Lecture 11

Clustering

$n, d, k \in \mathbb{N}_{\geq 0}$

Unknown: Centers $\mu_1^0, \cdots, \mu_k^0 \in \mathbb{R}^d$

assignment $X_1^0, \cdots, X_n^0 \in \{\mu_1^0, \cdots, \mu_k^0\}$

Observe: $Y_1, \cdots, Y_n$ with $Y_i = X_i^0 + W_i$

$W_1, \cdots, W_n \overset{iid}{\sim} N(0, I_d)$

* $\text{rank}(X^0) \leq k$

Matrix form: $Y = X^0 + W$, $W_{ij} \sim N(0,1)$, X^0 has at most k distinct columns

SVD: w.h.p. $\frac{1}{d \cdot n} \|\hat{X} - X^0\|_F^2 \lesssim \frac{d \cdot k + k \cdot n}{d \cdot n} = k \cdot \frac{d+n}{d \cdot n} \lesssim \frac{k}{d}$

$\longrightarrow \frac{1}{n} \|\hat{X} - X^0\|_F^2 \lesssim k$

$\Omega_k = \{X \in \mathbb{R}^{d \times n} \mid X \text{ has } \leq k \text{ columns}\}$

Polynomial system: $\vec{P}(X; Y) = (P_1(X;Y), \cdots, P_m(X;Y))$

Candidate estimator: given Y, choose $\hat{X} \in \{X \mid \vec{P}(X,Y) \geq 0\}$

design properties:

$\forall X^0 \in \Omega_k$, w.h.p. over $Y = X^0 + W$

- $\vec{P}(X^0; Y) \geq 0$
- $\|X^0 - X\|^2 \leq C$ holds over $\{X \mid \vec{P}(X,Y) \geq 0\}$

Theorem

If both properties satisfied, then estimator has error $\leq C$ w.h.p.

What polynomials can we use?

* $z_i^2 = z_i$, i.e., $z_i = \{0, 1\}$

Choose $\vec{P}(X, z; Y)$ such that (X, Y) extends to solution to $\vec{P}(X, z; Y) \geq 0$ iff

- $X \in \Omega_k$
- $W = Y - X$ "looks Gaussian"

$$\underbrace{\left\| \frac{1}{n} \sum_{i=1}^n w_i^{\otimes t} - \mathbb{E}_{W \sim N(0, I_d)}[W^{\otimes t}] \right\|_F^2 \leq \frac{1}{100}}_{\text{log-}2t \text{ poly in } X \text{ and } Y}$$

Polynomial inequalities $t \in \mathbb{N}$, even

- $(A+B)^t \leq 2^{t-1}(A^t + B^t)$
- $\left(\frac{a_1 + \cdots + a_n}{n}\right)^t \leq \frac{a_1^t + \cdots + a_n^t}{n}$

Theorem

If $(X^0, Y), (X, Y)$ satisfy the two conditions above, then

$$\frac{1}{n} \|X - X^0\|_F^2 \leq 4t \cdot k^{4/t}$$

Proof (theorem)

$U = X - X^0$

Claim 1: U looks Gaussian

$$\forall v, \underbrace{\frac{1}{n} \sum_i \langle u_i, v \rangle^t}_{\langle \frac{1}{n} \sum_i u_i^{\otimes t}, v^{\otimes t} \rangle} \leq (4t)^{t/2} \cdot \|v\|^t$$

Proof of claim 1: $W = Y - X$, $W^0 = Y - X^0 \longrightarrow W^0 - W = U$

$$\frac{1}{n} \sum_i \langle w_i, v \rangle^t = \left\langle \frac{1}{n} \sum_i w_i^{\otimes t}, v^{\otimes t} \right\rangle = \langle \mathbb{E}[w^{\otimes t}], v^{\otimes t} \rangle + \underbrace{\left\langle \frac{1}{n} \sum_i w_i^{\otimes t} - \mathbb{E}[w^{\otimes t}], v^{\otimes t} \right\rangle}_{\leq \underbrace{\| \cdots \|}_{\leq \sqrt{\frac{1}{100}}} \cdot \underbrace{\| v^{\otimes t} \|}_{= \|v\|^t}}$$

$$\le \|v\|^t \cdot (t)^{t/2}$$

$$= \mathbb{E}[\langle w, v\rangle^t] = \|v\|^t \cdot (t-1)\cdot(t-3)\cdots 1 \le \|v\|^t \cdot (t)^{t/2}$$

Similar $\frac{1}{n}\sum_i \langle w_i^\circ, v\rangle^t \le \|v\|^t \cdot (t)^{t/2}$

<u>together</u> $$\frac{1}{n}\sum_i \underbrace{\langle u_i, v\rangle^t}_{(\langle w_i^\circ, v\rangle - \langle w_i, v\rangle)^t} \le \frac{2^{t-1}}{n}\left(\left(\sum \langle w_i^\circ, v\rangle^t\right) + \left(\frac{1}{n}\sum_i \langle w_i, v\rangle^t\right)\right) \le 2^t \cdot \|v\|^t \cdot t^{t/2} = (4t)^{t/2} \cdot \|v\|^t$$ (no cancellation)

<u>Claim2:</u> "U looks clustered" $U \in \Omega_{k^2}$ $*$ col$(X)=k$, col$(X^\circ)=k$, at most $k\cdot k$ choices.

$\to \{u_1, \cdots, u_n\} = \{v_1, \cdots, v_K\}$ for $K = k^2$

<u>Proof of theorem from claims</u>

Let $C = [K]$, $t \ge 2$ even

$$(4t)^{t/2} \cdot \cancel{\|v_c\|^t} \overset{\text{claim1}}{\ge} \frac{1}{n}\sum_i \langle u_i, v_c\rangle^t$$
$$\ge \frac{1}{n}\sum_i [\![u_i = v_c]\!] \cdot \cancel{\|v_c\|^t} \cdot \|u_i\|^t$$
$$\to (4t)^{t/2} \ge \frac{1}{n}\sum_i [\![u_i = v_c]\!] \cdot \|u_i\|^t$$
$$\xrightarrow{\text{sum over } C} K \cdot (4t)^{t/2} \ge \frac{1}{n}\sum_i \|u_i\|^t \ge \left(\frac{1}{n}\sum_i \|u_i\|^2\right)^{t/2}$$
$$\to K^{2/t} \cdot 4t \ge \frac{1}{n}\sum_i \|u_i\|^2 = \frac{1}{n}\|X - X^\circ\|_F^2$$

<u>Theorem</u>

$\forall$ polynomial system $\vec{P}(X, Z; Y)$, $\exists$ poly-time estimator $\hat{X}$. $(N+m)^\ell$ N vars. in X and Z

Suppose $(X^\circ, Z^\circ, Y^\circ)$ satisfy

- $\vec{P}(X^\circ, Z^\circ; Y^\circ) \ge 0$
- $\|X - X^\circ\|_F^2 \le C$ over $\{\vec{P}(X, Z, Y^\circ) \ge 0\}$ via low-degree proof

Then, $\|\hat{X}(Y^\circ) - X^\circ\|_F^2 \le C$. $\vec{P}(X, Z, Y) \ge 0 \vdash_{\ell}^{X,Z} \|X - X^C\|_F^2 \le C$

SOS proof system

<u>polynomials</u> $P_1(X), \cdots, P_m(X), q(X)$

<u>definition</u> degree-ℓ SOS proof of statement

"$q(X) \ge 0$ over $\{P_1(X) \ge 0, \cdots, P_m(X) \ge 0\}$"

is any finite decomposition $q(X) = \sum_{r=1}^{R} S_r(X)^2 \cdot \bar{P}_r(X)$, where

- $\bar{P}_r$ is product of subset of $\{P_1(X), \cdots, P_m(X)\}$
- $\deg(S_r(X)^2 \cdot \bar{P}_r(X)) \le \ell$

<u>notation</u> $\vec{P}(X) \ge 0 \vdash_{\ell}^{X} q(X) \ge 0$

<u>examples of SOS proofs</u>

(a) $\vdash_2^{a,b} a \cdot b \le \frac{1}{2}a^2 + \frac{1}{2}b^2$

as $a \cdot b = \frac{1}{2}a^2 + \frac{1}{2}b^2 - \frac{1}{2}(a-b)^2$

(b) $\vdash_2^{a,b} \langle a, b\rangle \le \frac{1}{2}\|a\|^2 + \frac{1}{2}\|b\|^2$

(c) $\vdash_4^{a,b} \langle a, b\rangle^2 \le \|a\|^2 \cdot \|b\|^2$

Justification $\vdash^{a,b} \langle a, b\rangle^2 = \langle a \otimes b, b \otimes a\rangle \overset{(b)}{\le} \frac{1}{2}\|a \otimes b\|^2 + \frac{1}{2}\|b \otimes a\|^2 = \|a\|^2 \cdot \|b\|^2$

Justification: $\vdash_4 \langle a, b \rangle = \langle a\otimes b, b\otimes a\rangle \le \frac{1}{2}\|a\otimes b\| + \frac{1}{2}\|b\otimes a\| = \|a\|\cdot\|b\|$

(d) $\vdash_{\frac{x}{2}} \langle x, Mx\rangle \le \underbrace{\|M\|}_{\text{Spectral norm}} \cdot \|x\|^2$