CS 330 Autumn 2021/2022 Homework 0: Multitask Training for Recommender Systems

Kevin Lee (kelelee@stanford.edu)

Due on Monday September 27, 11:59 PM PST

1 Write-up

Figure 1: Tensorboard graphs

1. Does parameter sharing outperform having separate models? By comparing train and test set performance can you briefly explain these results?

Answer:
Parameter sharing (blue line) outperforms separate models (red line) for score prediction but separate models outperform parameter sharing for likelihood prediction. For fixed values of λ_F and λ_R, sharing the embeddings is a form of regularization that encourages the model to pick up on features that are important to both tasks. Since both tasks in this case are related, this actually ended up improving MSE by a lot with a small decrease in MRR as a tradeoff (due to model capacity).

2. In the shared model setting compare results for $\lambda_F = 0.99$ and $\lambda_R = 0.01$ and $\lambda_F = 0.5$ and $\lambda_R = 0.5$, can you explain the difference in performance?

Answer:
Both models (orange and blue lines) perform relatively the same on score prediction (similar MSE), but the model trained with $\lambda_F = 0.99$ and $\lambda_R = 0.01$ (orange line) performed significantly better on likelihood prediction (higher MRR). Since λ_F was a higher value, the model is encouraged to perform better on matrix factorization which led to better predictions for probabilities that users would watch movies, which ultimately resulted in a better MRR.