John Kenney jfk150030

Table:

max iterations = iter

training size = tr%

learning rate = lr

number of neurons in the hidden layer = neurons

training data set = tr

test data set = test

Mean Square Error = MSE

Accuracy = Acc

**function iter tr% lr neurons MSE tr MSE test Acc tr Acc test**

sigmoid 30 0.80 1e-05 2 0.20737 0.20823 0.79263 0.79177

sigmoid 30 0.80 1e-05 6 0.20681 0.21044 0.79319 0.78956

sigmoid 30 0.80 1e-05 10 0.03526 0.03544 0.96474 0.96456

sigmoid 200 0.80 1e-06 2 0.22296 0.22599 0.77704 0.77401

sigmoid 200 0.80 1e-07 2 0.20816 0.20505 0.79184 0.79495

sigmoid 200 0.80 1e-07 4 0.20756 0.20744 0.79244 0.79256

**sigmoid 200 0.80 1e-07 12 0.02883 0.02987 0.97117 0.97013**

sigmoid 200 0.80 1e-07 12 0.20844 0.20393 0.79156 0.79607

**sigmoid 200 0.80 1e-07 12 0.01923 0.01912 0.98077 0.98088**

sigmoid 1000 0.80 1e-08 4 0.20685 0.21030 0.79315 0.78970

tanh 100 0.80 1e-07 2 0.20749 0.20775 0.79251 0.79225

tanh 150 0.80 1e-08 5 0.71684 0.72188 0.28316 0.27812

tanh 200 0.80 1e-08 5 0.07032 0.07051 0.92968 0.92949

tanh 210 0.80 1e-08 5 0.10034 0.10081 0.89966 0.89919

tanh 200 0.80 1e-08 10 0.25177 0.25049 0.74823 0.74951

tanh 200 0.80 1e-07 5 0.08000 0.08096 0.92000 0.91904

**tanh 200 0.80 1e-07 12 0.03242 0.03387 0.96758 0.96613**

tanh 120 0.80 1e-07 12 0.06465 0.06564 0.93535 0.93436

tanh 120 0.80 1e-07 18 0.20776 0.20966 0.79224 0.79034

**function iter tr% lr neurons MSE tr MSE test Acc tr Acc test**

ReLu 100 0.80 1e-05 4 0.79011 0.78875 0.20989 0.21125

ReLu 100 0.80 1e-03 10 0.79245 0.79250 0.20755 0.20750

ReLu 150 0.80 1e-04 2 0.79172 0.79542 0.20828 0.20458

ReLu 150 0.80 1e-06 2 0.20780 0.20648 0.79220 0.79352

ReLu 150 0.80 1e-06 5 0.78689 0.78895 0.21311 0.21105

**ReLu 250 0.80 1e-07 5 0.06028 0.06127 0.93972 0.93873**

ReLu 250 0.80 1e-07 12 0.79136 0.79162 0.20864 0.20838

ReLu 250 0.80 1e-08 2 0.79147 0.79121 0.20853 0.20879

ReLu 200 0.80 1e-07 12 0.79274 0.79136 0.20726 0.20864

ReLu 80 0.80 1e-07 4 0.79184 0.79495 0.20816 0.20505

I found that using the activation function of sigmoid, max iterations = 200, training set = 0.80, learning rate = 1e-07, and hidden neurons = 12 gave me the smallest MSE and greatest accuracy value. The reason that these parameters gave the greatest accuracy is that a small learning rate gradually changed the weights to fit the data best, and that the number of neurons equal to 4 times the number of attributes gave a greater number of weights to calculate the best weights to get a more accurate result. In addition, sigmoid is the medium fit between tanh and ReLu giving a better fit for my data set. Lastly the number iterations used gave time for the parameters used to fit the weights to make accurate classifications of the data sets. I bolded the trials for each activation function that had the greatest accuracy in the table and highlighted the row that I specified in the previous sentence to differentiate it from the rest of the data. I am not to happy with the result because the results were not consistent with wide variance with the same parameters used sometimes. It is possible that this can be attributed to the dataset that was used.