- **Linear Activation**
  - The linear activation function, also known as "no activation," or "identity function" (multiplied x1.0), is where the activation is proportional to the input. **The function doesn't do anything to the weighted sum of the input, it simply spits out the value it was given.**

- **Generalized Delta Rule for updating weights**

$$\Delta w_{ij} = -c\,(\partial Error/\,\partial w_{ij}) = -c[-(d_j - O_j)\,.\,f'(act_j)\,.\,x_i]$$

- C is the learning rate
- F’ is the derivative of the activation function

Q / Part No. Rough Work

Delta Rule

Dataset:

$$\begin{array}{ccc} x_1 & x_2 & x_3 \\ \end{array} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{array}{c} b \text{ weight} \\ w_0 \end{array} \begin{bmatrix} 2 \end{bmatrix} \quad \text{label} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

unit step func
$x < 0$ 0
$x \geq 0$ 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 $(y-\hat{y}).b$	$(y-\hat{y}).x_1$	$(y-\hat{y}).x_2$	$(y-\hat{y}).x_3$
$0 - 1 = -1$	$1 \times (-1) = -1$	$-1 * 1 = -1$	$0 * -1 = 0$	$-1 * 1 = -1$
$1 - 1 = 0$	$1 \times 0 = 0$	$0 \times 0 = 0$	$0 * 0 = 0$	$0 * 0 = 0$
$\Sigma = -1$	$\Sigma = -1$	$\Sigma = -1$	$\Sigma\ 0$	$\Sigma\ -1$

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

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

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

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

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