4SL4 Assignment 3

Benjamin Li, 9 Dec 2020
400130878, lib83@mcmaster.ca

Experiment

This experiment addresses a binary classification problem through various neural networks. The dataset, "bank_authentication.txt" contains 1,372 4-feature-examples. 3 sub-experiments will be conducted, in which the number of features used for training/validation/testing data are [4, 3, 2] respectively. Within each sub-experiment, different 3-layer neural networks are evaluated — varying the size of each hidden layer in [2, 3, 4] for a total of 9 configurations.

Results

Using 3 of 4 features with $n_1=4$ $n_2=3$ produced the best model, while using all 4 features with $n_1=2$ $n_2=2$ produced the worst. The data for each sub-experiment is below, with the optimal model highlighted in green and the worst model in red:

Table 1: 4 features

Hidden layer 1	Hidden layer 2	Trial 1 error	Trial 2 error	Trial 3 error	Average Error
2	2	0.506625968631	1.416326096233	0.506395643466	0.809782569443
2	3	0.214285714285	0.470526891638	0.214285714285	0.299699440070
2	4	0.214285714285	0.222874349076	0.601869815658	0.346343293006
3	2	0.214285714285	0.488797621995	0.215554195549	0.306212510610
3	3	0.882327391526	1.131845026375	0.214285714285	0.742819377395
3	4	0.745054652845	1.209750230309	0.392704449262	0.782503110806 1
4	2	0.349582969890	0.220366042485	0.186284064961	0.252077692445
4	3	0.221293522870	0.214285714285	0.214285714285	0.216621650480
4	4	0.258210742324	0.398134994415	0.177643594868	0.277996443869

Optimal model misclassification rate: 0.466
There is a **positive relation between neurons and accuracy** here. This is the opposite of our findings in **Table 2**. This means the 4-feature dataset is more **suited to larger hidden layers**, while our smaller networks are underperforming due to underfitting.

Table 2: 3 features

Hidden layer 1	Hidden layer 2	Trial 1 error	Trial 2 error	Trial 3 error	Average Error
2	2	0.214285714285	0.212832853190	0.182399099022	0.203172555499
2	3	0.214285714285	0.214285714285	0.327382353426	0.251984593999
2	4	0.310060589595	0.347218756844	0.586425233920	0.414568193453
3	2	0.343003933402 09817	0.214285714285 71427	0.214285714285 71427	0.257191787324 50887
3	3	0.214285714285	0.214285714285	0.156350095304	0.194973841291
3	4	0.171779012448	0.280417686481	0.151570402474	0.201255700468
4	2	0.214285714285	0.214285714285	0.598185668078	0.342252365550
4	3	0.214285714285	0.178993284158	0.214442535127	0.202573844523
4	4	0.214986330147	0.214285714285	0.151324811419 4	0.193532285284

Optimal model misclassification rate: 0.332
The results of the 3-feature dataset show the trends seen in **Table 1** to an even greater degree. The largest model is the most performant while the smallest is the least. The models in-between fluctuate in accuracy.

Table 3: 2 features

Hidden layer 1	Hidden layer 2	Trial 1 error	Trial 2 error	Trial 3 error	Average Error
2	2	0.214285714285	0.214285714285	0.189438330466	0.206003253012
2	3	0.359741836623	0.214285714285	0.728094311708 6	0.434040620872
2	4	0.214285714285	0.163744716785	0.735646310459	0.371225580510
3	2	0.441766507026	0.412397920042	0.154767585418	0.336310670829
3	3	0.194596953459	0.214283648647	0.214285714285	0.207722105464
3	4	0.214285714285	0.202082153122	0.214459288408	0.210275718605
4	2	0.216688843739	0.214285714285	0.228088374728	0.219687644251
4	3	0.214266329435	1.125383028787	0.214285714285	0.517978357502
4	4	0.158801688480	1.204015661023	0.167808193835	0.510208514446

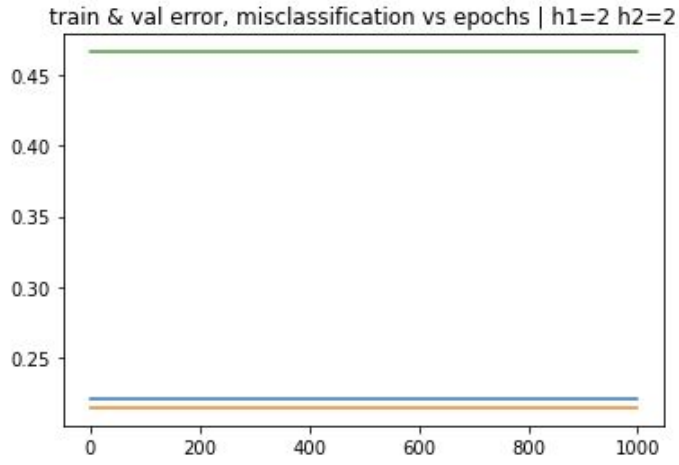
Optimal model misclassification rate: 0.422

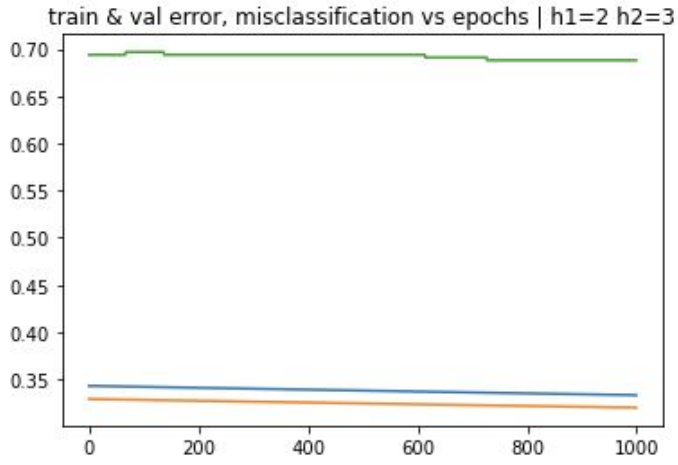
Too many neurons tends to cause **overfitting**. As the **input data is relatively small** with just 2 features, we easily overfit; hence, the negative correlation between neurons and accuracy. As you can see, the **smallest network** is the most performant as it avoids overfitting.

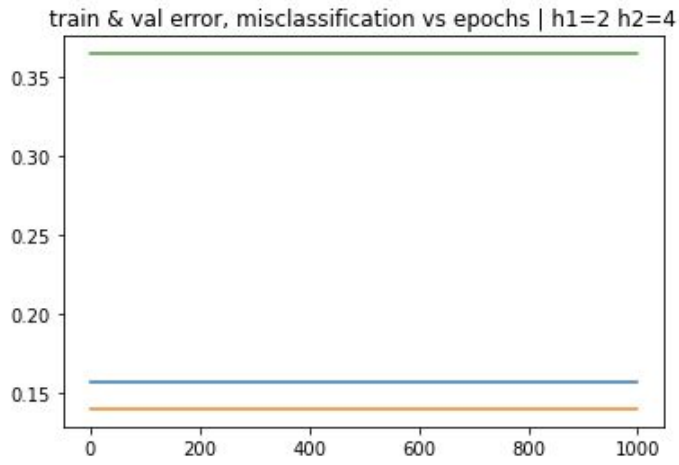
Plots of Training Error, Validation Error and Misclassification Rate vs Epoch
Green: misclassification rate
Orange: training error
Blue: validation error

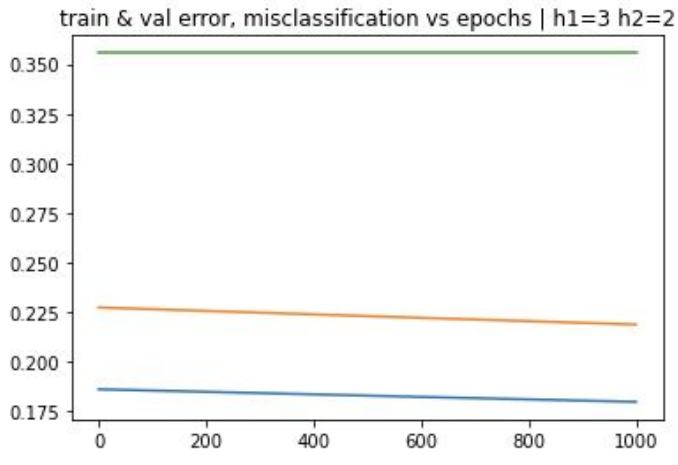
The errors and misclassification rate are generally flat. For binary classification problems, you have a 50% success rate from random guessing. Most models improve the success rate to 60%, with the best model increasing it to ~70% (h_1=4 h_2=2).

The error curves tended to have obvious downward slopes, while misclassification remained largely flat.

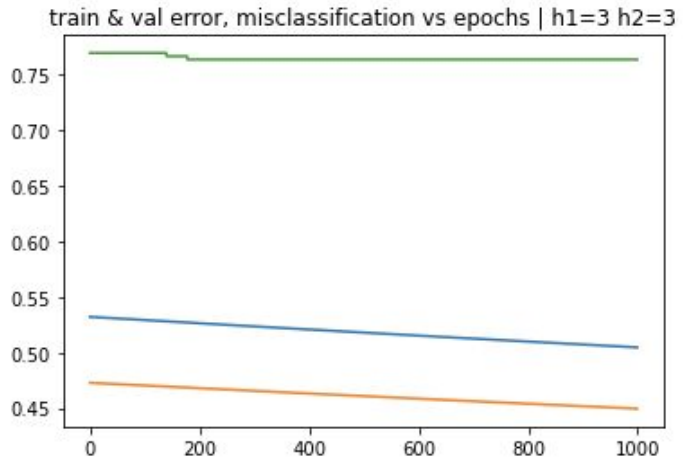




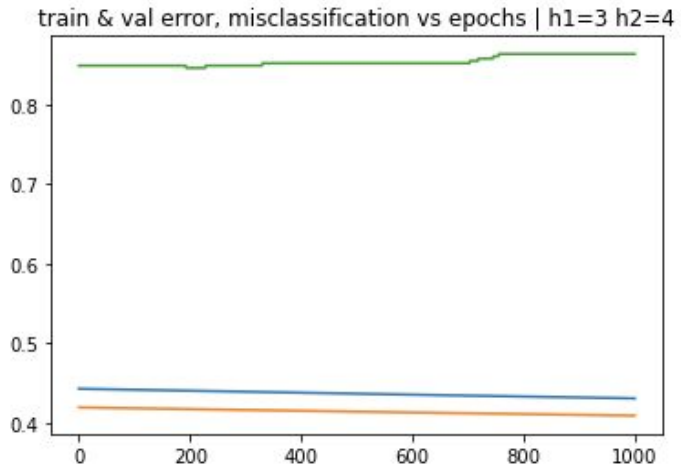




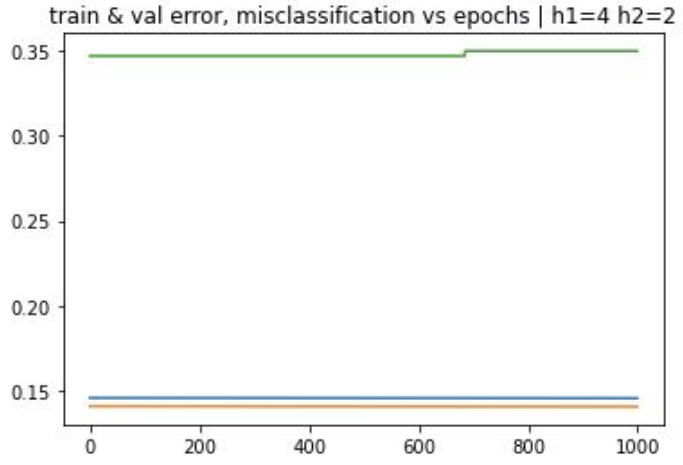
Models with ~30% misclassification, such as this one, were the top performers.

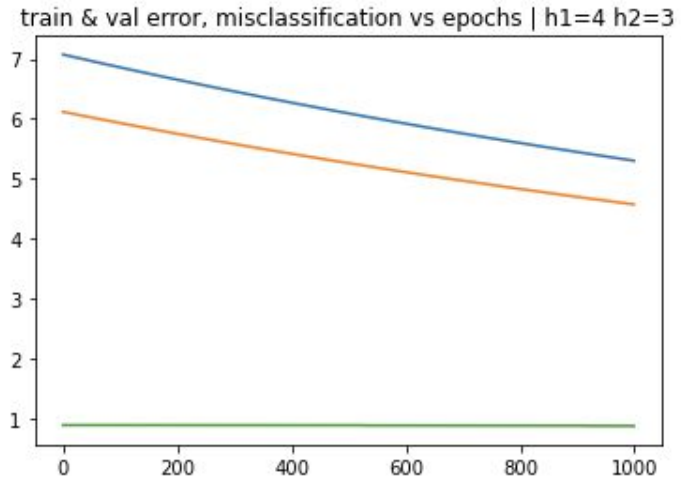


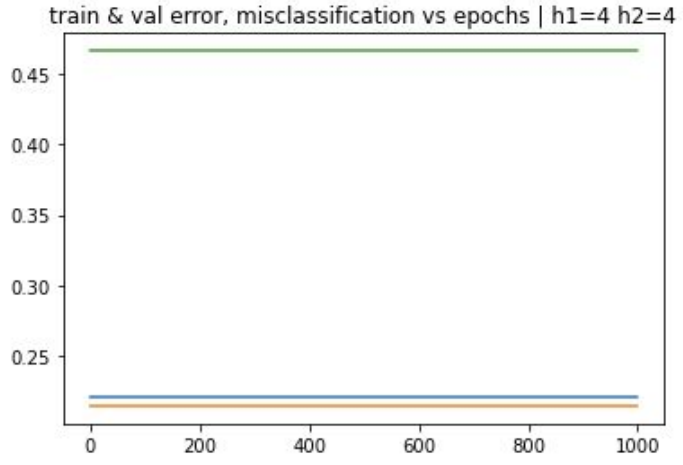
High misclassification rate (~70%), which is worse than random guessing.



Misclassification rate increases further into training and is high at over 80%.







Conclusion and Next Steps

For sub experiment 2 and 3 (2 features and 3 features respectively), we did not experiment with all possible feature-combinations. If the relationship of the unchosen features and labels differ significantly from the chosen feature set, results would vary. With that said, the top model from **this** experiment was obtained from using 3 of 4 features with $n_1=4$ $n_2=3$, while the worst was obtained from using all 4 features with $n_1=2$ $n_2=2$.

Furthermore, including **bias** parameters and exploring different cost functions should be done to achieve better results.