

Student 1: Mohammed Ibrahim
Student 2: Oluwatobi Oni-Orisan

PART 2

We modified the function that checked the values in the predicted set. If the value of every slot except the correct slot was less than 0.3 and the value in the correct slot was greater than 0.7 we declared that to be a "match". Else, we counted it as an error.

PART 3

Learning Rate	Number of Tests	Total probable Errors	Test Loss	Validation Loss
0.002	5000	5000	0.2632	0.2696
0.01	5000	5000	0.09939	0.0996
0.05	5000	4994	0.0868	0.0884
0.2	5000	4929	0.0752	0.0804
1.0	5000	3290	0.0372	0.0498
5.0	5000	1814	0.0078	0.0355
20.0	5000	1505	0.0038	0.0322

PART 4(Momentum)

Learning Rate	Number of Tests	Total probable Errors	Test Loss	Validation Loss
0.002	5000	5000	0.090634853	0.0924245
0.01	5000	4952	0.082565799	0.084867187
0.05	5000	4209	0.04762511	0.060104921
0.2	5000	2295	0.019010777	0.039789379
1.0	5000	2019	0.0086130574	0.036413942
5.0	5000	2308	0.029293643	0.050930604
20.0	5000	4564	0.092157245	0.092216231

Comments: We tweaked the learning rate (tried values greater than 1 and less than 2 and also between 2 and 3), the momentum (tried 8.0, 1.0 and 1.5), and the number of hidden layers (tried 100) and we concluded that the best learning rate we found was 1.0 (with the lowest error rate)

Fine Tuning: We kept testing our learning rate and we arrived at a final value of 1.4 with 10 hidden units and 0.94 momentum

PART 4(Generalization)

We added L2 regularization using the best parameters we had (10 hidden layers, learning rate of 1.4 and momentum of 0.94) and it did help. The number of errors, and loss were decreased. We were initially using our own definition of l2 regularization, but when we used the built-in function for regularization (`tf.nn.l2_loss()`), the errors and loss were decreased even further.

Without regularization, we got overfitting with at 30 hidden layers. There was a large deviation between the loss of the test set and that of the validation set.

Without regularization, we found the best hidden layer to be 10 which is the amount we were given from the start. Upon numerous test values (5,10, 15, 20, 25, 30, 35, 40,.....75) we discovered the best value to be to be 10.

With regularization we found the best hidden layer to be 17. We tested a bunch of values (10, 12, 15, 17, 38, 42, 47, 55) and 17 gave the best result with least amount of deviation between the loss of test and training sets, even when we tweaked the learning rate.

Best Settings:

- Hidden Layers: 17
- Learning Rate: 1.4

Test Data Classification Error: 0.00932