Assignment 4

1 (a).

We are given that the equation of our linear regression model is,

$$y = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + b$$

Since we are considering per example loss, our formula for criterion will be,

$$loss = (y_p^{(i)} - y^{(i)})^2$$

(since we are considering only a single example)

Let's find gradients for different weights and biase,

$$\frac{\partial loss}{\partial w_1} = \frac{\partial (y_p^{(i)} - y^{(i)})^2}{\partial w_1}$$

$$= \frac{\partial (y_p^{(i)} - y^{(i)})^2}{\partial (y_p^{(i)} - y^{(i)})} \times \frac{\partial (y_p^{(i)} - y^{(i)})}{\partial w_1}$$

$$= 2 \times (y_p^{(i)} - y^{(i)}) \times \frac{\partial (w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + b - y^{(i)})}{\partial w_1}$$

$$= 2 \times (y_p^{(i)} - y^{(i)}) \times x_1$$

$$= 2 \times (w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + b - y^{(i)}) \times w_1$$

,

$$\frac{\partial loss}{\partial w_2} = \frac{\partial (y_p^{(i)} - y^{(i)})^2}{\partial w_2}$$

$$= \frac{\partial (y_p^{(i)} - y^{(i)})^2}{\partial (y_p^{(i)} - y^{(i)})} \times \frac{\partial (y_p^{(i)} - y^{(i)})}{\partial w_2}$$

$$= 2 \times (y_p^{(i)} - y^{(i)}) \times \frac{\partial (w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 + w_5 x_5 + b - y^{(i)})}{\partial w_2}$$

$$= 2 \times (y_p^{(i)} - y^{(i)}) \times x_2$$

$$= 2 \times (w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 + w_5 x_5 + b - y^{(i)}) \times x_2$$

$$\frac{\partial loss}{\partial w_3} = \frac{\partial (y_p^{(i)} - y^{(i)})^2}{\partial w_3}$$

$$= \frac{\partial (y_p^{(i)} - y^{(i)})^2}{\partial (y_p^{(i)} - y^{(i)})} \times \frac{\partial (y_p^{(i)} - y^{(i)})}{\partial w_3}$$

$$= 2 \times (y_p^{(i)} - y^{(i)}) \times \frac{\partial (w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 + w_5 x_5 + b - y^{(i)})}{\partial w_3}$$

$$= 2 \times (y_p^{(i)} - y^{(i)}) \times x_3$$

$$= 2 \times (w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 + w_5 x_5 + b - y^{(i)}) \times x_3$$

$$\frac{\partial loss}{\partial w_4} = \frac{\partial (y_p^{(i)} - y^{(i)})^2}{\partial w_4}$$

$$= \frac{\partial (y_p^{(i)} - y^{(i)})^2}{\partial (y_p^{(i)} - y^{(i)})} \times \frac{\partial (y_p^{(i)} - y^{(i)})}{\partial w_4}$$

$$= 2 \times (y_p^{(i)} - y^{(i)}) \times \partial (w_1 x_1 + w_2 x_2 +$$

$$\frac{w_3x_3+w_4x_4+w_5x_5+b-y^{(i)})}{\partial w_4}$$

$$= 2\times(y_p^{(i)}-y^{(i)})\times x_4$$

$$= 2\times(w_1x_1+w_2x_2+w_3x_3+w_4x_4+w_5x_5+b-y^{(i)})\times x_4$$

$$\frac{\partial loss}{\partial w_5} = \frac{\partial(y_p^{(i)}-y^{(i)})^2}{\partial w_5}$$

$$= \frac{\partial(y_p^{(i)}-y^{(i)})^2}{\partial(y_p^{(i)}-y^{(i)})} \times \frac{\partial(y_p^{(i)}-y^{(i)})}{\partial w_5}$$

$$= 2\times(y_p^{(i)}-y^{(i)}) \times \frac{\partial(w_1x_1+w_2x_2+w_3x_3+w_4x_4+w_5x_5+b-y^{(i)})}{\partial w_5}$$

$$= 2\times(y_p^{(i)}-y^{(i)})\times x_5$$

$$= 2\times(w_1x_1+w_2x_2+w_3x_3+w_4x_4+w_5x_5+b-y^{(i)})\times x_5$$

Similarly gradients for other weights can be generalized by formula given below,

$$\frac{\partial loss}{\partial w_j} = 2\times(y_p^{(i)}-y^{(i)})\times x_j$$

Where w_j is the j^{th} weight
$y_p^{(i)}$ is prediction (predicted)
$y^{(i)}$ is the label of sample i (Actual)

Now let's calculate with respect to bias,

$$\frac{\partial loss}{\partial b} = \frac{\partial(y_p^{(i)}-y^{(i)})^2}{\partial b}$$

$$= \partial(y_p^{(i)}-y^{(i)})^2 \times \partial(w_1x_1+w_2x_2+w_3x_3+w_4x_4)$$

$$\frac{\partial (y_p^{(i)} - y^{(i)})}{} \qquad \frac{+w_5x_5 + b - y^{(i)}}{\partial b}$$

$$= 2\times(y_p^{(i)} - y^{(i)}) \times 1$$

$$= 2\times(w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + b - y^{(i)}) \times 1$$

let us write it in vector form,

$$\boxed{2\cdot(w^Tx^{(i)} - y^{(i)})\cdot x_j^{(i)}} \quad \text{—— (i)}$$

jth entry in vector

since we are considering only single example, else we can change

for bias,

$$\boxed{2\cdot(w^Tx^{(i)} - y^{(i)})} \quad \text{—— (ii)}$$

Let's write the update rule,

$$w_1 = w_1 - \alpha\cdot 2\cdot(w^Tx^{(i)} - y^{(i)})\cdot x_1$$
$$w_2 = w_2 - \alpha\cdot 2\cdot(w^Tx^{(i)} - y^{(i)})\cdot x_2$$
$$w_3 = w_3 - \alpha\cdot 2\cdot(w^Tx^{(i)} - y^{(i)})\cdot x_3$$
$$w_4 = w_4 - \alpha\cdot 2\cdot(w^Tx^{(i)} - y^{(i)})\cdot x_4$$
$$w_5 = w_5 - \alpha\cdot 2\cdot(w^Tx^{(i)} - y^{(i)})\cdot x_5$$
$$b = b - \alpha\cdot 2(w^Tx^{(i)} - y^{(i)})$$

(Here α is the learning rate)