CSE849 Project 3

Javen W. Zamojcin
zamojci1@msu.edu

2025, April 11

1 Question 1: Review Rating Prediction

Training/Validation Performance History

```
1/5 [00:46<03:04, 46.14s/it,
    Train Loss=1.141, Train Accuracy=48.81, Val Loss=0.978, Val Accuracy=56.86]
2/5 [01:31<02:17, 45.87s/it,
    Train Loss=0.953, Train Accuracy=58.06, Val Loss=0.942, Val Accuracy=58.41]
3/5 [02:17<01:31, 45.99s/it,
    Train Loss=0.913, Train Accuracy=59.94, Val Loss=0.924, Val Accuracy=59.43]
4/5 [03:04<00:46, 46.07s/it,
    Train Loss=0.892, Train Accuracy=60.93, Val Loss=0.906, Val Accuracy=60.08]
5/5 [03:48<00:00, 45.71s/it,
    Train Loss=0.879, Train Accuracy=61.63, Val Loss=0.902, Val Accuracy=60.38]
```

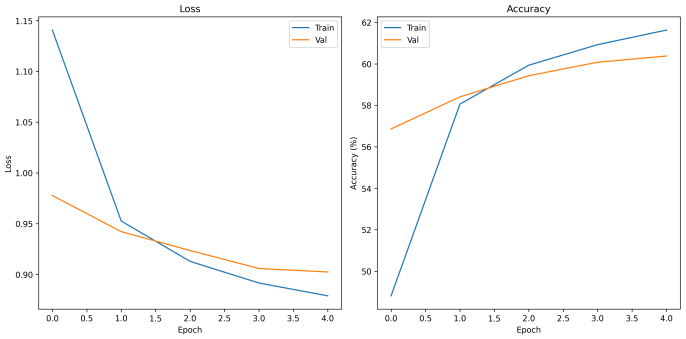

Figure 1: Q1 Training History

In Figure 1, we observe the training and validation performance over five epochs. Figure 2 depicts the respective trained model's validation prediction confusion matrix. After about five epochs of training, I began to experience validation accuracy loss, suggesting that overfitting was beginning to occur, so training was stopped after five epochs. It's also interesting that in the confusion

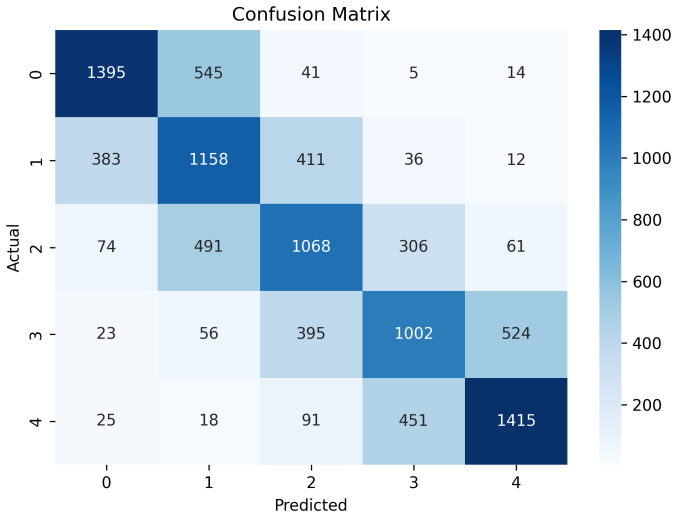

Figure 2: Q1 Confusion Matrix

matrix, we observe that the model had a significantly easier time in correctly predicting the two extreme ends of the ratings, 1 and 5 stars, which in my personal experience are usually the two default ratings for either having a bad or good experience, respectively, and contain the strongest language.

2 Question 2: Pig Latin Translation

Training/Validation Performance History

```
[1/50] Training MSE: 0.4556, CE: 0.2542, Accuracy: 30.85%
[1/50] Validation MSE: 0.1196, CE: 0.6941, Accuracy: 85.59%
[2/50] Training MSE: 0.1642, CE: 0.0418, Accuracy: 58.39%
[2/50] Validation MSE: 0.0715, CE: 0.5911, Accuracy: 88.57%
...
[48/50] Training MSE: 0.0824, CE: 0.0104, Accuracy: 88.33%
[48/50] Validation MSE: 0.0083, CE: 0.0212, Accuracy: 99.51%
[49/50] Training MSE: 0.0824, CE: 0.0105, Accuracy: 88.32%
[49/50] Validation MSE: 0.0081, CE: 0.0215, Accuracy: 99.49%
[50/50] Training MSE: 0.0821, CE: 0.0101, Accuracy: 88.81%
[50/50] Validation MSE: 0.0084, CE: 0.0182, Accuracy: 99.59%

Final Accuracies
Train: 88.81
Val: 99.59
```

```
Final Losses
Train MSE: 0.082
Train CE: 0.010
Val MSE: 0.008
Val CE: 0.018
```

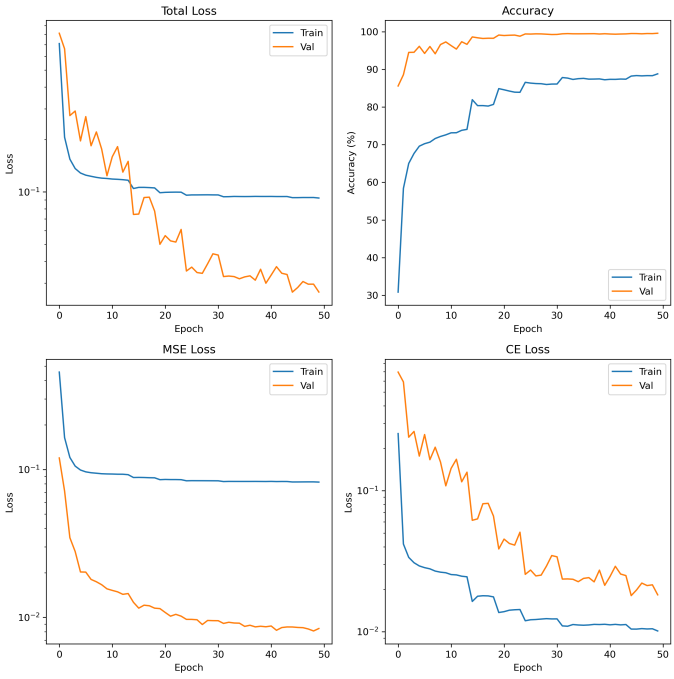

Figure 3: Q2 Training History

In Figure 3, we observe the training and validation performance history over 50 epochs. It is interesting that validation accuracy consistently remained well above training accuracy throughout the entire history. This suggests that there exists overlap between the training and validation dataset content, explaining why initially learning only a marginal amount of training data patterns resulted in an immediately high accuracy of the smaller validation dataset. It is also interesting that validation consistently had a smaller MSE loss but higher CE loss, and vise versa with the training history. The higher CE loss reflects the autoregressive nature of the validation task in predicting individual tokens with carryover error from previous predictions.