Case Study 4: Collaborative Filtering

Matrix Factorization and Probabilistic LFMNs for Network Modeling

Network Data

- Structure of network data

[Diagram of network data with nodes and edges, and an adjacency matrix]
Properties of Data Source

- Similarities to Netflix data:
  - Matrix-valued data (adj. matrix)
  - High-dimensional
  - Sparse

- Differences
  - Square
  - Binary

Differences:
- Square: same indices for rows and columns
- Binary: yes/no for link (other ext. possible... multiple)

If undirected, then matrix is symmetric

Matrix Factorization for Network Data

- Vanilla matrix factorization approach:

- What to return for link prediction?

- Slightly fancier:
Probabilistic Latent Space Models

- Assume features (covariates) of the user or relationship
- Each user has a “position” in a k-dimensional latent space
- Probability of link:

\[
\log \text{odds } p(r_{uv} = 1 \mid L_u, L_v, x_{uv}, \beta) = \beta_0 + \beta^T x_{uv} - |L_u - L_v|
\]

\[
\log \text{odds } p(r_{uv} = 1 \mid L_u, L_v, x_{uv}, \beta) = \beta_0 + \beta^T x_{uv} + |L_u^T L_v|
\]

- Bayesian approach:
  - Place prior on user factors and regression coefficients
  - Place hyperprior on user factor hyperparameters
- Many other options and extensions (e.g., can use GMM for \(L_u \rightarrow\) clustering of users in the latent space)
What you need to know...

- Representation of network data as a matrix
  - Adjacency matrix
- Similarities and differences between adjacency matrices and general matrix-valued data
- Matrix factorization approaches for network data
  - Just use standard MF and threshold output
  - Introduce link functions to constrain predicted values
- Probabilistic latent space models
  - Model link probabilities using distance between latent factors

Case Study 5: Mixed Membership Modeling

Clustering Documents Revisited,
Latent Dirichlet Allocation
Document Retrieval

- **Goal:** Retrieve documents of interest
- **Challenges:**
  - Tons of articles out there
  - How should we measure similarity?

### Task 1: Find Similar Documents

- **First considered:**
  - **Input:** Query article
  - **Output:** Set of k similar articles
Task 2: Cluster Documents

- **Then examined:**
  - Cluster documents based on topic

Document Representation

- Bag of words model

\[ X = \begin{bmatrix} \text{vector of word counts} \end{bmatrix} \]

\[ (e.g. \text{tf-idf}) \]

Previously performed operations on this vector

Now:

\[ X = \{ w_1, \ldots, w_N \} \]

unordered set of N words in document \( d \)
A Generative Model

- Documents: $x_1, \ldots, x_D$
- Associated topics: $z_1, \ldots, z_D$
- Parameters: $\theta = \{\pi, \beta\}$
- Generative model:
  
  Sample topic: $z_d \sim \pi$
  
  Sample words: $w_i^d \mid z_d \sim \beta_{z_d}$, $i = 1, \ldots, N_d$
  
  Given topic $z_d = k$ for doc $d$, draw each word from $\beta_k$
Model In Pictures

• Mixture weights (on topics)
  \( \pi \)

• Topic distributions (on words)
  \( \beta_1 \)  
  \( \beta_K \)

• For each document, 
  \( z^d \sim \pi \)  
  \( w^d_i \mid z^d \sim \beta_{z^d} \)

Bayesian Document Model

• Model parameters \( \pi, \{ \beta_k \} \) unknown

• Bayesian approach

• Need distribution on pmf’s
The Simplex in 3D

- The simplex defines the hyperplane of vectors that sum to 1

\[ 0 \leq \theta_k \leq 1 \]
\[ \sum_{k=1}^{3} \theta_k = 1 \]

Dirichlet Distributions

- The Dirichlet distribution is defined on the simplex

\[ p(\pi \mid \alpha) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_{k=1}^{K} \pi_k^{\alpha_k - 1} \]

Moments:

\[ E_\alpha [\pi_k] = \frac{\alpha_k}{\alpha_0} \]
\[ Var_\alpha [\pi_k] = \frac{K - 1}{K^2(\alpha_0 + 1)} \]
Dirichlet Probability Densities

$\mathbb{R}^K \rightarrow \mathbb{R}$

$\pi \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$

Dirichlet Distribution

Multinomial Distribution

Distribution over discrete outcomes

Represented by non-negative vector that sums to one

Picture representation

$(1,0,0)$

$(0,0,1)$

$(1/2,1/2,0)$

$(1/3,1/3,1/3)$

$(1/4,1/4,1/2)$

$(0,1,0)$

Come from a Dirichlet distribution

Dirichlet Samples

$\mathbb{E}_{\alpha}[\pi_k] = \frac{\alpha_k}{\alpha_0}$

• Samples are sparse for small values of $\alpha_i$

Dirichlet Samples

$\text{Dir}(\pi | 0.1, 0.1, 0.1, 0.1, 0.1)$

$\text{Dir}(\pi | 1.0, 1.0, 1.0, 1.0, 1.0)$
Model Summary

- Prior on model parameters
  - E.g., symmetric Dirichlet for $\pi$
  - Dirichlet prior for topic parameters $\beta_k$

- Sample observations as

\[
\begin{align*}
  z^d & \sim \pi \\
  w_i^d \mid z^d & \sim \beta_{z^d}
\end{align*}
\]

Posterior Inference via Sampling

- Iterate between sampling

- What form do these complete conditionals take?
  - First a look at statements of conditional independence in directed graphical models
Conditional Independence in Bayes Nets

- Consider 4 different junction configurations

- Conditional versus unconditional independence:

Bayes Ball Algorithm

- Consider 4 different junction configurations

- Bayes ball algorithm
Markov Blanket

- A node is conditionally independent of all other nodes in the graph given its Markov blanket

- Gibbs sampling iterates between full conditionals

→ simplify to

Unplated Document Model

- Recall that the plate notation is really indicating
Complete Conditional for $\pi$

- Recall conjugate Dirichlet prior
  
  \[ \pi \sim \text{Dir}(\alpha_1, \ldots, \alpha_K) \quad p(\pi \mid \alpha) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_k \pi_k^{\alpha_k-1} \]

- Likelihood:

- Dirichlet posterior
  - Count occurrences of
  - Then,

  \[ \text{Conjugacy: Posterior has same form as prior} \]

Complete Conditional for $\beta_k$

- Again, Dirichlet prior

- Consider docs $d$ such that
  - For these observations,
  - Do any other docs depend on $\beta_k$?

- Then,

  \[ \text{Again, posterior has same form as prior} \]
**Complete Conditional for** $z^d$

- We have $z^d \sim \pi$
- $w^d_i \mid z^d, \{\beta_k\} \sim \beta_{z^d}$
- Calculate the posterior for each value of $z^d$ ("responsibility" of each topic to the doc):
  
  $$r_{dk} = p(z^d = k \mid \{w^d_i\}, \pi, \beta) = \frac{\pi_k p\{w^d_i \mid \beta_k\}}{\sum_j \pi_j p\{w^d_i \mid \beta_j\}}$$
- Sample each cluster indicator as

---

**What you need to know...**

- Bayesian specification of document clustering model
- Rules of conditional and unconditional independence in directed graphical models (Bayes nets)
  - Bayes’ ball
  - Markov blanket
- Gibbs sampling for Bayesian document model
Reading for *Next Lecture*

- **Mixed Membership Models: KM Sec. 27.3**
  - Basic LDA:
  - Introduction:
  - Sampling: