Hidden Data Evaluation

Intro

We want to evaluate our recommender, using existing data.

But how do we do this?

This video: sketch the basic idea

Next module: more details and variations

Basic Idea

Use **existing data** collected from users

Use data to **simulate behavior**

Test whether recommender can **predict** behavior

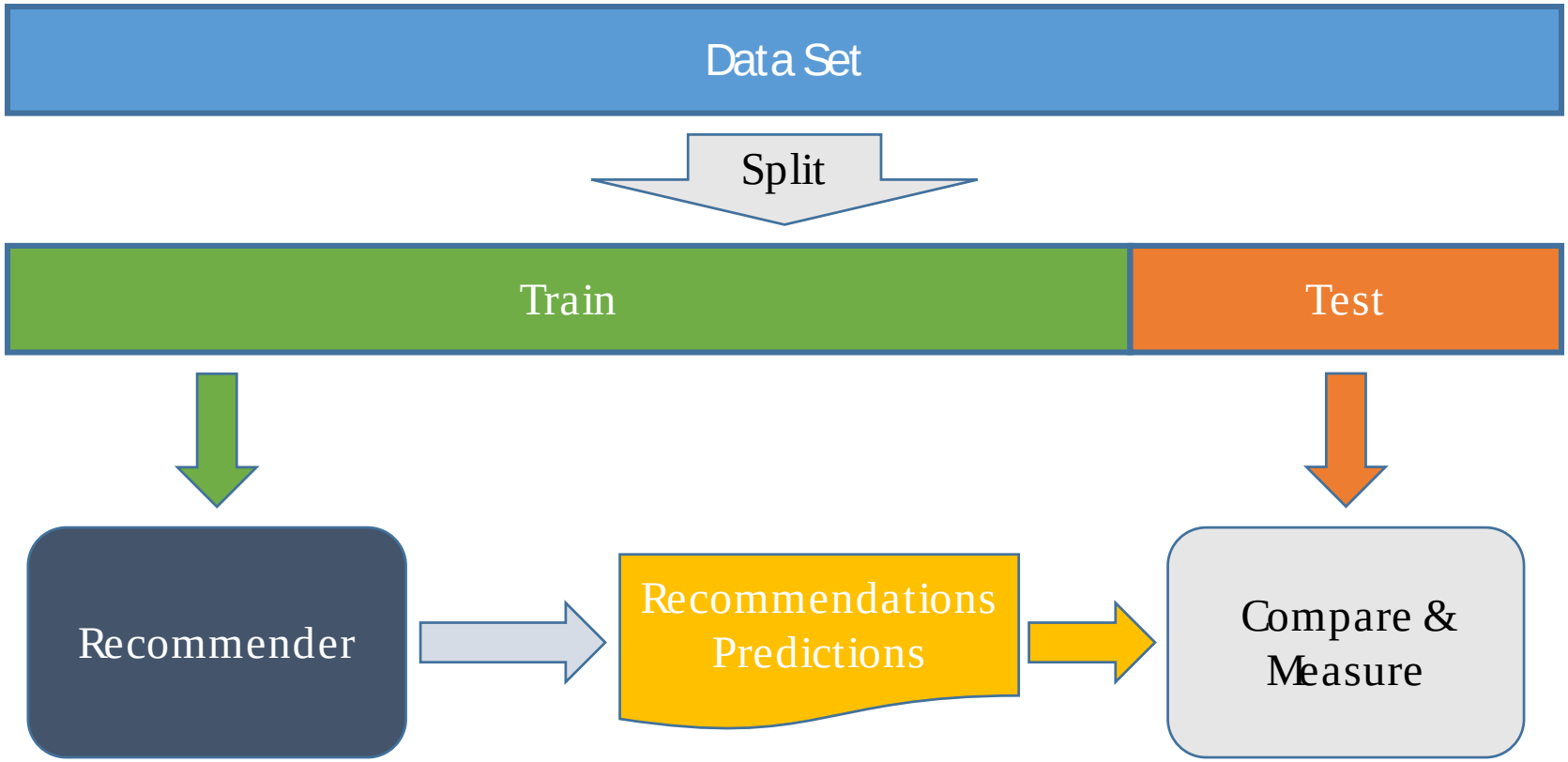
Prerequisites

1. Data

Some data set that is relevant to recommendation

- Ratings of movies
- Plays (or play counts) of songs
- Clicks of news articles

Evaluation Structure



What do we measure?

- How close is prediction to rating?
  - Aggregate
- Are purchased items in recommendation list?
- Where are purchased items ranked in recommendation list?

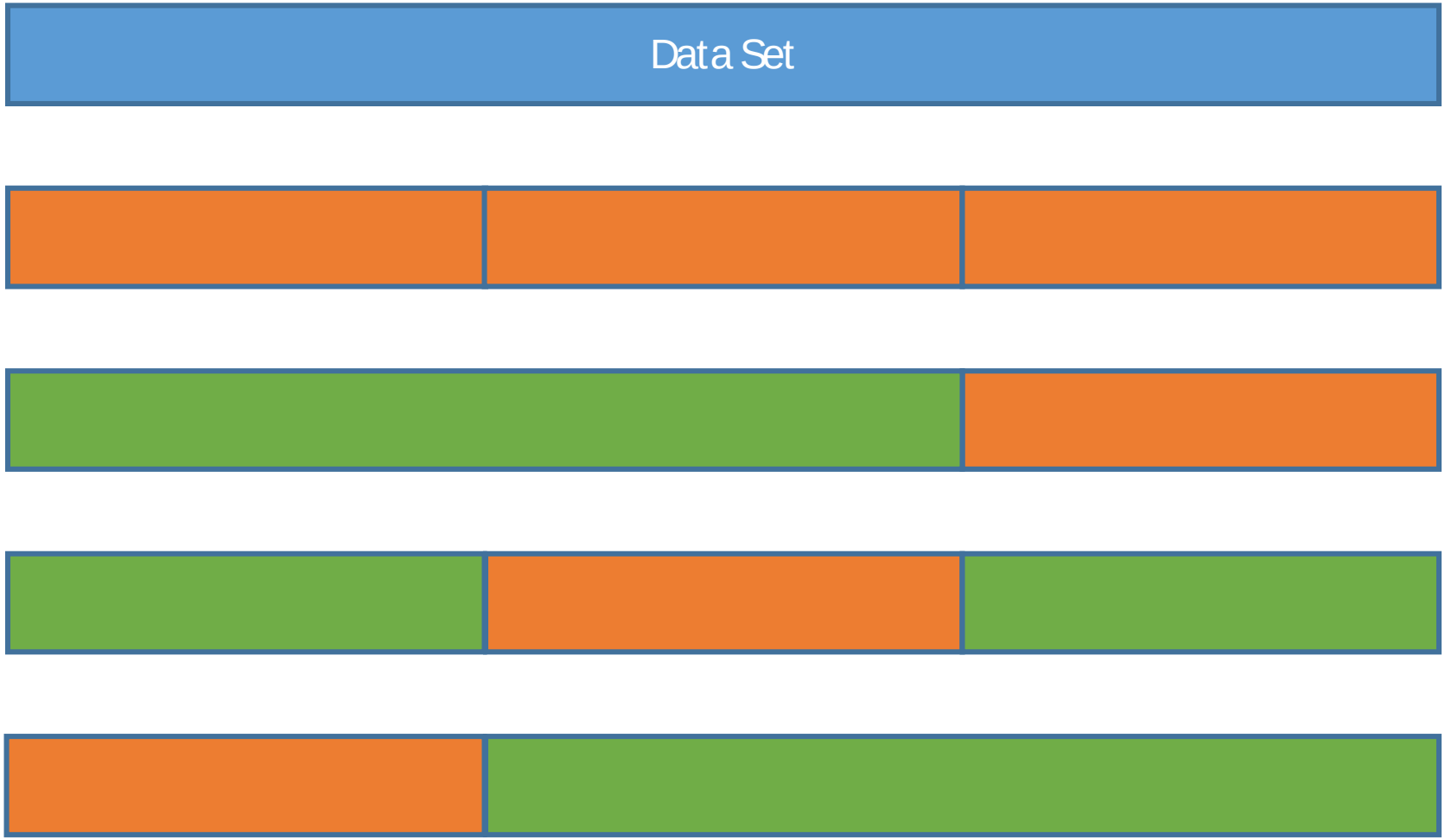
Cross-Validation

What if train-test split randomly selects mostly 'easy' or 'hard' users?

Solution: multiple splits

- Partition data into *k* partitions
- For each partition, train on others and test with it

Cross-Validation



Benefits

- Very efficient
- No users required

Drawbacks

- Does predictive accuracy matter?
- If user is shown rec list, would they still pick the item they bought?

More on these later.

Conclusion

- We can use data sets to estimate recommender performance
- Hide some data, ask recommender to predict it
- Compare to collected data

Hidden Data Evaluation