

Lecture 17 - scribbles

Friday, October 28, 2016
13:24

today: more exp. family
Gaussian networks
factor analysis

other example exp. family

UGM: let $p \in \mathcal{P}(G)$ G is undirected

with $\psi_c(x_c) > 0$

$$p(x) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \psi_c(x_c)$$

$$= \exp\left(\sum_c \ln \psi_c(x_c) - \log Z\right)$$

$$= \exp\left(\sum_{c \in \mathcal{C}} \sum_{y_c \in X_c} \underbrace{\mathbb{1}\{x_c = y_c\}}_{T(x)} \underbrace{\log \psi_c(x_c)}_{\eta(\psi_c)} - \log Z\right)$$

for every value of x_c & c ,
I have $T_{x_c, c}(x)$, $\eta_{x_c, c} = \log \psi_c(x_c)$

$$T(x) = \left(\mathbb{1}\{x_c = y_c\} \right)_{\substack{y_c \in X_c \\ c \in \mathcal{C}}} \\ \eta(\theta) = \left(\log \psi_c(y_c) \right)_{\dots}$$

notes:

a) Mult(1, π) is special case with complete graph (1 big clique)

b) feature perspective: instead of using all indicators $\mathbb{1}\{x_c = y_c\}$,
you could choose a subset

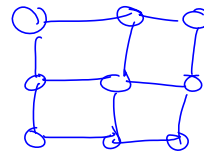
for example; suppose x_i is a word
feature on x_i & x_{i+1}
 $\mathbb{1}\{x_i \text{ is verb} \& x_{i+1} \text{ is noun}\}$

c) binary Ising model: $|C| \leq 2$

$x_i \in \{0, 1\}$ nodes & edges as cliques

\Rightarrow dimension of $T(x)$ is $2|V| + 4|E|$

'overparameterized'



a minimal representation

$$T(x) = \begin{pmatrix} (x_i)_{i \in V} \\ (x_i x_j)_{\{i, j\} \in E} \end{pmatrix} \rightarrow \dim |V| + |E|$$

$$\text{i.e. } p(x; \eta) = \frac{1}{Z} \exp\left(\sum_{i,j \in E} \eta_{ij} x_i x_j + \sum_{i \in V} \eta_i x_i\right)$$

def:

$p(x; \eta)$ is minimal exp family

(you say $\{T_j\}$ are a (finely dependent))

iff there does not exist $\{a_j\}$

$$\text{s.t. } \sum_{j=1}^P a_j T_j(x) + a_{P+1} = 0 \\ \forall x \text{ s.t. } h(x) \neq 0$$

properties of A :

- $A(\cdot)$ is C^∞ for $\eta \in \text{int}(\Omega)$
- $A(\cdot)$ is convex (Ω is convex set)
- $\eta \in \text{int}(\Omega) \triangleq \overset{\circ}{\Omega}$

$$\nabla_{\eta} A(\eta) = \mathbb{E}_{P_{\eta}}[T(x)] = \mu(\eta) \quad \text{"moment vector"}$$

$$\nabla^2 A(\eta) \text{ [Hessian]} = \mathbb{E}_{P_{\eta}}[(T(x) - \mu(\eta))(T(x) - \mu(\eta))^T] = \text{cov}(T(x))$$

proof (exercise)

* Z is the mgf for $p(x; \eta)$ $t \mapsto Z(\eta+t)$ for $t \in \mathbb{R}^P$
 $\hookrightarrow \mathbb{E} \exp(t^T x)$

$A(\eta) = \log \text{mgf} \rightarrow$ cumulant generating function

\rightarrow why derivatives of A give cumulants

1st derivative give mean
 2nd " give covariance
 " " give 3rd-cumulant etc...

Gaussian network:

$$X \sim N(\mu, \Sigma) \quad \mu \in \mathbb{R}^D, \Sigma \in \mathbb{R}^{D \times D} \quad \Sigma \succ 0$$

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

$$(x-\mu)^T \Sigma^{-1} (x-\mu) = \frac{x^T \Sigma^{-1} x}{1} - \frac{2\mu^T \Sigma^{-1} x}{1} + \frac{\mu^T \Sigma^{-1} \mu}{1}$$

$\Lambda \triangleq \Sigma^{-1}$
 define $\eta = \Sigma^{-1} \mu = \Lambda \mu$
 \Downarrow
 $\mu = \Lambda^{-1} \eta$

$\text{tr}(\Sigma^{-1} x x^T) \rightarrow$
 $\langle \Sigma^{-1}, x x^T \rangle - 2 \langle \Sigma^{-1} \mu, x \rangle + \mu^T \Sigma^{-1} \mu$
 $\rightarrow \langle \Lambda, x x^T \rangle - 2 \langle \eta, x \rangle + \eta^T \Lambda^{-1} \eta$

$p(x; \mu, \Sigma) = \exp(\eta^T x + \langle \Lambda, \frac{1}{2} x x^T \rangle - [\frac{1}{2} \eta^T \Lambda^{-1} \eta + \frac{1}{2} \log 2\pi - \frac{1}{2} \log |\Lambda|])$
 $A(\eta, \Lambda)$

$T(x) = \begin{pmatrix} x \\ \frac{1}{2} x x^T \end{pmatrix}$

canonical parameter $\tilde{\eta}(\theta) = \begin{pmatrix} \eta \\ \Lambda \end{pmatrix} = \begin{pmatrix} \Sigma^{-1} \mu \\ \Sigma^{-1} \end{pmatrix}$

$\Omega = \{ (\eta, \Lambda) : \eta \in \mathbb{R}^p, \Lambda \succ 0, \Lambda \in \mathbb{R}^{p \times p}, \Lambda = \Lambda^T \}$

$\nabla_{\eta} A(\eta, \Lambda) = \Lambda^{-1} \tilde{\eta} = \mu = \mathbb{E}[x]$

$\nabla_{\eta}^2 A(\eta, \Lambda) = \Lambda^{-1} = \Sigma = \text{cov}(x)$

UGM view:

$p(x; \eta, \Lambda) = \exp(-\frac{1}{2} \sum_{i,j} \Lambda_{ij} x_i x_j + \sum_i \eta_i x_i - A(\eta, \Lambda))$

$p \in \mathcal{F}(\theta)$ where $E = \{ \{i,j\} \text{ st. } \Lambda_{ij} \neq 0 \}$

zeros in precision matrix \Rightarrow cond. indep. properties

"Gaussian network"

Schur-complement disgression

block operations on matrices

$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$

to remove A_{21} ; subtract $A_{21} A_{11}^{-1}$ (1st row) to second row

$\begin{pmatrix} I & 0 \\ -A_{21} A_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} I & -A_{11}^{-1} A_{12} \\ 0 & T \end{pmatrix}$

$$= \begin{pmatrix} A_{11} & A_{12} \\ 0 & \underbrace{A_{22} - A_{21}A_{11}^{-1}A_{12}} \end{pmatrix} \begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix}$$

\triangleq Schur complement of A with respect A_{11}

$$= \begin{pmatrix} A_{11} & 0 \\ 0 & A/A_{11} \end{pmatrix}$$

$\stackrel{\text{red}}{=} A/A_{11}$

$$LAU = D$$

$$U^{-1}A^{-1}L^{-1} = D^{-1} = \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & (A/A_{11})^{-1} \end{pmatrix} \quad M \stackrel{\text{red}}{=} A/A_{11}$$

$$A^{-1} = U D^{-1} L = \begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & M^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix}$$

$$= \begin{pmatrix} A_{11}^{-1} & -A_{11}^{-1}A_{12}M^{-1} \\ 0 & M^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix}$$

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} A_{11}^{-1} & -A_{11}^{-1}A_{12}M^{-1} \\ -M^{-1}A_{21}A_{11}^{-1} & M^{-1} \end{pmatrix}$$

$$M = A/A_{11} = A_{22} - A_{21}A_{11}^{-1}A_{12}$$

If you do same process but using A_{22} to remove off-diagonal

$$A^{-1} = \begin{pmatrix} \underline{\underline{(A/A_{22})^{-1}}} & \underline{\underline{\quad}} \\ \underline{\underline{\quad}} & \underline{\underline{(A/A_{11})^{-1}}} \end{pmatrix}$$

"Woodbury-Sherman-Morrison" inversion formula

$$\text{bottom line: } \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} (A/A_{22})^{-1} & -A_{11}^{-1}A_{12}(A/A_{11})^{-1} \\ \underline{\underline{\quad}} & (A/A_{11})^{-1} \end{pmatrix}$$

$$[X = \begin{pmatrix} -x_1^T \\ \vdots \end{pmatrix}$$

$$X X^T = \text{vector} \cdot \text{vector}^T$$

$$\rightarrow (\text{matrix})_{ij} = x_i^T x_j \rightarrow \text{Gram matrix}$$

$$X^T X \quad (\text{vector}^T) \text{vector} = \sum_i x_i x_i^T$$

lemma (determinant lemma)

from $LAU = D$ get $|A| = |D| = |A_{11}| |A/A_{11}|$
 $|A| = |A_{11}| |A/A_{11}| = |A_{22}| |A/A_{22}|$

(positivity lemma) if A is symmetric
 $A > 0 \Leftrightarrow A_{11} > 0 \ \& \ A/A_{11} > 0$

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}$$

$$\Sigma_{12}^T = \Sigma_{21} \quad \Lambda_{22} = (\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})^{-1}$$

$$\begin{aligned} (x-\mu)^T \Sigma^{-1} (x-\mu) &= \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}^T \underbrace{\begin{pmatrix} I & -\Sigma_{12} \Sigma_{11}^{-1} \\ 0 & I \end{pmatrix}}_U \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & (\Sigma/\Sigma_{11})^{-1} \end{pmatrix} \underbrace{\begin{pmatrix} I & 0 \\ -\Sigma_{21}^T \Sigma_{11}^{-1} & I \end{pmatrix}}_{\begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 - b \end{pmatrix}} \\ &= \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 - b \end{pmatrix}^T \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & (\Sigma/\Sigma_{11})^{-1} \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 - b \end{pmatrix} \end{aligned}$$

$$p(x_1, x_2) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \exp\left((x_1 - \mu_1)^T \Sigma_{11}^{-1} (x_1 - \mu_1) \right) \} p(x_1)$$

$$\cdot \frac{1}{\sqrt{(2\pi)^2 |\Sigma/\Sigma_{11}|}} \exp\left((x_2 - \mu_2 - b)^T (\Sigma/\Sigma_{11})^{-1} (x_2 - \mu_2 - b) \right) \} p(x_2|x_1)$$

$$\begin{aligned} \mu_1^m &= \mu_1 \\ \Sigma_1^m &= \Sigma_{11} \end{aligned} \quad \left. \vphantom{\begin{aligned} \mu_1^m \\ \Sigma_1^m \end{aligned}} \right\} \text{marginal}$$

$$\begin{aligned} \mu_{2|1}^{\text{cond.}} &= \mu_2 + b = \mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (x_1 - \mu_1) \\ \Sigma_{21}^{\text{cond.}} &= \Sigma/\Sigma_{11} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \end{aligned} \quad \left. \vphantom{\begin{aligned} \mu_{2|1}^{\text{cond.}} \\ \Sigma_{21}^{\text{cond.}} \end{aligned}} \right\} \text{conditioning 2 on 1}$$

canonical parametrization: $\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}$

$$\begin{aligned} \Lambda_{22|1}^{\text{cond.}} &= \Lambda_{22} \\ \Lambda_{21}^{\text{cond.}} &= \Lambda_{21} - \Lambda_{21} \Lambda_{11}^{-1} \Lambda_{12} \end{aligned} \quad \rightarrow \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \Lambda \mu$$

$$\mu_1 = \mu_2 = \dots = \mu_{p-1}$$

$$\Rightarrow \mu_1 = \Lambda_{11}\mu_1 + \Lambda_{12}\mu_2$$

$$\mu_1^m = \mu_1 - \Lambda_{12}\Lambda_{22}^{-1}\mu_2$$

$$\Lambda_{11}^m = \Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21} = \Lambda_{11}/\Lambda_{22}$$

x_1, \dots, x_p Gaussian network,

block $\{i, j\}$ | rest

$$\text{cov}(X_I | X_{\text{rest}}) = \Sigma_{II|rest} = \Lambda_{II}^{-1} = \begin{pmatrix} \Lambda_{ii} & \Lambda_{ij} \\ \Lambda_{ji} & \Lambda_{jj} \end{pmatrix}^{-1}$$

$$\text{if } \Lambda_{ij} = 0 \Rightarrow \Sigma_{II|rest} = \begin{pmatrix} \Lambda_{ii}^{-1} & 0 \\ 0 & \Lambda_{jj}^{-1} \end{pmatrix}$$

$$\Rightarrow X_i \perp\!\!\!\perp X_j | X_{\text{rest}}$$

