1. Multivariate Linear Regression

1.1. Equation

1.1.1. Example

Independant / Input Variable / Feature : x

Number of features $x : n$, $n \in \mathbb{N}$

Output Variable / Dependant Variable / Target : y

Number of training examples $y : m$, $m \in \mathbb{N}$

the j^{th} feature : x_j for $j \in \{1, \dots, n\}$, where features are columns in a table

the i^{th} training set / vector : $x^{(i)}$ for $i \in \{1, \dots, m\}$, $x^{(i)} \in \mathbb{R}^n$, where indexes are rows in a table

value of feature j in i^{th} training example : $x_j^{(i)}$, in an $m \times n$ data table it is the value of the i^{th} row , j^{th} column

i^{th} training example : $(x_j^{(i)}, y^{(i)})$

Learning Rate : α

$$\vec{x} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} X = \begin{bmatrix} 1 & x_0 \\ 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix}, \Theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$$

$$\begin{aligned} h_\theta(x_0) &= \theta_0 + \theta_1 x_0 \\ h_\theta(x_1) &= \theta_0 + \theta_1 x_1 \\ h_\theta(x_2) &= \theta_0 + \theta_1 x_2 \\ h_\theta(x_3) &= \theta_0 + \theta_1 x_3 \end{aligned}$$

$$\begin{aligned} \begin{bmatrix} h_\theta(x_0) \\ h_\theta(x_1) \\ h_\theta(x_2) \\ h_\theta(x_3) \end{bmatrix} &= \begin{bmatrix} 1 & x_0 \\ 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix} \times \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} \\ &= \begin{bmatrix} (\theta_0 \cdot 1) + (\theta_1 \cdot x_0) \\ (\theta_0 \cdot 1) + (\theta_1 \cdot x_1) \\ (\theta_0 \cdot 1) + (\theta_1 \cdot x_2) \\ (\theta_0 \cdot 1) + (\theta_1 \cdot x_3) \end{bmatrix} \end{aligned}$$

$$\begin{aligned} A_{m \times n} & \times B_{n \times 1} = C_{m \times 1} \\ \begin{bmatrix} a_{0,n} & \cdots & a_{0,n} \\ \vdots & \ddots & \vdots \\ a_{m,0} & \cdots & a_{m,n} \end{bmatrix} & \times \begin{bmatrix} b_{0,n} \\ \vdots \\ b_{n,1} \end{bmatrix} = \begin{bmatrix} c_{0,1} \\ \vdots \\ c_{m,0} \end{bmatrix} \end{aligned}$$

1.1.2. General Form

$$h_\theta(x) = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \ldots + \theta_n \cdot x_n$$

$$\Theta_{(n+1 \times 1)} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \in \mathbb{R}^{n+1}$$

$$\Theta^T_{(1\times n+1)} = \left[\theta_0, \theta_1, \theta_2, \cdots, \theta_n\right]$$

$$\vec{x}_{(n+1\times 1)} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{n+1}$$

$$\begin{aligned} h_\theta(\vec{x}) &= \Theta^T_{(1\times n+1)} \bullet \vec{x}_{(n+1\times 1)} \\ &= \left[\theta_0, \theta_1, \theta_2, \cdots, \theta_n\right]_{(1\times n+1)} \bullet \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{(n+1\times 1)} \\ &= \left[x_0\theta_0 + x_1\theta_1 + x_2\theta_2 + \cdots + x_n\theta_n\right]_{(1\times 1)} \end{aligned}$$

$x^{(i)}_j$ = value of feature j in the i^{th} training example
$x^{(i)}$ = the input (features) of the i^{th} training example
m = the number of training examples
n = the number of features

$$\vec{x}^{(1)} = \begin{bmatrix} x^{(1)}_1 \\ x^{(1)}_2 \\ \vdots \\ x^{(1)}_j \\ \vdots \\ x^{(1)}_n \end{bmatrix}, \ldots, \vec{x}^{(i)} = \begin{bmatrix} x^{(i)}_1 \\ x^{(i)}_2 \\ \vdots \\ x^{(i)}_j \\ \vdots \\ x^{(i)}_n \end{bmatrix}, \ldots, \vec{x}^{(m)} = \begin{bmatrix} x^{(m)}_1 \\ x^{(m)}_2 \\ \vdots \\ x^{(m)}_j \\ \vdots \\ x^{(m)}_n \end{bmatrix}$$

$$\vec{x}^{(0)} = \begin{bmatrix} x^{(0)}_1 = 1 \\ x^{(0)}_2 = 1 \\ \vdots \\ x^{(0)}_j = 1 \\ \vdots \\ x^{(0)}_n = 1 \end{bmatrix}$$

$$\begin{aligned} X_{m\times n+1} &= \begin{bmatrix} \vec{x}^{(0)} & \vec{x}^{(1)} & \vec{x}^{(2)} & \cdots & \vec{x}^{(m)} \end{bmatrix} \\ &= \begin{bmatrix} x^{(0)}_1 & x^{(1)}_1 & x^{2)}_1 & \cdots & x^{(m)}_0 \\ x^{(0)}_2 & x^{(1)}_2 & x^{(2)}_2 & \cdots & x^{(m)}_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x^{(0)}_n & x^{(1)}_n & x^{(2)}_n & \cdots & x^{(m)}_n \end{bmatrix} \end{aligned}$$

$$
\begin{aligned}
h_\theta(X) \quad &= \Theta^T_{(1\times n+1)} \cdot X_{(n+1\times m)} \\
= \begin{bmatrix} h_\theta(x_0) \\ h_\theta(x_1) \\ h_\theta(x_2) \\ \vdots \\ h_\theta(x_m) \end{bmatrix}_{(1\times m)} \quad &= \begin{bmatrix}\theta_0, \theta_1, \theta_2, \cdots, \theta_n\end{bmatrix}_{(1\times n+1)} \cdot \begin{bmatrix} x_1^{(0)} & x_1^{(1)} & x_1^{(2)} & \cdots & x_0^{(m)} \\ x_2^{(0)} & x_2^{(1)} & x_2^{(2)} & \cdots & x_1^{(m)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n^{(0)} & x_n^{(1)} & x_n^{(2)} & \cdots & x_n^{(m)} \end{bmatrix}_{(n+1\times m)} \\
= \begin{bmatrix} h_\theta(x_0) \\ h_\theta(x_1) \\ h_\theta(x_2) \\ \vdots \\ h_\theta(x_j) \\ \vdots \\ h_\theta(x_m) \end{bmatrix}_{(1\times m)} \quad &= \begin{bmatrix} x_0^{(1)}\theta_0 + x_1^{(1)}\theta_1 + x_2^{(1)}\theta_2 + \cdots + x_n^{(1)}\theta_n \\ x_0^{(2)}\theta_0 + x_1^{(2)}\theta_1 + x_2^{(2)}\theta_2 + \cdots + x_n^{(2)}\theta_n \\ \vdots \\ x_j^{(i)}\theta_0 + x_j^{(i)}\theta_1 + x_j^{(i)}\theta_2 + \cdots + x_j^{(i)}\theta_n \\ \vdots \\ x_0^{(m)}\theta_0 + x_1^{(m)}\theta_1 + x_2^{(m)}\theta_2 + \cdots + x_n^{(m)}\theta_n \end{bmatrix}_{(1\times m)}
\end{aligned}
$$

$$0 \le i \le m \ , \ 0 \le j \le n \ , \ m, n \in \mathbb{N}$$

1.2. Cost Function

$$
\begin{aligned}
J(\Theta) &= J(\theta_0, \theta_1, \ldots, \theta_n) \\
J(\Theta) &= \frac{1}{2} \cdot \frac{1}{m} \cdot \sum_{i=1}^{m} \left\{ \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2 \right\} \\
&= \frac{1}{2} \cdot \frac{1}{m} \cdot \sum_{i=1}^{m} \left\{ \left(\sum_{j=0}^{n} \left(\theta_j x_j^{(i)} \right) - y^{(i)} \right)^2 \right\} \\
&= \frac{1}{2} \cdot \frac{1}{m} \cdot \sum_{i=1}^{m} \left\{ \left(\Theta^T x^{(i)} - y^{(i)} \right)^2 \right\}
\end{aligned} \tag{2}
$$

1.3. Gradient Descent

$$\nabla J(\Theta) = \left\langle \frac{\partial}{\partial \theta_0} J(\Theta), \frac{\partial}{\partial \theta_1} J(\Theta), \frac{\partial}{\partial \theta_2} J(\Theta) \ldots, \frac{\partial}{\partial \theta_m} J(\Theta) \right\rangle$$

$$
\begin{aligned}
\frac{\partial}{\partial \theta_j} J(\Theta) &= \frac{\partial}{\partial \theta_j} \left(\frac{1}{2} \cdot \frac{1}{m} \cdot \sum_{i=0}^{m} \left\{ \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2 \right\} \right) \\
&= \left(\cancel{2 \cdot \frac{1}{2}} \ \frac{1}{m} \cdot \sum_{i=0}^{m} \left\{ \left(h_\theta(x^{(i)}) - y^{(i)} \right) \cdot x_j^{(i)} \right\} \right)
\end{aligned}
$$

1.4. Update Rule

-§-