**OVERVIEW**

Purpose of project and business context

- Start with a sentence or two about how personalization is used everywhere nowadays.

- Why do we want to personalize?

• A company that can make high quality recommendations will sell more product.

• If users subsequently trust the company for making high quality recommendations, the company might sell even more.

Describe **this** project (in no particular order)

- What dataset are we using? Amazon film and TV

- Where did we get it? website

- Why did we choose it? chose it because it has a lot of ratings, resembles data we encounter in real life, real business problem than other datasets

- What are the most relevant statistics about the data?

- Explicit or implicit?

- Number of items

- Number of users

- Number of ratings

- Density

Describe **structure** of remainder of write-up.

**EXPLORATORY DATA ANALYSIS**

What are some interesting characteristics of the data?

What challenges will these characteristics pose?

Reason why the baseline is doing pretty well.

Look at anything else in notebook. Use plots.

**EVALUATION METRICS (in order of importance)**

• Precision

- What is precision?

- Why is it important in context of business problem?

• MAE

- What is mean absolute error

- Why is it important but less so than precision?

• Runtime

- Why is this important in context of business of business problem?

• Coverage

- What is coverage

- Why is this important

- Why are we only using it on small data

- What are the limits of measuring coverage only on small data?

**MODELS**

- Baseline average

- Baseline manual

- Baseline SGD

- Baseline ALS

Reason manual is worse than SGD and ALS is that we have two parameters: user biased, item biased—so learning these two parameters will be better

- SVD (unconstrained matrix factorization)

- NMF (nonnegative matrix factorization)

- KNN Item-based collaborative filtering

**METHODOLOGY**

Recap evaluation metrics.

**Sampling Methodology:** Describe goal of evaluating metrics across different dataset sizes.

• First sample a proportion of data randomly, then filter the data keeping only users and items with at least *k* ratings

• *k* thresholds and sample proportions chosen to maintain similar density for comparison purposes.

• All training and evaluations made on these sampled datasets rather than the original.

• Use table from Data Sampling notebook.

**Hyperparameter search/model tuning strategy:** Explain hyperparameter optimization using tree-parzen estimators (a form of Bayesian optimization)

* + Describe what TPE is.
  + Why did we choose TPE?
    - TPE > Random search > Grid search
    - Better hyperparmeters in same amount of time
  + What loss function were we minimizing when finding hyperparameters?
    - Minimize 3-fold cross-validated MAE
      * (Look up k-fold cross-validation for explanation)
  + What dataset was used for TPE optimization?
    - The largest sampled dataset.
  + For each model (that has hyperparameters that need tuning), mention which hyperparameters most affected MAE and at what values.
    - Use plots
    - Anything interesting or surprising?

**Evaluation Methodology:** How did we assess the performance of our tuned models?

* Perform 5-fold cross-validation on each dataset size. Record mean and standard deviation of main evaluation metrics (MAE, precision, recall).
* Due to computational expense, coverage is only computed from the largest dataset with no cross-validation (the largest dataset used to train the model)

**RESULTS**

Discuss how each model performs on each metric.

If we were to make a recommendation to the business, what would that be? Are there any caveats that need to be mentioned?

**CONCLUSION**

Discuss what we would like to do if we had more time (further avenues to explore).

Having observed these results, what would we have done differently for **this** project?