Q / Part No. Rough Work

Delta Rule

Dataset:

$$\begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \quad \begin{array}{c} b \end{array} \quad \text{label}$$

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} w_0 \\ 2 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

unit step func
$x < 0 \quad 0$
$x \geq 0 \quad 1$

weight
$$\begin{array}{ccc} w_1 & w_2 & w_3 \end{array}$$
$$\begin{bmatrix} -2 & 2 & 1 \end{bmatrix}$$

$= b_1 w_0 + x_1 w_1 + x_2 w_2 + x_3 w_3$

$= 1(2) + 1(-2) + 0(2) + 1(1)$

$= 2 - \cancel{2} + \cancel{2} + 1$

$= 3$

$\hat{y}_1 = act(3) = 1$

2nd sample

$b_2 w_0 + x_1 w_1 + x_2 w_2 + x_3 w_3$

$= 1(2) + 0(-2) + 0(2) + 0(1)$

$= 2$

$act(2) = 1$

$\hat{y}_2 = 1$

$\boxed{\alpha = 0.2}$

$y_i - \hat{y}_i$	for $(y-\hat{y}).b$	$(y-\hat{y}).x_1$	$(y-\hat{y}).x_2$	$(y-\hat{y}).x_3$
$0 - 1 = -1$	$1 \times (-1) = -1$	$-1 * 1 = -1$	$0 * -1 = 0$	$-1 * 1 = -1$
$1 - 1 = 0$	$1 \times 0 = 0$	$0 \times 0 = 0$	$0 * 0 = 0$	$0 * 0 = 0$
$\Sigma = -1$	$\Sigma = -1$	$\Sigma = -1$	$\Sigma\ 0$	$\Sigma = -1$

$w_{0_{new}} = w_0 + \Sigma\alpha(y_i - \hat{y}_i).b = 2 + 0.2(-1) = 1.2$

$w_{1_{new}} = w_1 + \Sigma\alpha(y_i - \hat{y}_i).x_1 = -2 + 0.2(-1) = -2.2$

$w_{2\,new} = w_2 + \Sigma\alpha(y_i - \hat{y}_i).x_2 = 2 + 0.2(0) = 2$

$w_{3\,new} = w_3 + \Sigma\alpha(y_i - \hat{y}_i).x_3 = 1 + 0.2(-1) = 0.8$

$$\begin{array}{c} w_0 \\ w_1 \\ w_2 \\ w_3 \end{array} = \begin{bmatrix} 1.2 \\ -2.2 \\ 2 \\ 0.8 \end{bmatrix}$$

Delta Rule: $\alpha . \Sigma (y-\hat{y}) . f'(act) . x_1$

Suppose no act function.

$(y-\hat{y}) . x_1$

cost function: $\frac{1}{2}(y-\hat{y})^2$

$$\frac{dc}{dw_0} = \frac{d\left(\frac{1}{2}(y-\hat{y})^2\right)}{dw_0}$$

now; $\hat{y} = w_0 b + w_1 x_1 + \ldots$

so $\hat{y}$ has many more variables and not only w_0

⇒ There fore we will take its Partial derivative.

⇒ Your cost is not directly dependent on W_0.

Cost depends on $\hat{y}$
$\hat{y}$ depends on W_0] This is called Chain rule

Therefore: $\frac{dc}{dw_0} = \frac{dc}{d\hat{y}} . \frac{d\hat{y}}{dw_0}$

Lets take the first term,

$$\frac{dc}{d\hat{y}} = \frac{\partial \frac{1}{2}(y-\hat{y})^2}{\partial(\hat{y})}$$

⇒ Since $\hat{y}$ has many variables hence taking its partial derivative.

$$= \frac{1}{\cancel{2}} . \cancel{2} (y-\hat{y})^{2-1} \left(\frac{dy}{d\hat{y}} - \frac{d\hat{y}}{d\hat{y}}\right)$$

$$(y-\hat{y})(0-1)$$

$$(y-\hat{y})(-1)$$

$$-(y-\hat{y})$$

Therefore: $\frac{dc}{dw_0} = \frac{dc}{d\hat{y}} \cdot \frac{d\hat{y}}{dw_0}$

Lets take the first term:

$$\frac{dc}{d\hat{y}} = \frac{\partial \frac{1}{2}(y-\hat{y})^2}{\partial(\hat{y})}$$

⇒ since $\hat{y}$ has many variables hence taking its partial derivative.

$$= \frac{1}{2} \cdot 2(y-\hat{y})^{2-1}\left(\frac{dy}{dy} - \frac{d\hat{y}}{d\hat{y}}\right)$$

$$(y-\hat{y})(0-1)$$

$$(y-\hat{y})(-1)$$

$$-(y-\hat{y})$$

So, $\frac{dc}{d\hat{y}} = -(y-\hat{y})$

$\hat{y} = w_0 b + w_1 x_1 + \cdots$

Now lets see; $\frac{d\hat{y}}{dw_0} = \frac{\partial(w_0 b + w_1 x_1 + \cdots)}{dw_0}$

$$= \frac{d(w_0 b)}{dw_0} + \frac{dw_1 x_1}{dw_0}$$

$$= 1 \cdot b + 0$$

$$= 1$$

Since bias is alway 1

$$\Rightarrow -(y-\hat{y}) \cdot 1$$

for weight update the equation is:

$$w_0 - \alpha \cdot \frac{\partial c}{\partial w}$$

$$w_0 - (-(y-\hat{y}) \cdot x_i)$$

$$w_0 + \alpha(y-\hat{y}) \cdot x_i$$