Graphical Models, Exponential Families and Variational Inference

4.3
Expectation Propagation

Vincent Adam
Alessandro Davide Ialongo

March 24, 2015

Expectation Propagation Algorithms

- What is EP?
  - Another 'message-passing' like algorithm
  - Sequence of moment matching operations

Expectation Propagation Algorithms

- What is EP?
  - Another 'message-passing' like algorithm
  - Sequence of moment matching operations

Outline

- EP as we know it
- EP and the variational principle

EP as we know it

- Approximation of $p(x) = \prod_i f_i(x_i)$ as $\tilde{p}(x) \approx \prod_i \tilde{f}_i(x_i)$ (sites)
  - each approximated factor $\tilde{f}_i$ has simple exp-fam parameterization with suff-stats $\phi_i(x_i)$

EP as we know it

- Approximation of $p(x) = \prod_i f_i(x_i)$ as $\tilde{p}(x) \approx \prod_i \tilde{f}_i(x_i)$ (sites)
  - each approximated factor $\tilde{f}_i$ has simple exp-fam parameterization with suff-stats $\phi_i(x_i)$
- Updating $\tilde{f}_i$ (pick an ordering),
  - $q_{\neg i}(x_i) = \int dx_{\setminus i} \frac{\tilde{p}(x)}{\tilde{f}_i(x_i)}$ (cavity)

EP as we know it

- Approximation of $p(x) = \prod_i f_i(x_i)$ as $\tilde{p}(x) \approx \prod_i \tilde{f}_i(x_i)$ (sites)
  - each approximated factor $\tilde{f}_i$ has simple exp-fam parameterization with suff-stats $\phi_i(x_i)$
- Updating $\tilde{f}_i$ (pick an ordering),
  - $q_{\neg i}(x_i) = \int dx_{\setminus i} \frac{\tilde{p}(x)}{\tilde{f}_i(x_i)}$ (cavity)
  - $\int dx_i \; q_{\neg i}(x_i) f_i(x_i) \phi_i(x_i) = \int dx_i \; q_{\neg i}(x_i) \tilde{f}_i(x_i) \phi_i(x_i)$
    $\mathbb{E}_{q_{\neg i} f_i(x_i)}[\phi_i] = \mathbb{E}_{q_{\neg i} \tilde{f}_i(x_i)}[\phi_i]$
    (moment matching)

EP as we know it

- Approximation of $p(x) = \prod_i f_i(x_i)$ as $\tilde{p}(x) \approx \prod_i \tilde{f}_i(x_i)$ (sites)
  - each approximated factor $\tilde{f}_i$ has simple exp-fam parameterization with suff-stats $\phi_i(x_i)$
- Updating $\tilde{f}_i$ (pick an ordering),
  - $q_{\neg i}(x_i) = \int dx_{\setminus i} \frac{\tilde{p}(x)}{\tilde{f}_i(x_i)}$ (cavity)
  - $\int dx_i \; q_{\neg i}(x_i) f_i(x_i) \phi_i(x_i) = \int dx_i \; q_{\neg i}(x_i) \tilde{f}_i(x_i) \phi_i(x_i)$
    $\mathbb{E}_{q_{\neg i} f_i(x_i)}[\phi_i] = \mathbb{E}_{q_{\neg i} \tilde{f}_i(x_i)}[\phi_i]$
    (moment matching)

Variational idea invoked locally:
$\mathbb{E}_{q_{\neg i} f_i(x_i)}[\phi_i] = \mathbb{E}_{q_{\neg i} \tilde{f}_i(x_i)}[\phi_i] \Leftrightarrow$
$\tilde{f}_i = \arg\min_{\tilde{f}_i} KL\left[q_{\neg i}(x_i) f_i(x_i) || q_{\neg i}(x_i) \tilde{f}_i(x_i)\right]$

Entropy Approximations Based on Term Decoupling (p111)

- $(X_1, \ldots, X_m) \in \mathbb{R}^m$
- $\underbrace{\phi = (\phi_1, \ldots, \phi_{d_T})}_{Tractable}$ and $\underbrace{\Phi = (\Phi^1, \ldots, \Phi^{d_I})}_{Intractable}$ sufficient statistics

Entropy Approximations Based on Term Decoupling (p111)

- $(X_1, ..., X_m) \in \mathbb{R}^m$
- $\underbrace{\phi = (\phi_1, ..., \phi_{d_T})}_{\textit{Tractable}}$ and $\underbrace{\Phi = (\Phi^1, ..., \Phi^{d_I})}_{\textit{Intractable}}$ sufficient statistics

The $(\phi, \Phi)-$Exponential Family

- parameters $\theta, \tilde{\theta} \leftrightarrow \phi, \Phi$
- $p(x; \theta, \tilde{\theta}) \propto f_0(x) \exp\left(\langle \theta, \phi(x) \rangle\right) \exp\left(\langle \tilde{\theta}, \Phi(x) \rangle\right)$
- base model $p(x; \theta, \overrightarrow{0}) \propto f_0(x) \exp\left(\langle \theta, \phi(x) \rangle\right)$
  (no intractable component)

Entropy Approximations Based on Term Decoupling (p111)

- $(X_1, ..., X_m) \in \mathbb{R}^m$
- $\underbrace{\phi = (\phi_1, ..., \phi_{d_T})}_{Tractable}$ and $\underbrace{\Phi = (\Phi^1, ..., \Phi^{d_I})}_{Intractable}$ sufficient statistics

The $(\phi, \Phi)-$Exponential Family

- parameters $\theta, \tilde{\theta} \leftrightarrow \phi, \Phi$
- $p(x; \theta, \tilde{\theta}) \propto f_0(x) \exp(\langle \theta, \phi(x) \rangle) \exp\left(\langle \tilde{\theta}, \Phi(x) \rangle\right)$
- base model $p(x; \theta, \overrightarrow{0}) \propto f_0(x) \exp(\langle \theta, \phi(x) \rangle)$ (no intractable component)

The $(\phi, \Phi^i)-$Exponential Family : "Φ^i-Augmented"

- $p(x; \theta, \tilde{\theta}^i) \propto f_0(x) \exp(\langle \theta, \phi(x) \rangle) \exp\left(\langle \tilde{\theta}^i, \Phi^i(x) \rangle\right)$

Example Tractable/Intractable Partitioning (p112)

Mixture Model

- Likelihood $p(y|X = x) = (1 - \alpha)\mathcal{N}(y; 0, \sigma_0^2\mathbb{I}) + \alpha\mathcal{N}(y; x, \sigma_1^2\mathbb{I})$
- Prior $X \sim \mathcal{N}(0, \Sigma)$

Example Tractable/Intractable Partitioning (p112)

Mixture Model

- Likelihood $p(y|X=x) = (1-\alpha)\mathcal{N}(y; 0, \sigma_0^2\mathbb{I}) + \alpha\mathcal{N}(y; x, \sigma_1^2\mathbb{I})$
- Prior $X \sim \mathcal{N}(0, \Sigma)$
- Posterior

$$
\begin{aligned}
p(x|y^1 ..., y^n) &\propto \exp\left(-\frac{1}{2}x^T\Sigma^{-1}x\right)\prod_i p(y^i|X=x) \\
&\propto \underbrace{\exp\left(-\frac{1}{2}x^T\Sigma^{-1}x\right)}_{Tractable=base}\underbrace{\exp\left\{\sum_i \log\ p(y^i|X=x)\right\}}_{Intractable,\, d_I=|\mathcal{Y}|}
\end{aligned}
$$

Example Tractable/Intractable Partitioning (p112)

Mixture Model

- Likelihood $p(y|X=x) = (1-\alpha)\mathcal{N}(y; 0, \sigma_0^2\mathbb{I}) + \alpha\mathcal{N}(y; x, \sigma_1^2\mathbb{I})$
- Prior $X \sim \mathcal{N}(0, \Sigma)$
- Posterior

$$
\begin{aligned}
p(x|y^1 ..., y^n) &\propto \exp\left(-\frac{1}{2}x^T\Sigma^{-1}x\right)\prod_i p(y^i|X=x) \\
&\propto \underbrace{\exp\left(-\frac{1}{2}x^T\Sigma^{-1}x\right)}_{Tractable=base}\underbrace{\exp\left\{\sum_i \log\ p(y^i|X=x)\right\}}_{Intractable,\, d_I=|\mathcal{Y}|}
\end{aligned}
$$

"Φ^i−Augmented" corresponds to having a single observation and is a tractable case (2 components, otherwise $2^{|\mathcal{Y}|}$)

“Φ^i−Augmented”, tractable

In the (ϕ, Φ^i)−Exponential Family

- marginals tractable
- Entropy tractable

"Φ^i-Augmented", tractable

In the $(\phi, \Phi^i)-$Exponential Family

- marginals tractable
- Entropy tractable

In what follows, use these 1-augmented families to

- approximate $\mathbb{M}(G)$
- approximate the entropy

Acceptable means for the (ϕ, Φ)−ExpFam (pp. 113-114)

Notation

- $\mu = \mathbb{E}[\phi(x)]$, $\tilde{\mu} = \mathbb{E}[\Phi(x)]$
- $\mathcal{M}(\phi, \Phi) = \{(\mu, \tilde{\mu}) \,|\, (\mu, \tilde{\mu}) = \mathbb{E}\left[(\phi(x), \Phi(x))\right] \text{ for some } p\}$
- Same for base (Φ empty) or "Φ^i−Augmented"

Acceptable means for the $(\phi, \Phi)-$ExpFam (pp. 113-114)

Notation

- $\mu = \mathbb{E}[\phi(x)]$, $\tilde{\mu} = \mathbb{E}[\Phi(x)]$
- $\mathcal{M}(\phi, \Phi) = \{(\mu, \tilde{\mu}) \,|\, (\mu, \tilde{\mu}) = \mathbb{E}\left[(\phi(x), \Phi(x))\right] \text{ for some } p\}$
- Same for base (Φ empty) or "Φ^i-Augmented"

Projection operator ('cropping') on acceptable means

- acceptable mean: $(\tau, \tilde{\tau}) \in \mathcal{M}(\phi, \Phi)$
- projection $(\tau, \tilde{\tau}) \stackrel{\Pi^i}{\rightarrow} (\tau, \tilde{\tau}^i)$

Acceptable means for the (ϕ, Φ)−ExpFam (pp. 113-114)

Notation

- $\mu = \mathbb{E}[\phi(x)]$, $\tilde{\mu} = \mathbb{E}[\Phi(x)]$
- $\mathcal{M}(\phi, \Phi) = \{(\mu, \tilde{\mu}) \,|\, (\mu, \tilde{\mu}) = \mathbb{E}\left[(\phi(x), \Phi(x))\right] \text{ for some } p\}$
- Same for base (Φ empty) or "Φ^i−Augmented"

Projection operator ('cropping') on acceptable means

- acceptable mean: $(\tau, \tilde{\tau}) \in \mathcal{M}(\phi, \Phi)$
- projection $(\tau, \tilde{\tau}) \overset{\Pi^i}{\to} (\tau, \tilde{\tau}^i)$

Approximating $\mathcal{M}(\phi, \Phi)$

$$\begin{aligned}\mathcal{L}(\phi, \Phi) &= \left\{(\tau, \tilde{\tau}) \,|\tau \in \mathcal{M}(\phi), \quad \Pi^i(\tau, \tilde{\tau}) \in \mathcal{M}(\phi, \Phi^i) \quad \forall i = 1, \ldots, d_I\right\} \\ &= \cap_i \left\{(\tau, \tilde{\tau}) \,|\tau \in \mathcal{M}(\phi), \quad \Pi^i(\tau, \tilde{\tau}) \in \mathcal{M}(\phi, \Phi^i)\right\}\end{aligned}$$

Acceptable means for the $(\phi, \Phi)-$ExpFam (pp. 113-114)

Notation

- $\mu = \mathbb{E}[\phi(x)]$, $\tilde{\mu} = \mathbb{E}[\Phi(x)]$
- $\mathcal{M}(\phi, \Phi) = \{(\mu, \tilde{\mu}) \,|\, (\mu, \tilde{\mu}) = \mathbb{E}\left[(\phi(x), \Phi(x))\right] \text{ for some } p\}$
- Same for base (Φ empty) or "Φ^i-Augmented"

Projection operator ('cropping') on acceptable means

- acceptable mean: $(\tau, \tilde{\tau}) \in \mathcal{M}(\phi, \Phi)$
- projection $(\tau, \tilde{\tau}) \stackrel{\Pi^i}{\rightarrow} (\tau, \tilde{\tau}^i)$

Approximating $\mathcal{M}(\phi, \Phi)$

$$\begin{aligned}
\mathcal{L}(\phi, \Phi) &= \left\{(\tau, \tilde{\tau}) \,|\tau \in \mathcal{M}(\phi), \quad \Pi^i(\tau, \tilde{\tau}) \in \mathcal{M}(\phi, \Phi^i) \quad \forall i = 1, ..., d_I\right\} \\
&= \cap_i \left\{(\tau, \tilde{\tau}) \,|\tau \in \mathcal{M}(\phi), \quad \Pi^i(\tau, \tilde{\tau}) \in \mathcal{M}(\phi, \Phi^i)\right\}
\end{aligned}$$

Remark:

- intersection of convex sets
- $\mathcal{M}(\phi, \Phi) \subseteq \mathcal{L}(\phi, \Phi)$

Approximating $\mathcal{M}$ and $H(\tau, \tilde{\tau})$ (pp. 114-115)

Approximating $\mathcal{M}$

$$\mathcal{L}(\phi, \Phi) \quad = \quad \left\{(\tau, \tilde{\tau})\,|\tau \in \mathcal{M}(\phi), \quad \Pi^{i}\,(\tau, \tilde{\tau}) \in \mathcal{M}(\phi, \Phi^{i}) \quad \forall i = 1, ..., d_I\right\}$$

Approximating $\mathcal{M}$ and $H(\tau, \tilde{\tau})$ (pp. 114-115)

Approximating $\mathcal{M}$

$$\mathcal{L}(\phi, \Phi) = \left\{(\tau, \tilde{\tau}) \,|\, \tau \in \mathcal{M}(\phi), \quad \Pi^{i}(\tau, \tilde{\tau}) \in \mathcal{M}(\phi, \Phi^{i}) \quad \forall i = 1, ..., d_I\right\}$$

Approximating $H(\tau, \tilde{\tau})$

- $H(\tau, \tilde{\tau})$ is not tractable , but $H(\tau, \tilde{\tau}^{I})$ tractable

$$\begin{aligned} H_{ep}(\tau, \tilde{\tau}) &= H(\tau) + \sum_{I} \left[H(\tau, \tilde{\tau}^{I}) - H(\tau) \right] \\ &= \sum_{I=1}^{d_I} H(\tau, \tilde{\tau}^{I}) - (d_I - 1)\, H(\tau) \end{aligned}$$

Approximating $\mathcal{M}$ and $H(\tau, \tilde{\tau})$ (pp. 114-115)

Approximating $\mathcal{M}$

$$\mathcal{L}(\phi, \Phi) \quad = \quad \left\{(\tau, \tilde{\tau})\,|\tau \in \mathcal{M}(\phi), \quad \Pi^{i}\,(\tau, \tilde{\tau}) \in \mathcal{M}(\phi, \Phi^{i}) \quad \forall i = 1, ..., d_I\right\}$$

Approximating $H(\tau, \tilde{\tau})$

- $H(\tau, \tilde{\tau})$ is not tractable , but $H(\tau, \tilde{\tau}^{I})$ tractable

$$\begin{aligned} H_{ep}(\tau, \tilde{\tau}) &= H(\tau) + \sum_{I} \left[H(\tau, \tilde{\tau}^{I}) - H(\tau)\right] \\ &= \sum_{I=1}^{d_I} H(\tau, \tilde{\tau}^{I}) - (d_I - 1)\, H(\tau) \end{aligned}$$

Final optimization problem

$$\max_{(\tau, \tau') \in \mathcal{L}(\phi, \Phi)} \left\{\langle \tau, \theta \rangle + \langle \tilde{\tau}, \tilde{\theta} \rangle + H_{ep}(\tau, \tau')\right\}, \text{ eq. } (4.69)$$

Understanding H_{ep}

$$
\begin{aligned}
H_{ep}(\tau, \tilde{\tau}) &= H(\tau) + \sum_{l} \left[H(\tau, \tilde{\tau}^{l}) - H(\tau) \right] \\
&= \sum_{l=1}^{d_l} H(\tau, \tilde{\tau}^{l}) - (d_l - 1)\, H(\tau)
\end{aligned}
$$

Understanding H_{ep}

$$
\begin{aligned}
H_{ep}(\tau, \tilde{\tau}) &= H(\tau) + \sum_{l} \left[H(\tau, \tilde{\tau}^{l}) - H(\tau) \right] \\
&= \sum_{l=1}^{d_l} H(\tau, \tilde{\tau}^{l}) - (d_l - 1)\, H(\tau)
\end{aligned}
$$

When is it exact?

- if $H(\tau, \tilde{\tau}) = H(\tau) + \sum_i \Delta H(\tilde{\tau}_i)$, it is exact
- if sufficient statistics have disjoint set of random variables

Understanding H_{ep}

$$
\begin{aligned}
H_{ep}(\tau, \tilde{\tau}) &= H(\tau) + \sum_{l} \left[H(\tau, \tilde{\tau}^{l}) - H(\tau) \right] \\
&= \sum_{l=1}^{d_l} H(\tau, \tilde{\tau}^{l}) - (d_l - 1)\, H(\tau)
\end{aligned}
$$

When is it exact?

- if $H(\tau, \tilde{\tau}) = H(\tau) + \sum_i \Delta H(\tilde{\tau}_i)$, it is exact
- if sufficient statistics have disjoint set of random variables

What it is not

- conditional entropy

Example 4.9 - Sum-Product and Bethe Approximation

Pairwise Markov random field on Graph $G = (V, E)$

$$
\begin{aligned}
\mathcal{L}(\phi, \Phi) &= \left\{ (\tau, \tilde{\tau}) \,|\, \underbrace{\tau \in \mathcal{M}(\phi)}_{\text{loc. norm.}}, \underbrace{(\tau, \tau_{uv}) \in \mathcal{M}(\phi, \Phi^{uv})}_{\text{loc. cons.}}, \forall (u, v) \in E \right\} \\
&= \mathbb{L}(G)
\end{aligned}
$$

4.3.2 Optimality in terms of Moment-Matching

Recall

$$\mathcal{L}(\phi, \Phi) \quad = \quad \cap_i \left\{ (\tau, \tilde{\tau}) \, | \tau \in \mathcal{M}(\phi), \quad \Pi^i \, (\tau, \tilde{\tau}) \in \mathcal{M}(\phi, \Phi^i) \right\}$$

4.3.2 Optimality in terms of Moment-Matching

Recall

$$\mathcal{L}(\phi, \Phi) = \cap_i \left\{ (\tau, \tilde{\tau}) \,|\tau \in \mathcal{M}(\phi), \quad \Pi^i \,(\tau, \tilde{\tau}) \in \mathcal{M}(\phi, \Phi^i) \right\}$$

Another construction

- 1-Expand (and decouple)
$$\{\tau \in \mathcal{M}(\phi)\} \otimes_i \left\{ \left(\eta^i, \tilde{\tau}^i\right) |\Pi^i \left(\eta^i, \tilde{\tau}^i\right) \in \mathcal{M}(\phi, \Phi^i) \right\}$$

4.3.2 Optimality in terms of Moment-Matching

Recall

$$\mathcal{L}(\phi, \Phi) \quad = \quad \cap_i \left\{ (\tau, \tilde{\tau}) \,|\tau \in \mathcal{M}(\phi), \quad \Pi^i \,(\tau, \tilde{\tau}) \in \mathcal{M}(\phi, \Phi^i) \right\}$$

Another construction

- 1-Expand (and decouple)
  $$\{\tau \in \mathcal{M}(\phi)\} \otimes_i \left\{ \left(\eta^i, \tilde{\tau}^i\right) |\Pi^i \left(\eta^i, \tilde{\tau}^i\right) \in \mathcal{M}(\phi, \Phi^i) \right\}$$

Expansion from $(\tau, \tilde{\tau}) \to \left\{ \tau, (\eta^i, \tilde{\tau}^i), i = 1..d_I \right\}$

4.3.2 Optimality in terms of Moment-Matching

Recall

$$\mathcal{L}(\phi, \Phi) = \cap_i \left\{ (\tau, \tilde{\tau}) \, | \tau \in \mathcal{M}(\phi), \quad \Pi^i (\tau, \tilde{\tau}) \in \mathcal{M}(\phi, \Phi^i) \right\}$$

Another construction

- 1-Expand (and decouple)

$$\{\tau \in \mathcal{M}(\phi)\} \otimes_i \left\{ \left(\eta^i, \tilde{\tau}^i\right) | \Pi^i \left(\eta^i, \tilde{\tau}^i\right) \in \mathcal{M}(\phi, \Phi^i) \right\}$$

Expansion from $(\tau, \tilde{\tau}) \to \left\{ \tau, (\eta^i, \tilde{\tau}^i), i = 1..d_l \right\}$

- 2-Couple back

$$\{\tau \in \mathcal{M}(\phi)\} \otimes_i \left\{ \left(\eta^i, \tilde{\tau}^i\right) | \Pi^i \left(\eta^i, \tilde{\tau}^i\right) \in \mathcal{M}(\phi, \Phi^i) \right\} \text{ and}$$
$$\forall i, j \quad (\tau_i, \tilde{\tau}_i) = (\tau_j, \tilde{\tau}_j)$$

No secret here, just more variables, coupled together.

4.3.2 Optimality in terms of Moment-Matching

Constrained optimization problem

$$\max_{\{\tau,(\eta^i,\tilde{\tau}^i)\}} \left\{ \langle \tau, \theta \rangle + \textstyle\sum_i \langle \tilde{\tau}^i, \tilde{\theta}^i \rangle + \underbrace{H(\tau) + \sum_i \left[H(\eta^i, \tilde{\tau}^i) - H(\eta^i) \right]}_{F(\tau,(\eta^i,\tilde{\tau}^i))} \right\}$$

$$\text{subject to } \left(\eta^i, \tilde{\tau}^i\right) \in \mathcal{M}(\phi, \Phi^i)$$
$$\text{and } \tau = \eta^i$$

4.3.2 Optimality in terms of Moment-Matching

Constrained optimization problem

$$\max_{\{\tau,(\eta^i,\tilde{\tau}^i)\}} \left\{ \langle \tau, \theta \rangle + \sum_i \langle \tilde{\tau}^i, \tilde{\theta}^i \rangle + F(\tau, (\eta^i, \tilde{\tau}^i)) \right\}$$

$$\text{subject to } \left(\eta^i, \tilde{\tau}^i\right) \in \mathcal{M}(\phi, \Phi^i)$$
$$\text{and } \tau \in \mathcal{M}(\phi)$$
$$\text{and } \tau = \eta^i$$

4.3.2 Optimality in terms of Moment-Matching

Constrained optimization problem

$$\max_{\{\tau,(\eta^i,\tilde{\tau}^i)\}} \left\{ \langle \tau, \theta \rangle + \sum_i \langle \tilde{\tau}^i, \tilde{\theta}^i \rangle + F(\tau, (\eta^i, \tilde{\tau}^i)) \right\}$$

$$\text{subject to } \left(\eta^i, \tilde{\tau}^i\right) \in \mathcal{M}(\phi, \Phi^i)$$
$$\text{and } \tau \in \mathcal{M}(\phi)$$
$$\text{and } \tau = \eta^i$$

Associated Partial Lagrangian

$$L(\tau; \lambda) = \langle \tau, \theta \rangle + \sum_i \langle \tilde{\tau}^i, \tilde{\theta}^i \rangle + F(\tau, (\eta^i, \tilde{\tau}^i)) + \sum_i \langle \lambda^i, \tau - \eta^i \rangle$$

$$\text{subject to } \left(\eta^i, \tilde{\tau}^i\right) \in \mathcal{M}(\phi, \Phi^i)$$
$$\text{and } \tau \in \mathcal{M}(\phi)$$

Solving the optimization problem

$$L(\tau;\lambda) = \langle \tau, \theta \rangle + \sum_i \langle \tilde{\tau}^i, \tilde{\theta}^i \rangle + F(\tau, (\eta^i, \tilde{\tau}^i)) + \sum_i \langle \lambda^i, \tau - \eta^i \rangle$$

$$\text{subject to } \left(\eta^i, \tilde{\tau}^i\right) \in \mathcal{M}(\phi, \Phi^i)$$
$$\text{and } \tau \in \mathcal{M}(\phi)$$

Solving the optimization problem

$$L(\tau;\lambda) = \langle \tau, \theta \rangle + \sum_i \langle \tilde{\tau}^i, \tilde{\theta}^i \rangle + F(\tau, (\eta^i, \tilde{\tau}^i)) + \sum_i \langle \lambda^i, \tau - \eta^i \rangle$$

$$\text{subject to } (\eta^i, \tilde{\tau}^i) \in \mathcal{M}(\phi, \Phi^i)$$
$$\text{and } \tau \in \mathcal{M}(\phi)$$

For an optimal solution $\{\tau, (\eta^i, \tilde{\tau}^i), i = 1..d_I\}$

$$\begin{aligned} \nabla_\tau L(\tau, \lambda) &= 0 \\ \nabla_{(\eta^i, \tilde{\tau}^i)} L(\tau, \lambda) &= 0, \quad \text{for } i = 1...d_I \\ \nabla_\lambda L(\tau, \lambda) &= 0 \quad \text{(constraint)} \end{aligned}$$

Solving the optimization problem

$$L(\tau;\lambda) = \langle \tau, \theta \rangle + \sum_i \langle \tilde{\tau}^i, \tilde{\theta}^i \rangle + F(\tau, (\eta^i, \tilde{\tau}^i)) + \sum_i \langle \lambda^i, \tau - \eta^i \rangle$$

subject to $\left(\eta^i, \tilde{\tau}^i\right) \in \mathcal{M}(\phi, \Phi^i)$ and $\tau \in \mathcal{M}(\phi)$

Solving the optimization problem

$$L(\tau;\lambda) = \langle \tau, \theta \rangle + \textstyle\sum_i \langle \tilde{\tau}^i, \tilde{\theta}^i \rangle + F(\tau, (\eta^i, \tilde{\tau}^i)) + \textstyle\sum_i \langle \lambda^i, \tau - \eta^i \rangle$$

$$\text{subject to } \left(\eta^i, \tilde{\tau}^i\right) \in \mathcal{M}(\phi, \Phi^i) \text{ and } \tau \in \mathcal{M}(\phi)$$

For an optimal solution $\left\{\tau, (\eta^i, \tilde{\tau}^i), i = 1..d_I\right\}$

Solving the optimization problem

$$L(\tau;\lambda) = \langle \tau, \theta \rangle + \sum_i \langle \tilde{\tau}^i, \tilde{\theta}^i \rangle + F(\tau, (\eta^i, \tilde{\tau}^i)) + \sum_i \langle \lambda^i, \tau - \eta^i \rangle$$

subject to $(\eta^i, \tilde{\tau}^i) \in \mathcal{M}(\phi, \Phi^i)$ and $\tau \in \mathcal{M}(\phi)$

For an optimal solution $\{\tau, (\eta^i, \tilde{\tau}^i), i = 1..d_I\}$

$$\nabla_\tau L(\tau, \lambda) = 0$$

$$\Rightarrow q(x;\theta,\lambda) \propto f_0(x) \exp\{\langle \theta + \sum_i \lambda_i, \phi(x) \rangle\} \in \mathcal{M}(\phi)$$

Solving the optimization problem

$$L(\tau;\lambda) = \langle \tau, \theta \rangle + \sum_i \langle \tilde{\tau}^i, \tilde{\theta}^i \rangle + F(\tau, (\eta^i, \tilde{\tau}^i)) + \sum_i \langle \lambda^i, \tau - \eta^i \rangle$$

$$\text{subject to } \left(\eta^i, \tilde{\tau}^i\right) \in \mathcal{M}(\phi, \Phi^i) \text{ and } \tau \in \mathcal{M}(\phi)$$

For an optimal solution $\left\{\tau, (\eta^i, \tilde{\tau}^i), i = 1..d_I\right\}$

$$\nabla_\tau L(\tau, \lambda) = 0$$
$$\Rightarrow q(x;\theta,\lambda) \propto f_0(x) \exp\{\langle \theta + \sum_i \lambda_i, \phi(x) \rangle\} \in \mathcal{M}(\phi)$$

$$\nabla_{(\eta^i, \tilde{\tau}^i)} L(\tau, \lambda) = 0$$
$$\Rightarrow q^i(x;\theta,\tilde{\theta}^i,\lambda) \propto f_0(x) \exp\left\{\langle \theta + \sum_{l \neq i} \lambda_i, \phi(x) \rangle + \langle \tilde{\theta}^i, \Phi^i(x) \rangle\right\} \in \mathcal{M}(\phi, \Phi^i)$$

Solving the optimization problem

$$L(\tau;\lambda) = \langle \tau, \theta \rangle + \sum_i \langle \tilde{\tau}^i, \tilde{\theta}^i \rangle + F(\tau, (\eta^i, \tilde{\tau}^i)) + \sum_i \langle \lambda^i, \tau - \eta^i \rangle$$

$$\text{subject to } (\eta^i, \tilde{\tau}^i) \in \mathcal{M}(\phi, \Phi^i) \text{ and } \tau \in \mathcal{M}(\phi)$$

For an optimal solution $\{\tau, (\eta^i, \tilde{\tau}^i), i = 1..d_I\}$

$$\nabla_\tau L(\tau, \lambda) = 0$$
$$\Rightarrow q(x; \theta, \lambda) \propto f_0(x) \exp\{\langle \theta + \sum_i \lambda_i, \phi(x) \rangle\} \in \mathcal{M}(\phi)$$

$$\nabla_{(\eta^i, \tilde{\tau}^i)} L(\tau, \lambda) = 0$$
$$\Rightarrow q^i(x; \theta, \tilde{\theta}^i, \lambda) \propto f_0(x) \exp\left\{\langle \theta + \sum_{l \neq i} \lambda_i, \phi(x) \rangle + \langle \tilde{\theta}^i, \Phi^i(x) \rangle\right\} \in \mathcal{M}(\phi, \Phi^i)$$

$$\nabla_\lambda L(\tau, \lambda) = 0 \Rightarrow \tau = \mathbb{E}_q[\phi(x)] \equiv \mathbb{E}_{q^i}[\phi(x)] = \eta^i$$

Understanding the updates

Analogy/connection with EP algorithm described earlier

- λ_i parameterizes the approximated factors ($\tilde{f}_i$)
- factor $\tilde{f}_i$ share the same suff-stats as the base model (i.e. tractable ϕ)

Understanding the updates

- $q(x; \theta, \lambda) \propto f_0(x) \exp \{\langle \theta + \sum_i \lambda_i, \phi(x) \rangle\}$
  This is the full posterior where all the 'intractable' factors were replaced by their approximation
- $q^i(x; \theta, \tilde{\theta}^i, \lambda) \propto f_0(x) \exp \left\{ \langle \theta + \sum_{l \neq i} \lambda_i, \phi(x) \rangle + \langle \tilde{\theta}^i, \Phi^i(x) \rangle \right\}$
  This is the full posterior where only one intractable factor is left unreplaced (others are all approximated)

EP Summary

Expectation-propagation (EP) updates:

(1) At iteration $n = 0$, initialize the Lagrange multiplier vectors $(\lambda^1, \ldots, \lambda^{d_I})$.

(2) At each iteration, $n = 1, 2, \ldots,$ choose some index $i(n) \in \{1, \ldots, d_I\}$, and

(a) Using Equation (4.78), form the augmented distribution $q^{i(n)}$ and compute the mean parameter

$$\eta^{i(n)} := \int q^{i(n)}(x)\phi(x)\nu(dx) = \mathbb{E}_{q^{i(n)}}[\phi(X)]. \tag{4.80}$$

(b) Using Equation (4.77), form the base distribution q and adjust $\lambda^{i(n)}$ to satisfy the moment-matching condition

$$\mathbb{E}_q[\phi(X)] = \eta^{i(n)}. \tag{4.81}$$

EP : Examples

Example 1:

- simple graph: (1)-(2)
- $$p(x_1, x_2) \propto \exp\left(\theta_1(x_1) + \theta_2(x_2) + \underbrace{\theta(x_1, x_2)}_{intractable}\right)$$

EP : Examples

Example 1:

- simple graph: (1)-(2)
- $p(x_1, x_2) \propto \exp\left(\theta_1(x_1) + \theta_2(x_2) + \underbrace{\theta(x_1, x_2)}_{intractable}\right)$

EP updates

- $q(x_1, x_2; \theta, \lambda) \propto \exp\left(\theta_1(x_1) + \lambda_{12}(x_1)\right) \exp\left(\theta_2(x_2) + \lambda_{12}(x_2)\right)$

EP : Examples

Example 1:

- simple graph: (1)-(2)
- $p(x_1, x_2) \propto \exp\left(\theta_1(x_1) + \theta_2(x_2) + \underbrace{\theta(x_1, x_2)}_{intractable}\right)$

EP updates

- $q(x_1, x_2; \theta, \lambda) \propto \exp\left(\theta_1(x_1) + \lambda_{12}(x_1)\right) \exp\left(\theta_2(x_2) + \lambda_{12}(x_2)\right)$
- $q^{12}(x_1, x_2; \theta, \lambda) \propto \exp\left(\theta_1(x_1) + \theta_2(x_2) + \theta(x_1, x_2)\right)$

EP : Examples

Example 1:

- simple graph: (1)-(2)
- $p(x_1, x_2) \propto \exp\left(\theta_1(x_1) + \theta_2(x_2) + \underbrace{\theta(x_1, x_2)}_{\textit{intractable}}\right)$

EP updates

- $q(x_1, x_2; \theta, \lambda) \propto \exp\left(\theta_1(x_1) + \lambda_{12}(x_1)\right) \exp\left(\theta_2(x_2) + \lambda_{12}(x_2)\right)$
- $q^{12}(x_1, x_2; \theta, \lambda) \propto \exp\left(\theta_1(x_1) + \theta_2(x_2) + \theta(x_1, x_2)\right)$
- $\mathbb{E}_{q^{12}(x_1)}(\theta_1(x_1)) = \mathbb{E}_{q(x_1)}(\theta_1(x_1))$ (update the λs)

EP : Examples

Example 2: Mixture of Gaussians

- $\mathcal{M}(\phi, \Phi) = \left\{ \mathbb{E}[X], \mathbb{E}[XX^T], \mathbb{E}\left[\log p(y^i|X)\right],\ i = 1..n \right\}$
- Lagrange multipliers $(\lambda^i, \Lambda^i) \in R^m \times R^{m \times m}$

EP : Examples

Example 2: Mixture of Gaussians

- $\mathcal{M}(\phi, \Phi) = \left\{\mathbb{E}[X], \mathbb{E}[XX^T], \mathbb{E}\left[\log p(y^i|X)\right], \ i = 1..n\right\}$
- Lagrange multipliers $(\lambda^i, \Lambda^i) \in R^m \times R^{m\times m}$

EP updates

- $q(x, \Sigma; (\lambda^i, \Lambda^i)) \propto \exp\left\{\langle \sum_i \lambda^i, x\rangle + \langle -\frac{1}{2}\Sigma^{-1} + \sum_i \Lambda^i, xx^T\rangle\right\}$

EP : Examples

Example 2: Mixture of Gaussians

- $\mathcal{M}(\phi, \Phi) = \left\{ \mathbb{E}[X], \mathbb{E}[XX^T], \mathbb{E}\left[\log\, p(y^i|X)\right],\ i = 1..n \right\}$
- Lagrange multipliers $(\lambda^i, \Lambda^i) \in R^m \times R^{m\times m}$

EP updates

- $q(x, \Sigma; (\lambda^i, \Lambda^i)) \propto \exp\left\{ \langle \sum_i \lambda^i, x\rangle + \langle -\frac{1}{2}\Sigma^{-1} + \sum_i \Lambda^i, xx^T \rangle \right\}$
- $q^i(x, \Sigma; (\lambda^i, \Lambda^i)) \propto$
  $\exp\left\{ \langle \sum_{l\neq i} \lambda^l, x\rangle + \langle -\frac{1}{2}\Sigma^{-1} + \sum_{l\neq i} \Lambda^l, xx^T \rangle + \langle \tilde{\theta}^i, \log\, p(y^i|x) \rangle \right\}$

That's it for today