RESULTS

The evaluation is performed on the following datasets:

1. Iris Dataset (Continuous, 4 attributes)
2. Car Dataset (Discrete, 6 attributes)
3. Breast Cancer Dataset (Continuous, 9 attributes)
4. Mushroom Dataset (Discrete, 22 attributes)
5. Pima Indian Diabetes Dataset (Continuous, 8 attributes)
6. Phishing Dataset (Discrete, 30 attributes)

Performance of Neural Network against Decision Tree:

We first run both the algorithms on the same splits of data, and compute the 10-fold accuracies for both. The accuracies, along with the Paired T-test scores for the same are reported below:

The following experiments are performed with the following specifications, using 10-fold validation, where in each fold, 70% is used for training, 20% for validation and 10% for testing.

1. Neural Network with 2 hidden layers, i.e [3,4]
2. Decision Tree with pruning
3. $\eta = 0.02$
4. $\alpha = 0.15$
5. Max Iterations = 20000
6. MSE Threshold = 0.025

Two algorithms are said to be statistically similar, when **0 falls in the confidence interval of their Paired T-test score.**

Dataset	Dataset Size	Neural Network			Decision Tree	Paired T-score	Statistically similar?
		Mean Iterations	Average MSE on Validation set	Accuracy	Accuracy		
Iris	150	5286	0.02	0.967 ± 0.025	0.940 ± 0.041	-0.027 ± 0.033	Yes
Mushroom	8124	5525	0.02	0.981 ± 0.006	1.0 ± 0.0	0.019 ± 0.006	No

Breast Cancer	699	18033	0.03	0.972 ± 0.012	0.935 ± 0.024	-0.038 ± 0.016	No
Pima Indians	768	20000	0.19	0.737 ± 0.052	0.742 ± 0.036	0.005 ± 0.041	Yes
Cars	1728	20000	0.10	0.881 ± 0.017	0.924 ± 0.012	0.044 ± 0.014	No
Phishing	11055	20000	0.05	0.932 ± 0.011	Timeout	N/A	No

Performance of neural network using different hidden layers

The performance of the neural network is shown for two datasets, using different combinations for the hidden layers specification. The following model parameters are used for all these experiments. The configuration that seem to learn the fastest and yield good accuracy are highlighted in yellow.

1. $\eta = 0.02$
2. $\alpha = 0.15$
3. Max Iterations = 20000
4. MSE Threshold = 0.025

(a) Iris Dataset (150 examples)

Hidden Layer Spec.	Average Number of Iterations	Average MSE	Accuracy
[] (empty)	20000	0.34	0.260 ± 0.105
[1]	20000	0.13	0.933 ± 0.044
[400]	19010	0.26	0.667 ± 0.177
[5,7]	4551	0.02	0.960 ± 0.040
[20, 20]	2670	0.02	0.947 ± 0.030
[100, 100]	7016	0.08	0.853 ± 0.119
[3,3,3]	13074	0.04	0.96 ± 0.033
[20,20,20,20]	3901	0.03	0.953 ± 0.045
[5,5,5,5,5]	18737	0.08	0.913 ± 0.083
[2,2,2,2,2,2,2,2]	20000	0.34	0.293 ± 0.107

(b) Mushroom Dataset (8124 examples)

Hidden Layer Spec.	Average Number of Iterations	Average MSE	Accuracy
[] (empty)	20000	0.25	0.506 ± 0.016
[1]	8450	0.02	0.986 ± 0.003
[400]	6549	0.04	0.942 ± 0.090
[5,7]	4550	0.02	0.979 ± 0.006
[20, 20]	3250	0.02	0.981 ±0.006
[100, 100]	1950	0.02	0.985 ± 0.004
[3,3,3]	10400	0.02	0.979 ± 0.005
[20,20,20,20]	4387	0.02	0.981 ± 0.004
[5,5,5,5,5]	10562	0.02	0.981 ± 0.004
[2,2,2,2,2,2,2,2]	20000	0.25	0.510 ± 0.017

(c) Pima Indians Diabetes (768)

Hidden Layer Spec.	Average Number of Iterations	Average MSE	Accuracy
[] (empty)	20000	0.24	0.654 ± 0.058
[1]	20000	0.19	0.739 ± 0.030
[400]	20000	0.25	0.780 ± 0.038
[5,7]	20000	0.17	0.745 ± 0.021
[20, 20]	20000	0.19	0.732 ± 0.018
[100, 100]	20000	0.21	0.737 ± 0.024
[3,3,3]	20000	0.17	0.761 ± 0.033
[20,20,20,20]	20000	0.19	0.751 ± 0.039