LEC10 Gaussian Graphical Model

1. Basic Recap

→ Joint dist $p(x_1, x_2, x_3) := \mathbb{P}(X_1 = x_1, X_2 = x_2, X_3 = x_3)$

→ Conditional dist $p(x_1, x_2 \mid x_3) := \mathbb{P}(X_1 = x_1, X_2 = x_2 \mid X_3 = x_3)$

$$= p(x_1 \mid x_2, x_3)\, p(x_2 \mid x_3)$$

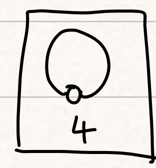
→ Conditional independent

x_1, x_2 is conditional indepedent given x_3

$$\Leftrightarrow \underline{p(x_1 . x_2 \mid x_3) = p(x_1 \mid x_3)\, p(x_2 \mid x_3)}$$

→ Undirected Graph $G = (V, E)$

(assume no self-loop) →



→ Directed Graph $G = (V, E)$

(assume no self-loop)

→ Markov Property $p(x_4 \mid x_3, x_2, x_1) = p(x_4 \mid x_3)$

↓

$$p(x_4, x_2 \mid x_3) = p(x_4 \mid x_2, x_3) \cdot p(x_2 \mid x_3)$$

$$= p(x_4 \mid x_3)\, p(x_2 \mid x_3)$$

$$\Rightarrow \boxed{x_{future} \perp\!\!\!\perp x_{past} \mid x_{now}}$$

→ Multi-variate Gaussian

$$P_G(x; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{p}{2}} (\det(\Sigma))^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$$

→ Gaussian Graphics Model

→ if $x \sim N(0, \Sigma)$, then $x_s \perp\!\!\!\perp x_t \mid x_{V-\{s,t\}}$

$\Leftrightarrow \Theta = \Sigma^{-1}, \ \Theta_{st} = 0$

$\Leftrightarrow (s,t) \notin E \quad s<t$

$\boxed{\begin{array}{l}\Sigma \to \text{covariance matrix} \\ \Theta \to \text{precision matrix}\end{array}}$

$\rightarrow$ MLE $\Longrightarrow$ estimate $\begin{cases} \mu \\ \Sigma \ (\Theta) \end{cases}$

$$\max_{\Theta \in S_{++}^p} \log\det(\Theta) - \langle S, \Theta \rangle \qquad S = \frac{1}{n}\sum_{i=1}^{n} X_i X_i^T$$

$$\Leftrightarrow \min_{\Theta \in S_{++}^p} \underbrace{\boxed{-\log\det(\Theta)}}_{h(\Theta)} + \langle S, \Theta \rangle$$

$$h(\Theta) = \begin{cases} -\sum_{j=1}^{p} \log(\lambda_j(\Theta)) & \text{if } \Theta \in S_{++}^p \\ +\infty & \text{o/w} \end{cases}$$

Property :

1. $h(\Theta) = -\log\det(\Theta)$ is convex on S_{++}^p — matrix derivative
2. $\boxed{\nabla h(\Theta) = -\Theta^{-1}}$ for any $\Theta \in S_{++}^p$

$\Downarrow$ imply

Corallary : if $\hat{\Theta}$ exists, then $\hat{\Theta} = S^{-1}$

$\rightarrow$ we need to make sure $S = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T$

$\longrightarrow$ a necessary condition is $n \geq p$

Limitation :

S is likely to be rank-deficient

$\Downarrow$

solution to MLE does not even exist

$\boxed{\text{regularization}}$ s.t $\begin{cases} \text{guarantee the existence of sol}^{\wedge} \\ \text{estimated } \Theta \text{ tend to be sparse} \end{cases}$

$\hookrightarrow$ $\boxed{\text{Graphical LASSO}}$ (add 1-norm $\|\Theta\|_{1,\text{off}}$)

$\downarrow$

orthogonal invariant (like $\|\cdot\|_*$)

$$\min_{\Theta \in S^p_{++}} \boxed{-\log\det(\Theta)} + \langle S, \Theta \rangle + \lambda \|\Theta\|_{1,\text{off}}$$

a) ADMM method

b) BCD

c) APG

d) Neighbourhood Selection

$\rightarrow$ idea: for each vertex j $\quad j = 1, 2, \ldots, p$

$\boxed{X_j \approx X_{-j} \cdot \beta}$ LR model + L1-norm

$$\hat{\beta}_j = \underset{\beta \in \mathbb{R}^{p-1}}{\arg\min} \ \frac{1}{2}\|X_j - X_{-j}\beta\|_2^2 + \lambda\|\beta\|_1$$

$\downarrow$

$\hat{\beta}_j$

use this sparse vector to determine the neighbour of j

$\Downarrow$ use rule (And / OR)

generate Edge Graph G