Homework 3

Start with the file `Example_MNIST_data.ipynb` on blackboard.
Remove the Section *Manipulate the Data* from this file.
Note in this file that validation is not used. You should do it this way as well.
You are to use only three layers (input, hidden, and output)
You are only allowed *five* epochs.
No more than 512 hidden neurons.
Give the accuracy during training, i.e., use `metrics=['accuracy']`

Use the `glorot_normal` weight initializer for the weights in each layer (see explanation on next page)
To use this weight initializer, in the section *Network Architecture* you will need to add

```
from tensorflow.keras import initializers
```

and for **each** layer you will need to add something like

```
network.add(layers.Dense(512, kernel_initializer='glorot_normal',
                    activation='relu', input_shape=(28 * 28,)))
```

See if you can improve the **test accuracy** of the above example using a **regularizer** and/or **dropout** and

In the section *Network Architecture* you will need to add

```
from keras import regularizers
from tensorflow.keras import initializers
```

The command for a quadratic regularizer (`l2`) with `lmda = 0.001` is

```
kernel_regularizer=regularizers.l2(0.001)
```

(See page 108 of Francois Chollet's book where to put this code)

The command for a dropout of 0.5 is

```
network.add(layers.Dropout(0,5))
```

(See page 110 of Francois Chollet's book where to put this code)

```
network.add(layers.Dense(512, kernel_regularizer=regularizers.l2(0.001),
                    kernel_initializer='glorot_uniform',
                    activation='relu', input_shape=(28 * 28,)))
network.add(layers.Dropout(0,5))
```

Hand in your `.ipynb` file
Hand in a page (and not more than one page) with the information about your network:

- number of layers with the number of neurons for each layer.
- Activation functions used on each layer.
- Weight initializer used for each layer.
- learning rate value (if used).
- Mini-batch size used.
- Optimizer used.
- Loss function used.
- dropout value used.
- lmda value for the regularizer.
- Give the final **test** accuracy.

This HW was designed to be done by just making some minor modifications to the original `.ipynb` file above.

Glorot normal initializer

Fully connected (dense) set of neurons from the $\ell - 1$ layer to the ℓ layer with weight matrix (tensor)

$$W \in \mathbb{R}^{n_\ell \times n_{\ell-1}}.$$

Each weight w_{ij} is initialized using an independent sample from a normal distribution $N(\mu; \sigma)$ where $\mu = 0$, $\sigma = \sqrt{\dfrac{2}{n_\ell + n_{\ell-1}}}$. This is similar to the `default_weight_initializer` in Nielsen's book.

The bias $b \in \mathbb{R}^{n_\ell}$ is initialized to 0.