**Neural networks**

**Softmax:**

Softmax is the function used to generate a probability function over all possible outputs in a classifier. For example, using the MNIST data, a Softmax output will have ten neurons each corresponding to the output number from 0 -9. The value of the neuron will be the probability that the input image is the corresponding number. This is done using the function

.

Here  is the output value of the th neuron in the Softmax output layer and is the sum of all the weights and bias input to the th neuron. It is clear to see from this equation that the sum over  of  will give a value of 1 like a probability distribution should.

**Cross Entropy Cost**

Cross entropy cost is one way in which a neural network can evaluate how bad the predicted probability distribution of a model  is compared to the true distribution . Using the MNIST example again,  will be the probability of the input image being the number  and  will be the one hot vector with a 1 at the index of the true value. The cross-entropy cost function is defined by the equation



Which will inform the network of how bad the prediction was and is the value the network tries to minimise during training. It is worth noting that if the network predicts an element to be zero, then the loss tends to infinity. It is also worth noting the minimum value of the loss is 0 so is always positive.