CS 596: Theoretical Machine Learning

Lecturer: Pranjal Awasthi
Scribe: Yiran Sun

Lecture 16
Nov 15, 2016

Matrix Reconstruction

Today we will see how to reconstruct M from a small set of entries. We are given as input a matrix $P_\Omega(M)$ where $\Omega \subseteq [n] \times [n]$, and

$$P_\Omega(M)_{i,j} = \begin{cases} M_{i,j} & i,j \in \Omega \\ 0 & otherwise \end{cases}$$

Our goal is to use $P_\Omega(M)$ to infer M either exactly or approximately. Ideally, we want $|\Omega| \ll n^2$.

Fundamental limits for exact reconstruction

If M is a random matrix then one cannot infer anything about the missing entries from $P_\Omega(M)$. Hence in this case we would need $|\Omega| = n^2$. However, in practice we are interested in learning structured or low-rank matrices. Hence we will assume that M is a low rank matrix, $M = U_{n\times r}\Sigma_{r\times r}V^T_{n\times r}$, where $r \ll n$. Let's first consider the case when M is a rank-1 matrix,

$$M = x \cdot y^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} * \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix} = \begin{bmatrix} x_1y_1 & x_1y_2 & \dots & x_1y_n \\ x_2y_1 & x_2y_2 & \dots & x_2y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_ny_1 & x_ny_2 & \dots & x_ny_n \end{bmatrix}$$

If there exists a row/column where none of the entries have been revealed, then we cannot say anything about the corresponding x_i or y_i value. Thus, in this case we need to see $|\Omega| \geqslant 2n-1$ entries. When M is a rank-r matrix, we can set up a system of polynomial equations in U, Σ, V such that $M = U\Sigma V^T, M_{ij} = (U\Sigma V^T)_{ij},\ i,j \in n$. The number of free variables is $2nr - r^2$ and hence we would need $|\Omega| \geqslant 2nr - r^2$ observations.

Is low rank + random sampling enough?

Consider the following rank 1 matrix

$$M = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} = e_ie_i^T$$

In this exact reconstruction is not possible unless one sees all the entries. Hence, we need to make further assumptions about the matrix in order to be able to exact reconstruction from much fewer entries. The standard assumption is called *incoherence* which implies that the singular vectors of M are not spiky.

Definition 1 (Incoherence)**.** *Let M be a rank r $n \times n$ matrix with $M = U\Sigma V^T$. We say that M is μ-incoherent if $\forall i \in [n], \|e_i^T U\|^2 \le \mu\frac{r}{n}$ and $\|e_j^T V\|^2 \le \mu\frac{r}{n}$.*

Incoherence implies that the useful information is uniformly spread out in the matrix and hence we can hope to reconstruct it from a small uniformly random sample. If U and V are random orthonormal basis of rank r then with high probability M will be $O(\log n)$-incoherent. For the rest of this lecture you should think of μ as a constant or at most $\log n$. There are three approaches to solve matrix completion

- Spectral approach [leads to a weak approximation]
- Spectral approach + Alternate Minimization [leads to a arbitrarily good approximations]
- Semi-definite programming (SDP) [does near optimal exact reconstruction]

Spectral Approach

We will see that simply doing r-SVD on a scaled version of $P_\Omega(M)$ leads to a weak approximation to M. Furthermore, we won't have to make any incoherence assumption. Let p be the sampling probability, i.e., the probability with which each entry of M is revealed. The algorithm is as follows

- Let $\tilde{M} = \frac{1}{p} P_\Omega(M)$.
- Let the SVD of $\tilde{M}$ be $\tilde{M} = \sum_{i=1}^{n} \sigma_i x_i y_i^T$.
- Output $M' = \sum_{i=1}^{r} \sigma_i x_i y_i^T$.

Theorem 1. *Let M be a rank r matrix with $M_{i,j} \in [0,1]$. If $p \geq c\frac{\log n}{n}$, then with probability $\geq 1 - \frac{1}{n^3}$, the spectral method outputs M' such that $1/n||M' - M||_F \leq O(\sqrt{\frac{s}{np}})$.*

Proof. Notice that $E[p\tilde{M}] = pM$. Let $R = p\tilde{M} - pM$. Then we have that each entry $R_{i,j}$ is a zero mean bounded random variable with variance $\sigma^2 \leq 4p(1-p)$. Hence, from previous lecture we know that with probability $\geq 1 - \frac{1}{n^3}$, $\|R\| \leq \sigma\sqrt{n} = O(\sqrt{pn})$. This implies that $\|\tilde{M} - M\| \leq O(\sqrt{\frac{n}{p}})$. Next we have that

$$
\begin{aligned}
\|M' - M\|_F &\leq \sqrt{2r}\|M' - M\| \ \{\text{This is because for any rank } r \text{ matrix } A,\ \|A\|_F \leq \sqrt{r}\|A\|.\} \\
&\leq \sqrt{2r}[\|M' - \tilde{M}\| + \|\tilde{M} - M\|] \ \{\text{triangle inequality.}\} \\
&\leq 2\sqrt{2r}\|\tilde{M} - M\|
\end{aligned}
$$

The last inequality follows from the fact that both M' and M are rank r matrices and M' is the best rank r approximation to $\tilde{M}$. Combining this with the spectral norm bound for $\|\tilde{M} - M\|$ we get the desired claim. □

In the next lecture we will see how to improve the above guarantee for incoherent matrices via Alternate Minimization.

Additional Readings

- The original paper of Candes and Recht that uses SDP for matrix completion. `https://statweb.stanford.edu/~candes/papers/MatrixCompletion.pdf`.