

# Linear Regression: Multivariate Linear Regression

## Hypothesis

Below is our *hypothesis* for *univariate* linear regression with a single feature value denoted by the variable  $x$ .

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

For multiple features we would represent them in our equation in the form of  $x_1, x_2, x_3$  all the way to  $x_n$ . For example, the below would be three features of  $x_1$  to  $x_3$ .

Features: *stock price* =  $x_1$ , *date sold* =  $x_2$ , *sale price* =  $x_3$

To denote a single value in example  $i$ , feature  $j$  we would write:

$$x_j^{(i)}$$

Therefore to get the 5th feature in the 3rd example we would write it as:

$$x_5^{(3)}$$

To support  $n$  features the hypothesis function has to change to the following.

$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots + \theta_n x_n$$

This can be written using vectors. Note, however, that  $x_0^{(i)}$  will be a constant of 1. This can be thought of as adding an additional 0 feature and our vectors are now *0-indexed*. The below is an example with 3 examples and 3+ features in each example:

$$\vec{X} = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & x_n^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & x_n^{(2)} \\ 1 & x_1^{(3)} & x_2^{(3)} & x_3^{(3)} & x_n^{(3)} \end{bmatrix}, \vec{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_n \end{bmatrix}$$

So, our hypothesis function can be written as the below where  $\theta_0 x_0 = 1 \cdot 1 = 1$

$$h_{\theta}(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots + \theta_n x_n$$

In order to multiply the two vectors,  $\theta$  and  $X$ , we need to *transpose* the theta vector which will then be labeled as  $\theta^T$  which is now an  $(n \times 1) \times 1$  matrix or *row vector*.

$$\theta^T = [\theta_0 \quad \theta_1 \quad \theta_2 \quad \theta_3 \quad \theta_n]$$

Now, the hypothesis function can be re-written as simply:

$$h_{\theta}(x) = \theta^T X$$

The function below visually illustrates our new hypothesis function.

$$h_{\theta}(x) = [\theta_0 \quad \theta_1 \quad \theta_2 \quad \theta_3 \quad \theta_n] \cdot \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & x_n^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & x_n^{(2)} \\ 1 & x_1^{(3)} & x_2^{(3)} & x_3^{(3)} & x_n^{(3)} \end{bmatrix}$$

# Gradient Descent

---

Our cost function for Gradient Descent is slightly different for multivariate linear regression. Here we are passing in not one, but multiple features, that range from  $\theta_0$  to  $\theta_n$ .

$$J(\theta_0, \theta_1, \dots, \theta_n) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

For simplicity we can just write the function as  $J(\theta)$ .

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

The Gradient Descent algorithm now looks like the following:

```
repeat until convergence {  
   $\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$   
} (Simultaneously updated for every  $j = 0, \dots, n$ )
```

Finding the partial derivative of the expression  $\alpha \frac{\partial}{\partial \theta_1} J(\theta)$  yields the below:

```
repeat until convergence {  
   $\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$   
} (Simultaneously updated for every  $j = 0, \dots, n$ )
```