

Gunnar Rätsch Benchmark Datasets

K S Ananth

Task 3

1 Overview:

Gunnar Rätsch's benchmark datasets are widely used in the machine learning community for evaluating algorithms. There are 13 datasets in total which are considered as the benchmark datasets. We shall use these datasets to evaluate neural networks and the back propagation algorithm. For doing so we shall train the neural network with 10 different instances of the dataset and figure the `test_set_error` of the trained model for all 10 instances and calculate the `average_test_error \pm standard_deviation`. The datasets are can be found at https://github.com/tdiethe/gunnar_raetsch_benchmark_datasets and is downloadable. All the datasets are in one single .mat file which can be extracted and used.

Some details before the start of the training and analysis: all the models use the mean square error function as the cost function (cross entropy needs data modification) and as all the datasets have output in the form of 1 or -1, Tanh function is used as the activation function for all layers and models.

2 Datasets:

2.1 Banana Dataset:

The banana dataset consists of 5300 total rows, from which 400 are considered as training set and the rest 4900 are considered as test set. The dataset has 2 features/inputs and one output which is either 1 or -1. The entire dataset here is shuffled for 10 times and the neural network is trained for each of the 10 times and the test error is calculated for each time. To analyse the algorithm we find the average test error and the standard deviation. The best performing neural architecture was of the form [2, 32, 24, 1], which implies 2 inputs, 32 neurons in 1st hidden layer, 24 neurons in 2nd hidden layer and 1 output. The training set loss vs epoch for one of the instance among the 10 is Figure: 1(a). The time elapsed for this training process is: 9.34375s. The learning rate used is: 0.1.

2.2 Breast Cancer Dataset:

The banana dataset consists of 277 total rows, from which 200 are considered as training set and the rest 77 are considered as test set. The dataset has 9 features/inputs and one output which is either 1 or -1. The entire dataset here is shuffled for 10 times and the neural network is trained for each of the 10 times and the test error is calculated for each time. To analyse the algorithm we find the average test error and the standard deviation. The best performing neural architecture was of the form [9, 27, 1], which implies 9 inputs, 27 neurons in hidden layer and 1 output. The training set loss vs epoch for one of the instance among the 10 is Figure: 1(b). The time elapsed for this training process is: 3.1875s. The learning rate used is: 0.2.

2.3 Diabetes Dataset:

The banana dataset consists of 768 total rows, from which 468 are considered as training

set and the rest 300 are considered as test set. The dataset has 9 features/inputs and one output which is either 1 or -1. The entire dataset here is shuffled for 10 times and the neural network is trained for each of the 10 times and the test error is calculated for each time. To analyse the algorithm we find the average test error and the standard deviation. The best performing neural architecture was of the form [8, 16, 1], which implies 8 inputs, 16 neurons in hidden layer and 1 output. The training set loss vs epoch for one of the instance among the 10 is Figure: 1(c). The time elapsed for this training process is: 7.4375s. The learning rate used is: 0.2.

2.4 German Dataset:

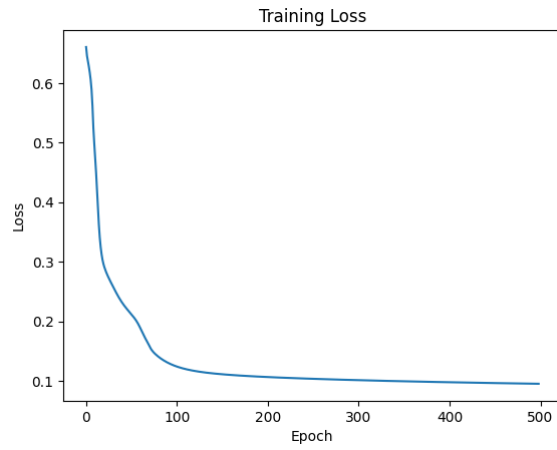
The banana dataset consists of 1000 total rows, from which 700 are considered as training set and the rest 300 are considered as test set. The dataset has 20 features/inputs and one output which is either 1 or -1. The entire dataset here is shuffled for 10 times and the neural network is trained for each of the 10 times and the test error is calculated for each time. To analyse the algorithm we find the average test error and the standard deviation. The best performing neural architecture was of the form [20, 16, 1], which implies 20 inputs, 16 neurons in hidden layer and 1 output. The training set loss vs epoch for one of the instance among the 10 is Figure: 1(d). The time elapsed for this training process is: 11.125s. The learning rate used is: 0.2.

2.5 Heart Dataset:

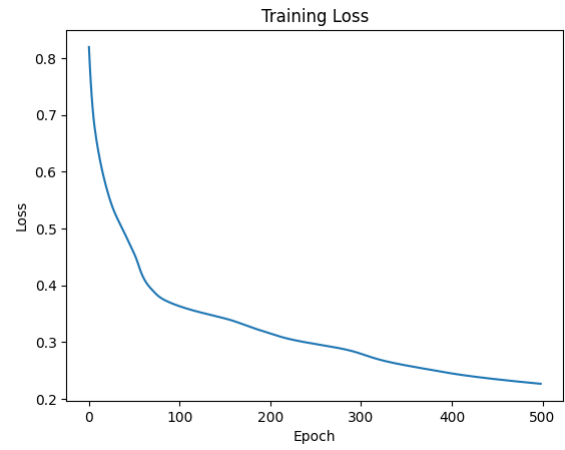
The banana dataset consists of 270 total rows, from which 170 are considered as training set and the rest 100 are considered as test set. The dataset has 13 features/inputs and one output which is either 1 or -1. The entire dataset here is shuffled for 10 times and the neural network is trained for each of the 10 times and the test error is calculated for each time. To analyse the algorithm we find the average test error and the standard deviation. The best performing neural architecture was of the form [13, 13, 1], which implies 13 inputs, 13 neurons in hidden layer and 1 output. The training set loss vs epoch for one of the instance among the 10 is Figure: 1(e). The time elapsed for this training process is: 0.21875s. The learning rate used is: 0.2.

2.6 Image Dataset:

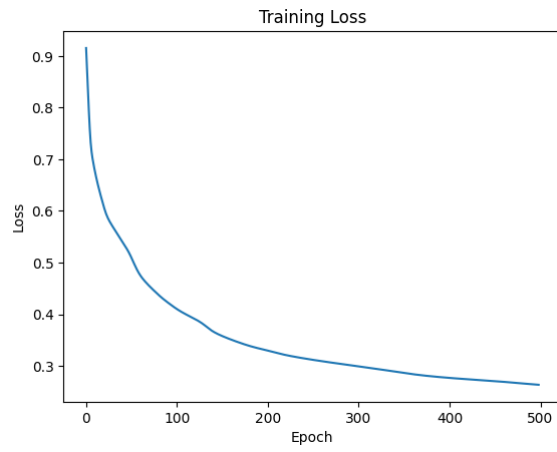
The banana dataset consists of 2310 total rows, from which 1300 are considered as training set and the rest 1010 are considered as test set. The dataset has 18 features/inputs and one output which is either 1 or -1. The entire dataset here is shuffled for 10 times and the neural network is trained for each of the 10 times and the test error is calculated for each time. To analyse the algorithm we find the average test error and the standard deviation. The best performing neural architecture was of the form [18, 18, 1], which implies 18 inputs, 18 neurons in hidden layer and 1 output. The training set loss vs epoch for one of the instance among the 10 is Figure: 1(f). The time elapsed for this training process is: 21.25s. The learning rate used is: 0.3.



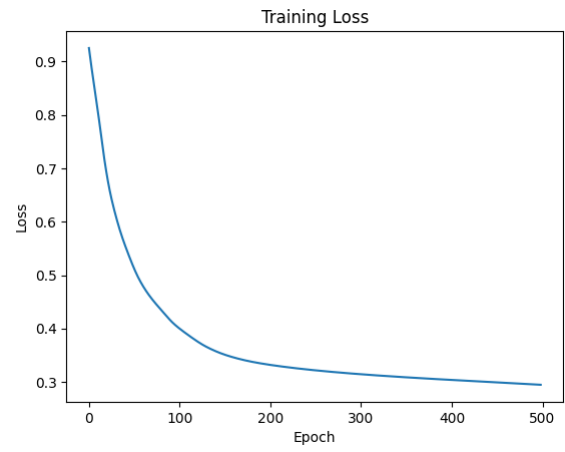
(a) Banana



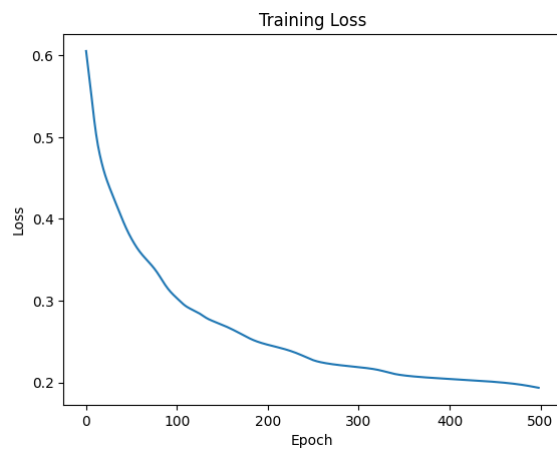
(b) Breast Cancer



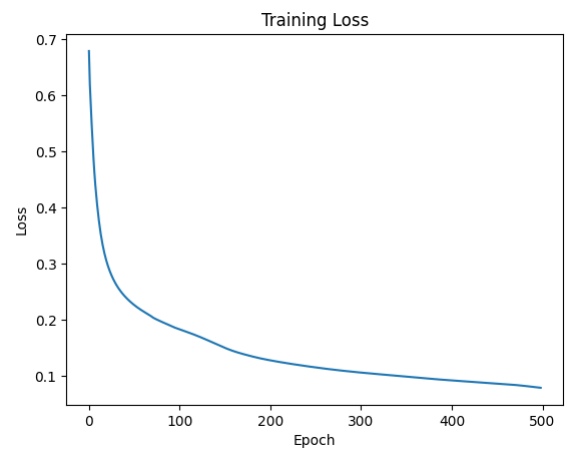
(c) Diabetes



(d) German



(e) Heart



(f) Image

Figure 1: Datasets

2.7 Ringnorm Dataset:

The banana dataset consists of 7400 total rows, from which 7000 are considered as training set and the rest 400 are considered as test set. The dataset has 20 features/inputs and one output which is either 1 or -1. The entire dataset here is shuffled for 10 times and the neural network is trained for each of the 10 times and the test error is calculated for each time. To analyse the algorithm we find the average test error and the standard deviation. The best performing neural architecture was of the form [20, 16, 1], which implies 20 inputs, 16 neurons in hidden layer and 1 output. The training set loss vs epoch for one of the instance among the 10 is Figure: 2(a). The time elapsed for this training process is: 3.9375s. The learning rate used is: 0.4.

2.8 Splice Dataset:

The banana dataset consists of 3175 total rows, from which 1000 are considered as training set and the rest 2175 are considered as test set. The dataset has 60 features/inputs and one output which is either 1 or -1. The entire dataset here is shuffled for 10 times and the neural network is trained for each of the 10 times and the test error is calculated for each time. To analyse the algorithm we find the average test error and the standard deviation. The best performing neural architecture was of the form [60, 12, 1], which implies 60 inputs, 12 neurons in hidden layer and 1 output. The training set loss vs epoch for one of the instance among the 10 is Figure: 2(b). The time elapsed for this training process is: 54.65625s. The learning rate used is: 0.4.

2.9 Thyroid Dataset:

The banana dataset consists of 215 total rows, from which 140 are considered as training set and the rest 75 are considered as test set. The dataset has 5 features/inputs and one output which is either 1 or -1. The entire dataset here is shuffled for 10 times and the neural network is trained for each of the 10 times and the test error is calculated for each time. To analyse the algorithm we find the average test error and the standard deviation. The best performing neural architecture was of the form [5, 15, 1], which implies 5 inputs, 15 neurons in hidden layer and 1 output. The training set loss vs epoch for one of the instance among the 10 is Figure: 2(c). The time elapsed for this training process is: 0.40625s. The learning rate used is: 0.1.

2.10 Titanic Dataset:

The banana dataset consists of 2201 total rows, from which 150 are considered as training set and the rest 2051 are considered as test set. The dataset has 3 features/inputs and one output which is either 1 or -1. The entire dataset here is shuffled for 10 times and the neural network is trained for each of the 10 times and the test error is calculated for each time. To analyse the algorithm we find the average test error and the standard deviation. The best performing neural architecture was of the form [3, 9, 1], which implies 3 inputs, 9 neurons in hidden layer and 1 output. The training set loss vs epoch for one of the instance among the 10 is Figure: 2(d). The time elapsed for this training process is: 2.0625s. The learning rate used is: 0.06.

2.11 Twonorm Dataset:

The banana dataset consists of 7400 total rows, from which 400 are considered as training set and the rest 7000 are considered as test set. The dataset has 20 features/inputs and one output which is either 1 or -1. The entire dataset here is shuffled for 10 times and the neural network is trained for each of the 10 times and the test error is calculated for each time. To analyse the algorithm we find the average test error and the standard deviation. The best performing neural architecture was of the form [20, 10, 1], which implies 20 inputs, 10 neurons in hidden layer and 1 output. The training set loss vs epoch for one

of the instance among the 10 is Figure: 2(e). The time elapsed for this training process is: 6.21875s. The learning rate used is: 0.3.

2.12 Waveform Dataset:

The banana dataset consists of 5000 total rows, from which 400 are considered as training set and the rest 4600 are considered as test set. The dataset has 21 features/inputs and one output which is either 1 or -1. The entire dataset here is shuffled for 10 times and the neural network is trained for each of the 10 times and the test error is calculated for each time. To analyse the algorithm we find the average test error and the standard deviation. The best performing neural architecture was of the form [21, 7, 1], which implies 21 inputs, 7 neurons in hidden layer and 1 output. The training set loss vs epoch for one of the instance among the 10 is Figure: 2(f). The time elapsed for this training process is: 3.15625s. The learning rate used is: 0.3.

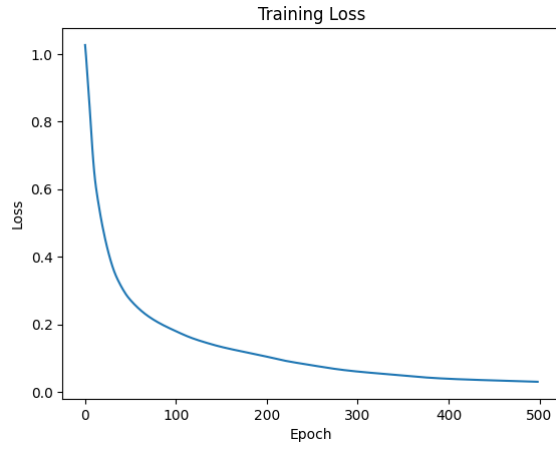
3 Test Set Error

All the average test set errors with it's standard deviation is given below in Table: 1 These values obtained are produced by choosing the best from repeated evaluation by

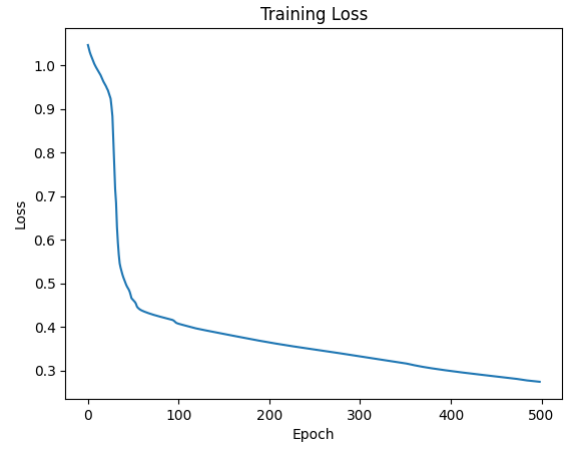
Dataset	Test Set Error	Neural Architecture	Learning Rate
Banana	10.63 ± 0.2	[2, 32, 24, 1]	0.1
Breast Cancer	26.75 ± 3.49	[9, 27, 1]	0.2
Diabetes	24.57 ± 1.72	[8, 16, 1]	0.2
German	24.73 ± 1.7	[20, 16, 1]	0.2
Heart	16.9 ± 1.3	[13, 13, 1]	0.2
Image	7.58 ± 1.23	[18, 18, 1]	0.3
Ringnorm	22.18 ± 2.02	[20, 16, 1]	0.4
Splice	34.97 ± 4.00	[60, 12, 1]	0.2
Thyroid	4.27 ± 1.96	[5, 15, 1]	0.1
Titanic	22.83 ± 1.95	[3, 9, 1]	0.06
Twonorm	5.02 ± 0.82	[20, 10, 1]	0.3
Waveform	14.49 ± 1.23	[21, 7, 1]	0.3

Table 1: Table with test set error

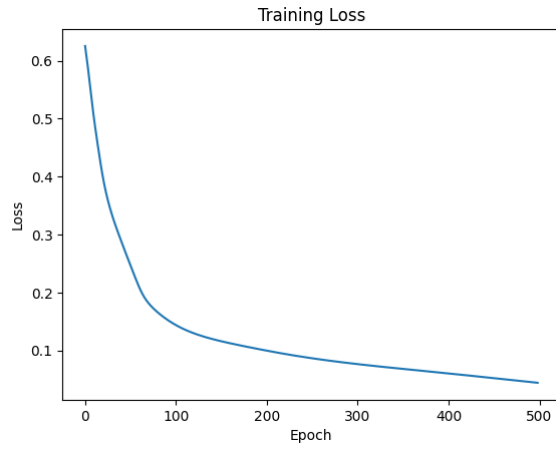
test and modifying the hyper parameters. I would be trying more of such modifications and will alter the values if any better found.



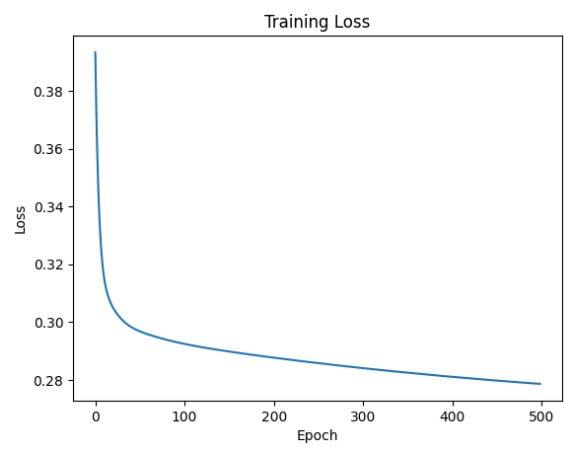
(a) Ringnorm



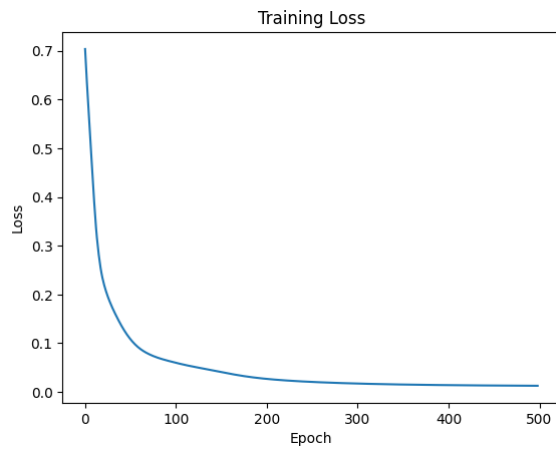
(b) Splice



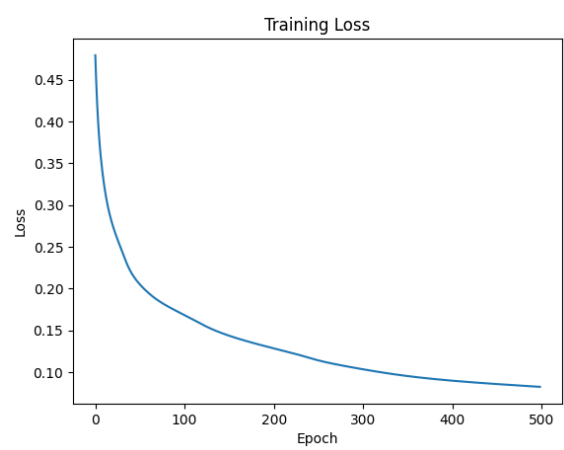
(c) Thyroid



(d) Titanic



(e) Twonorm



(f) Waveform

Figure 2: Datasets