

## **Theoretical questions**

### **PART 1: Fully connected layers**

#### **Task 1- NN with fully connected layers**

Loss value: 0.8021

Accuracy value: 0.6743

#### **Task 2- Activation functions**

The ReLU activation function allows fast convergence of the network, have a depth and allows backpropagation to learn, but when we have input that is negative or close to zero the gradient of the function tends changes to zero and the learning process stops.

However, the tanh function (non-linear) has the effect of centering the data at zero and this makes learning easier for the next layer. But it is subject to the ‘vanishing gradient’ issue (because the output runs between -1 and 1) with the optimization process the gradient goes smaller every time and the convergence goes slower.

#### **Task 3- Number of epochs**

Increasing the number of epochs, improves the model's performance. On the other hand, it required longer time and higher computational power required.

##### **25 epochs:**

Loss value: 1.0648

Accuracy value: 0.5714

##### **40 epochs:**

Loss value: 0.9677

Accuracy value: 0.6171

Both performance measures have improved with the larger number of epochs.

#### **Task 4- Mini-batches**

Minibatch training can be faster than training on single data points because it can take advantage of vectorized operations to process the entire minibatch at once. The stochastic nature of online/minibatch training can also make it possible to hop out of local minima that might otherwise trap batch training.

It is also more stable, contains less noise and its convergence is more accurate.

#### Task 4-Batch normalization

	loss	accuracy
new	0.8574	0.60
previous	0.8463	0.6229

It seems that the batch normalization did not improve the model's performance in this case.

### **Part 2: Convolutional Neural Network (CNN)**

#### Task 1- 2D CNN

- There are 8 layers.
- The first layer contains 64 filters, the second & third layers contains 128 filters, the fourth and fifth layers contains 256 layers and the last 3 layers does not include any filters.
- Conv2D CNN model uses max pooling (part of the filters) so the number of parameters will be lower than fully connected NN since there are much more parameters to estimate
- Yes, as it shows at the {kernel\_regularizer} argument, it means that the NN performs regularization of type "L2".

#### Task 2- Number of filters

Before: Test Loss: 7.817 - Test Accuracy : 0.3085

After (filters reduced by half): Test Loss: 8.1557 - Test Accuracy : 0.3085

The results were not as we expected ,they have not improved with the reduced number of filters. This may be caused due to bad data so that the filters number will not have a big influence or a neural network that is not suitable for this classification.