LAIS, Lecture #22

"Spectral Clustering"

$G = (V, E, W)$

$G_i = (V_i, E_i)$ $i \in [c]$

connected components:

$W_{ab} = 0$ whenever $a \in V_i$, $b \in V_j$ and $i \neq j$.

The G_i's are connected:

G_i: $\forall\, a, b \in V_i$ $\exists\, a_1, \dots, a_k \in V_i$ s.t.

$(a, a_1) \in E_i$, $(a_1, a_2) \in E_i$ $\dots$ $(a_{k-1}, a_k) \in E_i$, $(a_k, b) \in E_i$

Thm 1 $e_{G_1}, \dots, e_{G_c}$ are a basis of $N(L)$.

Prf $\alpha^T L \alpha = \sum_{i<j} w_{ij} (\alpha(i) - \alpha(j))^2$

$$= \sum_{\substack{i<j \\ i,j \in V_k \\ \text{for some } k}} w_{ij} (\alpha(i) - \alpha(j))^2$$

$$e_{G_k}^T L e_{G_k} = \sum_{\substack{i<j \\ i,j \in V_s \\ \text{for some } s}} w_{ij} (e_{G_k}(i) - e_{G_k}(j))^2$$

$$= \sum_{\substack{i<j \\ i,j \in V_k}} w_{ij} (e_{G_k}(i) - e_{G_k}(j))^2 = 0$$

$\xi \in \mathcal{N}(L)$

$$\sum_{\substack{i<j \\ i,j \in V_k \\ \text{for some } k}} w_{ij} \left(\xi(i) - \xi(j)\right)^2 = 0$$

Take $i,j \in V_s$. Since G_s is connected $\exists$ $i_1, \dots, i_\ell \in V_s$ s.t. $(i,i_1), (i_1,i_2), \dots, (i_{\ell-1}, i_\ell), (i_\ell, j) \in E_s$

$\Rightarrow w_{ii_1}, w_{i_1 i_2}, \dots, w_{i_{\ell-1} i_\ell}, w_{i_\ell j} > 0$

$\xi(i) = \xi(i_1)$, $\xi(i_1) = \xi(i_2)$, $\dots$, $\xi(i_\ell) = \xi(j)$

$\Rightarrow \xi(i) = \xi(j)$

So for any $i,j \in V_k$

$\xi(i) = \xi(j) = c_k$

$$\Rightarrow \xi = \sum_{k \in [c]} c_k e_{G_k}$$

"Subspace Clustering"

* special case of clustering
* generalized principal component analysis

Rem 2 [motivation]. In many cases the data $X \in \mathbb{R}^{D \times N}$ are heterogenous (come from different sources) and each source is modeled by a linear subspace.

Dfn 3 $\mathcal{X} = [\mathcal{X}_1 \cdots \mathcal{X}_c]\underbrace{\Gamma}_{\text{unknown permutation}}$

$\mathcal{X}_i = [x_{i1} \cdots x_{iN_i}] \in \mathbb{R}^{D \times N_i}$

$x_{ij} \in \underbrace{\mathcal{S}_i}_{\text{linear subspace of } \mathbb{R}^D} \quad \forall j \in [N_i]$ ⊟

Rem 4 $\mathcal{S} = \sum_{i \in [c]} \mathcal{S}_i$

$x_{ij} \in \mathcal{S} \quad \forall i \in [c], \forall j \in [N_i]$ and so we are back to the case of PCA. ⊟

Ex 5 $D = 10, \; c = 6$

$d_i = \dim \mathcal{S}_i = 2$. In this case it may happen that $\mathcal{S} = \mathbb{R}^{10}$. ⊟

* dimensionality reduction via PCA might not be possible

* need to cluster

"Sparse Subspace Clustering"

Dfn 6 [sparse representation, compressed sensing]

$b \in \mathbb{R}^D$, $A \in \mathbb{R}^{D \times N}$

find the sparsest $x \in \mathbb{R}^N$ s.t. $Ax = b$ (usually $N \gg D$)

$$\min_{x \in \mathbb{R}^N} \underbrace{\|x\|_0}_{\text{counts \# of non-zero entries}} \text{ s.t. } Ax = b$$

if noise $\min_{x \in \mathbb{R}^D} \|x\|_0$ s.t. $\|Ax - b\|_2 \le \epsilon$

→ NP hard

$$\min_{x \in \mathbb{R}^D} \|x\|_1 \text{ s.t. } \|Ax - b\|_2 \le \epsilon$$

→ convex, Lasso

min
$x \in \mathbb{R}^D$
→ basi…

Rem 7

i) dim $\mathcal{S}$…

Idea: …
point …
linea…
rest …

min
$c \in \mathbb{R}$…

unknown

* dimensionality reduction via PCA might not be possible

* need to cluster

"Sparse Subspace Clustering"

Dfn 6 [sparse representation, compressed sensing]

$b \in \mathbb{R}^D$, $A \in \mathbb{R}^{D \times N}$ find the sparsest $x \in \mathbb{R}^N$ s.t. $Ax = b$ (usually $N \gg D$)

$$\min_{x \in \mathbb{R}^N} \|x\|_0 \quad \text{s.t. } Ax = b$$

$\|x\|_0$ counts # of non-zero entries

if noise $\min_{x \in \mathbb{R}^D} \|x\|_0$ s.t. $\|Ax - b\|_2 \leq \epsilon$

NP hard

$$\min_{x \in \mathbb{R}^D} \|x\|_1 \quad \text{s.t. } \|Ax - b\|_2 \leq \epsilon$$

convex, Lasso

$$\min_{x \in \mathbb{R}^D} \|x\|_1 \quad \text{s.t. } Ax = b$$

basis pursuit

Rem 7 hypothesis:

i) $\dim S_i =: d_i \ll D$ ii) $\forall\, ij$ $\operatorname{rank}(X_i \setminus \{x_{ij}\}) = d_i$

Idea: express every point x_{ij} as a sparse linear combination of the rest of the points

$$\min_{c \in \mathbb{R}^N} \|c\|_0 \quad \text{s.t. } x_{ij} = Xc, \; c(ij) = 0$$

LAIS, Lecture #22

Dfn 8 The solution c is called subspace preserving if $c(kl)=0 \ \forall\, k \neq i$. ∎

Thm 9 Suppose the S_i's are independent. Then the solution c to (*) will be subspace preserving.

Prf $X c = [X_i \ \underbrace{X_{-i}}_{\text{all points outside } X_i}] \begin{bmatrix} c_i \\ c_{-i} \end{bmatrix}$

We will show $c_{-i}=0$. Suppose $c_{-i} \neq 0$.

$x_{ij} = Xc = X_i c_i + X_{-i} c_{-i}$

$\Rightarrow \underbrace{x_{ij} - X_i c_i}_{\in S_i} = \underbrace{X_{-i} c_{-i}}_{\in \sum_{j \neq i} S_j}$

By independence $x_{ij} = X_i c_i$ and $X_{-i} c_{-i} = 0$.

So $\begin{bmatrix} c_i \\ 0 \end{bmatrix}$ is a feasible solution to (*). But

$\left\| \begin{bmatrix} c_i \\ 0 \end{bmatrix} \right\| < \left\| \begin{bmatrix} c_i \\ c_{-i} \end{bmatrix} \right\| \Rightarrow \Leftarrow$ ∎

Dfn 10 [disjoint subspaces]

$S_i \cap S_j = 0 \ \forall\, i \neq j$

Thm 11 For each $i \in [c]$ let $\mathcal{L}_i$ be the set of all $\tilde{X}_i \in \mathbb{R}^{D \times d_i}$ submatrices of X_i of rank d_i.

θ_{ij}: first principal angle between S_i, S_j

$$\cos\theta_{ij} = \max_{\substack{u \in S_i \\ \|u\|_2 = 1 \\ v \in S_j \\ \|v\|_2 = 1}} |u^T v|$$

Then the solution to $(*)$ is subspace preserving if

$$\max_{\tilde{X}_i \in \mathcal{L}_i} \sigma_{d_i}(\tilde{X}_i) > \sqrt{d_i} \max_{\substack{k \neq i \\ \ell \in [N_k]}} \|x_{k\ell}\|_2 \cdot \max_{k \neq i} \cos\theta_{ik}$$

Thm 12 Auxiliary problems:

$$\min_{\alpha} \|\alpha\|_1 \ \text{s.t.}\ y = X_i \alpha \quad (1)$$

$$\min_{b} \|b\|_1 \ \text{s.t.}\ y = X_{-i} b \quad (2)$$

for $y \in S_i \cap \left(\sum_{j \neq i} S_j\right)$