Case Study 4: Collaborative Filtering

Probabilistic Models for Matrix Factorization

Cold-Start Problem

Interpreting Low-Rank Matrix Completion (aka Matrix Factorization)

\[
X = \underbrace{L}_{\text{movies}} \underbrace{R'}_{\text{topics}}
\]

©Carlos Guestrin 2013
Stochastic Gradient Descent

\[ \min_{L,R} \sum_{u,v} \left( \frac{1}{2} (L_u \cdot R_v - r_{uv})^2 + \frac{\lambda_u}{2} \|L_u\|_F^2 + \frac{\lambda_v}{2} \|R_v\|_F^2 \right) \]

- Observe one rating at a time \( r_{uv} \): \( \epsilon_t = L_t \cdot R_t - r_{uv} \)
- Gradient observing \( r_{uv} \):
  \[ \frac{\partial F}{\partial L_u} = \epsilon_t R_v + \lambda_u L_u \]
  \[ \frac{\partial F}{\partial R_v} = \epsilon_t L_u + \lambda_v R_v \]
- Updates:
  \[ \begin{bmatrix} L_{u(t+1)} \\ R_{v(t+1)} \end{bmatrix} \leftarrow \begin{bmatrix} (1 - \eta_t \lambda_u) L_{u(t)} - \eta_t \epsilon_t R_{v(t)} \\ (1 - \eta_t \lambda_v) R_{v(t)} - \eta_t \epsilon_t L_{u(t)} \end{bmatrix} \]

What’s Matrix Factorization Optimizing???

- A generative process:
  - Pick user factors
  - Pick movie factors
  - For each (user,movie) pair observed:
    - Pick rating as \( L_u R_v + \text{noise} \)
Maximum A Posteriori for Matrix Completion

\[
P(L, R | X) \propto P(L, R, X) \\
\propto e^{\frac{-1}{2\sigma_u^2} \sum_{u=1}^{n} \sum_{i=1}^{k} L_u^2} e^{\frac{-1}{2\sigma_v^2} \sum_{v=1}^{m} \sum_{i=1}^{k} R_v^2} e^{\frac{-1}{2\sigma_e^2} \sum_{r_{uv}} (L_u \cdot R_v - r_{uv})^2}
\]

MAP versus Regularized Least-Squares for Matrix Completion

- **MAP under Gaussian Model:**

  \[
P(L, R | X) \propto P(L, R, X) \\
  \propto e^{\frac{-1}{2\sigma_u^2} \sum_{u=1}^{n} \sum_{i=1}^{k} L_u^2} e^{\frac{-1}{2\sigma_v^2} \sum_{v=1}^{m} \sum_{i=1}^{k} R_v^2} e^{\frac{-1}{2\sigma_e^2} \sum_{r_{uv}} (L_u \cdot R_v - r_{uv})^2}
\]

- **Least-squares matrix completion with L\(_2\) regularization:**

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

- **Understanding as a probabilistic models is very useful! E.g.,**
  - Change priors
  - Incorporate other sources of information or dependencies
Cold-Start Problem

- **Challenge:** Cold-start problem (new movie or user)
- **Methods:** use features of movie/user

Cold-Start More Formally

- No observations about a particular user:
  \[ \min_{L,R} \frac{1}{2} \sum_{u,v} (L_u \cdot R_v - r_{uv})^2 + \frac{\lambda_u}{2} ||L||_F^2 + \frac{\lambda_v}{2} ||R||_F^2 \]

- A simpler model for collaborative filtering:
  - Observe ratings:
  - Given features of a movie:
  - Fit linear model:
  - Minimize:
Personalization

- If we don’t have any observations about a user, use wisdom of the crowd
  - Address cold-start problem

- But, as we gain more information about the user, forget the crowd:

- Graphically:

User Features…

- In addition to movie features, may have information user:

- Combine with features of movie:

- Unified linear model:
Feature-based Approach versus Matrix Factorization

- **Feature-based approach:**
  - Feature representation of user and movies fixed
  - Can address cold-start

- **Matrix factorization approach:**
  - Suffers from cold-start problem
  - User & movie features are learned from data

- **Unified model:**

MAP for Unified Collaborative Filtering via SGD

\[
\min_{L,R,w,\{w_u\}} \frac{1}{2} \sum_{u,v} (L_u \cdot R_v + (w + w_u) \cdot \phi(u, v) - r_{uv})^2 \\
+ \frac{\lambda_u}{2} ||L||_F^2 + \frac{\lambda_v}{2} ||R||_F^2 + \frac{\lambda_w}{2} ||w||_2^2 + \frac{\lambda_{wu}}{2} \sum_u ||w_u||_2^2
\]

- Gradient step observing \( r_{uv} \):
  - For \( L,R \)
    \[
    \begin{bmatrix}
    L_u^{(t+1)} \\
    R_v^{(t+1)}
    \end{bmatrix} 
    \leftarrow \begin{bmatrix}
    (1 - \eta_t \lambda_u) L_u^{(t)} - \eta_t \epsilon_t R_v^{(t)} \\
    (1 - \eta_t \lambda_v) R_v^{(t)} - \eta_t \epsilon_t L_u^{(t)}
    \end{bmatrix}
    \]

  - For \( w \) and \( w_u \):
What you need to know…

- Probabilistic model for collaborative filtering
  - Models, choice of priors
  - MAP equivalent to optimization for matrix completion

- Cold-start problem

- Feature-based methods for collaborative filtering
  - Help address cold-start problem

- Unified approach