Recommender Systems
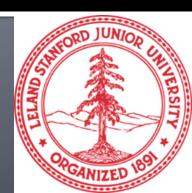
Evaluating recommender systems
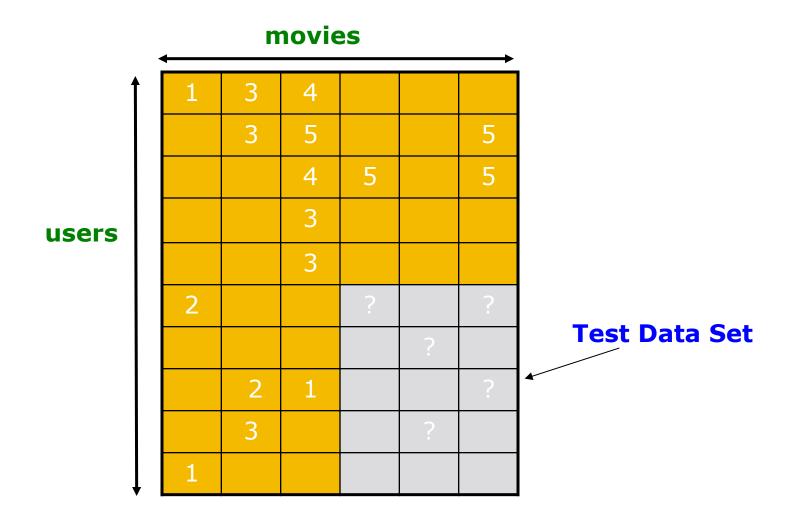
Mining of Massive Datasets
Leskovec, Rajaraman, and Ullman
Stanford University

Evaluation

movies

users

1	3	4			
	3	5			5
		4	5		5
		3			
		3			
2			2		2
				5	
	2	1			1
	3			3	
1					

Evaluation

movies

users

Test Data Set

Evaluating Predictions

- Compare predictions against withheld known ratings (test set *T*)

- Root-mean-square error (RMSE)

$$\sqrt{\frac{\sum_{(x,i) \in T}(r_{xi} - r^*_{xi})^2}{N}}$$

where *N* = |*T*|

*r*_{*xi*} is the predicted rating

*r*_{*xi*}* is the actual rating

Problems with Error Measures

- **Narrow focus on accuracy sometimes misses the point**
  - Prediction Diversity
  - Prediction Context
  - Order of predictions

- **In practice, we care only to predict high ratings**
  - RMSE might penalize a method that does well for high ratings and badly for others
  - Alternative: precision at top k
    - Percentage of predictions in the user's top k withheld ratings