

Recommending Products: Matrix Factorization

STAT/CSE 416: Intro to Machine Learning
Emily Fox
University of Washington
May 22, 2018

©2018 Emily Fox

Solution 0: Popularity

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Simplest approach: Popularity

What are people viewing now?

- Rank by global popularity

Limitation:

- No personalization

MOST POPULAR

E-MAILED | BLOGGED | SEARCHED

1. Really?: The Claim: Lack of Sleep Increases the Risk of Catching a Cold.
2. Magazine Preview: Coming Out in Middle School
3. Yes, We Speak Cupcake
4. Gossamer Silk, From Spiders Spun
5. Tie to Pets Has Germ Jumping to and Fro
6. Maureen Dowd: Where the Wild Thing Is
7. Maureen Dowd: Blue Is the New Black
8. The Holy Grail of the Unconscious
9. For Opening Night at the Metropolitan, a New Sound: Booming
10. Economic Scene: Medical Malpractice System Breeds More Waste

Go to Complete List »

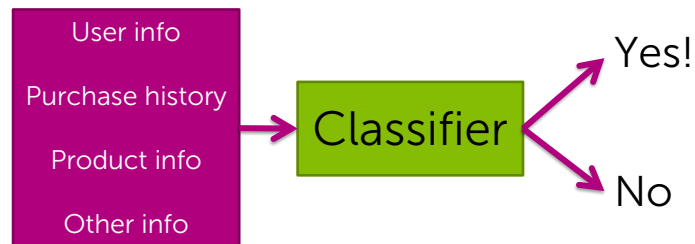
☒ CUSTOMIZE HEADLINES

Create a personalized list of headlines based on your interests. [Get Started »](#)



Solution 1: Classification model

What's the probability I'll buy this product?



Pros:

- **Personalized:**
Considers user info & purchase history
- **Features can capture context:**
Time of the day, what I just saw,...
- **Even handles limited user history:**
Age of user, ...

Cons:

- Features may not be available
- Often doesn't perform as well as **collaborative filtering** methods (next)

5

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

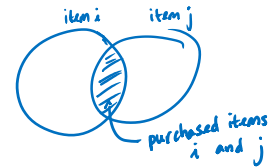
Solution 2: People who bought this
also bought...

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Co-occurrence matrix

$$\frac{\# \text{ purchased } i \text{ and } j}{\# \text{ purchased } i \text{ or } j}$$

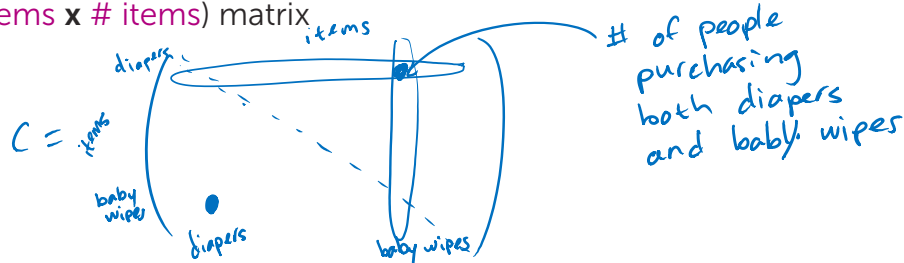


- People who bought *diapers* also bought *baby wipes*

- Matrix C:**

store # users who bought both items i & j

- (# items x # items) matrix



- **Symmetric:** # purchasing i & j same as # for j & i ($C_{ij} = C_{ji}$)

7

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

(Weighted) Average of purchased items

User  bought items {*diapers*, *milk*}

- Compute user-specific score for each item j in inventory by combining similarities:

$$\text{Score}(\text{stick figure}, \text{baby wipes}) = \frac{1}{2} (S_{\text{baby wipes}, \text{diapers}} + S_{\text{baby wipes}, \text{milk}})$$

- Could also weight recent purchases more

Sort $\text{Score}(\text{stick figure}, j)$ and find item j with highest similarity

8

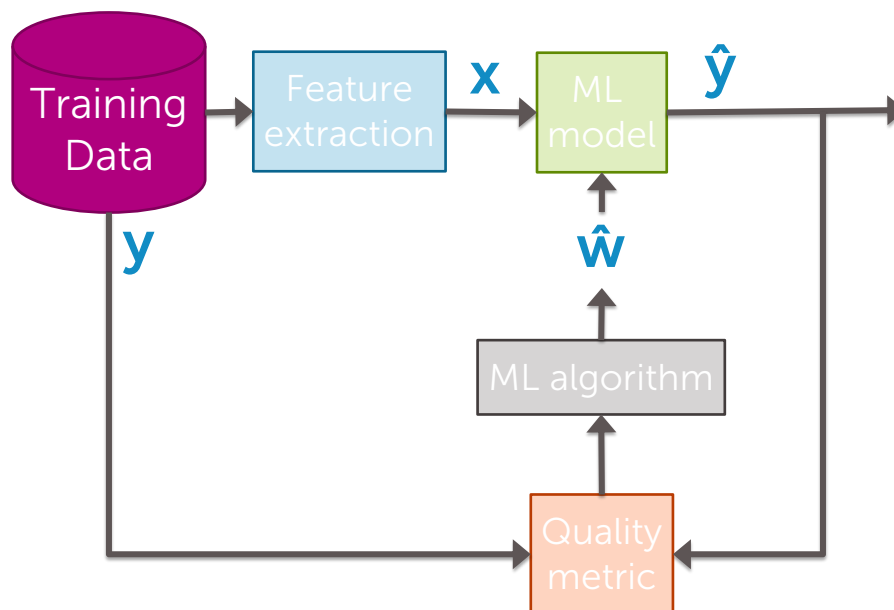
©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Solution 3: Discovering hidden structure by matrix factorization

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning



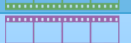
10

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Movie recommendation

Users watch movies and rate them

User	Movie	Rating
		
		
		
		
		
		
		
		
		

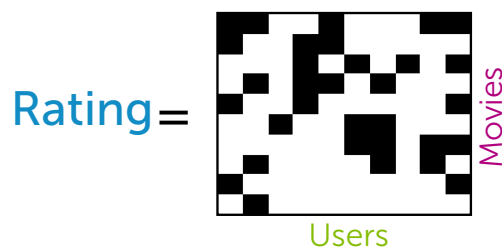
Each user only watches a few of the available movies

11

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Matrix completion problem



Data: Users score some movies

$Rating(u,v)$ known for black cells

$Rating(u,v)$ unknown for white cells

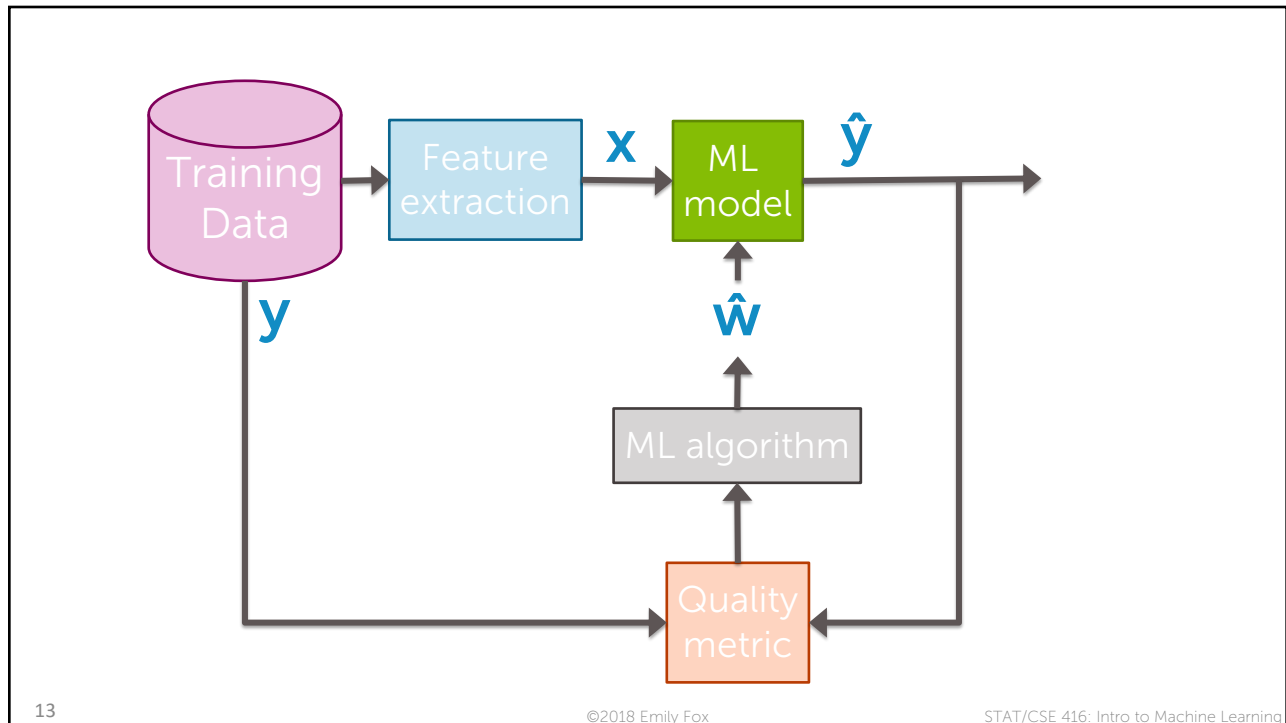
Goal: Filling missing data?



12

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning



Suppose we had d topics for each user & movie

- Describe movie v  with topics R_v
 - How much is it **action**, **romance**, **drama**,...
- Describe user u  with topics L_u
 - How much she likes **action**, **romance**, **drama**,...
- $\widehat{Rating}(u, v)$ is the product of the two vectors
- Recommendations:** sort movies user hasn't watched by $\widehat{Rating}(u, v)$

Predictions in matrix form

$$\text{Rating} = \text{Matrix} \approx \mathbf{L} \mathbf{R}'$$

But we don't know topics of users and movies...

15

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Matrix factorization model: Discovering topics from data

$$\text{Rating} = \text{Sparse Matrix} \approx \mathbf{L} \mathbf{R}'$$

Parameters of model

- Only use observed values to estimate “topic” vectors $\hat{\mathbf{L}}_u$ and $\hat{\mathbf{R}}_v$
- Use estimated $\hat{\mathbf{L}}_u$ and $\hat{\mathbf{R}}_v$ for recommendations

Many efficient algorithms for factorization

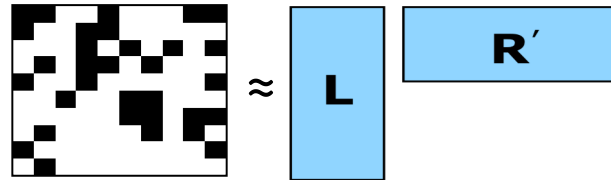
16

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Is the problem well posed?

Can we uniquely identify the latent factors?



If r_{uv} is described by L_u, R_v what happens if we redefine the “topics” as

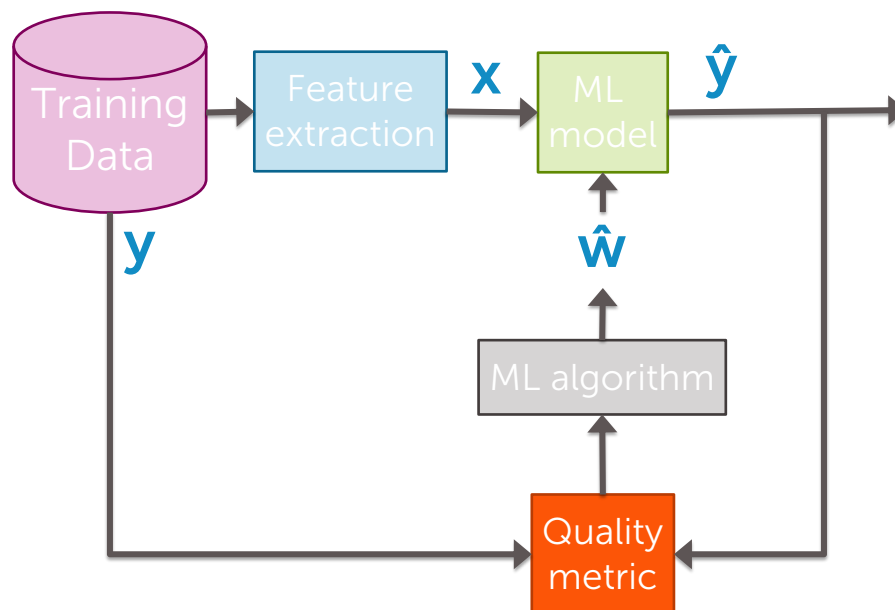
Then,

Other (orthonormal) transformations can have the same effect.

17

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

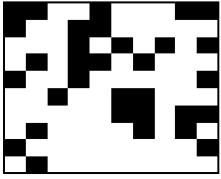


18

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Matrix factorization objective

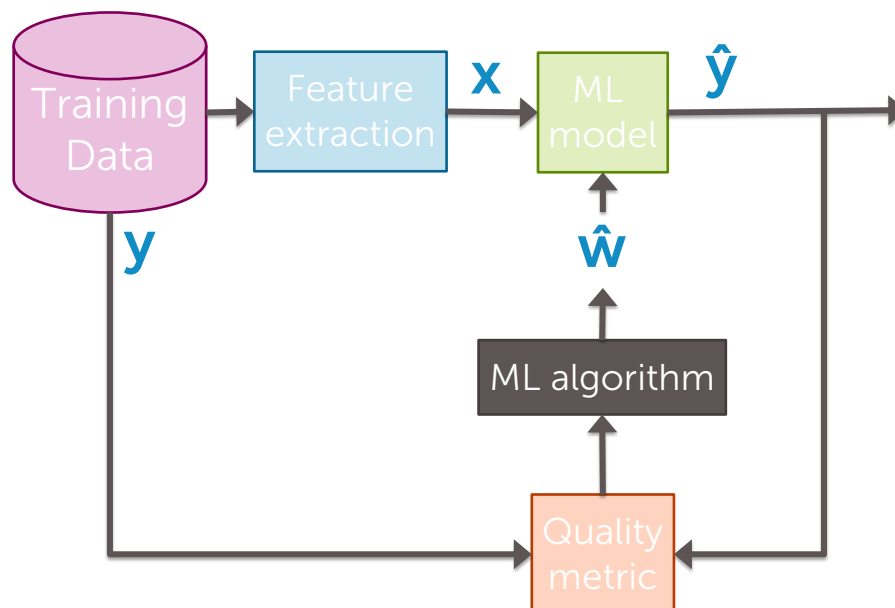
Rating =  $\approx$   $\nwarrow$ Parameters of model

- Minimize mean squared error:
 - (Other loss functions are possible)
- Non-convex objective

19

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning



20

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Coordinate descent

Goal: Minimize some function g

Often, hard to find minimum for all coordinates, but **easy for each coordinate**

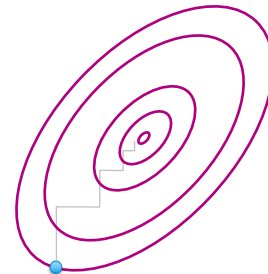
Coordinate descent:

Initialize $\hat{\mathbf{w}} = 0$ (or smartly...)

while not converged

pick a coordinate j

$\hat{w}_j \leftarrow$



21

©2017 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Comments on coordinate descent

How do we pick next coordinate?

- At random ("random" or "stochastic" coordinate descent), round robin, ...

No stepsize to choose!

Super useful approach for *many* problems

- Converges to optimum in some cases (e.g., "strongly convex")

22

©2017 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Coordinate descent for matrix factorization

$$\min_{L,R} \sum_{(u,v): r_{uv} \neq ?} (L_u \cdot R_v - r_{uv})^2$$

- Fix movie factors R_v , optimize for user factors L_u
- First key insight:

23

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Minimize objective separately for each user

- For each user u : $\min_{L_u} \sum_{v \in V_u} (L_u \cdot R_v - r_{uv})^2$
- Second key insight:

24

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Overall coordinate descent algorithm

$$\min_{L,R} \sum_{(u,v): r_{uv} \neq ?} (L_u \cdot R_v - r_{uv})^2$$

- Fix movie factors, optimize for user factors
 - Independent least-squares over users

$$\min_{L_u} \sum_{v \in V_u} (L_u \cdot R_v - r_{uv})^2$$

- Fix user factors, optimize for movie factors
 - Independent least-squares over movies

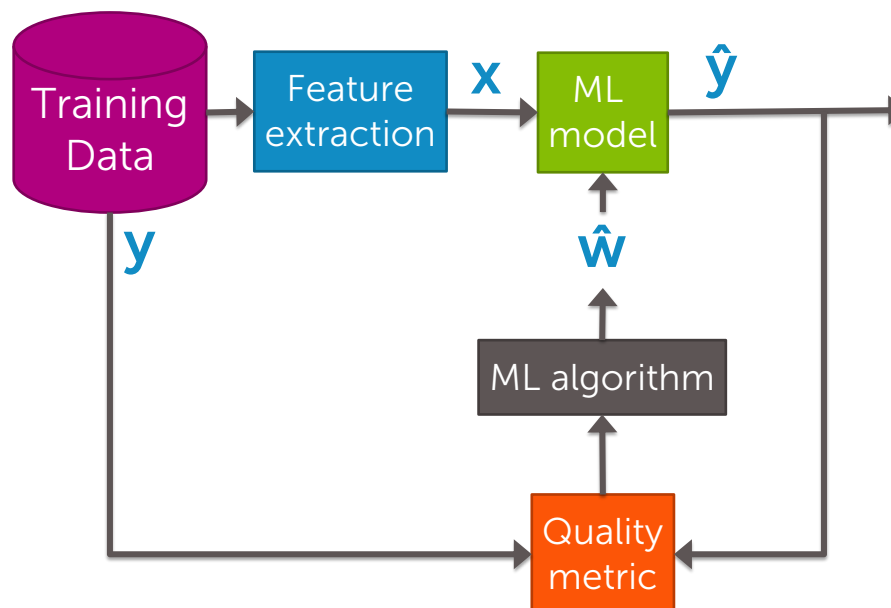
$$\min_{R_v} \sum_{u \in U_v} (L_u \cdot R_v - r_{uv})^2$$

- System may be underdetermined:
- Converges to
- Choices of regularizers and impact on algorithm:

25

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning



26

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Using the results of matrix factorization

- Discover “topics” R_v for each movie v
- Discover “topics” L_u for each user u
- $Score(u,v)$ is the product of the two vectors →
Predict how much a user will like a movie
- Recommendations: sort movies user hasn’t watched by $Score(u,v)$

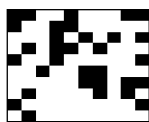
27

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Example topics discovered from Wikipedia

Application to text data:



≈

 R'

party/law
government
election court
president elected
council general minister
political national members
committee united office federal
member house parliament vote
public elections democratic held

son died
this
married family
king daughter john
death william father
born wife royal ireland
irish henry house lord
charles sir prince brother
children england queen duke
thomas years marriage george
earl edward english second

school students
university high college schools
education year program studies
campus community programs training
center members academic research peers
public systems association courses arts
educational include class institute
department teachers colleges classes
office activities universities classes
engineering learning founded faculty gis
sports children large international board
teaching academy secondary established

york county
american united
city washington john
texas served virginia
pennsylvania war moved ohio
chicago william carolina north
florida illinois george james died
massachusetts president
named jersey born boston south
john and carmine gennaro and jasper
indigenous first years philadelphia white

season team
game league games
played coach football
record teams baseball field year
second career play basketball
hockey three yards won
cavaliers lost season power threat audience
disappointment seasons players dark high
low ranked national led all third major
franchise stadium franchise head playing miss
history runs touchdown signed

album band
song released
music songs single records
recorded rock bands release
live tour video record albums
label group recording
over version tracks number featured time
about led us top performed double played
singles sound low pop artist sets of debut
single white members released early
second bass

century king
roman empire greek
production built engines
vehicle class models
speed vehicles designed
produced power front system
version type series motor rear
standard gun company
introduced range ford sold
since when cars first history machine
developed based replaced wheels time
concern small high weight electric body
being used easily

art museum work
works artists collection design
arts painting artist gallery
paintings exhibition style fine
including painted architecture york fashion
history contemporary collection years
musical as needed images time photography
figures academy exhibitions modern
period photography began studio drawing
include exhibited produced designed
period visual

radio station
news television
channel broadcast
stations network media tv
broadcasting time format local
program bbc programming live
on morning host began sports live at radio
and hosted coverage music live funding
daily channels digital also aired changed
current local communication
programme day broadcasts moved the
genre activity talk right

engine car
design model cars
production built engines
vehicle class models
speed vehicles designed
produced power front system
version type series motor rear
standard gun company
introduced range ford sold
since when cars first history machine
developed based replaced wheels time
concern small high weight electric body
being used easily

age 18 population
income average years
median living 65 males
females households 100 family
people families older town size
city household miles density
american township total area county
times census 2000 square 45 55 64
children 24 41 white female and including
white housing income individuals housing
poverty united village

music musical opera
festival orchestra dance
performed jazz piano theatre
performance works concert
symphony orchestra stage performance
instrumental musicians classical including
work composers major opera songs folk
instrument ballet contemporary composers
also performing currently playing stage
years include popular choir assemblies
sound alpha three acts had piano chamber
recordings string

war army military
forces battle force british
command general navy ship
division ships troops corps
service naval regiment
commander infantry attack men
officer naval soldiers units officers
operations unit june august brigade july the
leading month battleship and operation
captain september three enemy united
coldest new troop german machine major

white red
black blue called
color will head green gold side
small hand long arms top flag
horse wear silver common light
dog wood body type large
yellow worn worn dog not popular left
currently traditional hat front horses shoes
hat feet clock time coat three typically
modern face bones

28

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Bringing it all together: Featurized matrix factorization

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Limitations of matrix factorization

- Cold-start problem
 - This model still cannot handle a new user or movie



30

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Cold-start problem more formally

Consider a **new user** u' and predicting that user's ratings

- No previous observations
- Objective considered so far:

$$\min_{L,R} \frac{1}{2} \sum_{r_{uv}} (L_u \cdot R_v - r_{uv})^2 + \frac{\lambda_u}{2} \|L\|_F^2 + \frac{\lambda_v}{2} \|R\|_F^2$$

- Optimal user factor:
- Predicted user ratings:

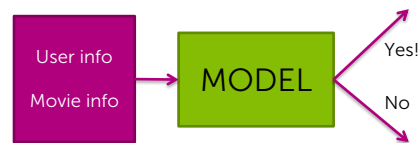
31

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Combining features and discovered topics

- Features capture **context**
 - *Time of day, what I just saw, user info, past purchases,...*
- Discovered topics from matrix factorization capture **groups of users** who behave similarly
 - *Women from Seattle who teach and have a baby*
- **Combine** to mitigate cold-start problem
 - Ratings for a new user from **features** only
 - As more information about user is discovered, matrix factorization **topics** become more relevant



32

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Collaborative filtering with specified features

- Create feature vector for each movie (often have this even for new movies):
- Define weights on these features for how much **all users** like each feature
- Fit linear model:
- Minimize:

33

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Building in personalization

- Of course, users do *not* have identical preferences
- Include a **user-specific deviation** from the global set of user weights:
- If we don't have any observations about a user, **use wisdom of the crowd**
- As we gain more information about the user, **forget the crowd**
- Can add in **user-specific features**, and cross-features, too

34

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Featurized matrix factorization— A combined approach

Feature-based approach:

- Feature representation of user and movies fixed
- Can address cold-start problem

Matrix factorization approach:

- Suffers from cold-start problem
- User & movie features are learned from data

A unified model:

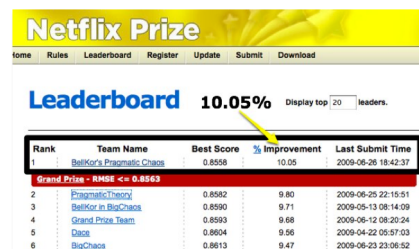
35

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Blending models

- 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**



The screenshot shows the Netflix Prize Leaderboard with a yellow header. Below the header, there are tabs for Home, Rules, Leaderboard, Register, Update, Submit, and Download. The Leaderboard tab is selected, showing a table of top teams. A yellow arrow points to the '10.05%' improvement mark. The table has columns for Rank, Team Name, Best Score, % Improvement, and Last Submit Time.

Rank	Team Name	Best Score	% Improvement	Last Submit Time
1	BellKor's Pragmatic Chaos	0.8556	10.05	2009-06-26 18:42:37
Grand Prize = RMSE < 0.8563				
2	Prizewinning Theory	0.8582	9.80	2009-06-25 22:16:51
3	Random 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	Dalco	0.8604	9.56	2009-04-22 05:57:03
6	BigChaos	0.8613	9.47	2009-06-23 23:06:52

36

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

A performance metric for
recommender systems

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

The world of all baby products



38

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Optimal recommender



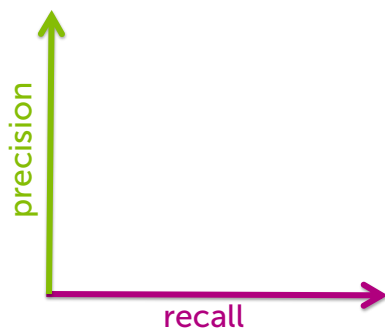
45

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

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**



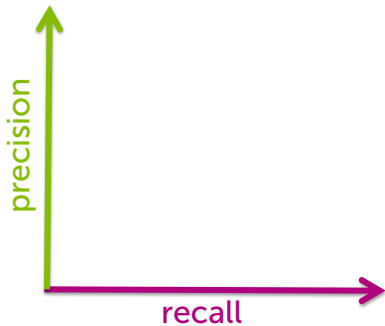
46

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

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**)



47

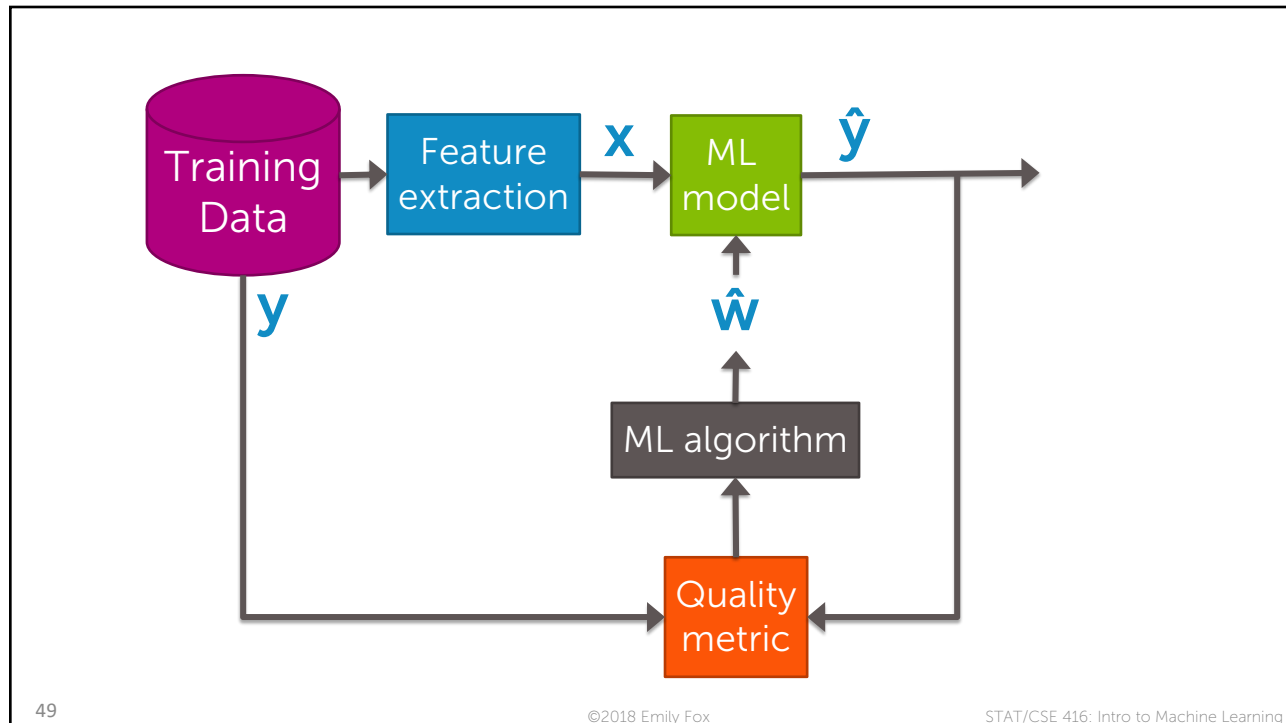
©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning

Summary of recommender systems

©2018 Emily Fox

STAT/CSE 416: Intro to Machine Learning



What you can do now...

- Describe the goal of a recommender system
- Provide examples of applications where recommender systems are useful
- Implement a co-occurrence based recommender system
- Describe the input (observations, number of “topics”) and output (“topic” vectors, predicted values) of a matrix factorization model
- Implement a coordinate descent algorithm for optimizing the matrix factorization objective presented
- Exploit estimated “topic” vectors to make recommendations
- Describe the cold-start problem and ways to handle it (e.g., incorporating features)
- Analyze performance of various recommender systems in terms of precision and recall
- Use AUC or precision-at-k to select amongst candidate algorithms