HỘI NGHỊ KHOA HỌC

Khoa CNTT, Trường ĐH Giao thông Vận tải Tp.HCM

Nguyễn Văn Diêu

HoChiMinh City University of Transport

dieu.nguyen@ut.edu.vn

2019

Least Squares Method for Linear Regression

Legendre (1805)

Gauss (1809)

Machine Learning (1959)

Least Squares Method for Linear Regression

Machine Learning

1. Supervised Learning
   - Classification
   - Regression
     - Linear Regression
       - Simple Linear Regression
       - Multiple Linear Regression
     - Polynomial Regression
     - ...
2. Unsupervised Learning
   - Clustering
   - Association Rule
3. Reinforcement Learning
   - Classification
   - Control

Least Squares Method for Linear Regression

Machine Learning

1. Supervised Learning
   - Classification
   - Regression
     - Linear Regression
       - Simple Linear Regression
       - Multiple Linear Regression
     - Polynomial Regression
     - ...
2. Unsupervised Learning
   - Clustering
   - Association Rule
3. Reinforcement Learning
   - Classification
   - Control

Least Squares Method for Linear Regression

Machine Learning

1. Supervised Learning
   - Classification
   - Regression

     Linear Regression

       Simple Linear Regression

       Multiple Linear Regression

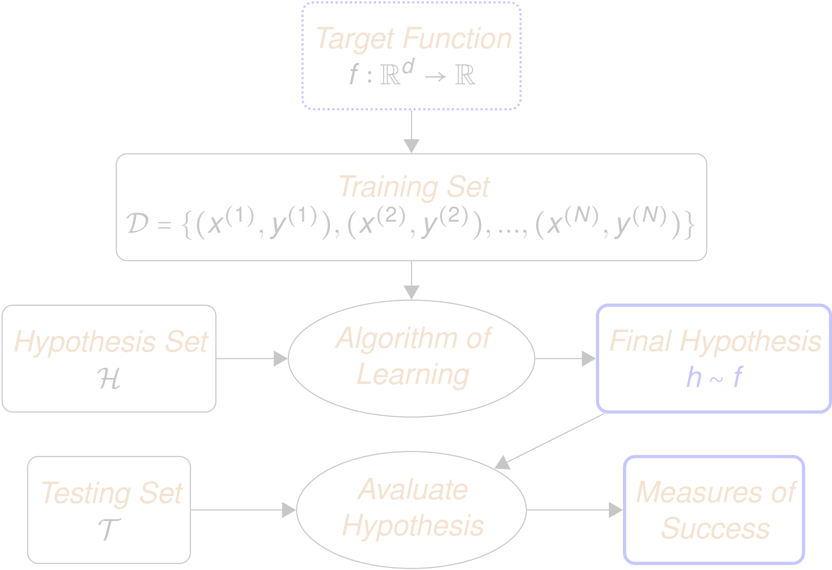
     Polynomial Regression

     ...
2. Unsupervised Learning
   - Clustering
   - Association Rule
3. Reinforcement Learning
   - Classification
   - Control

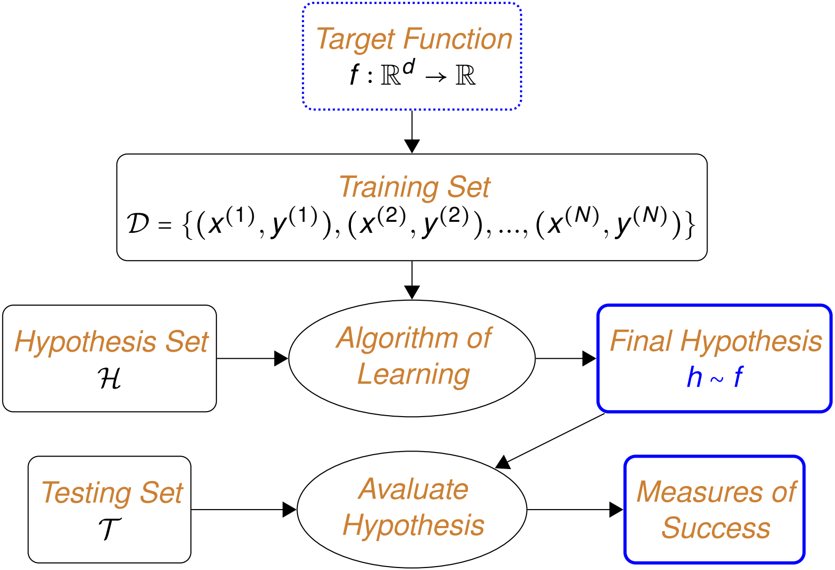
Machine Learning

1. Supervised Learning
   - Classification
   - Regression
     - Linear Regression
       - Simple Linear Regression
       - Multiple Linear Regression
     - Polynomial Regression
     - ...
2. Unsupervised Learning
   - Clustering
   - Association Rule
3. Reinforcement Learning
   - Classification
   - Control

Learning Components



Learning Components



Linear Regression examples

1. Simple Linear Regression.

   E.g. Blood Pressure ~ Age

2. Multiple Linear Regression.

   E.g. Blood Fat ~ (Age, Weight)

Linear Regression examples

1. Simple Linear Regression.

   E.g. Blood Pressure ~ Age

2. Multiple Linear Regression.

   E.g. Blood Fat ~ (Age, Weight)

Linear Regression examples

1. Simple Linear Regression.

   E.g. *Blood Pressure ~ Age*

2. Multiple Linear Regression.

   E.g. *Blood Fat ~ (Age, Weight)*

Linear Regression Algorithm

- Training set:

$$\mathcal{D} = \{(x^{(1)}, y^{(1)}), ..., (x^{(N)}, y^{(N)})\}$$

$$x^{(i)} \in \mathbb{R}^d, \quad y^{(i)} \in \mathbb{R}$$

- Hypothesis:

$$\begin{aligned} h(\theta, x) &= \theta_0 + \theta_1 x_1 + ... + \theta_d x_d \\ &= \theta_0 + \sum_{i=1}^{d} \theta_i x_i \\ &= \Theta^T \mathbf{x} \end{aligned}$$

$$\Theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_d \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_d \end{bmatrix}$$

Linear Regression Algorithm

- Training set:

$$\mathcal{D} = \{(x^{(1)}, y^{(1)}), ..., (x^{(N)}, y^{(N)})\}$$

$$x^{(i)} \in \mathbb{R}^d, \quad y^{(i)} \in \mathbb{R}$$

- Hypothesis:

$$\begin{aligned} h(\theta, x) &= \theta_0 + \theta_1 x_1 + ... + \theta_d x_d \\ &= \theta_0 + \sum_{i=1}^{d} \theta_i x_i \\ &= \Theta^T \mathbf{x} \end{aligned}$$

$$\Theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_d \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_d \end{bmatrix}$$

Linear Regression Algorithm

- Training set:

$$\mathcal{D} = \{(x^{(1)}, y^{(1)}), ..., (x^{(N)}, y^{(N)})\}$$

$$x^{(i)} \in \mathbb{R}^d, \quad y^{(i)} \in \mathbb{R}$$

- Hypothesis:

$$\begin{aligned} h(\theta, x) &= \theta_0 + \theta_1 x_1 + \ldots + \theta_d x_d \\ &= \theta_0 + \sum_{i=1}^{d} \theta_i x_i \\ &= \Theta^T \mathbf{x} \end{aligned}$$

$$\Theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_d \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_d \end{bmatrix}$$

Simple Linear Regression Algorithm

- Training set:

$$\mathcal{D} = \{(x^{(1)}, y^{(1)}), ..., (x^{(N)}, y^{(N)})\}$$

$$x^{(i)} \in \mathbb{R}, \quad y^{(i)} \in \mathbb{R}$$

- Hypothesis:

$$\begin{aligned} h(\theta, x) &= \theta_0 + \theta_1 x \\ &= \Theta^T \mathbf{x} \end{aligned}$$

$$\Theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} 1 \\ x \end{bmatrix}$$

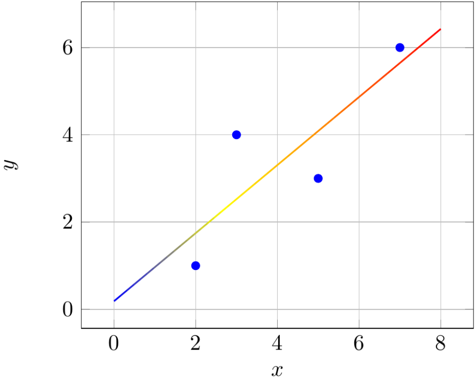
Loss Function

- Training set:

$$\mathcal{D} = \{(2,1),(3,4),(5,3),(7,6)\}$$

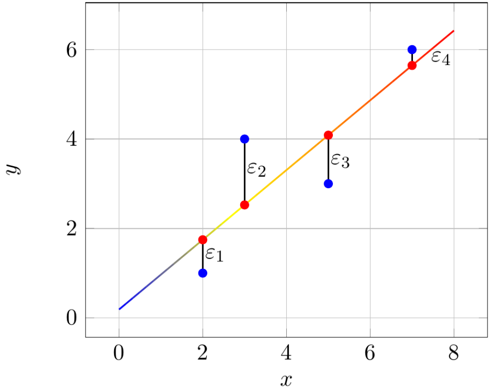
- Hypothesis:

$$h(\theta, x) = \theta_0 + \theta_1 x$$



Loss Function

$$\epsilon_i = (y^{(i)}, h(\theta, x^{(i)}))$$



$$\begin{aligned}\mathcal{L}_i(\theta_0, \theta_1) &= {\epsilon_i}^2 \\ &= (y^{(i)} - h(\theta, x^{(i)}))^2 \\ &= (y^{(i)} - (\theta_0 + \theta_1 x^{(i)}))^2\end{aligned}$$

Loss Function

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_i(\theta_0, \theta_1) \quad \Rightarrow \quad \underset{\theta_0,\theta_1}{argmin} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_i(\theta_0, \theta_1)$$

$$\underset{\theta_0,\theta_1}{argmin} \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - (\theta_0 + \theta_1 x^{(i)}))^2$$

$$\theta_1 = \frac{\frac{1}{N} \sum_{i=1}^{N} (x^{(i)} - \overline{x})(y^{(i)} - \overline{y})}{\frac{1}{N} \sum_{i=1}^{N} (x^{(i)} - \overline{x})^2}$$

$$\theta_0 = \overline{y} - \theta_1 \overline{x}$$

Loss Function

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_i(\theta_0, \theta_1) \quad \Rightarrow \quad \underset{\theta_0,\theta_1}{argmin} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_i(\theta_0, \theta_1)$$

$$\underset{\theta_0,\theta_1}{argmin} \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - (\theta_0 + \theta_1 x^{(i)}))^2$$

$$\theta_1 = \frac{\frac{1}{N} \sum_{i=1}^{N} (x^{(i)} - \overline{x})(y^{(i)} - \overline{y})}{\frac{1}{N} \sum_{i=1}^{N} (x^{(i)} - \overline{x})^2}$$

$$\theta_0 = \overline{y} - \theta_1 \overline{x}$$

Loss Function

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_i(\theta_0, \theta_1) \quad \Rightarrow \quad \underset{\theta_0, \theta_1}{argmin} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_i(\theta_0, \theta_1)$$

$$\underset{\theta_0, \theta_1}{argmin} \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - (\theta_0 + \theta_1 x^{(i)}))^2$$

$$\theta_1 = \frac{\frac{1}{N} \sum_{i=1}^{N} (x^{(i)} - \overline{x})(y^{(i)} - \overline{y})}{\frac{1}{N} \sum_{i=1}^{N} (x^{(i)} - \overline{x})^2}$$

$$\theta_0 = \overline{y} - \theta_1 \overline{x}$$

Multiple Linear Regression

$$\mathcal{D} = \{(x^{(1)}, y^{(1)}), \ldots, (x^{(N)}, y^{(N)})\}$$

$$x^{(i)} \in \mathbb{R}^d, \quad y^{(i)} \in \mathbb{R}$$

$$\Theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_d \end{bmatrix} \quad x^{(i)} = \begin{bmatrix} 1 \\ x_1^{(i)} \\ \vdots \\ x_d^{(i)} \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} (x^{(1)})^T \\ (x^{(2)})^T \\ \vdots \\ (x^{(N)})^T \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(N)} \end{bmatrix}$$

$$\mathcal{L}(\Theta) = \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - \Theta^T x^{(i)})^2$$

$$\Theta = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{y}$$

Polynomial Regression

$$\mathcal{D} = \{(x^{(1)}, y^{(1)}), ..., (x^{(N)}, y^{(N)})\}$$

$$x^{(i)} \in \mathbb{R}, \quad y^{(i)} \in \mathbb{R}$$

$$h(\theta, x) = \theta_0 + \theta_1 x + \theta_2 x^2 + ... + \theta_n x^k$$

Gradient Descent