HW4 – Report

Achyuthan Unni Krishnan, Kishor Sabarish

1. Section 1

The best result was observed to be 88.58% accuracy and the hyperparameters corresponding to this result was batch size of 16, 40 epochs, 3 hidden layers, 75 hidden units and 0.1 learning rate. The regulatization term was 0. It is to be noted that it was difficult to get good accuracies with higher regularization strengths. This might be a shortcoming in the code. The training snippet for the best hyper parameters is shown in the figure below.

```
epoch 16
current_loss= 0.36529861124353125
epoch 17
current_loss= 0.3503525163323803
epoch 18
current_loss= 0.34827361415000413
epoch 19
current_loss= 0.379138163540209
epoch 20
current_loss= 0.3765333443521579
epoch 21
current_loss= 0.383646191481485
epoch 22
current_loss= 0.3703018101531092
epoch 23
current_loss= 0.3680383047849805
epoch 24
current_loss= 0.39039843873421065
epoch 25
current_loss= 0.3738841656213189
epoch 26
current_loss= 0.40692092977440947
epoch 27
current_loss= 0.38405195147157445
epoch 28
current_loss= 0.3959290145122432
epoch 29
current_loss= 0.42813271815347936
epoch 30
current_loss= 0.3775051764538078
epoch 31
current_loss= 0.3842870822510857
epoch 32
current_loss= 0.39655154109993435
epoch 33
current_loss= 0.4021407495066881
epoch 34
current_loss= 0.431568876994918
epoch 35
current_loss= 0.4048582547461025
epoch 36
current_loss= 0.38075957616265904
epoch 37
current_loss= 0.4390575440523507
epoch 38
current_loss= 0.44121174994104193
epoch 39
current_loss= 0.39623116199436453
current_accuracy= 88.58
current best= combination= 20 Batch_size= 16 learning rate= 0.1 epochs= 40 best accuracy= 88.58 n_layers= 3 n_hidden= 75
```

Fig 1: The losses for the training of the model using the best hyperparameters

2. Section 2

The following plot of the cross entropy loss as a function of the neural network parameters were performed on 3500 randomly gathered datapoints from the training set. 3000 variants of the Weights and bias values were gathered in the trajectory list during training for the model that provided best results as mentioned above. Below is the corresponding surface plot and the descent of loss value.

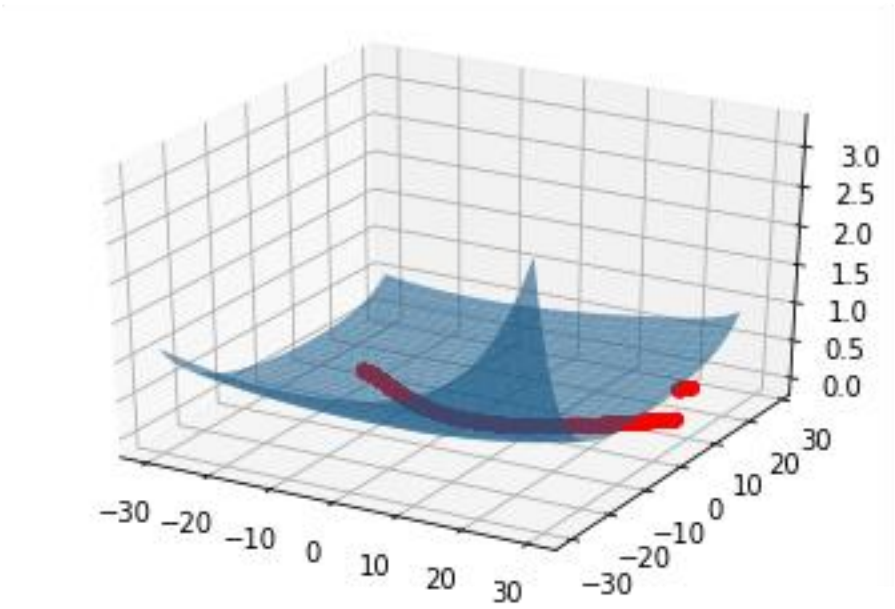

Fig 2: Cross entropy loss as a function of system parameters

For sub-optimal hyperparameters the descent in loss is more obvious as seen in the image below. For 20 epochs, 0.2 learning rate, 75 hidden units and 3 hidden layers we get the more obvious descent of loss.

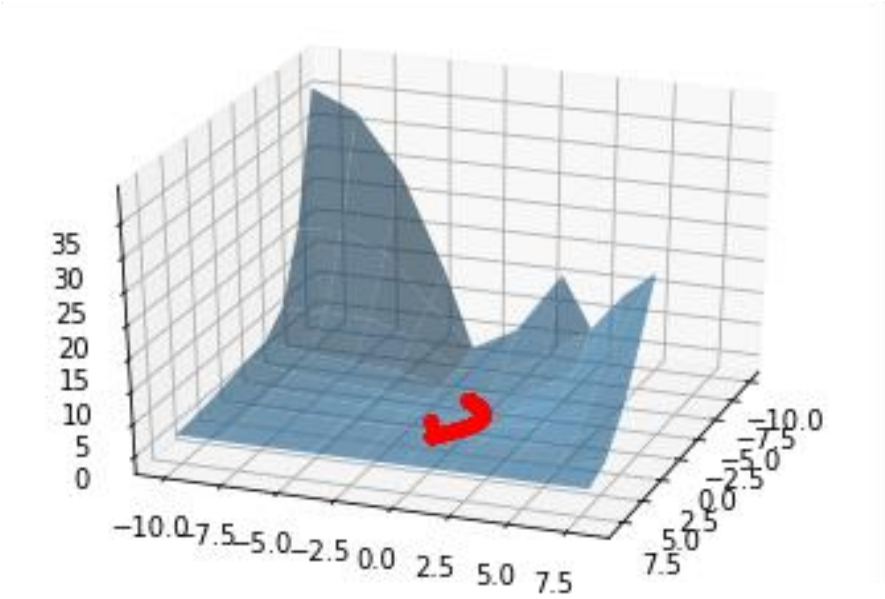

Fig 3: Cross entropy loss as a function of system parameters for suboptimal hyperparameters