Outline

- Matrix Completion with Features
  - Inductive Matrix Completion
  - Graph-based Regularization
Example I: Movie Recommendation with User Features

- Given: rating matrix $A \in \mathbb{R}^{m \times n}$, user feature matrix $X \in \mathbb{R}^{m \times d}$
- Goal: predict unknown ratings
Example II: Movie Recommendation with Features

- Given, rating matrix $A \in \mathbb{R}^{m \times n}$, user features $X \in \mathbb{R}^{m \times d_1}$, item features $Y \in \mathbb{R}^{n \times d_1}$
- Goal: predict unknown ratings
Applications

- Ad-word Recommendation
- Tag recommendation
- Disease-gene linkage prediction
- ...
Inductive Matrix Completion
Inductive Matrix Completion

- Assume the rating matrix $A$ is generated from $XMY^T$, where $X$ is the row feature matrix and $Y$ is the column feature matrix.
- Goal: Find the $M$ matrix to minimize the error.
Inductive Matrix Completion

- $M \in \mathbb{R}^{d_1 \times d_2}$ may be large
  - Further assume $M$ is a rank $k$ matrix
- Inductive matrix factorization:

$$
\min_{U \in \mathbb{R}^{d_1 \times k}, V \in \mathbb{R}^{d_2 \times k}} \sum_{i,j \in \Omega} ((XUV^TY^T)_{ij} - A_{ij})^2 + \lambda \|U\|_F^2 + \lambda \|V\|_F^2
$$

where $\Omega$ is the observed set, $A$ is the rating matrix, $X$, $Y$ are feature matrices.

\[ \begin{array}{cccc}
3 & 1 & 1 & 1 \\
2 & 3 & 3 & 3 \\
5 & 4 & 4 & 4 \\
4 & 2 & 2 & 2 \\
4 & 1 & 1 & 1 \\
1 & 3 & 3 & 3
\end{array} \sim
\begin{array}{c}
X \\
U \\
V^T \\
Y^T
\end{array} \]
Theoretical Guarantees

- Matrix Completion (without features):
  If the rating matrix $A \in \mathbb{R}^{m \times n}$ is sampled from a rank $k$ matrix, under certain condition we can recover $A$ with $O(kn \log^2(n))$ observed entries.

- Inductive Matrix Completion:
  If the rating matrix $A = XMY^T$ and $M$ is rank $k$, under certain condition we can recover $A$ with $O(dk \log k \log n)$ observed entries.

- Proved by the following papers:
  - Xu et al., “Speedup Matrix Completion with Side Information: Application to Multi-Label Learning”. In NIPS 2013.
  - Zhong et al., “Efficient Matrix Sensing using Rank-1 Gaussian Measurements”. In ALT 2015.
Applications

- Multi-label classification
- Yahoo Music challenge
- Biological application
- Semi-supervised clustering
The assumption $A = XMY^T$ means:

$$\text{col}(A) \subseteq \text{col}(X) \text{ and } \text{col}(A^T) \subseteq \text{col}(Y)$$

where $\text{col}(\cdot)$ denotes the column space of a matrix.

When features are not perfect:

$$\text{col}(A) \cap \text{col}(X)^T \neq \emptyset \text{ or } \text{col}(A^T) \cap \text{col}(Y)^T \neq \emptyset$$

Recovery by inductive matrix completion is impossible.
Graph-based Regularization
The similarity between users

- Assume we have $m$ users with feature vectors $\mathbf{x}_1, \ldots, \mathbf{x}_m$
- Similarity of rows: Define the similarity matrix $S \in \mathbb{R}^{m \times m}$

$$S_{ij} = \text{similarity}(\mathbf{x}_i, \mathbf{x}_j),$$

where $\mathbf{x}_i, \mathbf{x}_j$ are features for the $i(j)$-th rows.
- The similarity function can be defined by many ways, for example,

$$\text{similarity}(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \| \mathbf{x}_i - \mathbf{x}_j \|^2)$$
The similarity graph

- $S$ is a dense $m \times m$ matrix $\Rightarrow$ High computational cost
- Usually, a sparse similarity graph is preferred
- Several approaches:
  - K-nearest neighbor graph
  - Thresholding $S$ matrix
Matrix Completion with Graph Regularization

- Given, a partially observed rating matrix $A \in \mathbb{R}^{m \times n}$ and a similarity graph between users $S \in \mathbb{R}^{m \times m}$
- The similarity graph can be:
  - Constructed from features
  - Obtained by social activities
- Goal: predict unknown ratings
Matrix Completion with Graph Regularization

- Two approaches:
  - Average-based Regularization
  - Individual-based Regularization

- Discussed in
  - Ma et al., “Recommender Systems with Social Regularization”. In WSDM 2011
  - Rao et al., “Collaborative Filtering with Graph Information: Consistency and Scalable Methods”. In NIPS 2015
Average-based Regularization

- Compute matrices $U \in \mathbb{R}^{m \times k}$ and $V \in \mathbb{R}^{n \times k}$ by solving

\[
\arg\min_{U \in \mathbb{R}^{m \times k}, V \in \mathbb{R}^{n \times k}} \sum_{i,j \in \Omega} (A_{ij} - (UV^T)_{ij})^2 + \frac{\alpha}{2} \sum_{i=1}^{m} \|u_i - \frac{1}{\sum_k S_{ik}} \sum_j S_{ij} u_j\|^2 \\
+ \lambda \|U\|_F^2 + \lambda \|V\|_F^2
\]

where $u_i$ is the $i$-th row of $U$. 

Individual-based Regularization

- Compute matrices $U \in \mathbb{R}^{m \times k}$ and $V \in \mathbb{R}^{n \times k}$ by solving

$$
\text{argmin}_{U \in \mathbb{R}^{m \times k}, V \in \mathbb{R}^{n \times k}} \sum_{i,j \in \Omega} (A_{ij} - (UV^T)_{ij})^2 + \sum_{i,j} \frac{S_{ij}}{2} \|u_i - u_j\|^2 + \lambda \|U\|_F^2 + \lambda \|V\|_F^2
$$

where $u_i$ is the $i$-th row of $U$.

- Can be written as

$$
\text{argmin}_{U \in \mathbb{R}^{m \times k}, V \in \mathbb{R}^{n \times k}} \sum_{i,j \in \Omega} (A_{ij} - (UV^T)_{ij})^2 + \text{trace}(U^T(D - S)U) + \lambda \|U\|_F^2 + \lambda \|V\|_F^2
$$

where $D$ is a diagonal matrix with $D_{ii} = \sum_j S_{ij}$.

- The matrix $L = D - S$ is called the Laplacian matrix.
Coming up

- Homework due Thursday in class
- Next class: paper presentations on advanced optimization topics

Questions?