

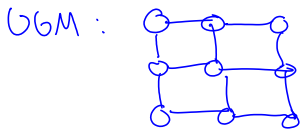
Lecture 21 - scribbles

Tuesday, November 15, 2016
14:38

- Today:
- Gibbs sampling example
 - variational methods
 - estimation in GM

Ising model - Gibbs sampling

Ising model $x_i \in \{-1, 1\}$



$$p(x) = \frac{1}{Z(n)} \exp\left(\sum_i \eta_i x_i + \sum_{\langle i, j \rangle \in E} \eta_{ij} x_i x_j\right)$$

for Gibbs, want compute $p(x_i | x_{-i}) \propto p(x_i, x_{-i})$

$$\propto \exp\left(\underbrace{\eta_i}_{\text{varies}} x_i + \sum_{\substack{j \in N(i) \\ \text{fixed}}} \eta_{ij} x_i x_j + \text{rest}\right)$$

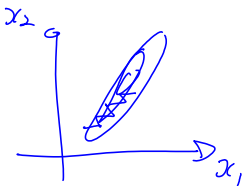
\Rightarrow renormalize to get conditional

$$p(x_i = 1 | x_{-i}) = \frac{\exp(\eta_i + \sum_{j \in N(i)} \eta_{ij} x_j)}{1 + \exp(\dots)}$$

$$= \sigma\left(\eta_i + \sum_{j \in N(i)} \eta_{ij} x_j\right)$$

where $\sigma(z) \triangleq \frac{1}{1 + \exp(-z)}$ Sigmoid

Gibbs sampling need to be able to sample from conditionals $p(x_i | x_{-i})$



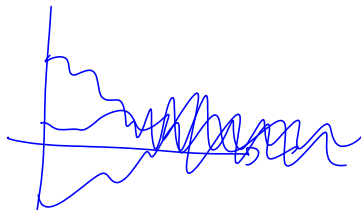
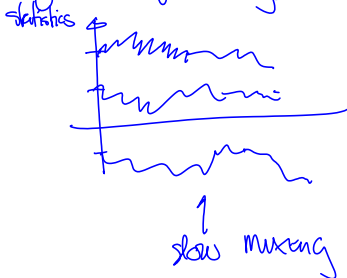
block Gibbs sampling

$$p(x_A | x_{-A})$$

↑
set

$\cup A_i = V$
↑
set of all variables

diagnostic of mixing



$$\frac{1}{Z} \exp\left(-\frac{1}{k_B T} E(x)\right)$$

↑ energy

Variational methods:

general idea: say we want to approximate Θ^*

then, express it as sol'n to optimization problem

$$\Theta^* = \underset{\theta \in \Theta}{\operatorname{argmin}} f(\theta) \quad] \text{OPT}$$

idea: approximate Θ^* by approximate OPT

linear algebra example: $\min_x \|Ax - b\|^2$
 $Ax = b \Rightarrow x = A^{-1}b$

Variational EM:

recall EM trick latent variable model $p(x, z)$ ^{unobserved}

$$\begin{aligned} \log p(x) &= \log \sum_z q(z) \frac{p(x, z)}{q(z)} \\ &= \log \mathbb{E}_q \left[\frac{p(x, z)}{q(z)} \right] \\ &\stackrel{\text{Jensen}}{\geq} \mathbb{E}_q \left[\log \frac{p(x, z)}{q(z)} \right] = J(q, \theta) \end{aligned}$$

$$\log p(x) - \mathbb{E}_q \left[\log \frac{p(x, z)}{q(z)} \right] = -\mathbb{E}_q \left[\log \frac{p(x, z)}{q(z)p(x)} \right]$$

↑ auxiliary fct. (lower bound)

$$= \mathbb{E}_q \left[\log \frac{q(z)}{p(z|x)} \right] = KL(q(\cdot) || p(\cdot|x))$$

$$\begin{aligned} \log p(x) & \\ \downarrow & \\ \underbrace{KL(q || p(z|x))}_{J(q, \theta)} & \end{aligned}$$

$$\begin{aligned} \text{E-step} &\rightarrow \underset{q \text{ over all dist}}{\operatorname{argmax}} J(q, \theta^{(t)}) \\ &\Leftrightarrow \underset{q}{\operatorname{argmin}} KL(q(\cdot) || p(\cdot|x)) \end{aligned}$$

variational approximation for the E-step is

$$\text{do } q_{\text{approx}} = \underset{q \in Q_{\text{simple}}}{\operatorname{argmin}} KL(q(\cdot) || p(\cdot|x))$$

[still get lower bound on $\log p(x | \theta^{(t)})$]

↳ to approximate $p(z|x)$

approximate M-step: $\underset{\theta}{\operatorname{argmax}} \mathbb{E}_{q_{\text{approx}}} [\log p(x, z | \theta)]$

more generally, using $\underset{q \in Q}{\operatorname{argmin}} KL(q || p)$ is variational approach

Let's suppose that $p(z)$ is exp family

$$z_1, \dots, z_p \quad p(z) = \exp(\eta^T T(z) - A(\eta))$$

mean field approximation: $\mathcal{Q} = \{q(z) = \prod_i q_i(z_i)\}$
 set of fully factored dist.

$$\begin{aligned} KL(q||p) &= \mathbb{E}_q \left[\log \frac{q(z)}{p(z)} \right] \\ &= -\eta^T \mathbb{E}_q [T(z)] + A(\eta) + \sum_z q(z) \log q(z) \\ &= \sum_i \underbrace{\left(\sum_{z_i} q_i(z_i) \log q_i(z_i) \right)} \end{aligned}$$

coordinate ascent fix q_j for $j \neq i$

maximize with respect to q_i : $KL(q_i: q_{-i} || p)$

$$= -\mathbb{E}_{q_i} \left[\underbrace{\eta^T \mathbb{E}_{q_{-i}} T(z)}_{f_i(z_i)} \right] + \text{const.} + \sum_{z_i} q_i(z_i) \log q_i(z_i)$$

add Lagrange multiplier for $\sum_{z_i} q_i(z_i) = 1$ $\rightarrow \lambda (1 - \sum_{z_i} q_i(z_i))$

$$\frac{\partial}{\partial q_i(z_i)} \rightarrow -f_i(z_i) + \log q_i(z_i) + 1 - \lambda = 0$$

$$q_i(z_i) \propto \exp(f_i(z_i))$$

general mean field update when target p is in exp family
 $q_i^{(t+1)}(z_i) \propto \exp(\eta^T \mathbb{E}_{q_{-i}^{(t)}} T(z))$

Ising model example: $T(z) = \begin{pmatrix} z_i \\ z_i z_j \end{pmatrix}_{i,j \in E}$

$$\mathbb{E}_{q_{-i}} [z_j] = q_j(z_j=1) \triangleq \mu_j$$

$$\mathbb{E}_{q_{-i}} [z_i z_j] = z_i \mu_j \mu_j^{(t)}$$

$$\begin{aligned} \eta^T \mathbb{E}_{q_{-i}^{(t)}} T(z) &= \eta_i z_i + \sum_{j \in E} \eta_j \underbrace{\mathbb{E}_{q_{-i}^{(t)}} [z_j]}_{\mu_j^{(t)}} \\ &+ \sum_{j \in E(i)} \eta_{ij} \underbrace{\mathbb{E}_{q_{-i}^{(t)}} [z_i z_j]}_{z_i \mu_j^{(t)}} + \text{rest (no } z_i) \end{aligned}$$

result: $q_i^{(t+1)}(z_i) \propto \exp(\eta_i z_i + \sum_{j \in E(i)} \eta_{ij} \mu_j^{(t)})$

$$\mu_i^{(t+1)} = \sigma \left(\eta_i + \sum_{j \in E(i)} \eta_{ij} \mu_j^{(t)} \right)$$

MF updates for $q_i(z_i)$ [with μ_i parameters]

here min $KL(q||p)$ over $q \in \mathcal{Q}$
 $KL(\cdot || p)$ is convex fun.
 but \mathcal{Q}_{MF} is non-convex constraint set

\Rightarrow MF can get stuck in local minimum
 but can monitor progress by $KL(q^{(t)} || p)$

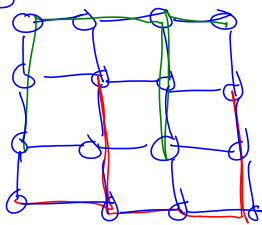
pros & cons of variational method vs sampling
 \oplus optimization based \rightarrow often faster & easier to debug
 \ominus noisy \Rightarrow hard to debug, causing problem for chains

\ominus biased estimate $\mathbb{E}_{q(z)} [f(z)] \neq \mathbb{E}_p [f(z)]$ \oplus unbiased estimate $\mathbb{E} [\mathbb{E}_{q(z)} [f(z)]] = \mathbb{E}_p [f(z)]$
 with respect to random sample

structured mean field:

\rightarrow use GM for Q with tractable treewidth

e.g. for Ising model



\rightarrow can compute $\mathbb{E}_q[\dots]$ efficiently etc...

suppose p is in exponential form.

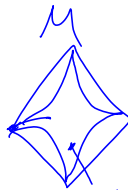
$$\begin{aligned}
 KL(q || p) &= \mathbb{E}_q \left[\log \frac{q(z)}{p(z)} \right] = \mathbb{E}_q [-\log p(z)] - (-\mathbb{E}_q [\log q(z)]) \\
 &= -n^T \underbrace{\mathbb{E}_q [T(x)]}_{\mu(q)} + A(n) - H(q)
 \end{aligned}$$

$$KL(q || p) \geq 0 \Rightarrow A(n) \geq n^T \mu(q) + H(q)$$

$$\begin{aligned}
 A(n) &= \sup_{q \text{ all dist.}} n^T \mu(q) + H(q) \\
 &= \sup_{\mu \in M} n^T \mu + H(\mu)
 \end{aligned}$$

recall that $\mu(q) \triangleq \mathbb{E}_q [T(x)]$

$M = \{ \mu : \exists \text{ a dist. } q \text{ s.t. } \mathbb{E}_q [T(x)] = \mu \}$
 \triangleq Marginal polytope



$$\mu(n) = \nabla_n A(n)$$

\mathcal{Q} = exponential family \rightarrow $\text{int}(M)$