### Problem 1: Neural Networks and Backpropagation

**1.1** Write a paragraph describing your design choices. In particular, specify all your parameter choices: your learning rate (or learning rate scheme), mini-batch size, initialization scheme.

**1.2** For each of the 6 networks trained (5 different values of nh and the ReLU network), plot the squared loss after every half epoch (starting with your initial squared error). Please label your axes in a readable way. Plot the loss of both the training and test losses on the same plot.

**1.3** Do the same as **1.2** for the 0/1 loss (i.e. 1 - accuracy), but this time start when the loss is below 7% (or when 23 of epochs have elapsed, whichever comes first) to make to plot more readable.

**1.4** What is your final squared loss and 0/1 loss for both the training and test sets for each network?

**1.5** How does using ReLUs compare to using the sigmoid function? Why?

**Problem 2: Ensemble Methods**

Consider an ensemble learning algorithm that uses simple majority voting among K learned hypotheses. Suppose that each hypothesis has error ϵ and that the errors made by each hypothesis are independent of the others'.

**2.1** Calculate a formula for the error of the ensemble algorithm in terms of K and ϵ, and evaluate it for the case where K = 5, 10 and 20 and ϵ = 0.1, 0.2 and 0.4.

The ensemble makes error when at least half the learned hypothesis make an error simultaneously. Probability that **x** learned hypothesis make an error from K :

P(x learned hypothesis make an error) =

If K is even, P(total error)

If K is odd, P(total error)

|  |  |  |  |
| --- | --- | --- | --- |
| K | 5 | 10 | 20 |
| ϵ = 0.1 | 0.00856 | 0.0016349374 | 4.12766449323329E-10 |
| ϵ = 0.2 | 0.05792 | 0.0327934976 | 1.55372226246155E-07 |
| ϵ = 0.4 | 0.31744 | 0.3668967424 | 0.000188972696041127 |

**2.2** If the independence assumption is removed, is it possible for the ensemble error to be worse than ϵ? Justify you answer.

Yes, the ensemble error will be worse than ϵ. Consider K = 3 and ϵ = 2/5 = 0.4 and the ensemble predicts 5 examples as follows :

Example 1 : K1 and K2 have error and out-vote K3, therefore example 1 is predicted incorrectly.

Example 2 : K1 and K3 have error and out-vote K2, therefore example 2 is predicted incorrectly.

Example 3 : K3 and K2 have error and out-vote K1, therefore example 3 is predicted incorrectly.

Example 4,5 : The predictions are correct by all learned hypothesis.

It is seen that each hypothesis have made error 2/5 = 0.4. However, the ensemble has incorrectly predicted 3 out of 5, thus has error 3/5 = 0.6 > ϵ = 2/5 = 0.4