Artificial Intelligence 2, Project 3

Georgios Nikolaou

April 12, 2023

Tuning

Tuning process & results

The progression of the tuning and some relative comments can be seen on the notebook.

Figure 1: Model: L32

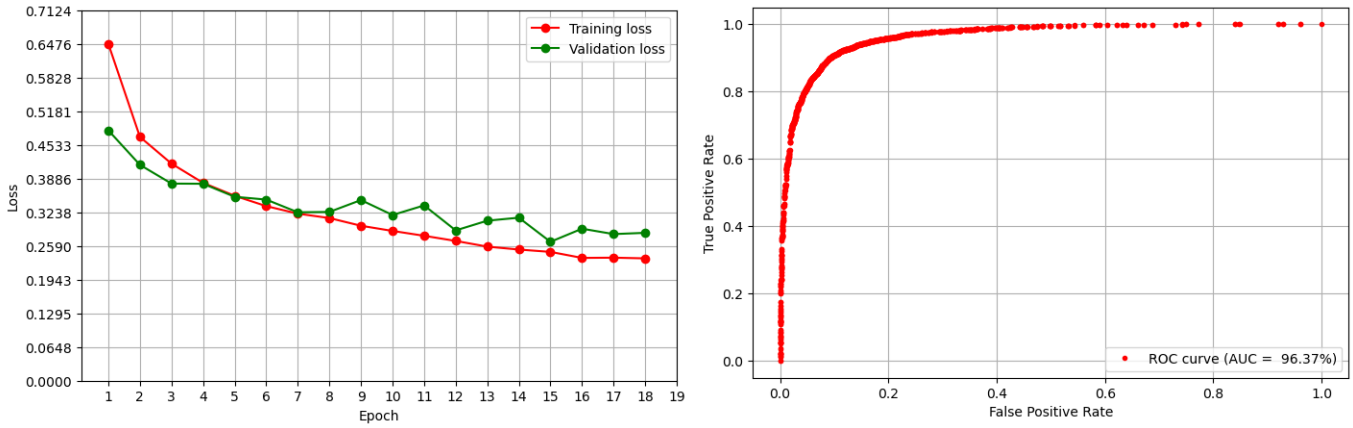


Figure 2: Model: L34

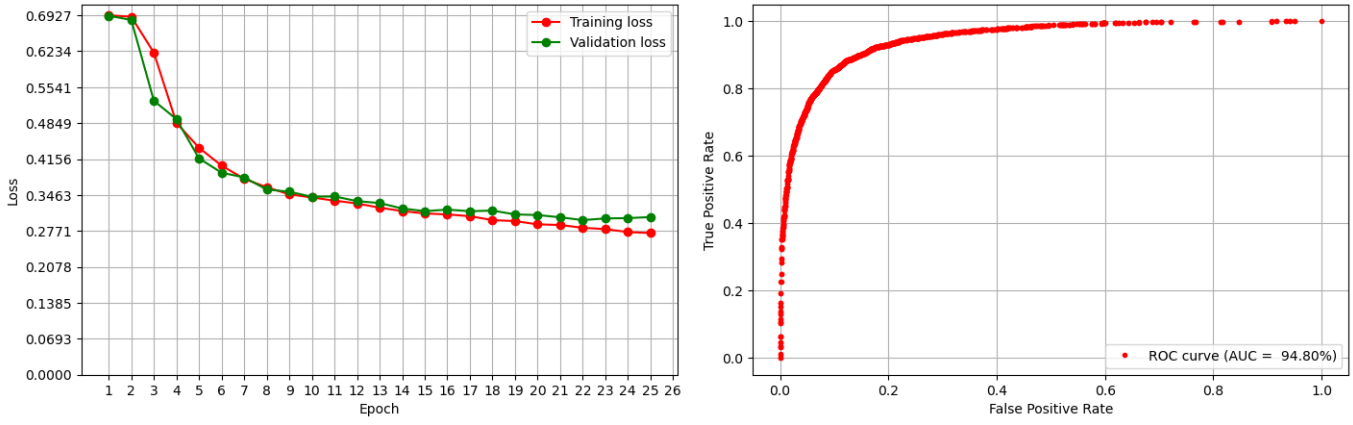


Figure 3: Model: L35

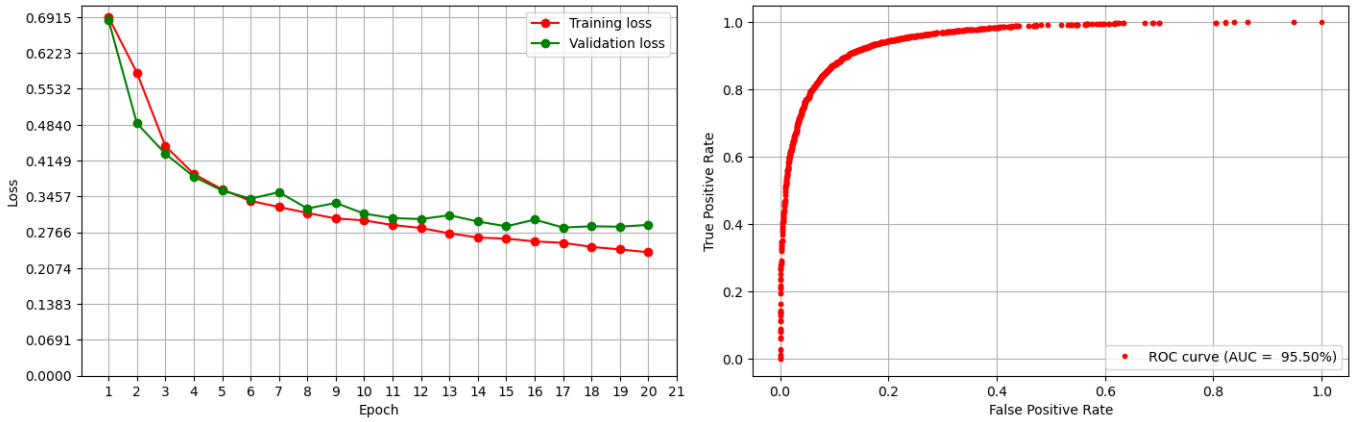
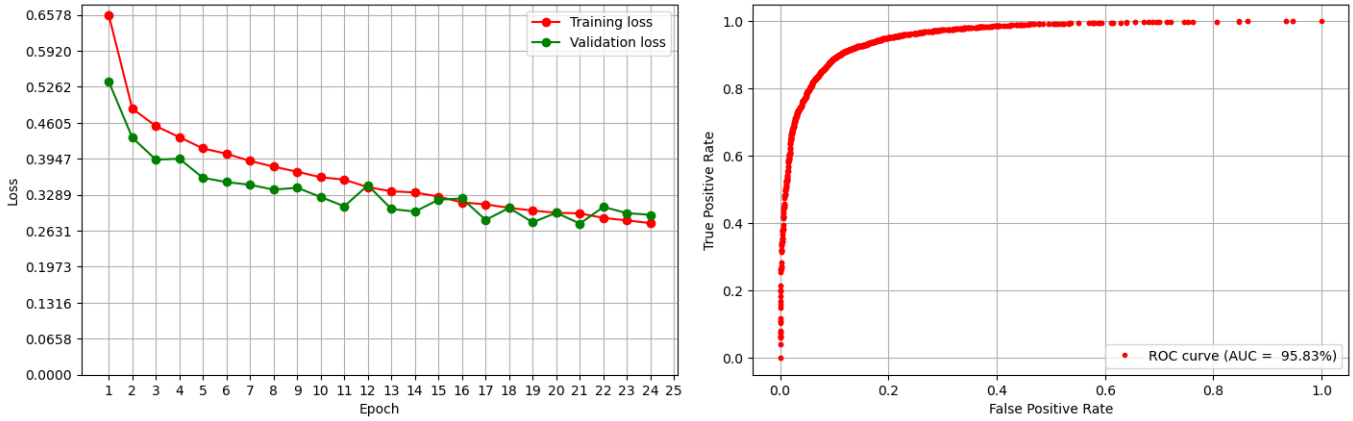


Figure 4: Model: G06



Although it was a little hard, we eventually started getting very good learning and roc curves.
We note that RNNs are complex models and therefore are more prone to overfitting than simple feed forward networks. This is apparent to the amount of experiments that were done to tune the model.

General notes:

- Learning rate was an important parameter to tune correctly and we tried various methods to do so.
- A loss penalty was important and (was tried and worked) if added manually produces excellent results. It was simulated by weight decay on the optimizer and worked very well for the GRU cells.
- Every method to reduce overfitting had to be applied massively.
- Skip connections didn't seem to make a difference.
- Attention was generally beneficial.

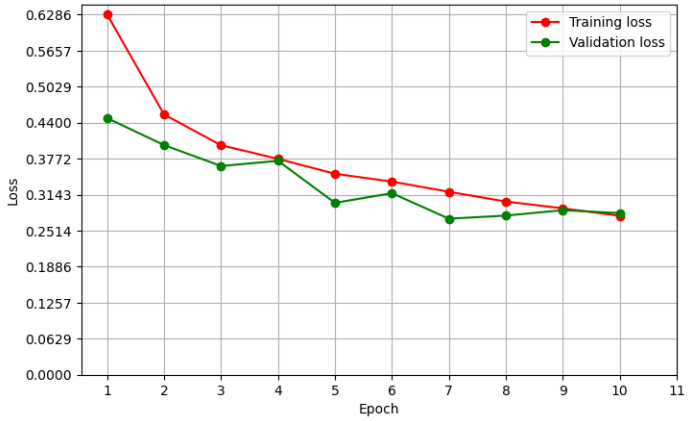
Testing

Unknown movies data split

Created a train, test set pair with the property that every review on the test set is for a movie that doesn't exist on the train set.

	Accuracy	Precision	Recall	F1-Score
L32	88.05%	88.04%	88.48%	88.26%
L34	85.31%	85.73%	89.49%	85.49%
L35	88.85%	88.66%	89.49%	89.07%
G06	89.14%	87.14%	92.22%	89.61%

Figure 5: L32



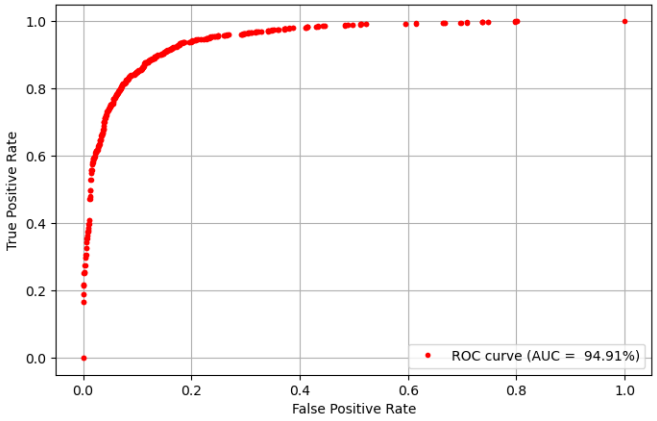
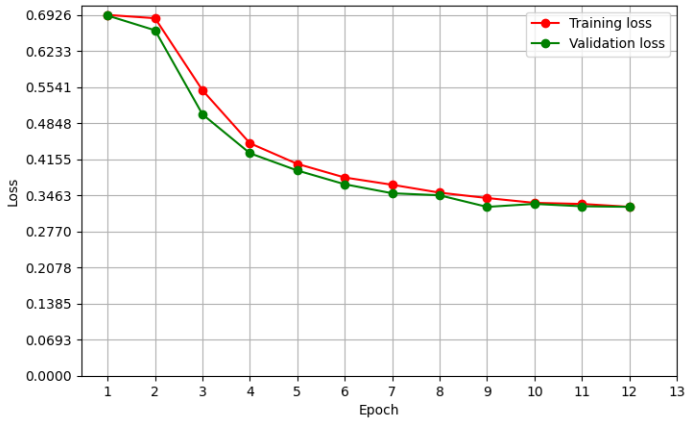


Figure 6: L34



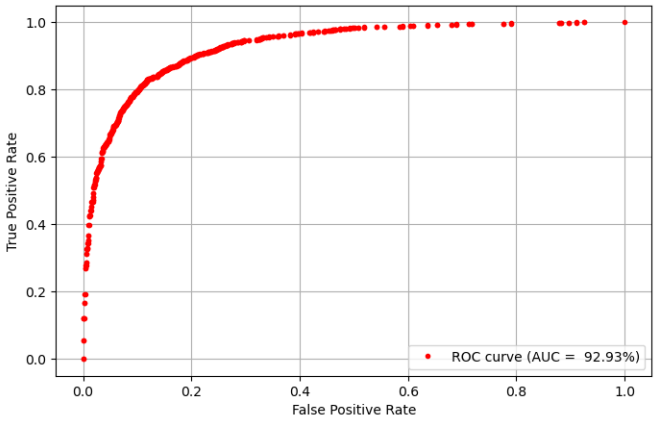
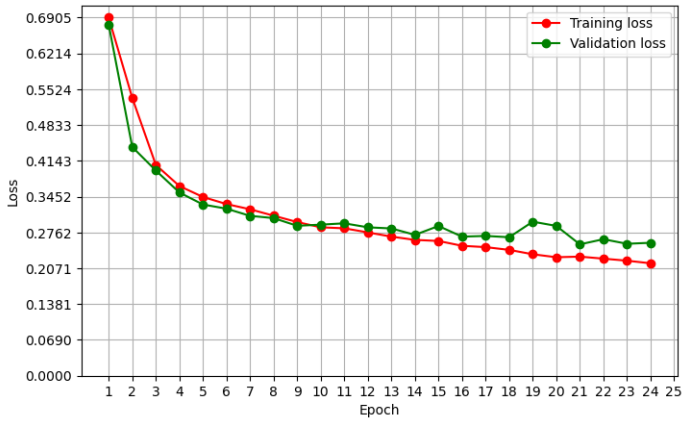


Figure 7: L35



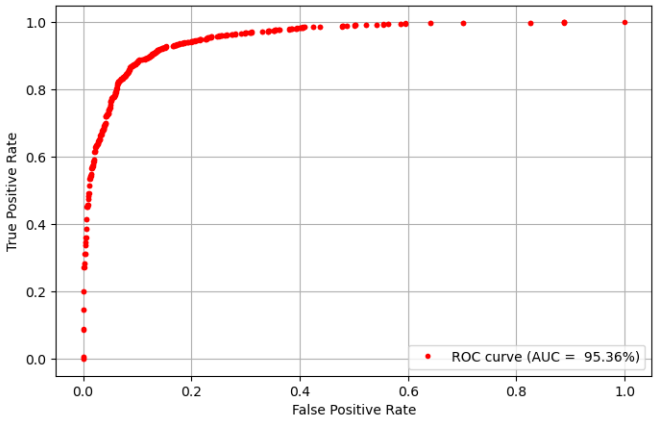
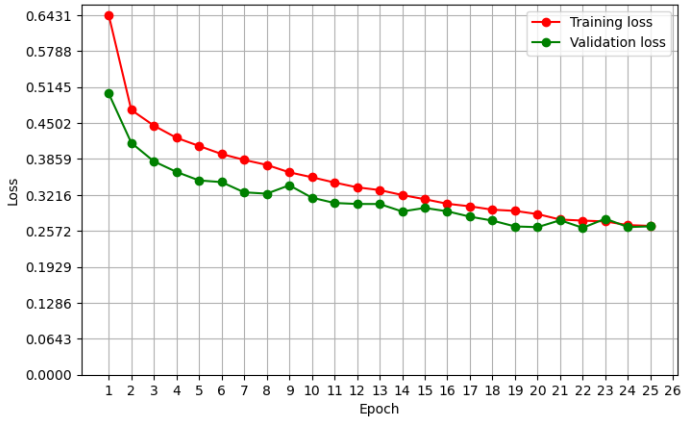
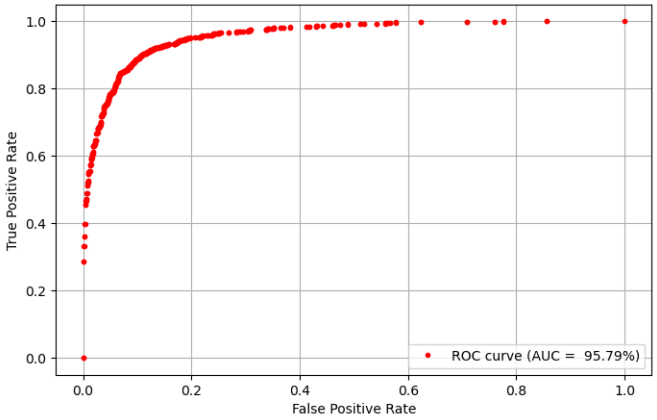


Figure 8: G06





We observe that the models chosen have not been overfit to any data set used during the tuning process. We get excellent learning curves and every metric on the test set is about where it was expected from the respective results during tuning.
$L34$ has significantly lower performance than the rest. This is expected because it learns the slowest and the patience value given during training was low. Changing it to a higher value, it would achive similarly good results.
We conclude that the models can generalize well and should have no problem on your hidden dataset.