HyperParameters

Setup:

In my setup, I have chosen gradient descent with early stopping. For early stopping, my patience counter is set to 10. Early stop works by checking if the validation loss has increased for the last 10 epochs; if so, it stops training. Each model is run 10 times, and their test accuracy mean, standard deviation and confidence interval are calculated.

After training all the models, the best model is retrained from scratch with data and validation as its data and test data as validation (for early stopping).

Hyperparameters and their Result:

Activation Function	Hidden Layers	Neurons	Learning Rate	Mean	Std	Confidence Interval	
Sigmoid	4	128	0.001	34.024	15.857	24.196	43.852
Sigmoid	4	128	0.0001	18.385	3.550	16.184	20.586
Sigmoid	4	256	0.001	52.787	14.633	43.717	61.857
Sigmoid	4	256	0.0001	27.605	9.204	21.900	33.310
Sigmoid	6	128	0.001	11.034	0.632	10.642	11.426
Sigmoid	6	128	0.0001	11.350	0.000	11.350	11.350
Sigmoid	6	256	0.001	13.500	0.000	11.350	11.350
Sigmoid	6	256	0.0001	11.350	0.000	11.350	11.350
Tanh	4	128	0.001	77.134	2.347	75.680	78.588
Tanh	4	128	0.0001	75.978	2.531	74.409	77.547
Tanh	**4**	**256**	**0.001**	**81.537**	**1.794**	**80.425**	**82.649**
Tanh	4	256	0.0001	81.151	1.167	80.428	81.874
Tanh	6	128	0.001	68.202	2.707	66.524	69.880
Tanh	6	128	0.0001	67.765	1.959	66.551	68.979
Tanh	6	256	0.001	76.767	2.214	75.395	78.139
Tanh	6	256	0.0001	77.157	1.769	76.061	78.253

The best Hyperparameter (bold) in the search was concluded as :

Activation Function: Tanh

Learning Rate: 0.001

Number of Hidden Layers: 4

Number of Neurons: 256

Test Result:

The best model is trained 10 times more, with data and validation combined, after which the test result are:

Mean: 91.73

Standard Deviation: 0.08

Confidence Interval: [91.67, 91.78]

Additional Questions:

– What type of measure or measures have you considered to prevent overfitting?

I have applied early stopping to prevent overfitting, which will stop training if the validation loss keeps on increasing for 10 epochs

– How could one understand that a model being trained starts to overfit?

If the validation loss keeps on increasing, it is a sign that the model is starting to overfit.

– Could we get rid of the search over the number of iterations (epochs) hyperparameter by setting it to a relatively high value and doing some additional work? What may this additional work be? (Hint: You can think of this question together with the first one.)

Yes, by applying early stop, we can reduce the number of epochs to train for.

– Is there a ”best” learning rate value that outperforms the other tested learning values in all hyperparameter configurations? (e.g it may always produce the smallest loss value and highest accuracy score among all of the tested hyperparameter configurations.).

Although not all of the time but 0.001 learning rate has given better results for most models.

– Is there a ”best” activation function that outperforms the other tested activation functions in all hyperparameter configurations? (e.g it may always produce the smallest loss value and highest accuracy score among all of the tested hyperparameter configurations.).

Yes, Tanh has always produced better results.

– What are the advantages and disadvantages of using a small learning rate?

Small learning rates can increase training time, and there is a possibility that they get stuck in a local minima since the steps are small, but it’s benefit is that it can train a model with higher accuracy.

– What are the advantages and disadvantages of using a big learning rate?

Big learning rates will take less time to learn, but they can jump over global minima (the best solution) and not provide the best result.

– Is it a good idea to use stochastic gradient descent learning with a very large dataset? What kind of problem or problems do you think could emerge?

Using stochastic gradient descent learning with a large dataset can be problematic since it relies on random jumps to avoid getting stuck in local minima. A large dataset can have many local minima, and the model will have to jump a lot of times, which will increase its training time.