

# DL Assignment – I

Om Vaknalli (18376)

## Part One

### For >300,000 trainable parameters

1. There are 3810 biases and 298300 weights.
2. The validation accuracy was 0.9667 while the training accuracy was 0.9884 at the end of 10 epochs. The accuracy for training data increased while the validation accuracy remained 0.9667 throughout. Hence, the model is **overfitting**.
3. A time of 3 seconds and 6 milliseconds was taken by each epoch to execute.

### For < 50,000 parameters

1. There are 48670 weights and 630 biases.
2. The model is a perfect one since the validation accuracies is actually higher than the training accuracies which means the model generalises pretty well and the validation accuracies actually rose after subsequent executions of the epochs.
3. Each epoch took a time of 1 second and 3 milliseconds to execute.

## Part Two

1. None of the models achieved accuracy score of more than 0.2. All of them had an accuracy score of 0.1050. The reason for the low accuracy score is the vanishing gradient problem which shows up when there are multiple neuron layers in the model. Since the sigmoid activation function has the derivative that approaches 0, this means that the corrections applied based on the loss function can become very small—causing the node to effectively stop learning. This is especially true of neural networks with several layers, because the vanishing gradient problem stacks across the layers.
2. The first model has 8840 trainable parameters, the second has 19,599 and the third model has 53,759 trainable parameters.
3. Usually, the accuracies increase with the increase in number of trainable parameters, however there's a risk of overfitting with a large number of parameters while training.

## Part Three

The maximum accuracy achieved was 91.12% after running the model with 3 hidden layers and 380 nodes per hidden layer. The model was run for 390 epochs.

If we compare this model with our earlier model which was run for just 10 epochs and had accuracies of over 93%, we could say our current model is computationally expensive as compared to the previous models and not that accurate either.