

Normal Equation

The normal equation is another method of solving for linear regression, but can be done in one step as opposed to the iterative method of gradient descent. However, this only works with linear regression.

Let's define the notation first.

Say we have a function $f(A)$, where A is a matrix in $\mathbb{R}^{m \times n}$. Let's define the function as a mapping of the entries of matrix A into some scalar output. An example could be $f(A) = A_{1,1} + A_{1,2}^2$ where A is a 2x2 matrix.

Now we define the gradient $\nabla_A f(A)$ as a matrix with the same dimensions as A where each element is the partial derivative of the function 'f' with respect to that element $A_{i,j}$ at the i -th row and j -th column. If we apply this to our example, then the result would be

$$\begin{bmatrix} 1 & 2A_{1,2} \\ 0 & 0 \end{bmatrix}$$

So now let's apply this to our cost function $J(\theta)$. We want to set $\nabla_{\theta} J(\theta)$ to 0 since we want to minimize the change in the cost. Then, we solve for the value of theta. Since the minimum or maximum of a function is where the derivative is 0, this will allow us to find the minimum for our linear regression.

Our cost function is defined as:

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

where theta is the set of parameters.

Now, we can simplify this a bit by defining the hypothesis function h to be the vector product of x and theta, where x is our feature matrix. $h_{\theta}(x) = x\theta$. Note that $h_{\theta}(x)$ is actually a vector with elements that represent $h_{\theta}(x^{(i)})$.

Now we can simplify the cost function into

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

Now we can take the gradient of this (we will actually refer to this notation as the derivative) which results in

$$\nabla_{\theta} J(\theta) = \frac{1}{m} (x^T x \theta - x^T y)$$

Set that to 0, and then solve for θ results in

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

Linear regression can also be applied to other functions of x . We can define x_2 to be x^2 or \sqrt{x} or even e^x .