Lecture #22: More Matrix Completion

Last time we studied generalization bounds:

$$X = \underset{\text{nuclear norm}}{\arg\min} \; \|X\|_* \quad \text{s.t.} \quad X_{ij} = M_{ij} \;\; \forall (i,j) \in \Omega$$

$$\text{and } \|X\|_\infty \le \mu \sqrt{\frac{r}{n_1 n_2}}$$

If $\|M\|_* \le r$ and $\|M\|_\infty \le \mu\sqrt{\frac{r}{n_1 n_2}}$ we know

there is a good solution. We proved:

simple solution + low training error = low test error

(Rademacher complexity) (zero)

In particular:

$$\frac{1}{n_1 n_2} \sum_{ij} |X_{ij} - M_{ij}| \le \underbrace{\frac{2r}{m} \mathbb{E}\big[\|\Sigma\|\big]}_{\text{Rademacher}} + \underbrace{2\mu\sqrt{\frac{r}{n_1 n_2}} \sqrt{\frac{\log 2/\delta}{2|\Omega|}}}_{\text{McDiarmid}}$$

with probability $\ge 1-\delta$.

Here Σ is a random matrix with $m = |\Omega|$ entrees chosen u.a.r, random ± 1s

Theorem: Let $X_1, X_2 \ldots X_k$ be independent, mean zero random matrices of size $d \times d$.

Let $\rho_i^2 = \max\left\{ \|\mathbb{E}[X_i X_i^T]\|, \|\mathbb{E}[X_i^T X_i]\| \right\}$ and $\|X_i\| \le L$

Then for any $\tau > 0$

$$\Pr\left[\left\| \sum_{i=1}^{k} X_i \right\| > \tau\right] \leq 2d\, e^{\frac{-\tau^2/2}{\sum p_i^2 + L\tau/3}}$$

Let's say we had a trivial bound:

$$\frac{1}{m}\mathbb{E}\left[\|\Sigma\|\right] \leq \frac{1}{m}\mathbb{E}\left[\|\Sigma\|_F\right] = \frac{1}{\sqrt{m}}$$

Now the average error on the l.h.s. is $(r, \mu = O(1))$

$$\sqrt{\frac{1}{n_1 n_2}} \underset{\text{to get nontrivial bound}}{\overset{\text{need}}{\gg}} \frac{1}{\sqrt{m}}$$

This is only true when $m \geq n_1 n_2$ and we see the entire matrix

Let's use matrix Bernstein to prove

$$\mathbb{E}\left[\|\Sigma\|\right] \sim \sqrt{\frac{m \log d}{\min(n_1, n_2)}}$$

in which case (again $r, \mu = O(1)$) we just need

$$\sqrt{\frac{1}{n_1 n_2}} \gg \sqrt{\frac{\log d}{m \min(n_1, n_2)}}$$

which happens when $m \gg \max(n_1, n_2) \log d$ which just requires seeing every row/column

In our setting

$$\mathbb{E}\big[\overset{n_1 \times n_2}{X_i} X_i^T\big] = \frac{1}{n_1} I \qquad \mathbb{E}\big[X_i^T X_i\big] = \frac{1}{n_2} I$$

random matrix with one ± 1 at random location

Hence $\rho_i^2 = \frac{1}{\min(n_1,n_2)}$, also $L=1$ so we have

$$\Pr\left[\Big\| \sum_{i=1}^{m} X_i \Big\| > \tau\right] \leq 2d\, e^{\frac{-\tau^2/2}{m/\min(n_1,n_2) + \tau/3}}$$

we want $\tau = c\sqrt{\frac{m \log d/\delta}{\min(n_1,n_2)}}$ which implies r.h.s. is at most δ provided

$$\frac{m}{\min(n_1,n_2)} \geq \sqrt{\frac{m \log d/\delta}{\min(n_1,n_2)}}$$

$$\Updownarrow$$

$$m \geq \min(n_1,n_2) \log d/\delta$$

An elegant aspect of these generalization bounds is that they still work when M is only close to low rank:

If we replace $X_{ij} = M_{ij} \ \forall (i,j) \in \Omega$ with

$\frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} |X_{ij} - M_{ij}| \leq \eta$ then we get same test error bound but $+\eta$

Later [Candes, Recht], [Candes, Tao], [Recht] studied exact matrix completion:

Let $n_1 = n_2 = n$ for simplicity and let

$$\mu_0 = \frac{n}{r} \max \left(\max_i \|P_u e_i\|^2, \max_i \|P_v e_i\|^2 \right)$$

left/right singular vectors of M

and $\|uv^T\|_\infty \le \frac{\mu_1 \sqrt{r}}{n}$

<u>Theorem:</u> If Ω is chosen u.a.r and $|\Omega| = m$ with

$$m \ge C \max(\mu_1^2, \mu_0)\, r n \log^2 n$$

then whp $X = M$

Again, X is the minimizer of $\|X\|_*$ s.t. $X_{ij} = M_{ij}$ $\forall (i,j) \in \Omega$

The approach is based on the following lemma

<u>Lemma:</u> $\|X\|_* = \max_{\|B\| \le 1} \langle X, B \rangle$

when $X = \mathrm{diag}(x)$ and $B = \mathrm{diag}(b)$ we get

$$\|X\|_* = \|x\|_1 \text{ and } \|B\| = \|b\|_\infty$$

Thus this is a matrix generalization of the more familiar $\ell_1 - \ell_\infty$ norm duality:

Lemma: $\|x\|_1 = \max_{\|b\|_\infty \le 1} \langle x, b \rangle$

Their approach is given M and another candidate solution $X = M + Z$ find a B s.t.

① $\|B\| \le 1$

② $\langle M+Z, B \rangle \ge \|M\|_* + c\|P_{U^\perp} Z P_{V^\perp}\|_*$

Thus $\|M+Z\|_* > \|M\|_*$ and it is a worse solution

Let U' and V' be singular vectors of $P_{U^\perp} Z P_{V^\perp}$

Then they set $B = [U, U'] \begin{bmatrix} V^T \\ V'^T \end{bmatrix}$

For this choice of B we have

$$\langle M+Z, B \rangle = \underbrace{\langle M, UV^T \rangle}_{= \|M\|_*} + \underbrace{\langle P_{U^\perp} Z P_{V^\perp}, U'V'^T \rangle}_{= \|P_{U^\perp} Z P_{V^\perp}\|_*}$$

$$+ \underbrace{\langle P_U Z + Z P_V - P_U Z P_V, U'V'^T \rangle}_{\text{long argument to show this is small, quantum golfing}}$$

This is a more complex version of other arguments we've seen that use convex programs for exact recovery e.g. [Feige, Kilian]

What about tensor completion:

low-rank, incoherent T is $n_1 \times n_2 \times n_3$

We can try:

$$\min \|X\|_* \quad \text{s.t.} \quad T_{ijk} = X_{ijk} \quad \forall (i,j,k) \in \Omega$$

tensor nuclear norm,

convex comb.

$$\text{unit ball} = \left\{ X \;\middle|\; X = \sum \lambda_i u_i \otimes v_i \otimes w_i \text{ where } \|u_i\|, \|v_i\|, \|w_i\| \leq 1 \right\}$$

Theorem [Gurvits] Computing the tensor nuclear norm is hard

We can also flatten it into a matrix

$$\text{flat}(T) \text{ is } n_1 n_2 \times n_3$$

But then we need $m \geq \max(n_1 n_2, n_3)$ observations

Following [Srebro, Shraibman] approach, [Barak, Moitra]:

If $n_1 = n_2 = n_3 = n$ then can estimate the entries when $m \geq n^{3/2} r$

Is there an algorithm for exact tensor completion?

Also [Barak, Moitra]: any relaxation of tensor nuclear norm that has nontrivial rademacher complexity for $m = n^{3/2-\delta}$ cannot come from SOS, and would imply better algorithms for refuting random CSPs

Other linear inverse problems that seem to have computational vs. statistical tradeoffs?

Practical algorithms for tensor completion?