$$\nabla f(x,y) = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right]^T$$

$$f(x_1, x_2, \ldots, x_n) = \nabla f = \left[\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \ldots, \frac{\partial}{\partial x_n}\right]$$

$$Err = \frac{1}{N}\sum_{\mu}(\underbrace{y^{\mu}}_{\text{Ground truth}} - \underbrace{(W_1 X^{\mu} + W_0)}_{\text{Predicted values}})^2$$

MSE

With this error signal, we can calculate the partial derivatives of the error and get the Gradient.

$$\frac{\partial}{\partial W_0} = -\frac{2}{N}\sum_{\mu}(y^{\mu} - (W_1 X^{\mu} + W_0))$$

$$\frac{\partial}{\partial W_1} = -\frac{2}{N}\sum_{\mu} X \times (y^{\mu} - (W_1 \times X^{\mu} + W_0))$$

$$\nabla Err = \left[\frac{\partial}{\partial W_0}, \frac{\partial}{\partial W_1}\right]^T$$

Where Err is the MSE Error Function

We want to update the weights so they can push the error down in the next iteration.We need to make them follow the opposite direction of each respective gradient signal.

$$W0 = W0 - \eta(\frac{\partial}{\partial W0})$$
$$W1 = W1 - \eta(\frac{\partial}{\partial W1})$$

step size η is the learning rate