PS1 WRITTEN

(a) [2 points (Written)] Consider the case with F = 0.99 and R = 0.01. Based on the train/test loss curves, does parameter sharing outperform having separate models?

- There is a stark difference in the Mean Reciprocal Rank(MRR) highlighting the fact the shared model does not do a better job of returning more accurate movie suggestions.
- However, the shared representation does not recommend movies that are unlikely to be recommended to the user. The shared representation recommends more likely movies and avoids recommending unlikely movies.

(b) [2 points (Written)] Now consider the case with F = 0.5 and R = 0.5. Based on the train/test loss curves, does parameter sharing outperform having separate models?

- I see very identical MSE curves in the test set. However, the MRR couldn't be more alike. We can clearly see that the shared representation does not do a better job of suggesting movies to users as they rank movies as more likely to be popular.
- Also, there is a stark difference shared in the factorization loss function in training; the shared loss function has negative linear behavior while the separate representation establishes positive concavity and quickly reduces loss. Sharing parameters demonstrates less dramatic classifications.

(c) [2 points (Written)] In the shared model setting compare results for F = 0.99 and R = 0.01 and F = 0.5 and R = 0.5, can you explain the difference in performance?

- Shared representation models yield better results in classification problems than regression problems.
- Shared representation models yield much higher initial loss and seem to be computationally inefficient; however, eventually converge to non-shared representations.
- Shared representations were not better able to predict more accurate suggestions with the MRR algorithm compared to non-shared representations.