In next lecture (Step 3 - Building the Model) we have taken below values to add the first fully connected hidden layer,

units=128, activation='relu', input_shape=(784,))

Let us understand why we have these values,

1) units (No of neurons) = 128

The artificial neuron receives one or more inputs and sums them to produce an output, then this output is passed throught the activation function

Now there are some rules to select the number of neurons but every dataset is different, every problem is different, So a set of rules may not work properly

To select an appropriate number of units (neurons), we have to try different number of neurons and select number of neurons that works better for the model performance

We have to use systematic experimentation to discover what works best for a specific dataset

Here i have selected number on neurons=128 because it works better, and this is on the basis of various experiments

You can try different number of neurons at different layer and check the result, its fun !

2) activation='relu'

The activation function calculates the weighted sum of its input, adds a bias and then decides wheather a neuron should be activated or not

In simple word we can say that, 'An activation function decides a signal should be passed forward or not'

The activation function adds non linearity to the output of a neuron

With the non linear transformation a neural network is capable of learning and performing more complex tasks.

We are using 'relu' function in first layer because,

It is Non-linear, easily backpropagates the errors and multiple layers of neuron being activated by ReLU function

As well as, it is computationally less expensive than tanh and sigmoid functions, RELU learns much faster than sigmoid and Tanh function

3) input_shape=(784,)

In previous lecture we flattened the dataset, before flattening the dataset shape for x_train is (60000, 28, 28) and shape for x_text is (10000, 28, 28)

60000 and 1000 are the number of images (batch size)

After flattening the dataset shape for x_train is (60000, 784) and shape for x_text is (10000, 784)

60000 and 1000 are the number of images (batch size)

The first dimension is the batch size, it's None because it can vary depending on how many examples we give for training

hence input shape is (784,)

Hope you have understood the reasons behing the selection of parameters

Regards,

Vijay Gadhave