CS 525: Deep Neural Networks
Homework 3 - Problem 1

We are given the following cost function:

$$J(w) = \frac{1}{2}\sum(w^T x - y)^2 = \frac{1}{2}(Xw - y)^T(Xw - y)$$

Using the formula given in class[1], we find:

$$\nabla_w J = \frac{1}{2} * 2X^T(Xw - y) = X^T Xw - X^T y \tag{1}$$

We take the gradient of this function to find the Hessian:

$$H(J)(w) = X^T X \tag{2}$$

Finally, given equations (1) and (2), we can compute Newton's method and find w^* in a single iteration:

$$w^* = w^{(0)} - [H(J)(w^{(0)})]^{-1}\nabla_w J(w^{(0)})$$

$$= w^{(0)} - (X^T X)^{-1}(X^T X w^{(0)} - X^T y)$$

$$= w^{(0)} - (X^T X)^{-1}(X^T X)w^{(0)} + (X^T X)^{-1}(X^T y)$$

$$= w^{(0)} - Iw^{(0)} + (X^T X)^{-1}(X^T y)$$

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

We have derived the correct expression for w^*, and clearly it is not dependent on our choice of $w^{(0)}$.

[1] If $f(x) = [Ax + b]^T[Ax + b]$ then $\nabla_x f = 2A^T(Ax + b)$