CSE – 426 Homework 1

Griffin Kent

1) Let the training examples be $\boldsymbol{x}^{(1)} = [1, 1]^T$, $\boldsymbol{x}^{(2)} = [1,2]^T$, $y^{(1)} = 1$, $y^{(2)} = 5$. Write the MSE loss function $J(\boldsymbol{\theta})$ with parameters $\boldsymbol{\theta}$ using the linear hypothesis $h_{\boldsymbol{\theta}}(\boldsymbol{x}) = \boldsymbol{\theta}^T\boldsymbol{x}$.

(Solution): The MSE loss function is defined as follows:

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

Using the linear hypothesis, we can write the loss function as

$$J(\boldsymbol{\theta}) = \frac{1}{2m}\left[\left(\boldsymbol{\theta}^T\,\boldsymbol{x}^{(1)} - y^{(1)}\right)^2 + \left(\boldsymbol{\theta}^T\,\boldsymbol{x}^{(2)} - y^{(2)}\right)^2\right]$$

$$= \frac{1}{2\cdot 2}\left[\left([\theta_0 \quad \theta_1]\begin{bmatrix}1\\1\end{bmatrix} - 1\right)^2 + \left([\theta_0 \quad \theta_1]\begin{bmatrix}1\\2\end{bmatrix} - 5\right)^2\right]$$

$$= \frac{1}{4}[(\theta_0 + \theta_1 - 1)^2 + (\theta_0 + 2\theta_1 - 5)^2]$$

$$= \frac{1}{4}[(\theta_0^2 + \theta_0\theta_1 - \theta_0 + \theta_0\theta_1 + \theta_1^2 - \theta_1 - \theta_0 - \theta_1 + 1)$$
$$+ (\theta_0^2 + 2\theta_0\theta_1 - 5\theta_0 + 2\theta_0\theta_1 + 4\theta_1^2 - 10\theta_1 - 5\theta_0 - 10\theta_1 + 25)]$$

$$= \frac{1}{4}[2\theta_0^2 + 5\theta_1^2 + 6\theta_0\theta_1 - 12\theta_0 - 22\theta_1 + 26]$$

$$= \frac{1}{2}\theta_0^2 + \frac{5}{4}\theta_1^2 + \frac{3}{2}\theta_0\theta_1 - 3\theta_0 - \frac{11}{2}\theta_1 + \frac{13}{2}. \blacksquare$$

2) Show that the above loss function is a quadratic function in θ_1.

(Proof): Since the majority of the work was done above, we can simply look and see that a θ_1^2 term exists in the final expression of $J(\boldsymbol{\theta})$. Thus, it is a quadratic function in θ_1. $\blacksquare$

3) Find the gradient of $J(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = [0,0]^T$. That is, compute $\nabla_{\boldsymbol{\theta}}J(\boldsymbol{\theta})|_{\boldsymbol{\theta}=[0,0]^T} = \frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=[0,0]^T}$. Your answer should be a column vector of two entries.

(Solution): First, we can find a general expression for the gradient of the loss function above:

$$\nabla_{\boldsymbol{\theta}}J(\boldsymbol{\theta}) = \begin{bmatrix}\dfrac{\partial J(\boldsymbol{\theta})}{\partial \theta_0}\\[2ex] \dfrac{\partial J(\boldsymbol{\theta})}{\partial \theta_1}\end{bmatrix} = \begin{bmatrix}\theta_0 + \dfrac{3}{2}\theta_1 - 3\\[2ex] \dfrac{3}{2}\theta_0 + \dfrac{5}{2}\theta_1 - \dfrac{11}{2}\end{bmatrix}.$$

Now, evaluating the gradient at $\boldsymbol{\theta} = [0,0]^T$, we obtain

$$\nabla_{\boldsymbol{\theta}}J(\boldsymbol{\theta})|_{\boldsymbol{\theta}=[0,0]^T} = \begin{bmatrix}-3\\ -\dfrac{11}{2}\end{bmatrix}. \blacksquare$$

4) Use the normal equation to compute the optimal parameter θ^*. Leave the matrix inversion as it is.

(Solution): To start, let's define the matrix $X = \begin{bmatrix} \boldsymbol{x}^{(1)^T} \\ \boldsymbol{x}^{(2)^T} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$ as the training examples and the column vector $\boldsymbol{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$ as the true-value vector. The normal equation is then defined as follows:

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

Thus, we can see that

$$\boldsymbol{\theta}^* = \left(\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 6 \\ 11 \end{bmatrix}$$

To find the inverse, we can calculate the determinant to be:

$$det\left(\begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix} \right) = 10 - 9 = 1.$$

Thus,

$$\begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}^{-1} = \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix}$$

This will give the solution

$$\boldsymbol{\theta}^* = \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 11 \end{bmatrix} = \begin{bmatrix} -3 \\ 4 \end{bmatrix}. \blacksquare$$

5) (Graduate Only) We had you find the gradient of J with respect to θ. Now you can take the derivative of the gradient with respect to θ again to find the second-order derivative of J with respect to θ. Your result should be a 2x2 matrix (the Hessian matrix).

(Solution): Taking the derivative of the gradient, we have the hessian matrix

$$\nabla_\theta^2 J(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial^2 J(\boldsymbol{\theta})}{\partial \theta_0^2} & \frac{\partial^2 J(\boldsymbol{\theta})}{\partial \theta_0 \partial \theta_1} \\ \frac{\partial^2 J(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_0} & \frac{\partial^2 J(\boldsymbol{\theta})}{\partial \theta_1^2} \end{bmatrix} = \begin{bmatrix} 1 & \frac{3}{2} \\ \frac{3}{2} & \frac{5}{2} \end{bmatrix}. \blacksquare$$