The Question

Hints for NPTEL MLESA assignment-3/question-7, by `http://perwad.in` on 2019/02/18 at 22:00

Consider

$$J(\mathbf{w}) = \frac{1}{10}\sum_{i=1}^{5}(y^{(i)} - w_1 x^{(i)} - w_0)^2 \qquad \text{where } \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

and the constants $x^{(i)}$ and $y^{(i)}$ are provided in the table below:

i	**x**	**y**
1	0.00	0.8822
2	0.25	1.2165
3	0.50	1.3171
4	0.75	1.7930
5	1.00	1.9826

Find the **w** which minimize the J(**w**).

$$\text{i.e. } \operatorname*{argmin}_{\mathbf{w}} J(\mathbf{w})$$

Variables in the function J

The J($\mathbf{w}$) is a multivariable function. There are only two variables in the function: w_0 and w_1. The $x^{(i)}$ and $y^{(i)}$ are constants not variables. Let us expand the J($\mathbf{w}$) in order to get clarified.

$$\begin{aligned}
J(\mathbf{w}) &= \frac{1}{10}\sum_{i=1}^{5}(y^{(i)} - w_1 x^{(i)} - w_0)^2 \\
&= \frac{1}{10}((y^{(1)} - x^{(1)}w_1 - w_0)^2 + (y^{(2)} - x^{(2)}w_1 - w_0)^2) + \\
&\quad (y^{(3)} - x^{(3)}w_1 - w_0)^2) + (y^{(4)} - x^{(4)}w_1 - w_0)^2) + (y^{(5)} - x^{(5)}w_1 - w_0)^2)
\end{aligned}$$

x and y are constants. Let us replace them with corresponding values.

$$\begin{aligned}
J(\mathbf{w}) &= \frac{1}{10}(\ (0.8822 - 0.00w_1 - w_0)^2 + (1.2165 - 0.25w_1 - w_0)^2 + \\
&\quad (1.3171 - 0.50w_1 - w_0)^2 + (1.7930 - 0.75w_1 - w_0)^2 + (1.9826 - 1.00w_1 - w_0)^2\ ) \\
&\approx 5w_0^2 + 5w_0w_1 - 14.3828w_0 + 1.875w_1^2 - 8.58005w_1 + 11.1385
\end{aligned}$$

It proves that there are only two variables in the function. i.e. w_0 and w_1. Please cross-check the approximation if you want to use it.

Gradient of the function J

Let us see ∇J (gradient of J).

$$\nabla J = \begin{bmatrix} \frac{\partial J}{\partial w_0} \\ \\ \frac{\partial J}{\partial w_1} \end{bmatrix}$$

The partial differentiation of J with respect to w_0 and w_1:

$$\begin{aligned}
\frac{\partial J}{\partial w_0} &= \\
&= \frac{1}{10}\sum_{i=1}^{5}(2 * (y^{(i)} - w_1 x^{(i)} - w_0) * -1 \\
&= \frac{-2}{10}\sum_{i=1}^{5}(y^{(i)} - w_1 x^{(i)} - w_0)
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial J}{\partial w_1} &= \\
&= \frac{1}{10}\sum_{i=1}^{5}(2 * (y^{(i)} - w_1 x^{(i)} - w_0) * -x^{(i)} \\
&= \frac{-2}{10}\sum_{i=1}^{5}((y^{(i)} - w_1 x^{(i)} - w_0) * x^{(i)})
\end{aligned}$$

Computing $\mathbf{w}^{new}$ from $\mathbf{w}^{old}$

Each iteration starts with a $\mathbf{w}$, say $\mathbf{w}^{old}$, and at the end of the iteration will have the updated $\mathbf{w}$, say $\mathbf{w}^{new}$.

$$\mathbf{w}^{new} = \mathbf{w}^{old} - \alpha * \nabla_{\mathbf{w}^{old}} J \qquad \text{where } \alpha \text{ is learning rate.}$$

Let us expand it.

$$\begin{bmatrix} w_0^{new} \\ \\ w_1^{new} \end{bmatrix} = \begin{bmatrix} w_0^{old} \\ \\ w_1^{old} \end{bmatrix} - \alpha * \begin{bmatrix} \frac{\partial J}{\partial w_0^{old}} \\ \\ \frac{\partial J}{\partial w_1^{old}} \end{bmatrix}$$

$$= \begin{bmatrix} w_0^{old} - \alpha * \frac{\partial J}{\partial w_0^{old}} \\ \\ w_1^{old} - \alpha * \frac{\partial J}{\partial w_1^{old}} \end{bmatrix}$$

Have a look at the w_0^{new} and w_1^{new} separately.

$$w_0^{new} = w_0^{old} - \alpha * \frac{\partial J}{\partial w_0^{old}}$$

and

$$w_1^{new} = w_1^{old} - \alpha * \frac{\partial J}{\partial w_1^{old}}$$