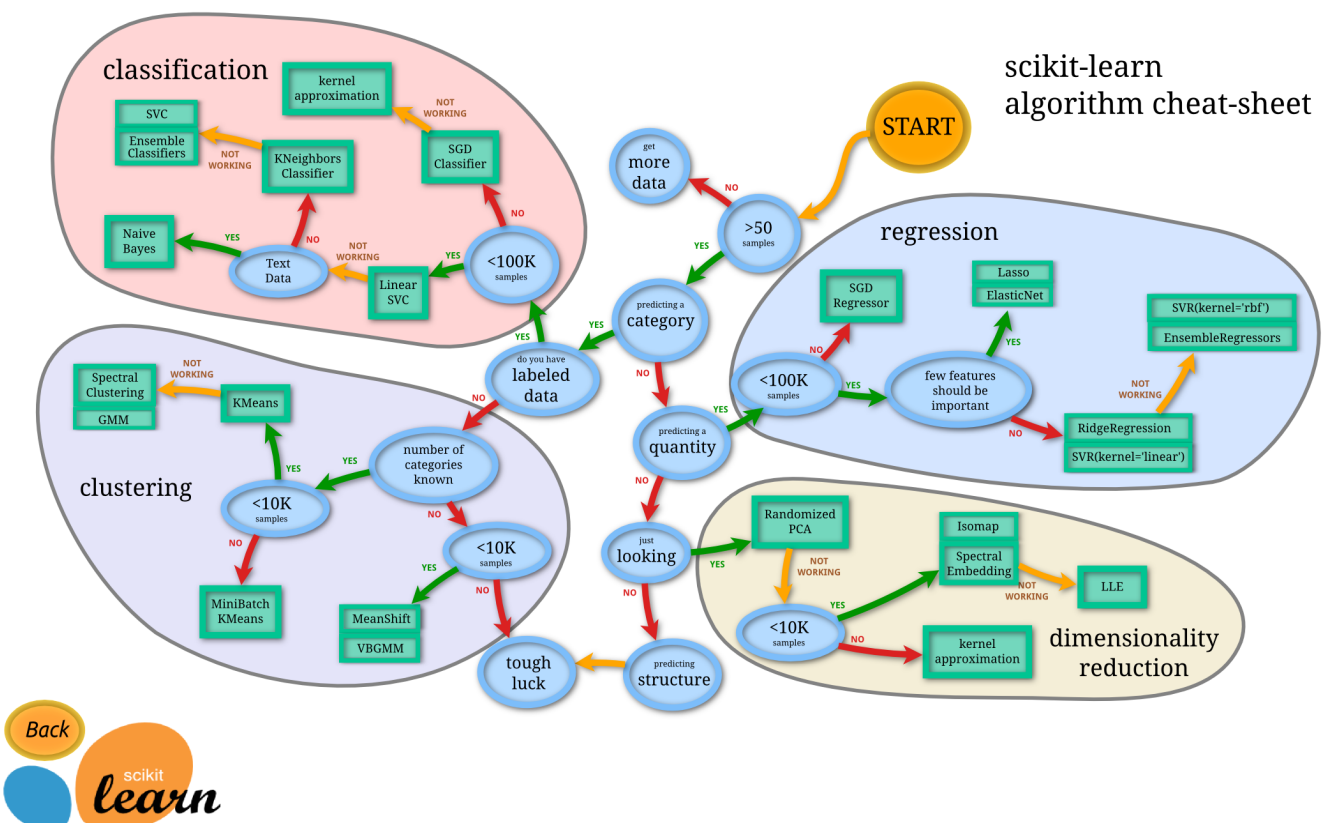
by Prof. Seungchul Lee  
iSystems Design Lab  
<http://isystems.unist.ac.kr/>  
UNIST

## Table of Contents

- I. 1. What is Machine Learning
- II. 2. Supervised vs. Unsupervised Learning
  - I. 2.1. Supervised Learning
  - II. 2.2. Unsupervised Learning
- III. 3. Machine Learning and Deep Learning Tools

# 1. What is Machine Learning

- Draw a meaningful conclusion, given a set of data (observation, measurement)
- In 1959, Arthur Samuel defined machine learning ([https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning)) as a "Field of study that gives computers the ability to learn without being explicitly programmed"
  - Often hand programming not possible
  - Solution? Get the computer to program itself, by showing it examples of the behavior we want! This is the *learning* approach of AI
  - Really, we write the structure of the program and the computer tunes many internal parameters
- Many related terms:
  - Pattern recognition
  - Neural networks → Deep learning
  - Data mining
  - Adaptive control
  - Statistical modeling
  - Data analytics / data science
  - Artificial intelligence
  - Machine learning

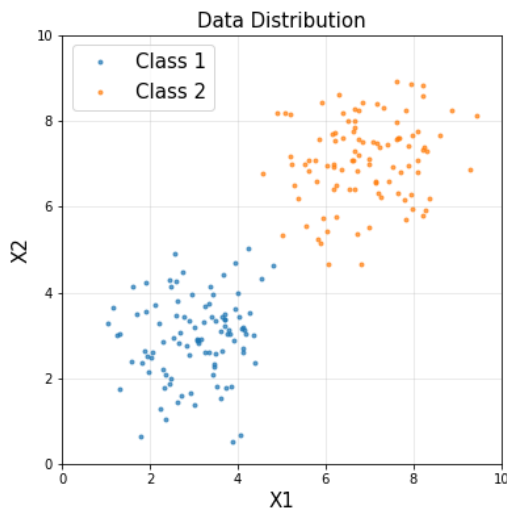


## 2. Supervised vs. Unsupervised Learning

- Supervised: building a model from labeled data
- Unsupervised: clustering from unlabeled data

## 2.1. Supervised Learning

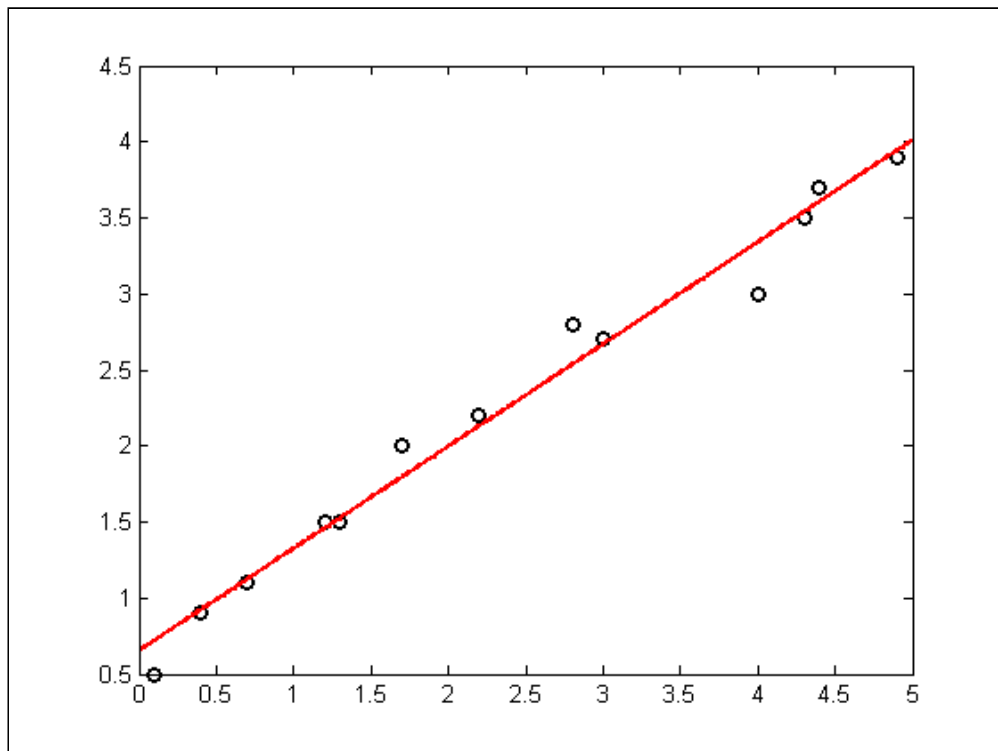
- Regression
  - linear, nonlinear (kernel), lasso ( $L_1$  norm regularization), ridge ( $L_2$  norm regularization)
- Classification
  - perceptron, logistic regression, Support Vector Machine, Bayesian classifier



$$\begin{aligned} \{x^{(1)}, x^{(2)}, \dots, x^{(m)}\} \\ \{y^{(1)}, y^{(2)}, \dots, y^{(m)}\} \end{aligned} \Rightarrow \text{Classification}$$

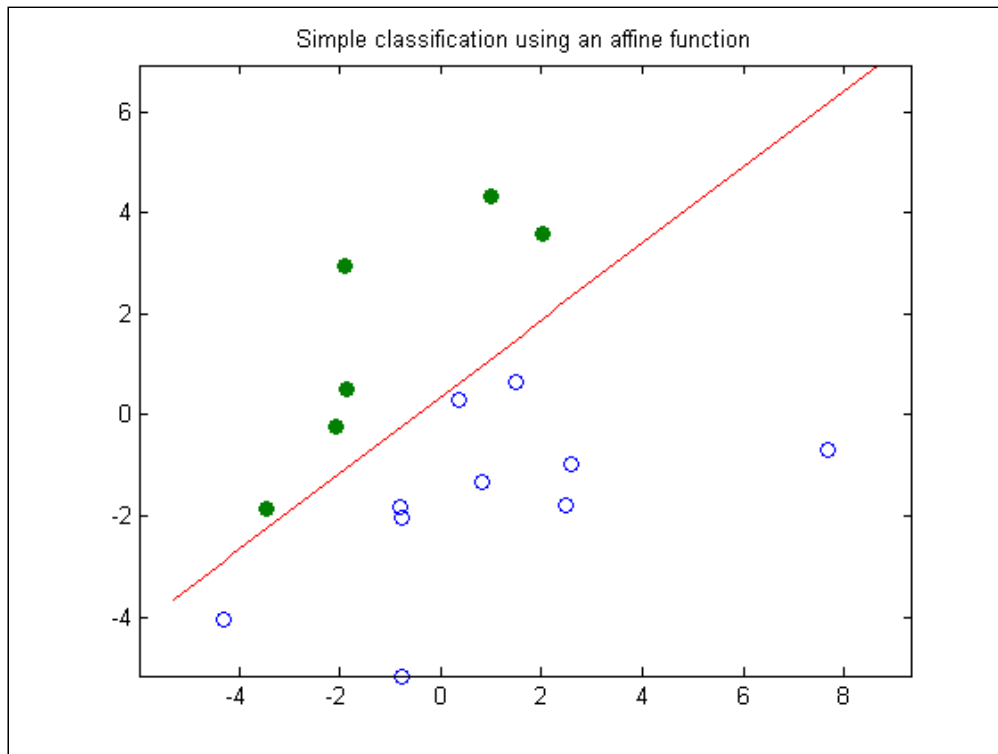
### Data Fitting or Approximation (Regression)

- a statistical process for estimating the relationships among variables (source: [wikipedia](https://en.wikipedia.org/wiki/Regression_analysis) ([https://en.wikipedia.org/wiki/Regression\\_analysis](https://en.wikipedia.org/wiki/Regression_analysis)))



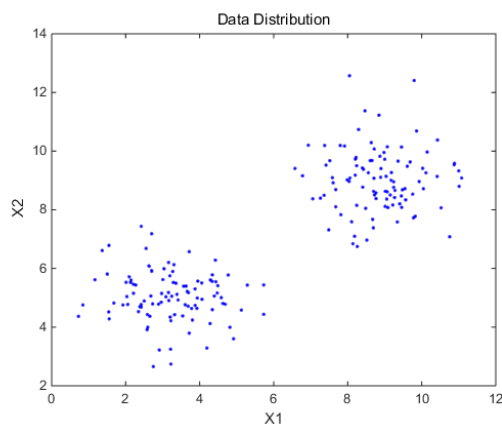
## Classification

- the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known (source: [wikipedia \(https://en.wikipedia.org/wiki/Statistical\\_classification\)](https://en.wikipedia.org/wiki/Statistical_classification))



## 2.2. Unsupervised Learning

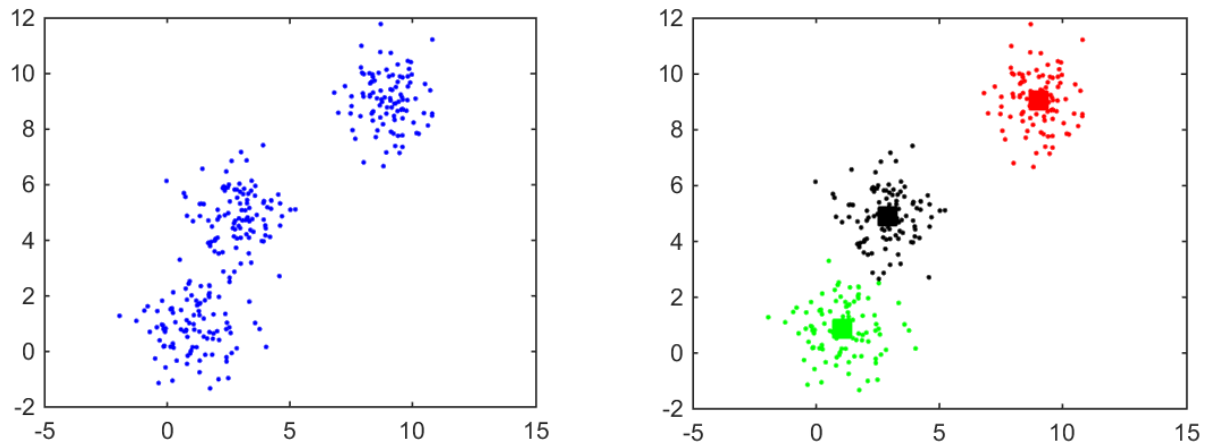
- Clustering
  - k-means, Gaussian Mixture Model
  - graph partitioning (spectral clustering)
- Dimension Reduction
  - PCA



$$\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\} \Rightarrow \text{Clustering}$$

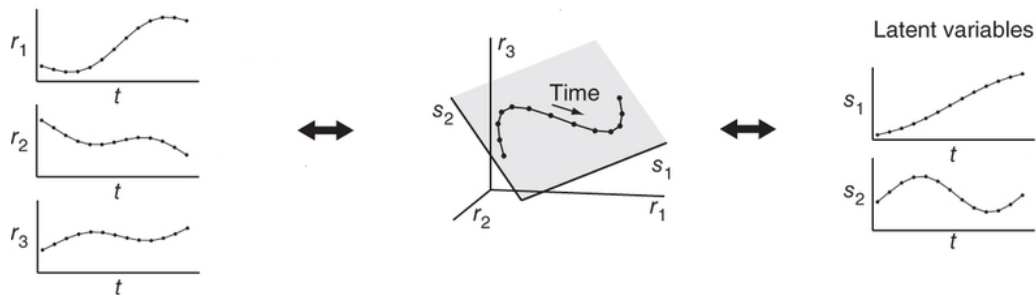
## Clustering

- clustering groups examples based of their mutual similarities



## Dimension Reduction (Multiple Senses + Principal Components)

- the process of reducing the number of random variables under consideration, and can be divided into feature selection and feature extraction.



## 3. Machine Learning and Deep Learning Tools



- 본래는 웹개발, 그래픽인터페이스, 스크립트, 계산과학등 다양한 용도로 사용되는 고급 프로그래밍 언어
- 5만개가 넘는 확장패키지 (eg. Scikit-learn, Numpy 등) 을 제공하는 것이 특징
- 직관적이고 유연한 구조로 접근성이 용이

### Package

- Python은 데이터분석을 위한 다양한 패키지를 제공
- 관련된 클래스 혹은 함수들을 미리 묶어놓은 것
  - 다양한 용도로 파이썬을 활용할 수 있도록 해줌
- Numpy, Scikit-learn (sklearn), TensorFlow 등이 데이터분석을 위한 대표적인 패키지

### Numpy



- 과학계산을 위한 패키지
- 배열간의 수학계산을 수행하는 함수, 선형대수의 계산, 푸리에 변환, 난수 발생기 같은 수치와 관련된 다양한 기능을 제공

### Scipy



- 미분방정식 해석기, 방정식의 근을 구하는 알고리즘, 표준 연속/이산 확률분포와 다양한 통계관련 도구 등을 제공

### Pandas



- 일반 데이터베이스처럼 데이터를 합치고 관계연산을 수행
- 자동적으로 혹은 명시적으로 축의 이름에 따라 데이터를 정렬할 수 있는 자료구조 제공

### Scikit-learn (sklearn)



- Regression, Classification, Clustering 과 같은 머신러닝을 위한 다양한 모형 제공
- 전처리, 모형평가 등을 위한 서브 패키지 제공

### TensorFlow for Deep Learning



- 기계 학습과 딥러닝을 위한 함수, 클래스를 제공

In [1]:

```
%%javascript  
$.getScript('https://kmahelona.github.io/ipython_notebook_goodies/ipython_notebook_toc.  
js')
```

