

AI Assignment 1

Nurseiit Abdimomyn – 20172001

26/09/2021

1.1 We have $\theta = [\theta_0, \theta_1, \theta_2]^T$ and $x_i = [1, x_{i0}, x_{i1}]^T$ for $h_\theta(x) = \theta^T x$.

So we derive for the gradient:

$$\begin{aligned}\frac{\partial L(\theta)}{\partial \theta_0} &= \frac{\partial}{\partial \theta_0} \sum_i (\theta_0 + \theta_1 x_{i0} + \theta_2 x_{i1} - y_i)^2 \\ &= \sum_i \frac{\partial}{\partial \theta_0} (\theta_0 + \theta_1 x_{i0} + \theta_2 x_{i1} - y_i)^2 \\ &= \sum_i 2(\theta_0 + \theta_1 x_{i0} + \theta_2 x_{i1} - y_i)\end{aligned}$$

Doing the same for θ_1 and θ_2 we get:

$$\begin{aligned}\frac{\partial L(\theta)}{\partial \theta_1} &= \sum_i 2(\theta_0 + \theta_1 x_{i0} + \theta_2 x_{i1} - y_i) * x_{i0} \\ \frac{\partial L(\theta)}{\partial \theta_2} &= \sum_i 2(\theta_0 + \theta_1 x_{i0} + \theta_2 x_{i1} - y_i) * x_{i1}\end{aligned}$$

To find a minimum by gradient descent we use:

$$\theta^{k+1} \leftarrow \theta^k - \eta_k \frac{\partial L(\theta)}{\partial \theta} \text{ where we have } \eta = 0.01 \text{ for all } k.$$

$$\text{Thus, } \theta^{k+1} \leftarrow \theta^k - 0.01 * \frac{\partial L(\theta)}{\partial \theta}.$$

Initial parameter $\theta^0 = [0, 0, 0]^T$, so compute for θ^1 simultaneously:

$$\begin{aligned}\theta_0^1 &= \theta_0^0 - 0.01 * \sum_i 2(\theta_0^0 + \theta_1^0 x_{i0} + \theta_2^0 x_{i1} - y_i) \\ &= 0 - 0.01 * \sum_i 2(0 + 0 * x_{i0} + 0 * x_{i1} - y_i) \\ &= 0.02 * \sum_i y_i = 0.02 * 37 = 0.74.\end{aligned}$$

$$\begin{aligned}
\theta_1^1 &= \theta_1^0 - 0.01 * \sum_i 2(\theta_0^0 + \theta_1^0 x_{i0} + \theta_2^0 x_{i1} - y_i) * x_{i0} \\
&= 0 - 0.01 * \sum_i 2(0 + 0 * x_{i0} + 0 * x_{i1} - y_i) * x_{i0} \\
&= 0.02 * \sum_i (y_i * x_{i0}) \\
&= 0.02 * (6 * 1 + 4 * 2 + 7 * 3 + 10 * 3 + 10 * 4) = 0.02 * 105 = 2.1.
\end{aligned}$$

$$\begin{aligned}
\theta_2^1 &= \theta_2^0 - 0.01 * \sum_i 2(\theta_0^0 + \theta_1^0 x_{i0} + \theta_2^0 x_{i1} - y_i) * x_{i1} \\
&= 0 - 0.01 * \sum_i 2(0 + 0 * x_{i0} + 0 * x_{i1} - y_i) * x_{i1} \\
&= 0.02 * \sum_i (y_i * x_{i1}) \\
&= 0.02 * (6 * 2 + 4 * 1 + 7 * 2 + 10 * 3 + 10 * 5) = 0.02 * 110 = 2.2.
\end{aligned}$$

So for the first iteration we have $\theta^1 = [0.74, 2.1, 2.2]^T$.

Second iteration:

$$\begin{aligned}
\theta_0^2 &= \theta_0^1 - 0.01 * \sum_i 2(\theta_0^1 + \theta_1^1 x_{i0} + \theta_2^1 x_{i1} - y_i) \\
&= 0.74 - 0.01 * \sum_i 2(0.74 + 2.1x_{i0} + 2.2x_{i1} - y_i) \\
&= 0.74 - 0.02 * 22.6 = 0.288.
\end{aligned}$$

$$\begin{aligned}
\theta_1^2 &= \theta_1^1 - 0.01 * \sum_i 2(\theta_0^1 + \theta_1^1 x_{i0} + \theta_2^1 x_{i1} - y_i) * x_{i0} \\
&= 2.1 - 0.01 * \sum_i 2(0.74 + 2.1x_{i0} + 2.2x_{i1} - y_i) * x_{i0} \\
&= 2.1 - 0.02 * \sum_i (0.74 + 2.1x_{i0} + 2.2x_{i1} - y_i) * x_{i0} = 0.6536
\end{aligned}$$

$$\begin{aligned}
\theta_2^2 &= \theta_2^1 - 0.01 * \sum_i 2(\theta_0^1 + \theta_1^1 x_{i0} + \theta_2^1 x_{i1} - y_i) * x_{i1} \\
&= \theta_2^1 - 0.02 * \sum_i (\theta_0^1 + \theta_1^1 x_{i0} + \theta_2^1 x_{i1} - y_i) * x_{i1}
\end{aligned}$$

$$= 2.2 - 0.02 * \sum_i (0.74 + 2.1x_{i0} + 2.2x_{i1} - y_i) * x_{i1} = 0.6776.$$

$$\text{Thus, } \theta^2 = [0.288, 0.6536, 0.6776]^T.$$

$$1.2 \quad \theta^* = (X^T X)^{-1} X^T y$$

$$\begin{aligned}
&= \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 1 & 2 & 3 & 5 \\ 1 & 2 & 3 & 3 & 4 \end{bmatrix} * \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 3 & 2 \\ 1 & 3 & 3 \\ 1 & 4 & 5 \end{bmatrix} \right)^{-1} * \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 1 & 2 & 3 & 5 \\ 1 & 2 & 3 & 3 & 4 \end{bmatrix} * \begin{bmatrix} 6 \\ 4 \\ 7 \\ 10 \\ 10 \end{bmatrix} \\
&= \left(\begin{bmatrix} 5 & 13 & 13 \\ 13 & 39 & 43 \\ 13 & 39 & 39 \end{bmatrix} \right)^{-1} * \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 1 & 2 & 3 & 5 \\ 1 & 2 & 3 & 3 & 4 \end{bmatrix} * \begin{bmatrix} 6 \\ 4 \\ 7 \\ 10 \\ 10 \end{bmatrix} \\
&= \begin{bmatrix} 3/2 & 0 & -1/2 \\ -1/2 & -1/4 & 23/52 \\ 0 & 1/4 & -1/4 \end{bmatrix} * \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 1 & 2 & 3 & 5 \\ 1 & 2 & 3 & 3 & 4 \end{bmatrix} * \begin{bmatrix} 6 \\ 4 \\ 7 \\ 10 \\ 10 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1/2 & 0 & 0 & -1/2 \\ -20/52 & 7/52 & 17/52 & 1/13 & 1/52 \\ 1/4 & -1/4 & -1/4 & 0 & 1/4 \end{bmatrix} * \begin{bmatrix} 6 \\ 4 \\ 7 \\ 10 \\ 10 \end{bmatrix} \\
&= \begin{bmatrix} 3 \\ 23/52 \\ 5/4 \end{bmatrix}
\end{aligned}$$