Hyper-parameters:

The data is already splitted. For the input and hidden layer, we use the RELU as an activation function, and the last layer uses the SOFTMAX. The used optimizer is ADAM, and for loss function, my choice was sparse_categorical_crossentropy. The metric is accuracy. The number of epochs is 10.

Model:

The model is Sequential. I used 4 layers. The first one is the input layer. The second and third layers are the hidden layers. Each of these layers consists of 128 neurons. I added a dropout layer before the output with 0.25 probability. The last layer is the output, and the number of neurons is 10.

Results:

The achieved accuracy for trained data is 99.7%. The accuracy for test data is 98.2%.

The changes:

I added 2 more layers to the model with the RELU activation function. I added a dropout function which would increase the variation of the input data while training the data because it will drop some neurons in some epochs. The ADAM optimizer leads to a higher result in comparison with the SGD optimizer. I increased the number of epochs to give the model the chance to go over the network more than one time and learn from the previous mistakes.