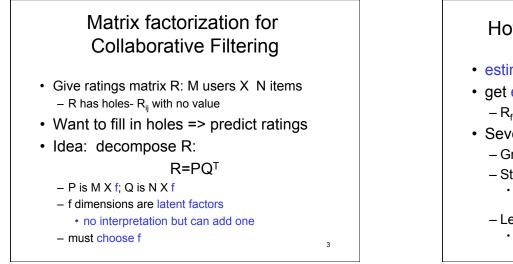


Matrix factorization motivation

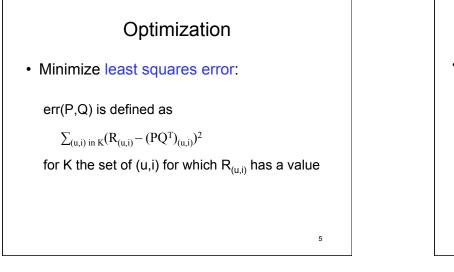
- Matrix representation
 - users X items
 - documents X terms
- Discover/use latent factors – attributes, topics, features
- · Factor matrices to uncover latent factors
- Don't know what latent factors represent
 can conjecture
- For recommenders, matrix has holes – use factorization to fill in

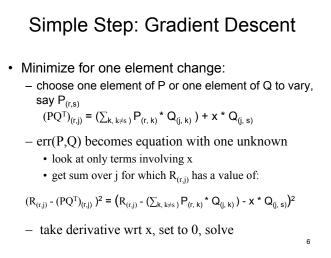


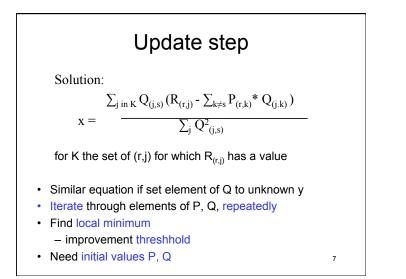
How does decomposition help?

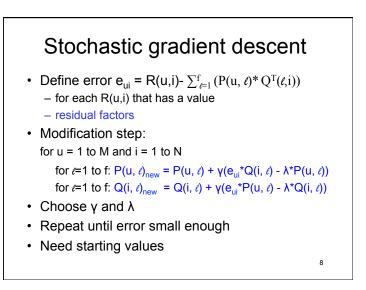
2

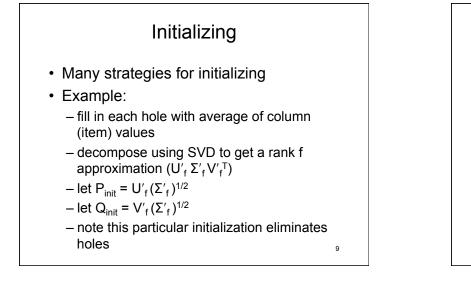
- estimate P and Q, leaving no holes
- get estimate of R as R_f = PQ^T
 R_f has holes of R filled in
- Several methods for estimation, e.g.
 - Gradient descent
 - Stochastic gradient descent
 - Koren et al. Matrix Factorization Techniques for Recommender Systems, IEEE Computer, Aug 2009
 - Least squares based calculations
 Bell et al Modeling Relat'ships at Multiple Scales to
 - Bell et al Modeling Relat ships at Multiple Scales to Improve Accuracy of Large Recom. Sys., KDD Aug 2007. 4











Matrix factorization: summary

- · Very effective method
- · Issues:
 - Iteration is costly
 - Wait for local optimum?
 - Must choose initial values
- · Subject of ongoing research

High-level issues for Collaborative Filtering: Global effects

Effects over many or all of ratings

- ✓ different users have different rating scales
- metadata (attributes) for items and/or users
 hybrid content/collaborative
- date of rating
- · trend of user's ratings over time
- · trend of item's ratings over time

Reference: Scalable Collaborative Filtering w/ Jointly Derived Neighborhood Interpolation Weights, Bell and Koren, *IEEE Intern. Conf. Data Mining* (part of winning Netflix contest team)¹

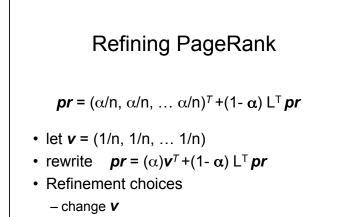
Final thought

All techniques we've seen behavior or topic oriented

What about links? What about PageRank?

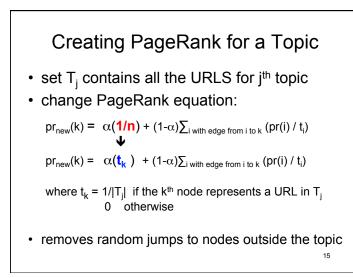
12

10



13

– change L



"Topic Sensitive" PageRank Haveliwala Use pre-defined topics Open Directory Project (DMOZ) "the largest, most comprehensive human-edited directory of the Web." 16 top-level topics 16 top-level topics Degree to which page is part of topic Calculate similarity of query to each topic Use linear combination of topic PageRanks based on similarity values query to topic

Personalized PageRank Kamvar et. al. Random leaps are biased by personal interests – change v Combined with use of block structure to make more efficient: Divide Web graph into blocks (clusters) Use high-level domains (e.g. princeton.edu) Calc. local PageRank within each block Collapse each block into 1 node – new graph Weighted edges between nodes Calc. PageRank with biased leaps for block structure Weight local PageRanks with block PageRank Use to initialize power calculation

Refinement & Personalization Summary

- · Looked at several techniques to modify search
- explicit user feedback
- user behavior: history
 - user history
 - crowd history
 - collaborative history: "people like you"
- role of social networks
 - general analysis
 - relationships
- models of recommender systems

17