Theory

Q1 a. Given expanded eqn form:

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

Now let g be the gradient and the derivation for linear regression will be given as:-

$$y = w_1x_1 + w_2x_2 + \dots + w_nx_n + c$$

Here y is predicted output ~~as~~ and ~~x_i~~ x_i are the input ~~function~~ features, w_i are the weights and c_i is the bais term and lastly n is no of features

Now MSE (Mean squared error) loss function is ~~go~~ given as:-

$$L(w, c_i \{x^i, y^i\}) = \frac{1}{N} \sum_{i=1}^{n} (x^i w + c - y^i)^2$$

Now to find the derivates w.r.t w_i and the bias term

DATE

Gradient w.r.t w_i:

$$\frac{\partial L}{\partial w_i} = \frac{2}{N} \sum_{i=1}^{5} x_i^i (x_i^{T} w + c - y^i)$$

Gradient w.r.t c

$$\frac{\partial L}{\partial c} = \frac{2}{N} \sum_{i=1}^{5} (x_i w + c - y^i)$$

Gradient Descent Update ~~Rule~~ Rule:-

Now the rule for weights and bais using the MSE cost function will be given as:

$$w_i = w_i - \alpha \frac{\partial L}{\partial w_i}$$

$$c_i = c - \alpha \frac{\partial L}{\partial c}$$

α is learning rate

This rule applied to the iteration of y to minimise the MSE and improve model's predictions

DATE

Now for Vector form :

$w = [w_1, w_2, w_3, w_4, w_5]^T$

$x = [x_1, x_2, x_3, x_4, x_5]^T$,

$y = w^T x + b$

~~w~~ ~~as~~ W and b are weights and offset

~~Let~~ Given

$$y = \sum_{i=1}^{5} x_i w_i + C$$

in vector form $\sum_{i=1}^{5} x_i w_i = \cancel{x} X W$

Here $X \Rightarrow [x_1, x_2 ... x_5]$, $W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_5 \end{bmatrix}$

$$MSE = \frac{1}{N} \sum_{i=1}^{5} \cancel{(\alpha} (w_i^T x_i - \bar{y})^2$$

This can be written as $\frac{1}{N} (XW - \bar{y})^2$

Hence, $\cancel{x}(XW - \bar{y})^2 = (XW - \bar{y})^T (XW - \bar{y})$

$$MSE = \frac{1}{N} (XW - \bar{y})^T (XW - \bar{y})$$

DATE

Expanding ~~all of~~ this

$$(XW - \bar{y})^T (XW - \bar{y})$$

$$(W^T X^T - \bar{y}^T)(XW - \bar{y})$$

$$(W^T X^T)(XW) - (W^T X^T)\bar{y} - \bar{y}^T (XW) + \bar{y}^T y$$

$$N \times \frac{d\,MSE}{dW} = \frac{d}{dW}(W^T X^T)(XW) - \frac{d}{dW}(W^T X^T)\bar{y}$$

$$- \frac{d}{dW}(\bar{y}^T) XW + \cancel{\frac{d}{dW} \bar{y}^T y}$$

After solving we get

$$\frac{d\,MSE}{dW} = \frac{2}{N}(X^T X W - X^T \bar{y})$$

For minima $\frac{d\,MSE}{dW} = 0$

$$\therefore X^T X W - X^T \bar{y} = 0$$

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