$(x^{(m)}, y^{(m)})$; n features

⇒ Derivation of Normal equation linear regression.

Suppose we are given m examples (datapoints) and n features

So let our hypothesis function be :→

$$y(x) = m_0x_0 + m_1x_1 + \text{——} + m_nx_n.$$

Note

$x^i \Rightarrow$ i th sample.

$y^i_{actual} \Rightarrow$ i th expected result.

$y_{actual} \Rightarrow$ label vector.

So our objective is to minimize.

$$J(x_0, x_1, \ldots, x_n) = \frac{1}{2m} \sum_{i=1}^{m} \left[y(x^{(i)}) - y^{(i)}_{actual}\right]^2$$

So for our convenience add a another colum vector x_0 having 1

$$x_0 = \begin{bmatrix} 1 \\ \vdots \\ \vdots \\ 1 \end{bmatrix}_{m \times 1}$$

[which means we are adding a extra feature to our dataset for all datapoints]

We can write

$$y(x^i) \Rightarrow A x^i.$$

Where $A \Rightarrow m \times n{+}1$ dimension matrix represent feature vector. i.e. each colum represent a feature vector.

$x^i \Rightarrow n{+}1 \times 1$ dimension matrix represent ith datapoint from dataset.

So we can write

$$J(x) = \frac{1}{2m} (Ax - y_{actual})^T (Ax - y_{actual}).$$

$$J(x) = \frac{1}{2m} \left[\left((Ax)^T - y_{actual}^T\right)(Ax - y_{actual})\right]$$

$$\Rightarrow \frac{1}{2m} \left[(Ax)^T (Ax) - (Ax)^T y_{actual} - y_{actual}^T (Ax) + (y_{actual})(y_{actual})\right]$$

$$J(x) \Rightarrow \frac{1}{2m}\left[x^T A^T A x - 2(Ax)^T y_{actual} + y^T_{actual}\, y_{actual}\right].$$

Since $(Ax)^T y_{actual}$ and $y^T_{actual}(Ax)$ are finally going to be scalar so we can club them, because dimension works out.

Now find partial derivative wrt x.

$$\frac{\partial J}{\partial x} = 2A^T A x - 2A^T y_{actual} = 0$$

$$\text{So, } \cancel{2}A^T A x = \cancel{2}A^T y_{actual}$$

multiply both side by $(A^T A)^{-1}$.

$$(A^T A)^{-1}(A^T A)\, x = (A^T A)^{-1}(A^T \cdot y_{actual})$$

So $[\because (A^T A)^{-1}(A^T \cdot A) = I$ and $I \cdot x = x]$.

$$x = (A^T A)^{-1}(A^T y_{actual}) \quad \left[\text{Assuming } |A^T A| \neq 0\right]$$

So, x is the minimized value for $J(x)$ which is $(A^T A)^{-1}(A^T y_{actual})$. □