INTRODUCTION TO DATA SCIENCE

This lecture is
based on course by E. Fox and C. Guestrin, Univ of Washington

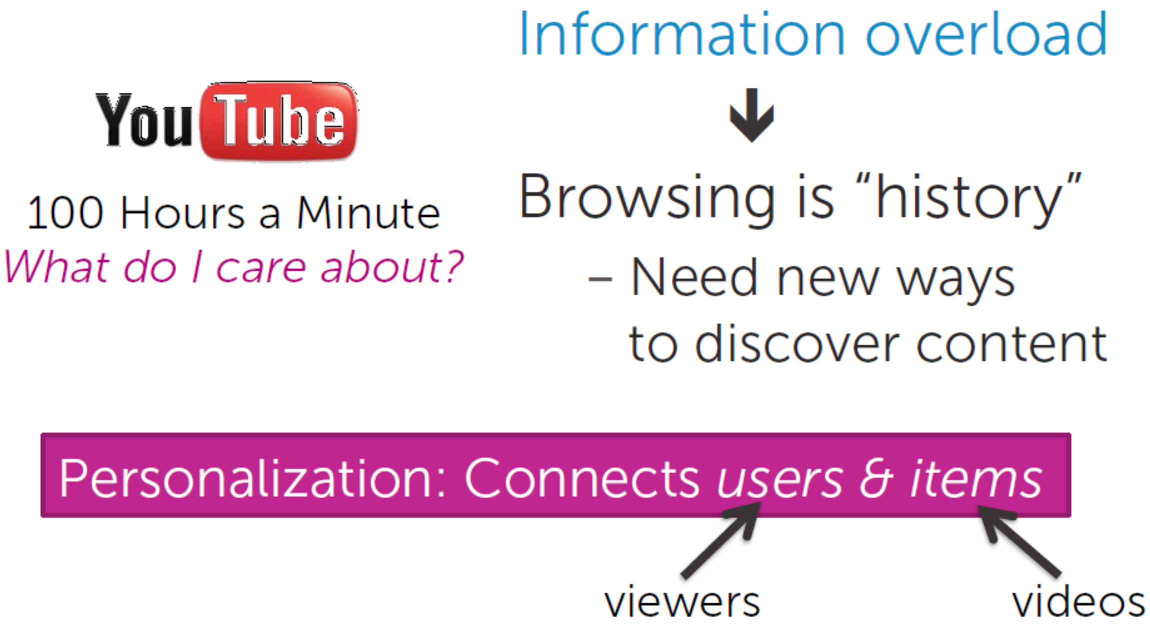
4/01/2023

WFAiS UJ, Informatyka Stosowana
I stopień studiów

Recommending system: films

Machine learning: recommending system

- □ Personalizacja



100 Hours a Minute

What do I care about?

Information overload

↓

Browsing is "history"

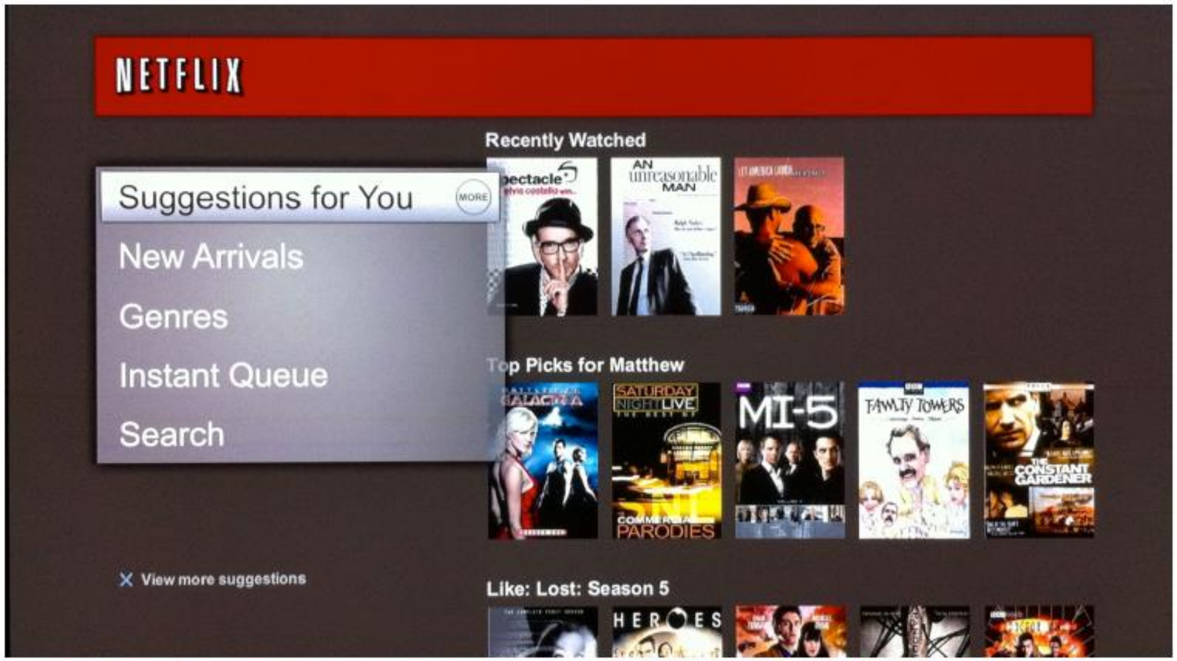
– Need new ways to discover content

Personalization: Connects *users & items*

viewers

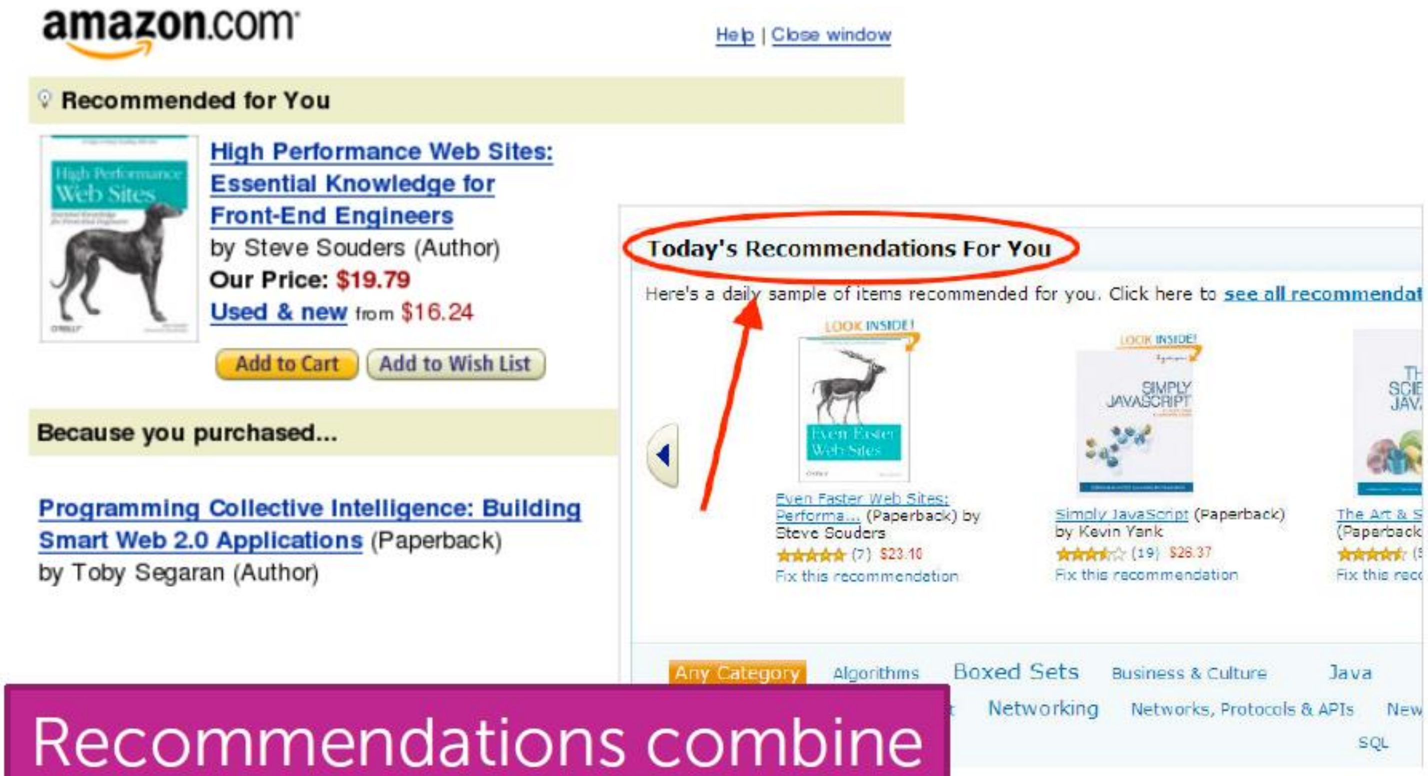
videos

Recomending system:



Connect users with movies they may want to watch

Recomending system:



amazon.com

Help | Close window

Recommended for You

High Performance Web Sites: Essential Knowledge for Front-End Engineers by Steve Souders (Author)
Our Price: $19.79
Used & new from $16.24

Because you purchased...

Programming Collective Intelligence: Building Smart Web 2.0 Applications (Paperback) by Toby Segaran (Author)

Today's Recommendations For You

Here's a daily sample of items recommended for you. Click here to see all recommendat

Even Faster Web Sites: Performa... (Paperback) by Steve Souders
(7) $23.10
Fix this recommendation

Simply JavaScript (Paperback) by Kevin Yank
(19) $26.37
Fix this recommendation

The Art & S (Paperback
Fix this reco

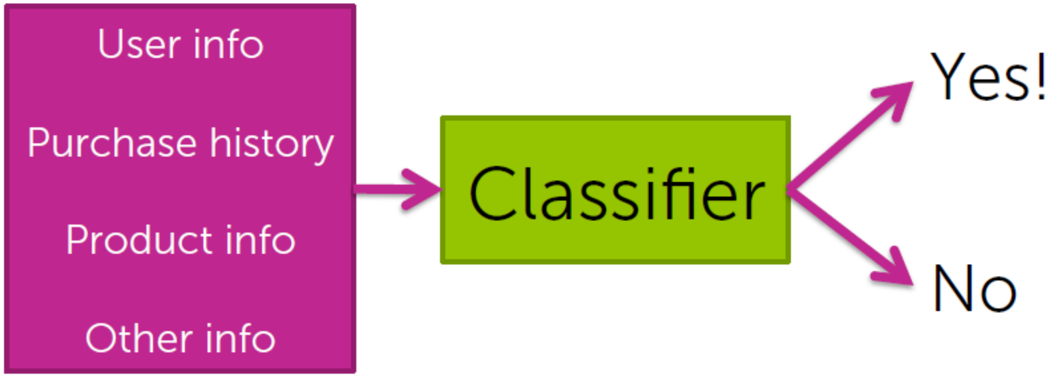
Recommendations combine global & session interests

Recommending system: popularity?

- Popularity?
  - Ranking vs number of downloading?
  - No personalisation in this case

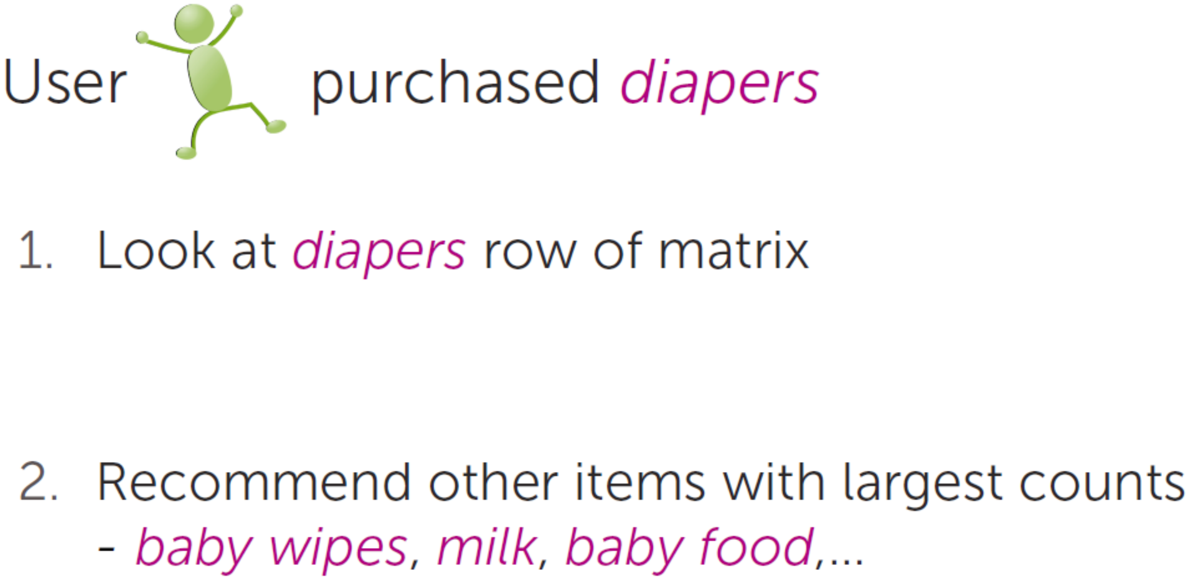
Recommending system: classification

- **Classification?**
  - **What is probability that I will buy this product?**
  - **Personalisation: purhase history, monthly and yearly trends, etc.**



Recommending system: correlations

- **Analyse correlations. Customers who bought product A also bought product B**
  - **Correlation matrix**



User purchased *diapers*

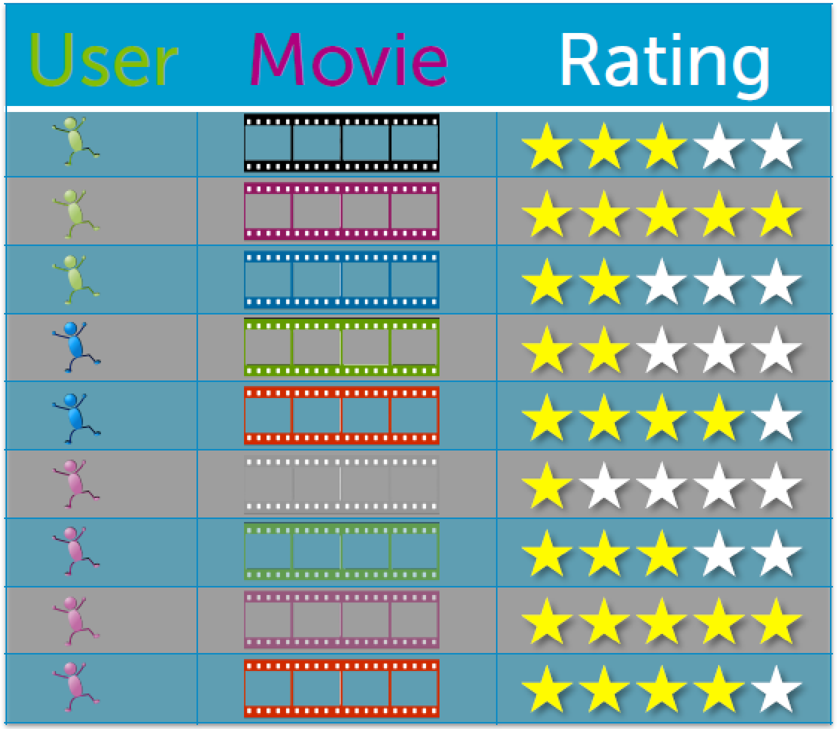
1. Look at *diapers* row of matrix
2. Recommend other items with largest counts
   - *baby wipes*, *milk*, *baby food*,...

Recommending system: correlations

- Analyse correlations. Customers who bought product A also bought product B
  - Should we normalise correlation matrix?
  - How to quantify that products are „products”?
- Limitation of correlationss:
  - It is not looking at the purhasing history (trends in time)
  - How to add a new customer (no info on correlations)?

Recommmending system: films

- Users watch movies and rate them

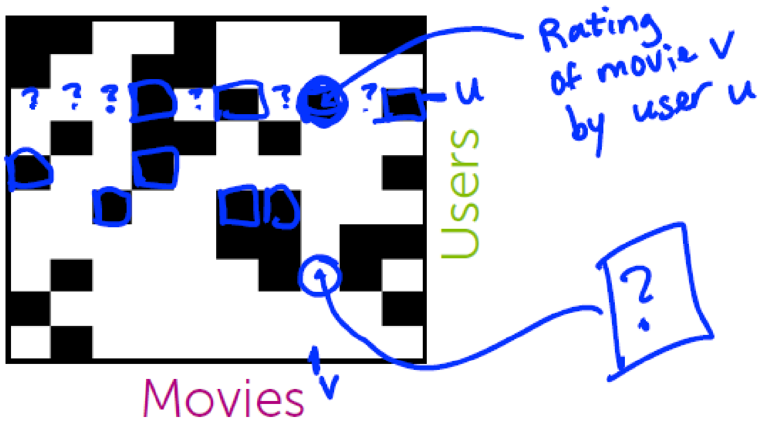


User	Movie	Rating
🟩	⬛	★★★☆☆
🟩	🟪	★★★★★
🟩	🟦	★★☆☆☆
🟦	🟩	★★☆☆☆
🟦	🟥	★★★★☆
🟪	⬜	★☆☆☆☆
🟪	🟩	★★★☆☆
🟪	🟪	★★★★★
🟪	🟥	★★★★☆

Each user only watches a few of the available movies

Recommending system: films

Rating =

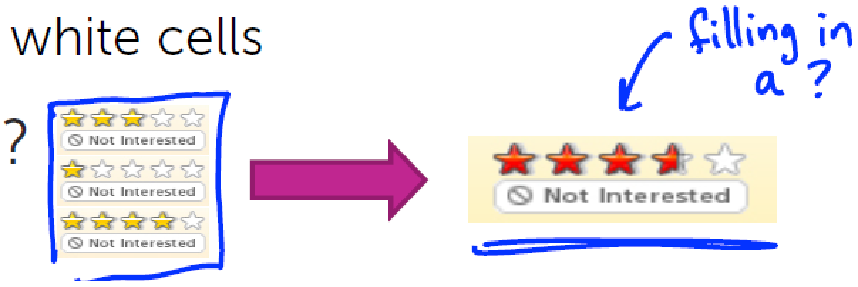


- **Data:** Users score some movies

  Rating(u,v) known for black cells

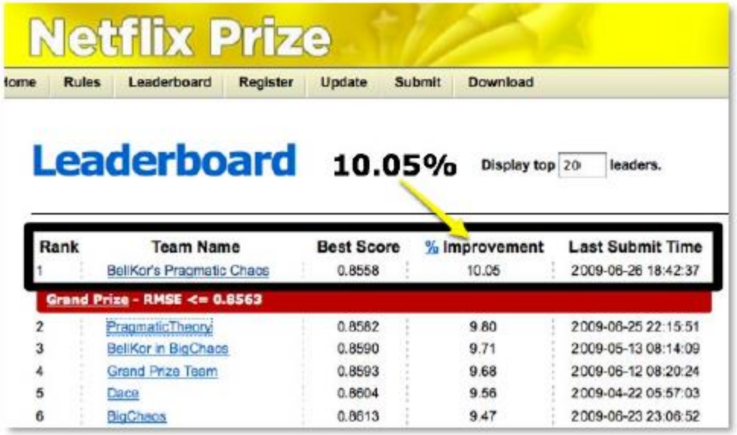
  Rating(u,v) unknown for white cells

- **Goal:** Filling missing data?



Recommending system: optimisation

- Squeezing last bit of accuracy by blending models
- Netflix Prize 2006-2009
  - 100M ratings
  - 17,770 movies
  - 480,189 users
  - Predict 3 million ratings to highest accuracy
  - Winning team blended over 100 models



Leaderboard 10.05% Display top 20 leaders.

Rank	Team Name	Best Score	% Improvement	Last Submit Time
1	BellKor's Pragmatic Chaos	0.8558	10.05	2009-06-26 18:42:37
Grand Prize - RMSE <= 0.8563				
2	PragmaticTheory	0.8582	9.80	2009-06-25 22:15:51
3	BellKor in BigChaos	0.8590	9.71	2009-05-13 08:14:09
4	Grand Prize Team	0.8593	9.68	2009-06-12 08:20:24
5	Dace	0.8604	9.56	2009-04-22 05:57:03
6	BigChaos	0.8613	9.47	2009-06-23 23:06:52

Recommending system: how effective?

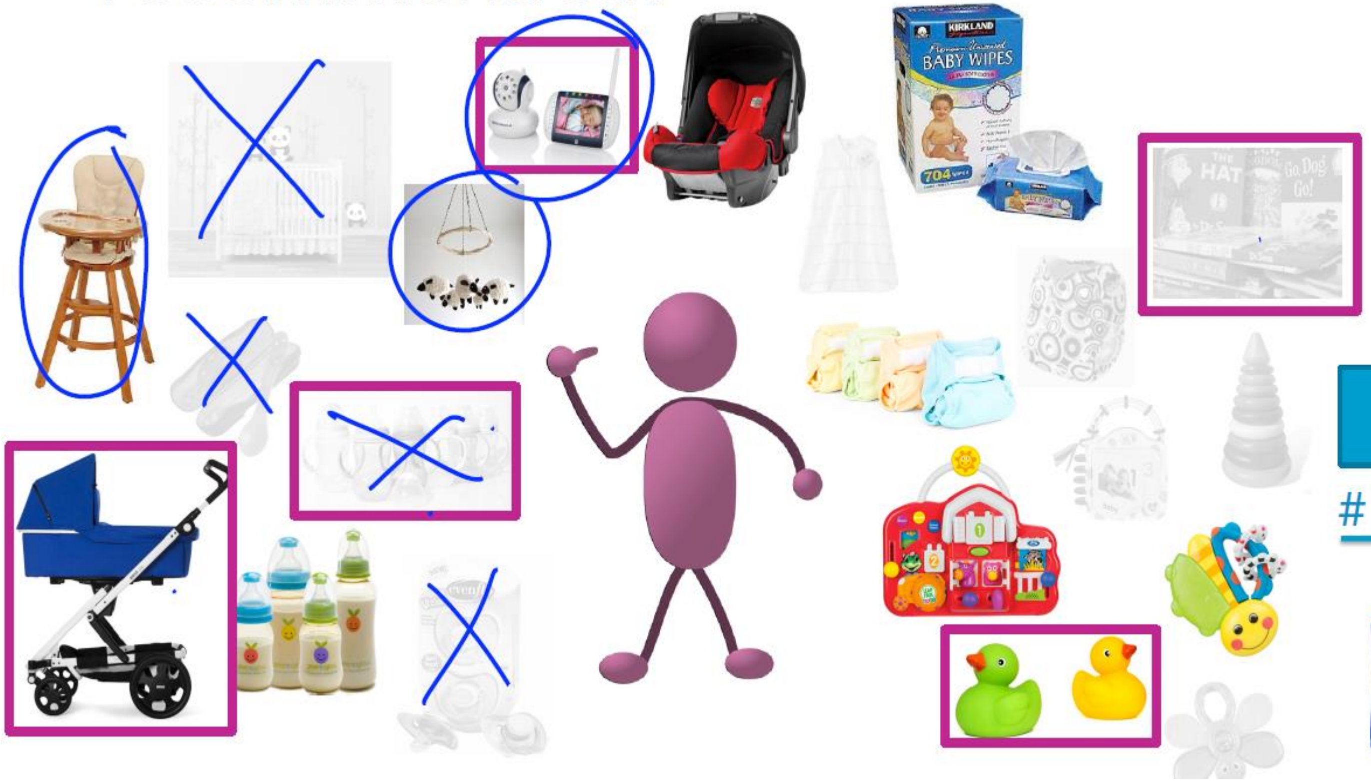
The world of all baby products

Recommending system: how effective?

User likes subset of items

Recommending system: how effective?

How many liked items were recommended?

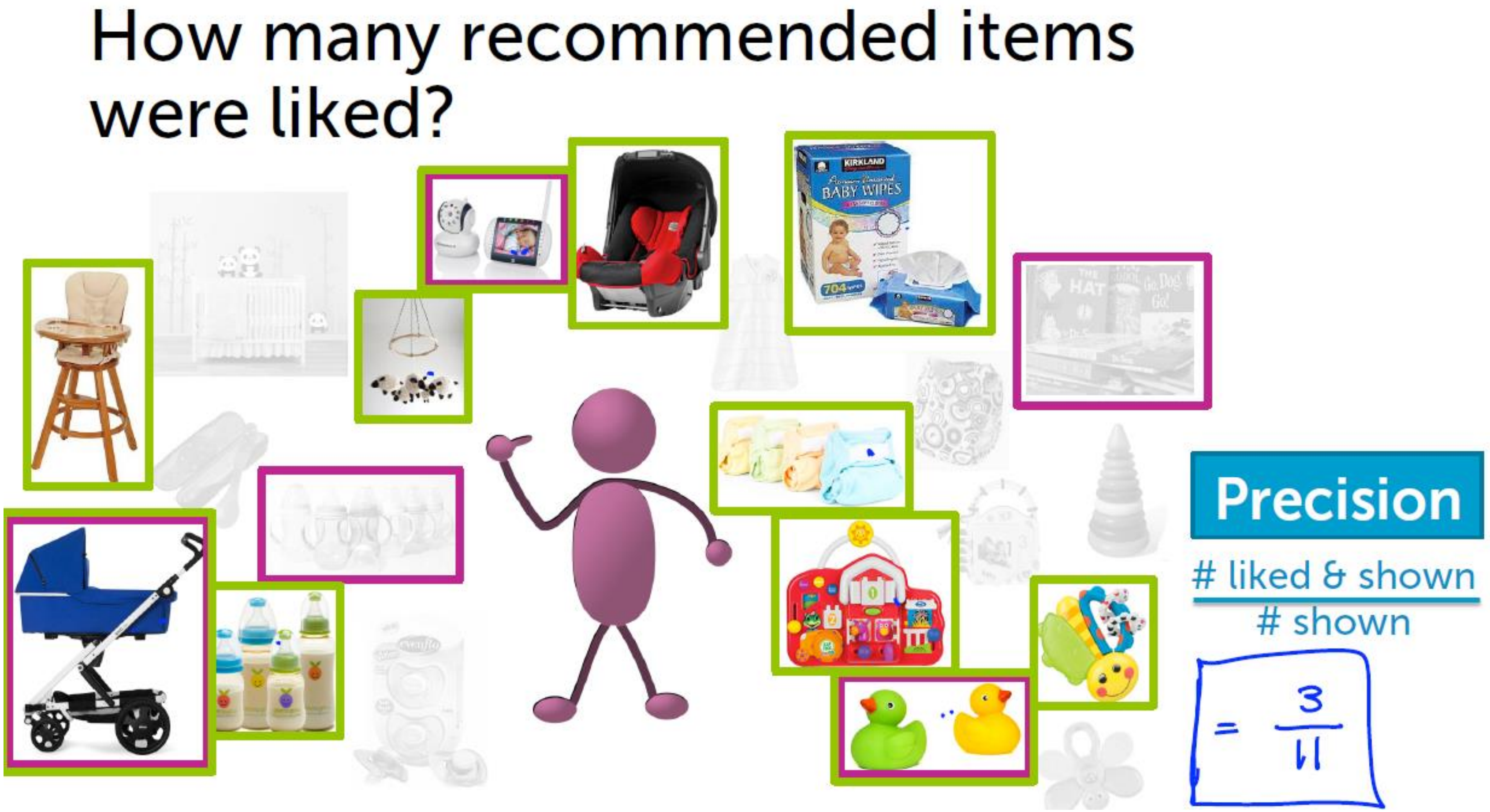


Recall

$$\frac{\text{\# liked \& shown}}{\text{\# liked}}$$

$$= \frac{3}{5}$$

Recommending system: how effective?



How many recommended items were liked?

Precision

$$\frac{\#\text{ liked \& shown}}{\#\text{ shown}}$$

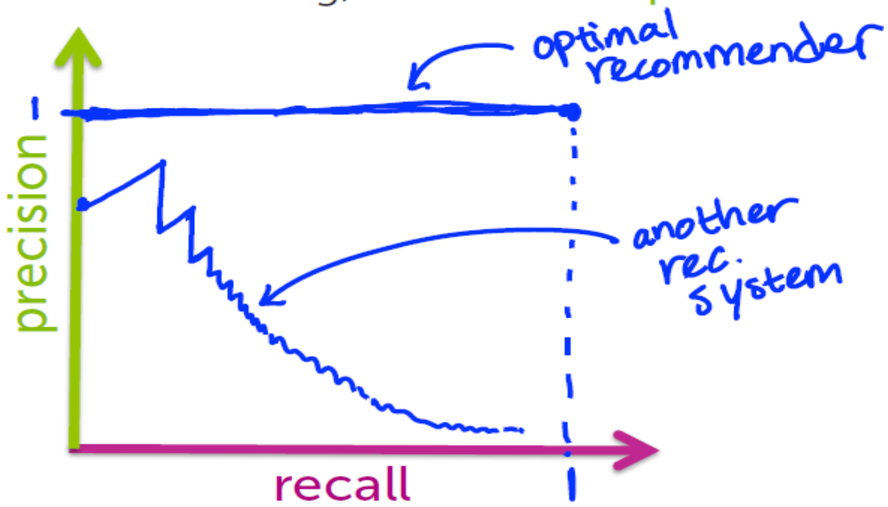
$$\boxed{= \frac{3}{11}}$$

Recommending system: how effective?

Precision-recall curve

- **Input:** A specific recommender system
- **Output:** Algorithm-specific precision-recall curve

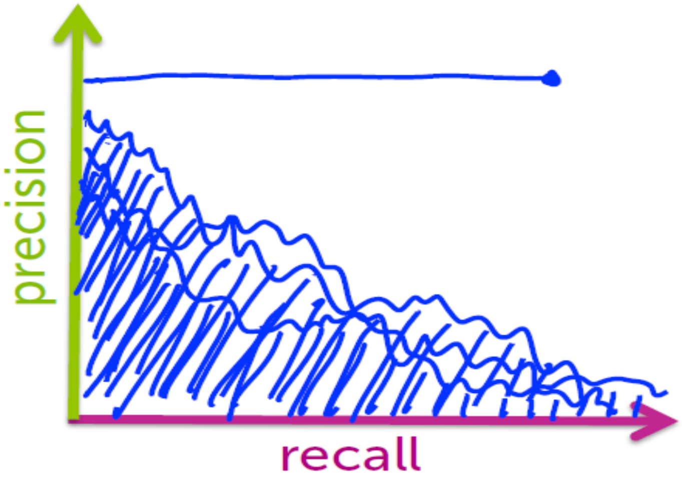
- To draw curve, vary threshold on # items recommended
  - For each setting, calculate the precision and recall



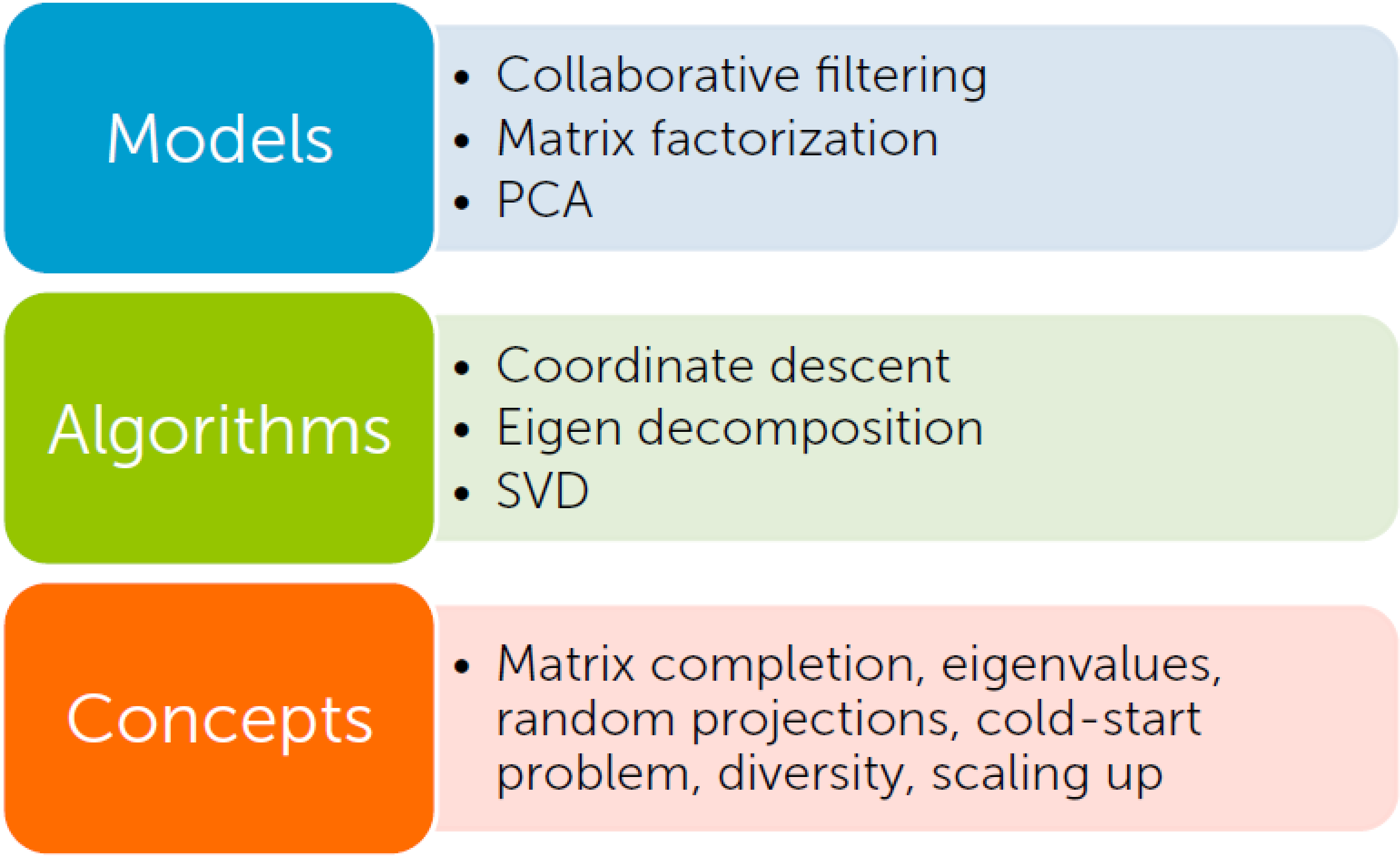
Recommending system: how effective?

Which Algorithm is Best?

- For a given precision, want recall as large as possible (or vice versa)
- One metric: largest area under the curve (AUC) ☆
- Another: set desired recall and maximize precision (precision at k)



Recommending system



Models

- Collaborative filtering
- Matrix factorization
- PCA

Algorithms

- Coordinate descent
- Eigen decomposition
- SVD

Concepts

- Matrix completion, eigenvalues, random projections, cold-start problem, diversity, scaling up