Machine Learning

Nguyen Quoc Huy

Seoul National University of Science and Technology

nqhuy2205@gmail.com

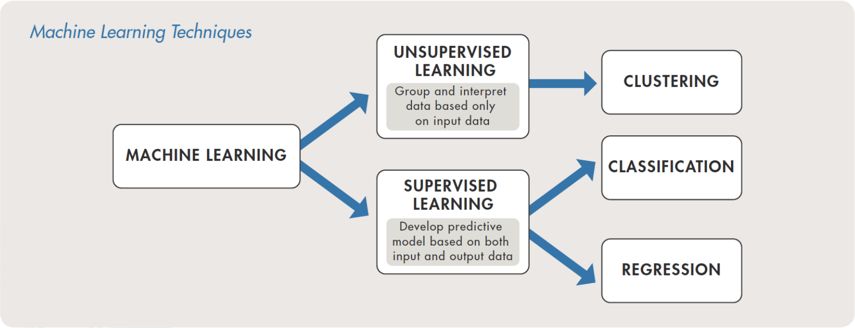
June 28, 2017

Contents

What is machine learning

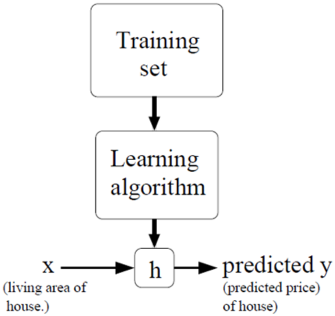
Machine learning is the idea that there are generic algorithms that can tell you something interesting about a set of data without you having to write any custom code specific to the problem. Instead of writing code, you feed data to the generic algorithm and it builds its own logic based on the data.

Machine Learning Techniques

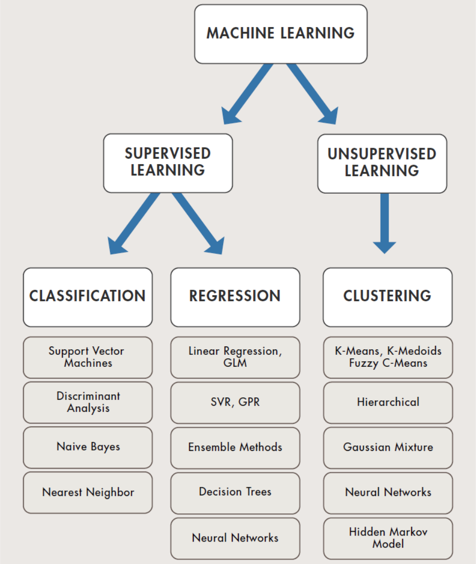


Machine Learning Model

- function h is called hypothesis
- hypothesis h: predict function
- hypothesis h: classifier



Machine Learning Algorithm



Regresstion

- Linear Regression
  - hyphothesis h is a linear function

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \tag{1}$$

  - θ_i's are called **Weights**. Let $x_0 = 1$

$$h(x) = \sum_{i=0}^{n} \theta_i x_i = \theta^T x \tag{2}$$

- Logistic Regression
  - hypothesis h is a sigmoid function

$$h_\theta(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}} \tag{3}$$

- Generalized Linear Models

Linear Regression

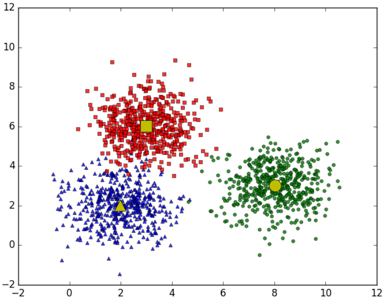
- hyphothesis h

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \tag{4}$$

- Cost function (Lost function)
  - $||h_\theta(x^{(i)}) - y^{(i)}||$
  - $J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x_{(i)}) - y^{(i)})^2$
    so we want $\theta = \arg\min(J(\theta))$
- Gradient Descent

The k-means clustering

We are given a training set $\mathbf{X} = [x^{(1)}, x^{(2)}, ..., x^{(N)}]$, and we want to group data into K **clusters**



The k-means clustering: Describe label matrix

Introduce $y^{(i)} = [y_1^{(i)}, y_2^{(i)}, ..., y_K^{(i)},]$ is the label vector of data point $x^{(i)}$ so that if data point $x^{(i)}$ is assigned to cluster k then $y_k^{(i)} = 1$ and $y_j^{(i)} = 0$ for $j \neq k$

$$y_k^{(i)} \in \{0, 1\}, \sum_{k=1}^{K} y_k^{(i)} = 1 \tag{5}$$

Ex:

$x^{(1)}$ belong to 1^{st} cluster, $y^{(1)} = [1, 0, 0, ...]$
$x^{(2)}$ belong to 2^{nd} cluster, $y^{(2)} = [0, 1, 0, ...]$

The k-means clustering: Loss function

Introduce a set of D_Dimensional vectors m_k, where $k = 1, ..., K$, in which m_k is a prototype associate with the k^{th} cluster. For each point data $x^{(i)}$ assigned to cluster k, has error is:

$$||x^{(i)} - m_k||_2^2 \tag{6}$$

Because $x^{(i)}$ was assigned to cluster k, $y_k^{(i)} = 1$, $y_j^{(i)} = 0$ for $j \neq k$. Thus (6) can be rewrite:

$$y_k^{(i)}||x^{(i)} - m_k||_2^2 = \sum_{j=1}^{K} y_j^{(i)}||x^{(i)} - m_j||_2^2 \tag{7}$$

Thus Error for whole data set:

$$\mathcal{L}(\mathbf{Y}, \mathbf{M}) = \sum_{i=1}^{N}\sum_{j=1}^{K} y_j^{(i)}\|\mathbf{x}^{(i)} - \mathbf{m}_j\|_2^2 \tag{8}$$

Where $\mathbf{Y} = [\mathbf{y^{(1)}}; ...; \mathbf{y^{(N)}}]$, $\mathbf{M} = [\mathbf{m_1}, ..., \mathbf{m_K}]$ is label vector and centers

The k-means clustering: Optimal Problem

Optimal Problem

$$\mathbf{Y}, \mathbf{M} = \arg\min_{\mathbf{Y},\mathbf{M}} \sum_{i=1}^{N} \sum_{j=1}^{K} y_j^{(i)} \|\mathbf{x}^{(i)} - \mathbf{m}_j\|_2^2 \tag{9}$$

subject to $y_i^{(i)} \in \{0, 1\} \;\; \forall i, j; \;\; \sum_{j=1}^{K} = 1 \;\; \forall i$

The k-means clustering: Optimal Problem

K-means Cluster Algorithm:

1. Initialize the centers in $\mathbf{m}_k$ using random sampling
2. Assign data point to close cluster by calculate minimun distance from data point to centers
3. if there are no change in cluster centers, stop algorithm
4. Update new centers by calculate mean of new cluster of step 2
5. return step 2