Gradient Descent: The Code

From before we saw that one weight update can be calculated as:

$$\Delta w_i = \eta \, \delta x_i$$

with the error term δ as

$$\delta = (y - \hat{y}) f'(h) = (y - \hat{y}) f'\left(\sum w_i x_i\right)$$

Remember, in the above equation $(y - \hat{y})$ is the output error, and $f'(h)$ refers to the derivative of the activation function, $f(h)$. We'll call that derivative the output gradient.

Now I'll write this out in code for the case of only one output unit. We'll also be using the sigmoid as the activation function $f(h)$.

```python
# Defining the sigmoid function for activations
def sigmoid(x):
    return 1/(1+np.exp(-x))
```