Assignment 1: Neural Networks

Troy Toth

Introduction:

The point of this assignment is to explore the performances of multiple different neural network configurations across an IDMB movie review dataset. I set out to see how changing the standard 2 hidden layers that was previously used would affect validation and test accuracy. I did the same thing with number of hidden units. Furthermore, I investigated how utilizing the MSE loss function changed the accuracy compared to the normal binary_crossentropy. The final thing I did was to switch out tanh activation instead of relu and then explored the dropout and regularization techniques.

Configurations Tested:

Configuration	Validation Accuracy	Test Accuracy
1 layer, 32 units	.8956	.8642
2 layers, 32 units	.8896	.8594
3 layers, 32 units	.8924	.8626
2 layers, 64 units	.8850	.8664
2 layers, 128 units	.8916	.8511
2 layers, 32 units, MSE Loss	.8912	.8678
2 layers, 32 units, Tanh Activation	.8910	.8507
2 layers, 32 units w/ Dropout	.8920	.8759
2 layers, 32 units, Dropout and L2 regularization	.8948	.8674

Key Findings:

- Best performing model
    - 2 layers, 32 units, with dropout
    - Accuracy of .8759
    - It appeared to significantly improve generalization
    - 1 layer, 32 units had the highest validation accuracy at .8956
- Extra layers and units
    - Adding more layers did not directly lead to better validation or test accuracy as the best performing one with 32 units was the one layer

- Adding more units also did not mean better accuracy. The 2 layer, 64 units out performed the 2 layer, 32 units but the 2 layer, 128 units was the least accurate.
- Regularization plus activation and loss functions
    - Dropout and L2 regularization both improved accuracy of test and validation
    - Dropout alone was the best overall result for test accuracy
    - Using tanh activation greatly reduced accuracy, giving the lowest result
    - MSE loss function improved the accuracy and gave the second best result and a solid validation accuracy of .8912

Conclusion:

The results showed the shallower models performing well. A simple 1-2 layers with 32-64 units tested the best. The MSE loss function greatly improved the test accuracy of the 2 layer, 32 units compared to binary cross entropy. While tanh activation greatly reduced test accuracy. The models that used dropout alone and dropout plus L2 both performed toward the top of the pool. The best overall result was the 2 layers, 32 units dropout by balancing simplicity and performance. The highest validation score was the most simple 1 layer and 32 units followed by the 2 layers, 32 units, Dropout and L2 regularization. In conclusion, the simple shallow and low unit ones utilizing the dropout technique were the most precise overall.