CSE 8803RS: Recommendation Systems

Lecture 9: Low-Rank Matrix Completion

Hongyuan Zha

School of Computational Science & Engineering
College of Computing
Georgia Institute of Technology

Matrix Completion

A much broader problem: arbitrary matrices can't be completed, need some structure/constraints on the matrices

- Positive semi-definite (PSD) matrix completion
  — a partial PSD matrix with diagonal entries all ones
  — it has PSD completion iff it's graph is *chordal*
- Chordal graph: no minimal cycles of length 4 or more, *triangulated* graph

Matrix Completion Under Rank Constraints

$A \in R^{N \times N}$ is rank-k, known partially

- Rank-k matrices parameterized by $(2N - k)k$ degree of freedom
- No hope to recover an *arbitrary* low-rank matrix from a sample of its entries

$$A = e_1 e_N^T$$

  has one 1 in $(1, N)$ entry, everywhere else is zero. Clearly this matrix cannot be recovered from a sampling of its entries unless we pretty much see all the entries.

- For instance, if we were to see 90% of the entries selected at random, then 10% of the time we would only get to see zeroes

$$(1 - 1/N^2)^{\alpha N^2} \approx 1 - \alpha$$

Simple Model of Low-Rank Matrices

It is therefore impossible to recover *all* low-rank matrices from a set of sampled entries but can one recover *most* of them?

- Rank-k matrices parameterized by SVD

$$A = \sum_{i=1}^{k} \sigma_i u_i v_i^T = U_k \Sigma_k V_k^T$$

- $U_k \equiv [u_1, \ldots, u_k]$ and $V_k \equiv [v_1, \ldots, v_k]$ selected *uniformly* at random among all families of k orthonormal vectors $\Rightarrow$ random orthogonal model

Which Sampling Sets?

- If $A = uv^T, A_{ij} = u_i v_j$, if we don't sample any element from row one, we won't know u_1
- Can one recover a low-rank matrix from almost all sampling sets of cardinality big enough?
- If the number of known entries is sufficiently large, and if the entries are sufficiently uniformly distributed $\Rightarrow$ only one low-rank matrix with these entries.

Which Algorithm?

- Optimization problem,

$$\min \operatorname{rank}(X), \text{subject to } \; X_{ij} = A_{ij}, (i, i) \in \mathcal{O}$$

- Convex relaxation,

$$\min \|X\|_*, \text{subject to } \; X_{ij} = A_{ij}, (i, i) \in \mathcal{O}$$

where $\| \cdot \|_*$ is the nuclear norm

Main Theorem

- Let $A \in R^{N \times N}$ sampled from random orthogonal model. We also observe m entries of A with locations sampled uniformly at random.
- Then there are constants C and c such that if

$$m \geq CN^{5/4} k \log N$$

- The minimizer to the nuclear norm optimization is unique and $= A$ with probability $\geq 1 - cN^{-3}$
- A surprisingly small number of entries are sufficient to complete a generic low-rank matrix.

Improvement

Keshavan et. al.

- $m \geq CNk \max\{\log N, k\}$.
- $\log N$ is related to coupon collector's problem

Coherence of an Orthonormal Basis

- Sampling operator $A = P_{\mathcal{O}}(A) + P_{\mathcal{O}}^{\perp}(A)$
- If $P_{\mathcal{O}}(A)$ we gain little, this happens to $A = e_1 e_N^T$, also for

$$A = \sigma_1 u_1 u_1^T + \sigma_2 u_2 u_2^T, \quad u_1 = (e_1 + e_2)/\sqrt{2}, u_2 = (e_1 - e_2)/\sqrt{2}$$

  only the leading 2×2 submatrix is nonzero
- The singular vectors need to be sufficiently spread to minimize the number of observations needed to recover a low-rank matrix

Coherence of an Orthonormal Basis

- Definition: U_k orthonormal matrix

$$\mu(U_k) \equiv \frac{N}{k} \max_{1 \leq i \leq N} \| U_k^T e_i \|^2$$

- Some extreme care
  — $k = 1$, $U_k = [1, \ldots, 1]/\sqrt{N}$ gives $\mu(U_k) = 1$
  — $k = 1$, $U_k = e_i$ gives $\mu(U_k) = N$
  — More generally, $1 \leq \mu(U_k) \leq N/k$
- For the random orthogonal model, $\mu = O(1)$