

CS 486/686 Spring 2023

Assignment 3 Solutions

July 8, 2023

Question 1

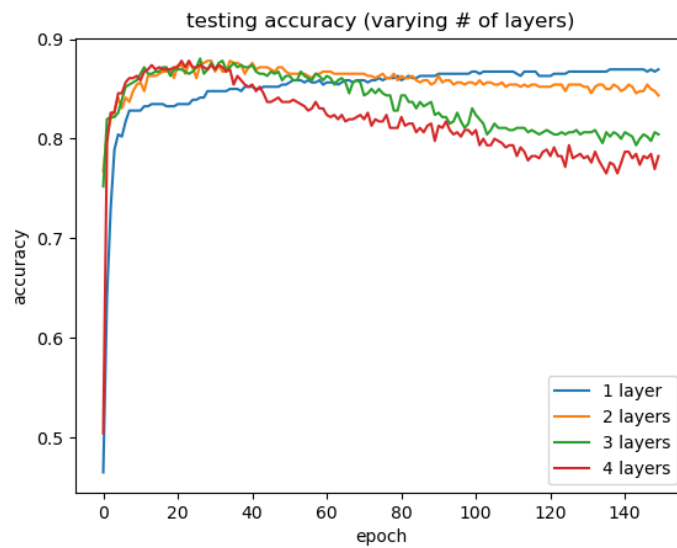
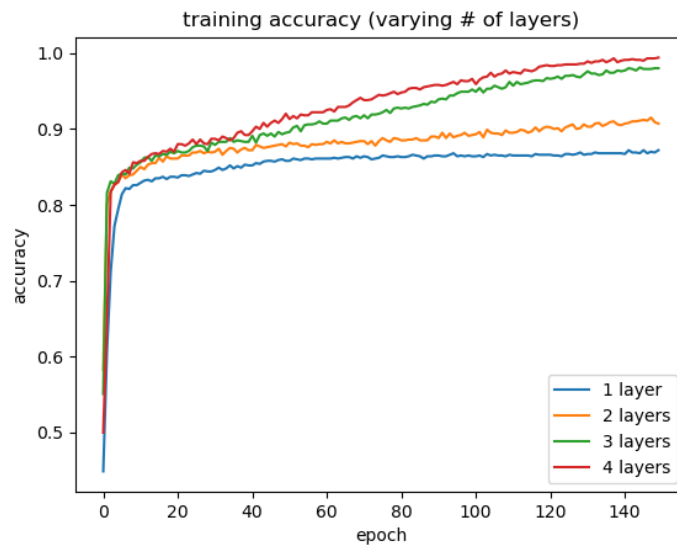
- Threshold 10: average validation accuracy 0.778
- Threshold 20: average validation accuracy 0.8089999999999999
- Threshold 40: average validation accuracy 0.821
- Threshold 80: average validation accuracy 0.8150000000000001
- Threshold 160: average validation accuracy 0.8009999999999999

Best threshold 40: test accuracy 0.8239130434782609

Explanation of the results: As the threshold increases, the size of the tree decreases. When the threshold is small, the tree tends to overfit. When the threshold is large, the tree tends to underfit. Hence the best results are obtained with a threshold of 40.

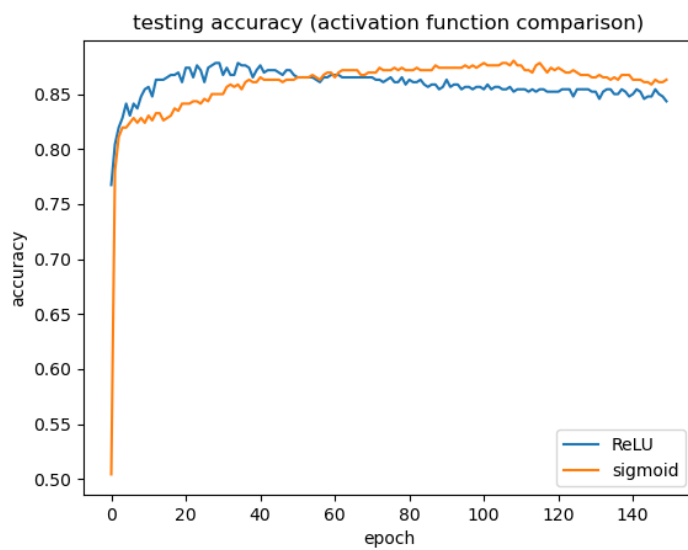
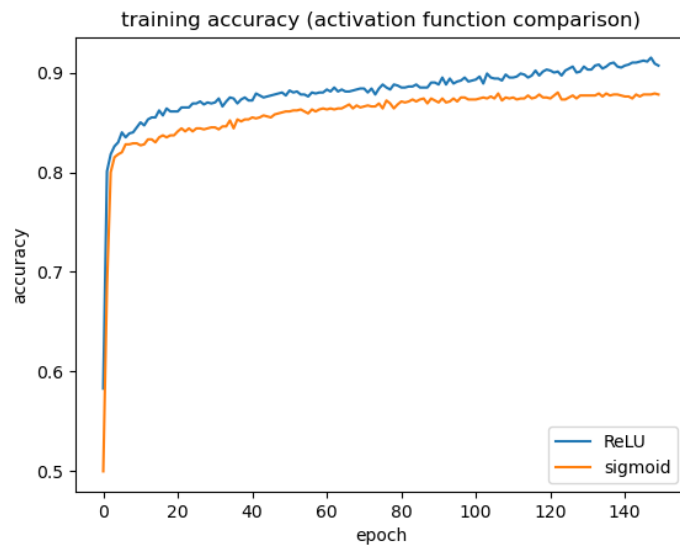
Question 2

Architecture comparison



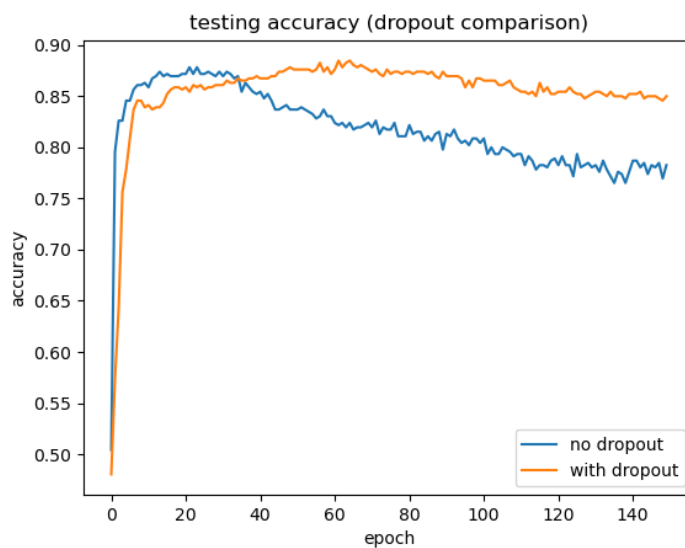
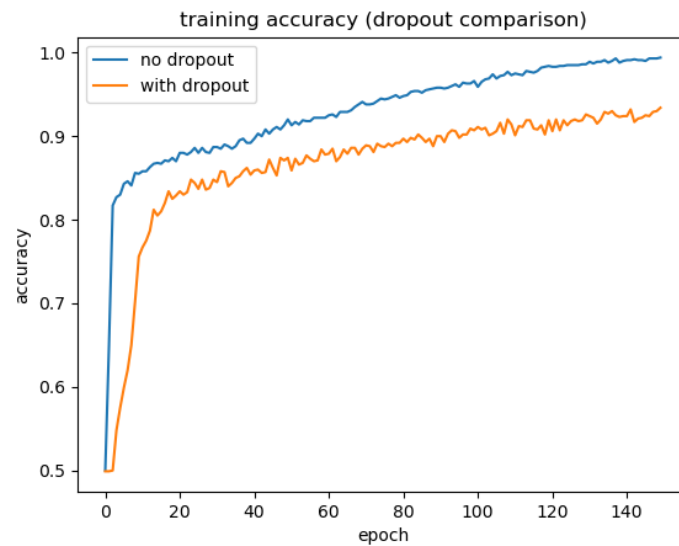
Explanation of the results: Training accuracy increases with the number of layers, but roughly the same best testing accuracy can be achieved with each architecture. However, the best testing accuracy is achieved faster with more layers and overfitting increases with more layers. Only the network with one layer does not seem to overfit.

Activation function comparison



Explanation of the results: ReLU yields better training accuracy than sigmoid. ReLU and sigmoid achieve roughly the same best testing accuracy, but ReLU achieves a peak faster and overfits more.

Dropout evaluation



Explanation of the results: The neural network with dropout yields lower training accuracy than the one without dropout. However, the gap between training and testing accuracy is reduced with dropout. The neural networks with and without dropout yield roughly the same best accuracy, but dropout mitigates the amount of overfitting.