

## Lecture 14 Recommender Systems

- Goals recommend something to users based off data
- Challenges
  - Scale : millions of users and things to rate
  - Cold Start: change in content or users interest
  - Sparse Data : not a lot of users take the time to rate things
- How to predict ratings (using example of movies)
  - Data exists for both users and movies
    - Neighborhood method
      - (user, user) similarity measure
        - i.e. recommend same movies to similar users (requires info about users)
      - (item, item) similarity measure
        - recommend movies that are similar (requires info about movies)
      - Classification tools using user features to predict movie ratings
      - Pros:
        - Intuitive /easy to explain
        - No training required
      - Cons:
        - Users rate differently
        - Ratings change over time
        - Bias
  - Data only exists for movies
    - Content based filtering
      - Assume you have features for movies and learn the features for the users
      - Category, genre ,
      - Use these to get a feature-to-movie similarity matrix and a user-to-feature similarity matrix
      - Multiply them to get the rating
        - $\text{user-to-feature} \times \text{feature-to-movie} = \text{user-to-movie} = R_{ij}$
  - Only have access to ratings
    - Collaborative filtering
      - Challenge: how do we get the correct features?
      - Formulate optimization problem to solve
        - 1. Start with random Q
        - 2. Get P
        - 3. Improve Q
        - 4. Repeat 2 & 3