Question 1. A simple neural network in tensorflow.

(a) After the introduction of a hidden layer to the linear regression model (please refer to Homework 4), we have

$$\hat{y} = \boldsymbol{S}\left(f\left(\boldsymbol{x}\boldsymbol{W_1} + \boldsymbol{b_1}\right)\boldsymbol{W_2} + \boldsymbol{b_2}\right)$$

where $f(x) = max(0, x)$, i.e. ReLU.

Setting the size of the hidden layer to 50 units, we obtain the results that are shown in Figure 1.

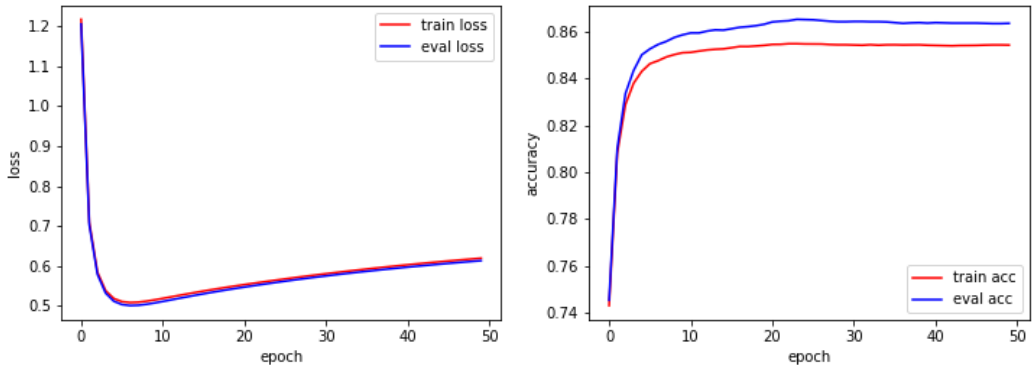

Figure 1: Loss and accuracy plots for a neural network with a single 50-unit hidden layer. The network was trained over 50 epochs. The achieved test loss and accuracy were 0.575 and 0.864, respectively.

(b) Number of units in the hidden layer was changed to 20 and 70 and the model was evaluated as in (a). The results appear in Figures 2 and 3, respectively.

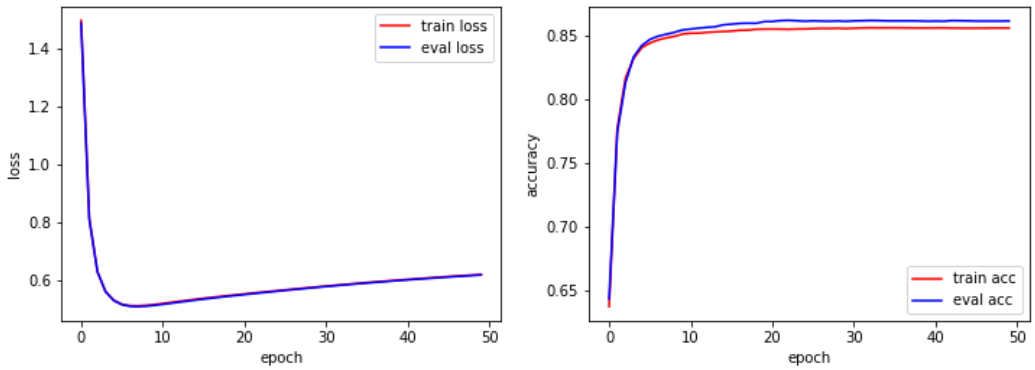

Figure 2: Loss and accuracy plots for a neural network with a single 20-unit hidden layer. The network was trained over 50 epochs. The achieved test loss and accuracy were 0.584 and 0.864, respectively.

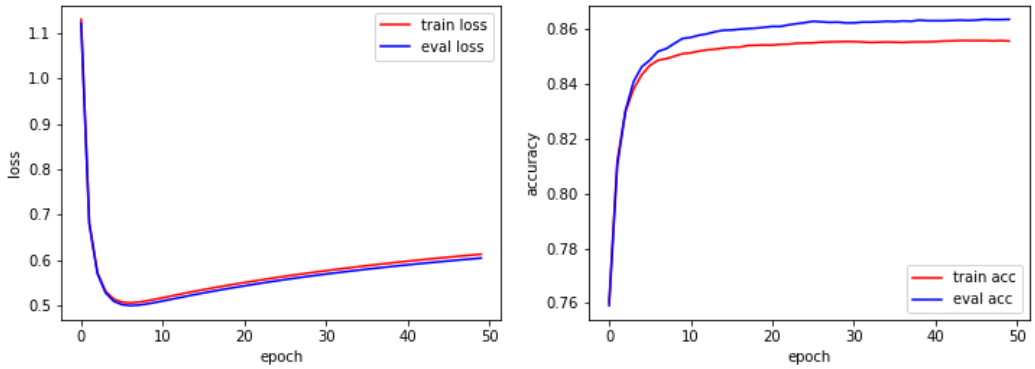

Figure 3: Loss and accuracy plots for a neural network with a single 70-unit hidden layer. The network was trained over 50 epochs. The achieved test loss and accuracy were 0.569 and 0.864, respectively.

(c) Next, a second hidden layer was introduced and the model performance was evaluated for different layer sizes. The results for the best model appear in Figure 4.

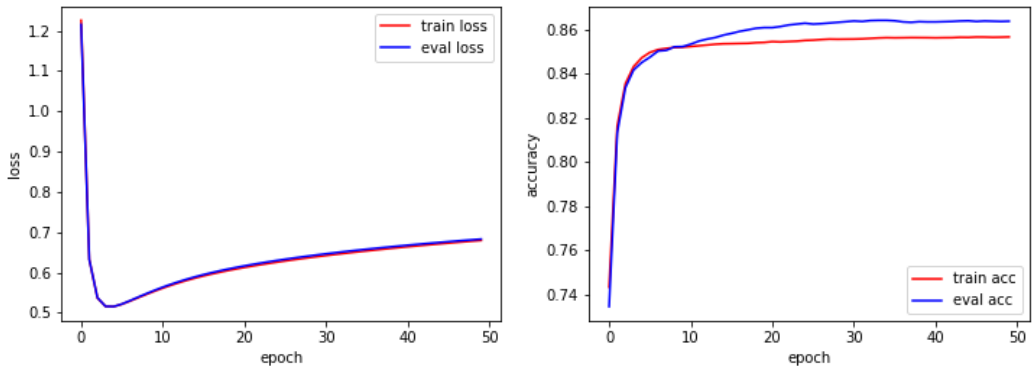

Figure 4: Loss and accuracy plots for a neural network with two hidden layers (300 and 100 units, repspectively). The network was trained over 50 epochs. The achieved test loss and accuracy were 0.635 and 0.865, respectively.

(d) Now, we evaluate the performance of a model with 3 hidden layers containing 500, 300, and 100 units, respectively. The results are shown in Figure 5. A model with the same architecture was trained with a smaller learning rate (0.001 instead of 0.01), and the results can be seen in Figure 6.

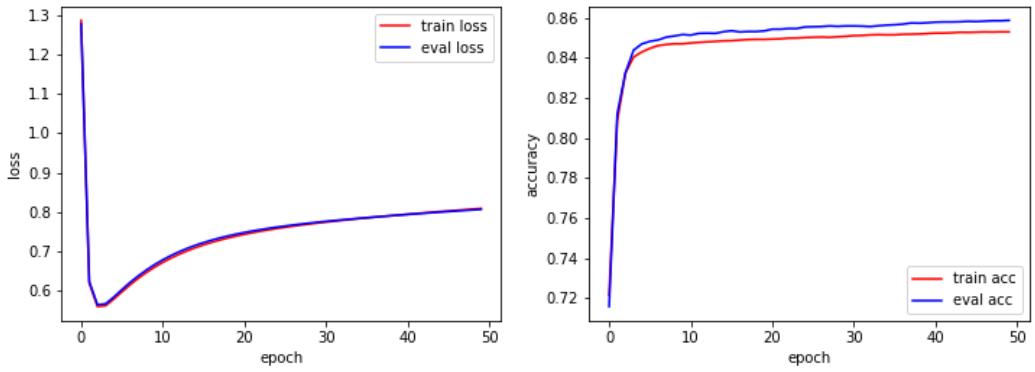

Figure 5: Loss and accuracy plots for a neural network with three hidden layers (500, 300, and 100 units, repspectively). The network was trained over 50 epochs. The achieved test loss and accuracy were 0.765 and 0.862, respectively.

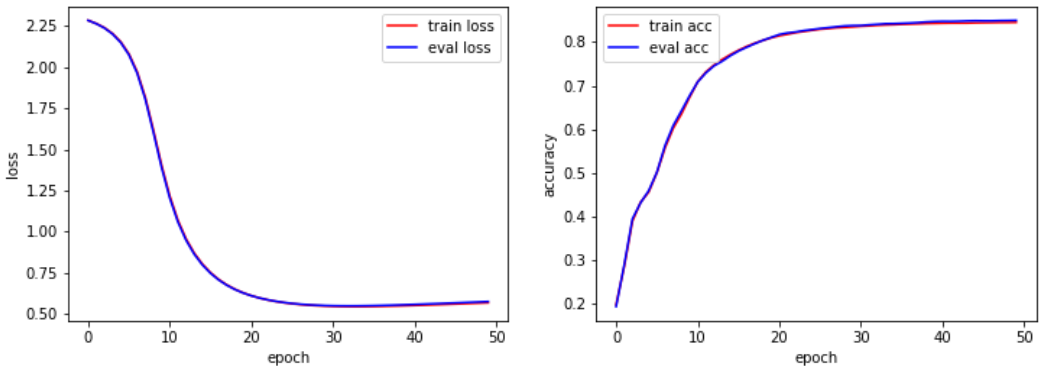

Figure 6: Same as in Figure 5 but with a smaller learning rate (0.001 instead of 0.01). The achieved test loss and accuracy were 0.544 and 0.855, respectively.