CSC 820
HOMEWORK 13
REPORT
Andrew Dahlstrom
5/13/24

Comparison of Sentiment Classification Model Performance

Four observations stood out in particular about the differences in the two classification models architecture and the impact on their performance. The first implementation has two hidden layers with 300 nodes each and uses the ReLU activation function, while the second implementation has four hidden layers with 200 nodes each and uses the Tanh activation function. The change in the number of hidden layers and the type of activation function can affect the learning capability and efficiency of the network.

Both models achieve an overall accuracy of close to 74%. However, a closer look at the precision, recall, and f1-score metrics shows slight differences in their performance across the three classes. For instance, the first model has a slightly better recall for class 2 (0.86 vs. 0.82) but performs similarly on the other classes. The variations in performance metrics suggest that while both models are competitive, the additional layers in the second implementation provide marginal improvements in recall for certain classes, potentially due to the deeper architecture capturing more complex relationships.

The precision, recall, and f1-score metrics indicate that both models handle class imbalances reasonably well, but the second model shows a balanced precision and recall for class 1 (0.66). This could be attributed to the more complex architecture with additional layers and nodes, which helps in better representation and classification of the classes.

Increasing the number of hidden layers from two to four and adjusting the number of nodes per layer changes the network's capacity to learn from the data. The second model, with its deeper architecture, might take longer to train due to the increased number of parameters, but it can also capture more intricate patterns in the data. Despite the increased complexity, both models achieve similar overall accuracy, suggesting that for this specific dataset, the added depth did not drastically improve the generalization ability.

The plot below shows a performance comparison of the loss between the two models with the first implementation in blue and the second implementation in red.