$$D = \{(1/2, 3), (0, 3/2), (-1, 1)\}$$

- Given the above dataset, want to fit a linear model to the data: $y(x) = w_0 + w_1 x$

  of degree 1

- What are the model parameters $\vec{w} = (w_0, w_1)$?

- Solution (Stochastic Gradient Descent)

Step 0: Initialize $\vec{w}^{(0)} = (0, 0)$

- process $(1/2, 3)$ to update the weight vector
- Error function $E(\vec{w}) = [3 - (w_0 + 1/2\, w_1)]^2$

$$\Rightarrow \nabla E(\vec{w}) = 2(3 - w_0 - 1/2\, w_1) \begin{bmatrix} -1 \\ -1/2 \end{bmatrix}$$

Step 1: if $(w_0, w_1) = (0, 0) \Rightarrow \nabla E(\vec{w})\Big|_{\vec{w}=(0,0)} = 2(3) \begin{bmatrix} -1 \\ -1/2 \end{bmatrix} = \begin{bmatrix} -6 \\ -3 \end{bmatrix}$

- Now update the weight vector:

$$\vec{w}^{(1)} = \vec{w}^{(0)} - \eta \nabla E(\vec{w})\Big|_{\vec{w}^{(0)}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 1 \times \begin{bmatrix} -6 \\ -3 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}$$

- We have assumed the learning rate is set to 1 but it can be other values.

- process $(0, 3/2)$ to get updated weight vector

- Error function:

$$E(\vec{w}) = \left[3/2 - (w_0 + w_1 \times 0)\right]^2 = (3/2 - w_0)^2$$

$$\Rightarrow \nabla E(\vec{w}) = \begin{bmatrix} \frac{\partial E(\vec{w})}{\partial w_0} \\ \frac{\partial E(\vec{w})}{\partial w_1} \end{bmatrix} = 2(3/2 - w_0) \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

Step 2

$$\Rightarrow \nabla E(\vec{w})\Big|_{\vec{w}^{(1)} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}} = 2 \times (3/2 - 6) \begin{bmatrix} -1 \\ 0 \end{bmatrix} = (3 - 12) \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 9 \\ 0 \end{bmatrix}$$

- Now update the weight vector:

$$\vec{w}^{(2)} = \vec{w}^{(1)} - \eta \nabla E(\vec{w})\Big|_{\vec{w}^{(1)}}$$

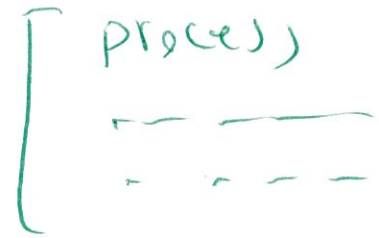
$$= \begin{bmatrix} 6 \\ 3 \end{bmatrix} - 1 \times \begin{bmatrix} 9 \\ 0 \end{bmatrix} = \begin{bmatrix} 6-9 \\ 3-0 \end{bmatrix} = \begin{bmatrix} -3 \\ 3 \end{bmatrix}$$

Note that the learning rate η was set to 1.

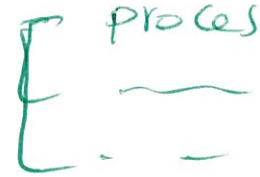
Step 3 — Process $(-1, 1)$ to update the weights

Step 4 — Process $(1/2, 3)$ to update the weights

Step 5 — Process $(0, 3/2)$ to update the weights



Step 6 — Process $(-1, 1)$ to update the weights



Step 7 — Process $(1/2, 3)$ to update weights

until convergence ~~(i.e. weights do not change)~~