

1 Activation Functions

1.1 Linear

$$\hat{y} = mx + b \quad (1.1.1)$$

The functional form of equation 1.1.1 is shown by figure 1.

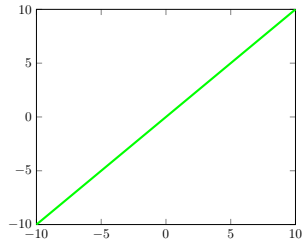


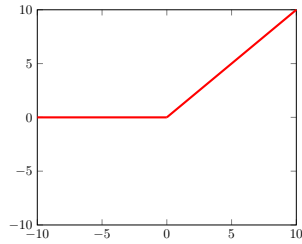
Figure 1: *Linear*.

1.2 ReLU

Rectified Linear Unit (ReLU) is defined by:

$$\hat{y} = \begin{cases} wx + b & \text{if } wx + b > 0 \\ 0 & \text{Otherwise.} \end{cases} \quad (1.2.1)$$

From this we see that negative outputs are set to 0. Using ReLU in hidden layers adds efficient non-linearity.



1.3 Sigmoid

Using a Sigmoid function at the end/output generates a probability for each digit (in the case of mnist data). Using the *softmax* function causes all sigmoid probabilities to be re-weighted, such that the sum to unity.

$$\hat{y} = \frac{1}{1 + e^{-(wx+b)}} \quad (1.3.1)$$

