SC4/SM8 Advanced Topics in Statistical Machine Learning

Collaborative Filtering

Dino Sejdinovic
Department of Statistics
Oxford

Slides and other materials available at:
http://www.stats.ox.ac.uk/~sejdinov/atsml19/
Ratings and Recommendations

<table>
<thead>
<tr>
<th>movie \ user</th>
<th>Alice</th>
<th>Bob</th>
<th>Chuck</th>
<th>Dan</th>
<th>Eve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy Gilmore</td>
<td>?</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Click</td>
<td>1</td>
<td>?</td>
<td>4</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Ex Machina</td>
<td>?</td>
<td>4</td>
<td>?</td>
<td>?</td>
<td>2</td>
</tr>
<tr>
<td>Blade Runner</td>
<td>5</td>
<td>?</td>
<td>1</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>The Matrix</td>
<td>5</td>
<td>5</td>
<td>?</td>
<td>?</td>
<td>4</td>
</tr>
</tbody>
</table>

- **Data:** a partially observed matrix \( \mathbf{Y} \in \mathbb{R}^{n_1 \times n_2} \) where \( y_{i,j} \) is the rating (e.g. between 1 and 5) of item \( i \) by user \( j \).
- Most entries will be missing/unknown since most users will not have rated most movies.
- **Exposures:** \( e_{i,j} = 1 \) if the user \( j \) has rated movie \( i \) and \( e_{i,j} = 0 \) otherwise.
Features of items and users
Content based recommendation

- Each item $i$ has a feature vector $\phi_i = [\phi_{i1}, \ldots, \phi_{ik}]^T \in \mathbb{R}^k$
- If $\phi_i$ observed, simply solve one linear model per user:
  \[
  \min_{\psi_j} \sum_{i: e_{i,j} = 1} (y_{i,j} - \phi_i^T \psi_j)^2 + \lambda_{\psi} \|\psi_j\|_2^2, \quad j = 1, \ldots, n_2.
  \]

  - $\psi_j$ is the corresponding vector of coefficients in the linear model corresponding to user $j$ - can be treated as a feature (preference) vector of user $j$.
- If $\psi_j$ observed, but $\phi_i$ is hidden, solve one linear model per item:
  \[
  \min_{\phi_i} \sum_{j: e_{i,j} = 1} (y_{i,j} - \phi_i^T \psi_j)^2 + \lambda_{\phi} \|\phi_i\|_2^2, \quad i = 1, \ldots, n_1.
  \]
Collaborative Filtering

Alternating linear regressions

- Assume neither features are observed.
- Formulate recommendations solely based on the ratings matrix: alternating regression model.
- Often simply use stochastic gradient descent (SGD) updates: as soon as new rating becomes available:

\[
\begin{align*}
\phi_i & \leftarrow (1 - \epsilon_t \lambda_{\phi})\phi_i + \epsilon_t \psi_j (y_{ij} - \phi_i^T \psi_j), \\
\psi_j & \leftarrow (1 - \epsilon_t \lambda_{\psi})\psi_j + \epsilon_t \phi_i (y_{ij} - \phi_i^T \psi_j).
\end{align*}
\]

- **Collaborative**: predictions for each user can potentially depend on ratings of all other users.
- Potentially results in features/preferences which do not have a readily interpretable meaning.
Probabilistic Matrix Factorization

Introduced in [Salakhutdinov and Mnih, 2007], the generative model corresponding to CF can be described as follows:

- For each movie $i = 1, \ldots, n_1$, sample independently the latent vector of features $\phi_i \sim \mathcal{N}(0, \sigma^2_\phi I_k)$ from a $k$-dimensional normal distribution,

- For each user $j = 1, \ldots, n_2$, sample independently the latent vector of preferences $\psi_j \sim \mathcal{N}(0, \sigma^2_\psi I_k)$ from a $k$-dimensional normal distribution,

- For each movie-user pair $(i, j)$, sample $e_{i,j} \sim \text{Bernoulli}(p)$ independently and if $e_{i,j} = 1$, sample $y_{i,j}|\phi_i, \psi_j \sim \mathcal{N}(\phi_i^\top \psi_j, \sigma^2_y)$. 
Beyond Gaussian “ratings” likelihood

Binary ratings:
- Logistic link: \( p(y_{i,j}|\phi_i, \psi_j) \sim \sigma(y_{i,j}\phi_i^\top \psi_j) \)
- Probit link: \( p(y_{i,j}|\phi_i, \psi_j) \sim \Phi(y_{i,j}\phi_i^\top \psi_j) \)

“Count” ratings:
- Poisson link: \( y_{i,j} \sim \text{Poisson}(\exp(\phi_i^\top \psi_j)) \)

Example of using logistic link CF for analysing UK parliament data.