Deriving the normal equation for linear regression ¶

Jay Urbain

References:

- https://en.wikipedia.org/wiki/Linear_regression (https://en.wikipedia.org/wiki/Linear_regression) [show]
- Eli Bendersky

Given the hypothesis function:

$$h_\theta(x) = \theta_1 x_1 + \ldots + \theta_n x_n$$

We would like to minimize the least-squared error cost function:

$$J(\theta_{0 \ldots n}) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

Representing the hypothesis function in matrix form, where θ_0 and x are vectors:

$$h_\theta(x) = \theta_0^T x$$

Updating the cost function to be in matrix form:

$$J(\theta) = \frac{1}{2m} (\theta_0 X - y)^T (\theta_0 X - y)$$

Dropping $\frac{1}{2m}$ since we're comparing a derivitive of 0, and rearranging terms:

$$J(\theta) = ((\theta_0 X)^T - y^T)(X\theta_0 - y)$$

Note: Since ($X\theta$ and y) are vectors, we can multiply them in different orders provided the dimensions are correct.

$$J(\theta) = (X\theta_0)^T X\theta - 2(X\theta_0)^T y + y^T y$$

θ is what we are solving for, to find a minimum for our cost function $J(\theta)$, we need to take the derivitives of J with respect to θ.

$$\frac{\partial J}{\partial \theta} = 2X^T X\theta = 2X^T y = 0$$

or

$$X^T X\theta = X^T y$$

Multiply both sides by $(X^T X)^{-1}$

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

In [ ]: