18/3/22

Robust Subspace Clustering

Problem formulation: there are L subspaces, $\{S_1, \dots, S_L\} \in \mathbb{R}^n$ and have dimensions $d_1, \dots, d_L$ $\{L, d_i, S_i\}$ are unknown.

given $Y \in \mathbb{R}^n$, $|Y| = N$, $Y = Y_1 \cup Y_2 \cup \dots \cup Y_L$ and Y_ℓ is collection of N_L vectors from S_ℓ (or close to)

Goal: obtain $S_1, \dots, S_L$ from Y.

Eqn:
$$y = x + z \quad , \quad [\text{there are } N \text{ such vectors}]$$
such that $x = S_k a_k$ for some k, z independent "noise".

Preliminary assumptions: (a) $E\|z\|_2^2 = \sigma^2 \|x\|_2^2$, $\sigma < \sigma^* < 1$

(b) $\max_\ell d_\ell \leq C_0 \dfrac{n}{\log^2 N}$

Challenges: (a) how "far" are subspaces S_ℓ?

(b) How big are # sample N_ℓ?

Def^n: $\boxed{\text{Principal angles}}$ $\cos(\theta^i) = \max_{u \in S} \max_{v \in S'} \dfrac{u^T v}{\|u\|_2 \|v\|_2} = \dfrac{u_i^T v_i}{\|u_i\|_2 \|v_i\|_2}$; $i = 1, \dots \min(d, d')$

$$u^T u_j = 0, \; v^T v_j = 0, \; j = 0, \dots i-1$$

- singular values of $U^T V$

Def^n: <u>Normalized affinity</u>: $\text{aff}^2(S, S') = \dfrac{\sum_{i=1}^{\min(d,d')} \cos^2\theta^i}{\min(d,d')}$

- $\text{aff} \approx 0 \Rightarrow$ ~~hard~~ easy prob $\Rightarrow U^\perp \approx V$ (span-wise)
- $\text{aff} \approx 1 \Rightarrow$ hard prob, $\Rightarrow U \approx V$ (span-wise)

Def^n: <u>Sampling Density</u>: P_ℓ of subspace S_ℓ is defined as $\cancel{P_\ell = d_\ell / N_\ell}$

$P_\ell = \dfrac{N_\ell}{d_\ell}$. If $P_\ell < 1$, no hope of recovering S.S

<u>NORMALIZED MODEL</u> : $Y = X + Z$, $\quad Y, X, Z \in \mathbb{R}^{n \times N}$

assume that $\|x\|_2 = 1 \Rightarrow \|y_i\|^2 \approx \sqrt{1+\sigma^2}\, \|x_i\|_2^2$ (in expectation)

$Z_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2/n)$

<u>SOLUTION</u> : the Sparse Subspace Clustering scheme!

A. Compute affinity matrix encoding similarities b/w samples, and construct weighted graph W

B. Construct clusters using spectral clustering

C. Apply PCA to each cluster

<u>INTUITION</u> (for $Z=0$ case) : Express x_i as a sparse lin. comb of x_j, $j \neq i$. (must hold since x_j would be from same S.S as x_i)

- So, solve $\min_{\beta \in \mathbb{R}^n} \|\beta\|_1 \;\; \text{s.t.} \;\; x_i = X\beta, \; \beta_i = 0.$

- stack β's as columns of matrix B, $W \triangleq |B| + |B|^T$

for noisy case, $x_i = XB \iff y_i - z_i = (Y-Z)\beta$

$$= y_i = Y\beta + \underbrace{(z_i - Z\beta)} \rightarrow \text{noise! (perturbation)}$$

Use LASSO

$$\min_{\beta \in \mathbb{R}^N} \frac{1}{2}\|y_i - Y\beta\|_2^2 + \lambda\|\beta\|_1 \quad s.t \ \beta_i = 0$$

Performance Metrics: Def^n: False disovery, (i,j) obeying $B_{ij} \neq 0$ is false disovery if y_i and y_j don't belong to same SS

Def^n: True disovery: (i,j) obeying $B_{ij} \neq 0$ is T.D if $y_i, y_j \in$ same SS.

- if there are no F.D's, then "subspace detection property" holds.
- then B is permutation similar to block-diagonal matrix.

Data-dependent regularization: (consider noiseless case)

$$L.F: \quad K(\beta,\lambda) = \frac{1}{2}\|x_i - X\beta\|_2^2 + \lambda\|\beta\|_1 \quad (1)$$

minimizer of (1) - $\hat{\beta}$ and let $\hat{\beta}_{eq}$ be solution with equality constraints - ie sol^n of (1) with $\lambda \rightarrow 0^+$. then

$$\frac{1}{2}\|x - X\hat{\beta}\|_2^2 \leq K(\hat{\beta},\lambda) \leq K(\hat{\beta}_{eq},\lambda) = \lambda\|\hat{\beta}_{eq}\|_1$$

(A) If $\|\hat{\beta}\|_0 \in [0.5d, 0.8d]$, $\|x - X\hat{\beta}\|_2^2 \geq C$ since $\|\hat{\beta}\|_0 \leq 0.8d$.

(B) $\|\hat{\beta}_{eq}\|_1 = O(\sqrt{d})$ $\left[d \times \frac{1}{\sqrt{d}}\right]$ (if also has no false discoveries!)

$$\Rightarrow \lambda = \Omega\left(\frac{1}{\sqrt{d}}\right) \quad \text{and if } \underbrace{\lambda > \|X^T x\|_\infty,\ \hat{\beta} = 0}$$

THEORETICAL RESULTS :

[A] Assumptions :

- affinity. S_ℓ obeys affinity condition if

$$\max_{k: k \neq \ell} \operatorname{aff}(S_k, S_\ell) \leq \frac{K_0}{\log N} \qquad K_0 \text{ - fixed constant}$$

- S_ℓ obeys sampling condition if $\rho_\ell \geq \rho^*$ $\qquad \rho^*$ - fixed constant

[B] Main Results :

Theorem (No false discoveries) : Assume that the subspace attached to i^{th} col. obeys A.C, S.C above and σ^* is a small numerical constant. In Algorithm 2, take $\tau = 2\sigma$ and $f(t) \geq \frac{\sigma}{\sqrt{2t}}$. Then, with probability at least $1 - 2e^{-\gamma_1 n} - 6e^{-\gamma_2 d(i)} - e^{-\sqrt{N(i) d(i)}} - \frac{23}{N^2}$ there is no false discovery in the i^{th} col of B_0

Theorem (Many true discoveries) : Under assumptions of theorem 3.1, with $f(t)$ also obeying $f(t) \leq \frac{\alpha_0}{t}$ for some α_0, with prob. at least $1 - 2e^{-\gamma_1 n} - 6e^{-\gamma_2 d(i)} - e^{-\sqrt{N(i) d(i)}} - \frac{23}{N^2}$ there are at least $\frac{C_0 d(i)}{\log \rho(i)}$ true discoveries in the i^{th} column.

[C] Proofs :

Proof of Theorem 3.1 :

Lemma 8.2 : Let $\mathrm{val}(\mathrm{step}\,1) \rightarrow \|\beta^*\|_1$ be optimal value of Algo. 2 [data driven regularization] with $\tau = 2\sigma$. Assume $p_1 > p^*$. Then, w.h.p.

$$\frac{1}{10}\sqrt{\frac{d_1}{\log p_1}} \leq \mathrm{val}(\mathrm{step}\,1) \leq 2\sqrt{d_1}$$

- upper bnd holds w.p $\geq 1 - e^{-\gamma_1 n} - e^{-\gamma_2 d_1}$
- lower bnd holds w.p $\geq 1 - e^{-\gamma_3 d_1} - 10/N^2$

using Lem. 8.2, $\lambda = f(\|\beta^*\|_1) \geq \frac{\sqrt{2}\,\sigma}{2\|\beta^*\|_1} \geq \frac{\sigma}{\sqrt{8 d_1}}$

now, need to show that there are no false discoveries. First, let us use $Y = \{Y^{(1)}, \ldots, Y^{(L)}\}$ remove y_i from $Y^{(1)}$ and so step 1 is of the form

$$\min_{\beta \in \mathbb{R}^{N-1}} \frac{1}{2}\|y - Y^{(1)}\beta^{(1)} - Y^{(2)}\beta^{(2)} \cdots - Y^{(L)}\beta^{(L)}\|_2^2 + \lambda\left(\|\beta^{(1)}\|_1 + \cdots + \|\beta^{(L)}\|_1\right)$$

if there are no false discoveries, above is equivalent to

$$\min_{\beta^{(1)} \in \mathbb{R}^{d_1 - 1}} \frac{1}{2}\|y - Y^{(1)}\beta^{(1)}\| + \lambda\|\beta^{(1)}\|_1$$. let arg min be $\hat{\beta}^{(1)}$

now, it suffices to show $\hat{\beta}^{(1)}$ obeys, $\|Y^{(l)^T}(y - Y^{(1)}\hat{\beta}^{(1)})\|_\infty < \lambda$, $\forall l \neq 1$

Lemma 8.6 : Fix $A \in \mathbb{R}^{d \times N}$ and $T \subseteq \{1, \ldots, N\}$. Suppose there is a sol^n x^* to

$$\min \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1 \quad \text{s.t } x_{T^c} = 0$$

obeying $\|A_{T^c}^T(y - Ax^*)\|_\infty \leq \lambda$. Then any optimal $\hat{x}$ to $\min \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1$ satisfies $\hat{x}_{T^c} = 0$

now, $Y^{(l)^T}(y - Y^{(1)}\hat{\beta}^{(1)}) = X^{(l)^T}(y_\parallel - Y_\parallel^{(1)}\hat{\beta}^{(1)}) + X^{(l)^T}(y_\perp - Y_\perp^{(1)}\hat{\beta}^{(1)})$

$$+ Z^{(l)^T}(y_\parallel - Y_\parallel^{(1)}\hat{\beta}^{(1)}) + Z^{(l)^T}(y_\perp - Y_\perp^{(1)}\hat{\beta}^{(1)})$$

~~Remark~~:

Lemma 8.1: Let $A \in \mathbb{R}^{d_1 \times N_1}$ be a matrix with columns sampled uniformly at random from the unit sphere of $\mathbb{R}^{d_1}$, $w \in \mathbb{R}^{d_2}$ be a vector sampled u.a.r. from unit sphere of $\mathbb{R}^{d_2}$ and independent of A. $\Sigma \in \mathbb{R}^{d_1 \times d_2}$ be deterministic matrix. We have w.p. $1 - \frac{2}{\sqrt{a}} - \frac{2N_1}{\sqrt{b}}$

$$\|A^T \Sigma w\|_\infty \leqslant \sqrt{\log a \log b}\ \frac{\|\Sigma\|_F}{\sqrt{d_1}\sqrt{d_2}}$$

Term 1: $\|X^{(1)^T}(y_{11} - Y^{(1)}_{11}\hat{\beta}^{(1)})\|_\infty$; $A = X^{(1)^T} \in \mathbb{R}^{d_2 \times N}$,

$w = \left(y_{11} - Y^{(1)}_{11}\hat{\beta}_{(1)}\right) \in \mathbb{R}^n$, w.p. $1 - \frac{4}{N^2}$

~~w.p.~~ $a = \sqrt{4\log N}$, $b = \sqrt{8\log N}$ $\Rightarrow$ $\|\cdot\|_\infty \leqslant \sqrt{32}\log N\ \frac{aff(S_1, S_2)}{\sqrt{d_1}} \|y_{11} - Y^{(1)}_{11}\hat{\beta}^{(1)}\|_2$

and $\|X^{(1)^T}(y_{11} - Y^{(1)}_{11}\hat{\beta}^{(1)})\|_\infty \leqslant \lambda\ aff(S_1, S_2) \triangleq \lambda I_1$

from lm 8.5, $\|y_{11} - Y^{(1)}_{11}\hat{\beta}^{(1)}\|_2 \leqslant C\lambda\sqrt{d_1}$

Term 2: $\|X^{(1)^T}(y_1 - Y^{(1)}_1\hat{\beta}^{(1)})\|_\infty \leqslant \sqrt{32}\log N\ \frac{1}{\sqrt{n d_1}} \|y_1 - Y^{(1)}_1\hat{\beta}^{(1)}\|_2$

and from lm 8.5, $\|X^{(1)^T}(y_1 - Y^{(1)}_1\hat{\beta}^{(1)})\|_\infty \leqslant L_n \frac{\sigma}{\sqrt{n}} \triangleq I_2$

Term 3: $\|Z^{(1)^T}(y_{11} - Y^{(1)}_{11}\hat{\beta}_{(1)})\|_\infty \leqslant C\lambda\sigma\sqrt{\frac{d_1 \log N}{n}} \triangleq \lambda I_3$

Term 4: $\|Z^{(1)^T}(y_1 - Y^{(1)}_1\hat{\beta}^{(1)})\|_\infty \leqslant \sigma^2\sqrt{\frac{L_n}{n}} \triangleq I_4$

$\Rightarrow$ we need $(I_1 + I_3)\lambda + I_4 + I_4 < \lambda \Rightarrow$ it suffices to have

$\lambda > L_N \frac{\sigma}{\sqrt{n}}$.

we also need $\lambda \ngtr \frac{\sigma}{\sqrt{8d_1}}$ $\Rightarrow$ $\lambda > \sqrt{\frac{\sigma}{8\sqrt{d_1}}}\max\left(1, L_n\sqrt{\frac{d_1}{n}}\right)$

and using $d_1 \leqslant \frac{n}{L_n^2}$ completes proof $\boxtimes$

Term ①: $X^{(1)^T}(\underline{y_{11}} - Y_{11}^{(1)}\hat{\beta}^{(1)}) = X^{(1)^T} y_{11} - X^{(1)^T} Y_{11}^{(1)} \hat{\beta}^{(1)}$

$X^{(1)^T} y_{11} = A^T \Sigma w$, , A - colums sampled from S_{d_2}, $\Sigma = U^{(1)^T} U^{(1)}$, $w = y_{11} / \|y_{11}\|$

$X^{(1)^T} Y_{11}^{(1)} \hat{\beta}^{(1)} = A^T \Sigma w$, A - " , $\Sigma = U^{(1)^T} U^{(1)}$, $w = Y_{11}^{(1)} \hat{\beta}^{(1)} / \|Y_{11}^{(1)} \hat{\beta}^{(1)}\|$

$$\Rightarrow \|X^{(1)^T}(y_{11} - Y_{11}^{(1)}\hat{\beta}^{(1)})\|_\infty \leq \underbrace{\|A^T \Sigma w\|}\ \|y_{11} - Y_{11}^{(1)}\hat{\beta}^{(1)}\|_2$$

$$\leq 32 \log n \frac{\text{aff}(S_1, S_2)}{\sqrt{d_1}} \|y_{11} - Y_{11}^{(1)}\hat{\beta}_{(1)}\|_2$$

from Lem 8.1

Proof of Lemma 8.2: Let $\beta_0 = X^{(1)^T}(X^{(1)} X^{(1)^T})^{-1} x$ be min l_2 norm sol^n to $X\beta_0 = x$. We first show β_0 is feasible for (2.4).

$$y - Y\beta_0 = (x - X\beta_0) + (z - Z\beta_0) = z - Z\beta_0$$

$$\mathcal{L}(z - Z\beta_0 \mid X, x) \sim \mathcal{N}\left(0, \frac{(1 + \|\beta_0\|_2^2)}{n} \sigma^2 I\right)$$

$$\|(X^{(1)} X^{(1)^T})^{-1}\|_2 \leq \frac{1}{\sigma_{min}^2(X^{(1)})}$$

$\Rightarrow \|z - Z\beta_0\|_2^2$ is ~~chi~~ χ^2 and so, w.p. $\geq 1 - e^{-\gamma_1 n}$,

$$\|z - Z\beta_0\|_2^2 \leq 2(1 + \|\beta_0\|_2^2)\sigma \qquad ④$$

but $\|\beta_0\|_2 = \|X^{(1)^T}(X^{(1)} X^{(1)^T})^{-1} x\|_2 \leq \|X^{(1)}\|_2 \|(X^{(1)} X^{(1)^T})^{-1}\|_2 \|x\|_2 \leq \frac{\|x\|_2}{\sigma_{min}(X^{(1)})}$

$$\|\beta_0\|_2 \leq \frac{1}{\sqrt{\frac{N_1}{d_1}} - 2} \quad \text{w.p.} \quad 1 - e^{-\gamma_2 d_1} !$$

$$\frac{\sigma_{max}(X^{(1)})}{\sigma_{min}^2(X^{(1)})}$$

if $N_1 > 9d_1$, $\|\beta_0\|_2 < 1$ and thus β_0 is feasible. (due to strong duality)

$$\|\beta^*\|_1 \leq \|\beta_0\|_1 \leq \sqrt{N_1} \|\beta_0\|_2 \leq \frac{\sqrt{d_1}}{1 - 2\sqrt{\frac{d_1}{N_1}}} \leq 2\sqrt{d_1}$$

<u>lower bnd</u>: let β^* be ~~the~~ an optimal solution. Notice that

$\|y_u - Y_u\beta^*\|_2 \leq \|y - Y\beta^*\|_2 \leq 2\sigma$, so, β^* is feasible for

(P) $\min \|\beta\|_1$ s.t $\|y_u - Y_u\beta^*\|_2 \leq 2\sigma$

(D) $\max \langle y_u, v\rangle - 2\sigma\|v\|_2$ s.t $\|Y_u^T v\|_\infty \leq 1$

let $Y_u^{(1)} = A$. Define $v^* \in \arg\max \langle y_u, v\rangle$ s.t $\|A^T v\|_\infty \leq 1$

w.h.p $\|Y_u^{(l)T} v^*\|_\infty \leq 1$ for $l \neq 1$ $\Rightarrow$ v^* is feasible for (D).

$$\langle y_u, v^*\rangle \geq \frac{1}{\sqrt{2\pi}\, e}\sqrt{\frac{d_1}{\log p_1}} \qquad [\text{ random direction} - y_u, v^*]$$

$$\|v^*\|_2 \leq \frac{16}{3}\sqrt{\frac{d_1}{\log p_1}} \Rightarrow \langle y_u, v^*\rangle - 2\sigma\|v\|_2$$

$$\geq \sqrt{\frac{d}{\log p_1}}\left[\frac{1}{\sqrt{2\pi}e} - \frac{32\sigma}{3}\right] \geq \frac{1}{10}\sqrt{\frac{d_1}{\log p_1}} \quad \text{if } \sigma \leq 0.015$$

$\boxtimes$

<u>Lemma 8.4</u>: The projection of the residual vector $r = y - Y^{(1)}\beta^{(1)}$ onto either S_1 or $S_1^\perp$ has uniform orientation.