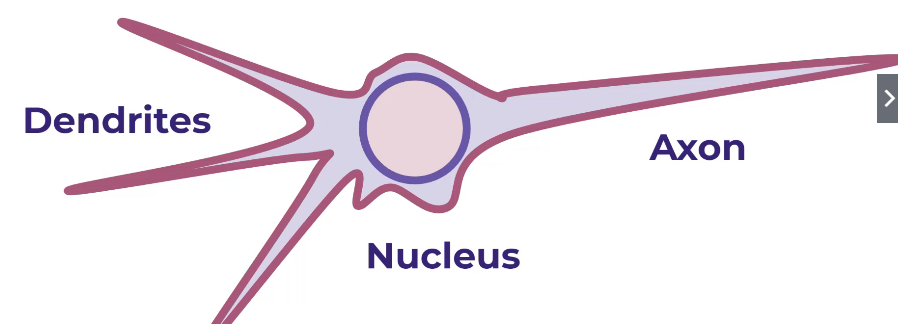
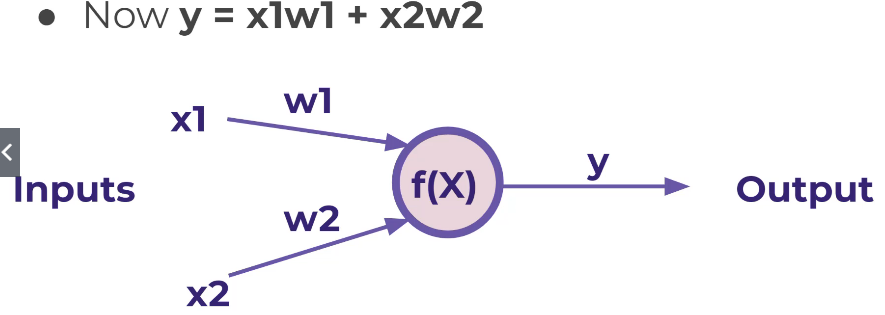
## Perceptron Model

### What are the important parts of the neuron you need to understand?



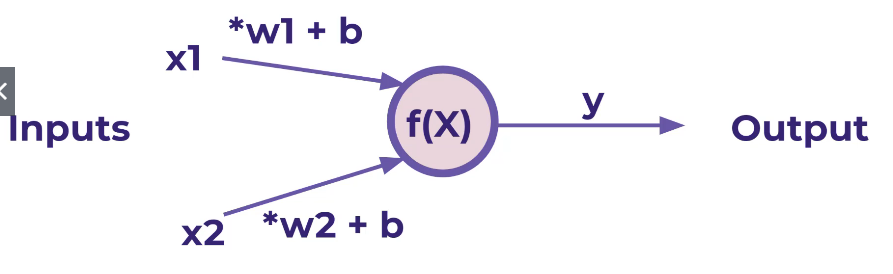
### What a perceptron model?

It was a form of neural network

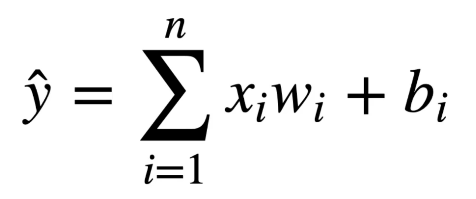


### What happens if x1 is 0?

Add a bias to the weights

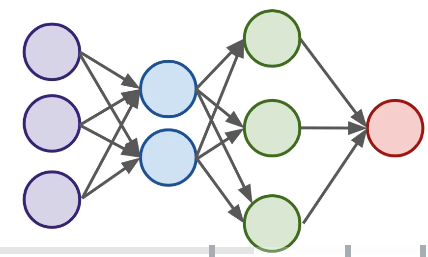


### What is the equation for the output value?

  
Just the sum of input values multiplied by the weights and addition of the bias.

## Neutral Networks

### What is multi-layer perceptron model?



This example is a fully connected model.

### What are the 3 types of layers in a neural network?

Input, hidden and output

### When is a neural network a deep network?

When there are 2 or more hidden layers

## Activation Functions

### What is an activation function?

Set boundaries so that the summation of the input values, weight and biases can be changed to a specific range.

### What gets passed through the activation function?



The input value multiplied by the weight and plus the bias. So here, z would get passed into the activation function.

## Multi-Class Classification Considerations

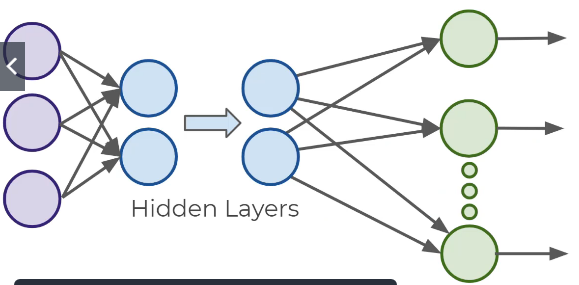
### What are the 2 main types of multi-class situations?

Non-exclusive – data point can have multiple class

Mutually exclusive – a data point can only have one class/category assigned to it.

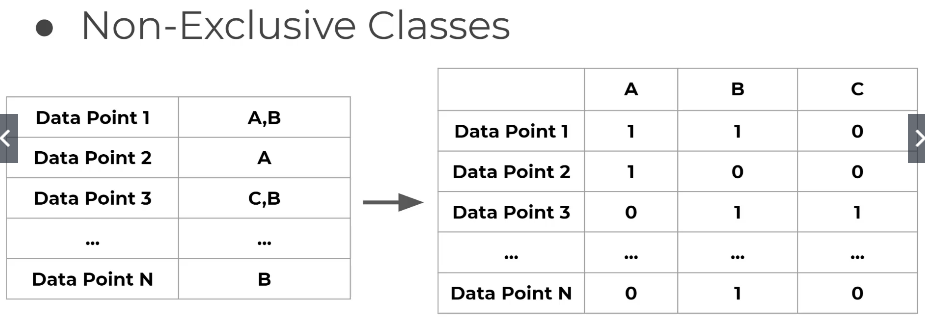
### How can we have a neural network handle multiple classes?

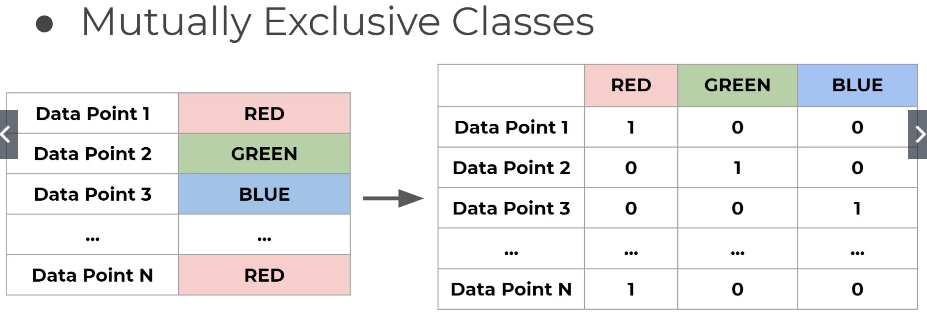
Have a output neuron for each class



### How do we transform our classes so that it becomes variables?

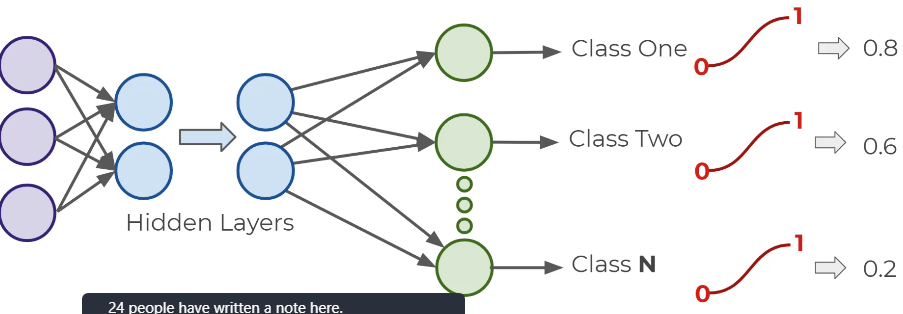
One-hot encoding or creating dummy variables.





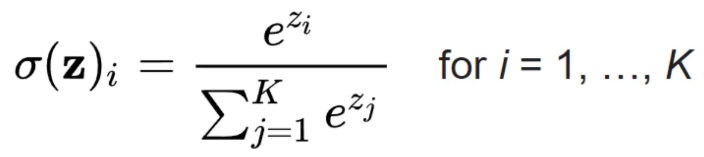
### Which activation would you use if dealing with non-exclusive classes?

Sigmoid function

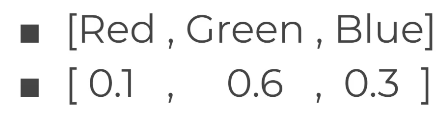


### What activation function is used for mutually exclusive classes?

Softmax function, where K is the categories



The sum of all the probabilities is equal to 1. The target class is chosen based on the highest probability.



## Cost Functions and Gradient Descent

### What are other terms for the cost function?

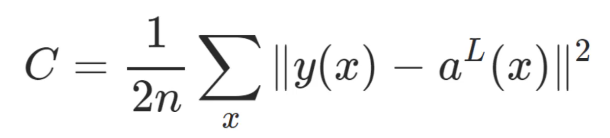
Loss functions or error functions

### What is the point of the cost function?

To determine how well our output value of the model matches with the real value. Which should be an average.

### What is a common cost function?

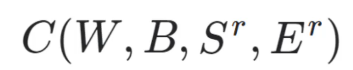
Quadratic function



L is the last layer

a is the weights and biases and the activation function, so basically the output of a neuron

### What are the 4 main functions of the cost function?



W is weights

B is biases

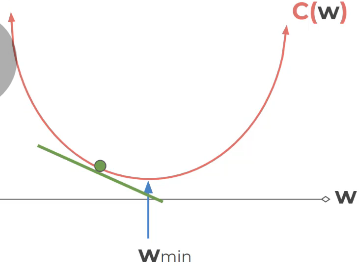
Sr is the input of a single training sample

Er is the desired output of that training sample

So we want to change the weights and biases to minimize the cost function

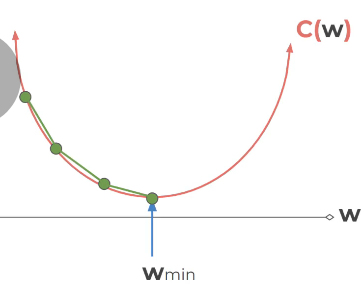
### How is the cost function minimised?

We use stochastic gradient descent. We see what the cost function is at a particular weight and bias and find the gradient and move downwards in the direction of the slope until you converge to the minimum.



### What is the learning rate?

The step size of where the model tests for the next minimum cost function.



### What is adaptive gradient descent?

This is where the step size is changed depending how close you are to the minimum.

### What is the most common adaptive gradient descent optimization?

Adam.

### What is the symbol used for gradient in equations?



Upside down triangle.

### What cost function is used for classification problems?

Cross entropy

## Backpropagation

### What is backpropagation?

Using the cost function, the weight and biases are updated throughout the network.

### How do we label the layers?



Last layer is L

## Tensorflow vs Keras