Mean Field Asymptotics in Statistical Learning.

Apr 12

Lecture 21. Derivation of AMP. II.

① From Belief propagation to message passing algorithm

Wrong intuition: the beliefs are approximately Gaussian.

Intuition: In the update rule, only the mean and var of input beliefs are important.

Input beliefs can be approximated by Gaussian dist. in the update rule.

Real belief is still non-Gaussian.

Def (Message passing algorithm) mean and variance of beliefs.

$$\{ m_{i\to a}^k, v_{i\to a}^k, \hat{m}_{a\to i}^k, \hat{v}_{a\to i}^k \}_{a\in F,\, i\in V,\, k\geq 0} \subseteq \mathbb{R}. \quad \text{Messages.}$$

Update rule: Calculating $\{ m_{i\to a}^{k+1}, v_{i\to a}^{k+1}, \hat{m}_{a\to i}^{k+1}, \hat{v}_{a\to i}^{k+1} \}$ using $\{ m_{i\to a}^k, v_{i\to a}^k, \hat{m}_{a\to i}^k, \hat{v}_{a\to i}^k \}$.

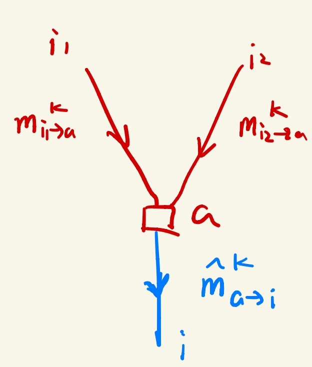
$$p_{i\to a}^k(x_i) = \frac{1}{\sqrt{2\pi v_{i\to a}^k}} \exp\left\{ -\frac{(x_i - m_{i\to a}^k)^2}{2 v_{i\to a}^k} \right\}. \quad \text{density of } \mathcal{N}(m_{i\to a}^k, v_{i\to a}^k)$$

$$\hat{p}_{a\to i}^k(x_i) = \frac{1}{\sqrt{2\pi \hat{v}_{a\to i}^k}} \exp\left\{ -\frac{(x_i - \hat{m}_{a\to i}^k)^2}{2 \hat{v}_{a\to i}^k} \right\}. \quad \text{density of } \mathcal{N}(\hat{m}_{a\to i}^k, \hat{v}_{a\to i}^k)$$

$$r_{i\to a}^{k+1}(x_i) \propto \psi_i(x_i) \prod_{b\in\partial i\setminus a} \hat{p}_{b\to i}^k(x_i)$$

Non-backtracking

$$\hat{r}_{a\to i}^k(x_i) \propto \sum_{x_{\partial a\setminus i}} \psi_a(x_{\partial a}) \prod_{j\in\partial a\setminus i} p_{j\to a}^k(x_j)$$



$$(m_{i\to a}^{k+1}, v_{i\to a}^{k+1}) = (\text{mean}, \text{variance}) \text{ under } r_{i\to a}^{k+1}(x_i)$$

$$(\hat{m}_{a\to i}^k, \hat{v}_{a\to i}^k) = (\text{mean}, \text{variance}) \text{ under } \hat{r}_{a\to i}^k(x_i)$$

Extract marginals:

$$r_i^k(x_i) \propto \psi_i(x_i) \prod_{b\in\partial i} \hat{p}_{b\to i}^{k-1}(x_i)$$

$\{ p_{i\to a}^k, \hat{p}_{a\to i}^k \}$ input beliefs, assumed to be Gaussian

$\{ r_{i\to a}^k, \hat{r}_{a\to i}^k \}$ output beliefs, possibly non-Gaussian

② Example: LASSO with finite temperature β

⊗ $$r^{k+1}_{i\to a}(x_i) \propto \exp\{-\beta\lambda|x_i|\} \times \exp\left\{-\sum_{b\neq a}\frac{(x_i-\hat{m}^k_{b\to i})^2}{2\hat{v}^k_{b\to i}}\right\}$$

$$\propto \exp\left\{-\beta\left(\frac{(x_i-\theta^k_{i\to a})^2}{2\varsigma^k_{i\to a}}+\lambda|x_i|\right)\right\}.$$

$$\boxed{\begin{aligned}
\beta(\varsigma^k_{i\to a})^{-1} &= \sum_{b\neq a}(\hat{v}^k_{b\to i})^{-1}\\
\beta(\varsigma^k_{i\to a})^{-1}\theta^k_{i\to a} &= \sum_{b\neq a}(\hat{v}^k_{b\to i})^{-1}\,\hat{m}^k_{b\to i}\\
m^{k+1}_{i\to a} &= \mathbb{E}_{X_i\sim\pi(\beta,\lambda,\theta^k_{i\to a},\varsigma^k_{i\to a})}[X_i]\\
v^{k+1}_{i\to a} &= \mathrm{Var}_{X_i\sim\pi(\beta,\lambda,\theta^k_{i\to a},\varsigma^k_{i\to a})}[X_i]
\end{aligned}}$$

LASSO MP Update rule

where $\pi(\beta,\lambda,\theta,\varsigma) \propto \exp\left\{-\beta\left(\frac{(x-\theta)^2}{2\varsigma^2}+\lambda|x|\right)\right\}.$

⊗ $$\hat{r}^k_{a\to i}(x) \propto \int_{\mathbb{R}^{d-1}} \exp\left\{-\frac{\beta}{2}(y_a-\langle A_a,x\rangle)^2\right\}$$

$$\times \exp\left\{-\sum_{j\neq i}\frac{(x_j-m^k_{j\to a})^2}{2v^k_{j\to a}}\right\}dx_{\setminus i}.$$

This is a Gaussian integration, which gives a Gaussian density.

$$\propto \exp\left\{-\frac{(x_i-\hat{m}^{k+1}_{a\to i})^2}{2\hat{v}^{k+1}_{a\to i}}\right\}.$$

$$\boxed{\begin{aligned}
A_{ai}\cdot\hat{m}^k_{a\to i} &= y_a-\sum_{j\neq i}A_{aj}m^k_{j\to a}\\
A_{ai}^2\cdot\hat{v}^k_{a\to i} &= \sum_{j\neq i}A_{aj}^2 v^k_{j\to a}+\frac{1}{\beta}
\end{aligned}}$$

⊗ Simplification as $\beta\to\infty$. (This step is unnecessary if $\beta=$ finite)

$$\lim_{\beta\to\infty}\mathbb{E}_{X\sim\pi(\beta,\lambda,\theta,\varsigma)}[X] = \arg\min_x\left\{\frac{(x-\theta)^2}{2\varsigma}+\lambda|x|\right\}$$

$$= \eta(\theta;\lambda\varsigma)$$ Soft thresholding.

$$\lim_{\beta\to\infty}\beta\cdot\mathrm{Var}_{X\sim\pi(\beta,\lambda,\theta,\varsigma)}[X] = \begin{cases}\varsigma, & \text{if } |\theta|>\lambda\varsigma,\\ 0, & \text{if } |\theta|<\lambda\varsigma.\end{cases}$$

$$= \varsigma\cdot\eta'(\theta;\lambda\varsigma).$$

We are short of alphabets. Abuse notation:

$$m^k_{i\to a} \longleftarrow \lim_{\beta\to\infty} m^k_{i\to a}. \qquad \hat{m}^k_{a\to i} \longleftarrow \lim_{\beta\to\infty} \hat{m}^k_{a\to i}.$$

$$\nu^k_{i\to a} \longleftarrow \lim_{\beta\to\infty} \beta\cdot \nu^k_{i\to a}, \qquad \hat{\nu}^k_{a\to i} \longleftarrow \lim_{\beta\to\infty} \beta\cdot \nu^k_{a\to i}.$$

$$\boxed{\begin{aligned}
(S^k_{i\to a})^{-1} &= \sum_{b\neq a} (\hat{\nu}^k_{b\to i})^{-1} \\
(S^k_{i\to a})^{-1}\theta^k_{i\to a} &= \sum_{b\neq a} (\hat{\nu}^k_{b\to i})^{-1}\, \hat{m}^k_{b\to i} \\
m^{k+1}_{i\to a} &= \eta(\theta^k_{i\to a};\lambda S^k_{i\to a}) \\
\nu^{k+1}_{i\to a} &= S^k_{i\to a}\cdot \eta'(\theta^k_{i\to a},\lambda S^k_{i\to a}). \\
A_{ai}\cdot \hat{m}^k_{a\to i} &= y_a - \sum_{j\neq i} A_{aj}\, m^k_{j\to a} \\
A_{ai}^2\cdot \hat{\nu}^k_{a\to i} &= \sum_{j\neq i} A_{aj}^2\, \nu^k_{j\to a} + 1
\end{aligned}}$$

Change of variable: $z^k_{a\to i} \equiv A_{ai}\cdot \hat{m}^k_{a\to i} = y_a - \sum_{j\neq i} A_{aj}\, m^k_{j\to a}$

$$\tau^k_{a\to i} \equiv A_{ai}^2\cdot \hat{\nu}^k_{a\to i} = \sum_{j\neq i} A_{aj}^2\, \nu^k_{j\to a} + 1$$

Assume $A_{ai} \sim$ iid $\mathrm{Unif}(\{\pm \frac{1}{\sqrt{n}}\})$. Can replace A_{ai}^2 by $\frac{1}{n}$.

$$\Rightarrow \boxed{\begin{aligned}
\theta^k_{i\to a} &= \frac{\sum_{b\neq a} A_{bi}\, z^k_{b\to i} / \tau^k_{b\to i}}{\frac{1}{n}\sum_{b\neq a} 1/\tau^k_{b\to i}} \\
m^{k+1}_{i\to a} &= \eta(\theta^k_{i\to a};\lambda S^k_{i\to a}) \\
z^k_{a\to i} &= y_a - \sum_{j\neq i} A_{aj}\, m^k_{j\to a} \\
S^k_{i\to a} &= \left(\frac{1}{n}\sum_{b\neq a} 1/\tau^k_{b\to i}\right)^{-1} \\
\nu^{k+1}_{i\to a} &= S^k_{i\to a}\cdot \eta'(\theta^k_{i\to a},\lambda S^k_{i\to a}). \\
\tau^k_{a\to i} &= \frac{1}{n}\sum_{j\neq i} \nu^k_{j\to a} + 1
\end{aligned}}$$

$i \in V$

$a \in F$

there are

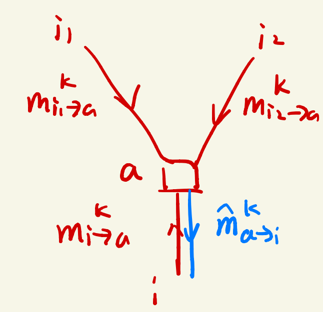
$2\times n\times d$ messages

③ From message passing to approximate message passing.

> Goal: Simplify MP ($2 \times n \times d$ # of messages)
>
> after simplification, there are only d messages.

A crude approximation: throw away non-backtracking property.

$$\theta_i^k = \frac{\sum_b A_{bi} z_b^k / \tau_b^k}{\frac{1}{n}\sum_b 1/\tau_b^k}$$



$$m_i^{k+1} = \eta(\theta_i^k ; \lambda \varsigma_i^k)$$

$$z_a^k = y_a - \sum_j A_{aj} m_j^k$$

$$\varsigma_i^k = \left(\frac{1}{n}\sum_b 1/\tau_b^k\right)^{-1} \Rightarrow \varsigma^k = \tau^k$$

$$v_i^{k+1} = \varsigma_i^k \cdot \eta'(\theta_i^k , \lambda \varsigma_i^k).$$

$$\tau_a^k = \frac{1}{n}\sum_j v_j^k + 1 \Rightarrow \tau^k = \frac{1}{n}\sum_j v_j^k + 1$$

Simplifying this gives

$$m^{k+1} = \eta(m^k + A^T z^k ; \lambda \varsigma^k) \quad \in \mathbb{R}^d.$$

$$z^k = y - A m^k \quad (\underline{\text{without Onsager term}}) \in \mathbb{R}^n$$

$$\varsigma^{k+1} = \varsigma^k \times \frac{1}{n}\sum_{i\in[n]} \eta'(m^k + Az^k ; \lambda\varsigma^k)_i + 1 \in \mathbb{R}$$

Doesn't give a correct approx. Needs a correction.

The equations for ς, v and τ are fine.

The equations for θ, m, z are off by $O(1)$.

Derivation of AMP.

⊗ Replace $\nu_{i\to a}^k$ by ν^k, $\varsigma_{i\to a}^k$ by ς^k.

$$\boxed{\begin{aligned}
\theta_{i\to a}^k &= \sum_{b\neq a} A_{bi}\, z_{b\to i}^k && \text{ⓐ}\\
m_{i\to a}^{k+1} &= \eta(\theta_{i\to a}^k ; \lambda\varsigma^k) && \text{ⓑ}\\
z_{a\to i}^k &= y_a - \sum_{j\neq i} A_{aj}\, m_{j\to a}^k && \text{ⓒ}\\
\varsigma^{k+1} &= \varsigma^k \times \frac{1}{n}\sum_{i\in[n]} \eta'(m^k + Az^k ; \lambda\varsigma^k)_i + 1
\end{aligned}}$$

⊗ Let

$$\theta_{i\to a}^k = \theta_i^k + \delta\theta_{i\to a}^k$$

$$m_{i\to a}^k = m_i^k + \delta m_{i\to a}^k$$

$$z_{a\to i}^k = z_a^k + \delta z_{a\to i}^k$$

ⓐ :

$$\theta_{i\to a}^k = \underbrace{\sum_b A_{bi}\, z_{b\to i}^k}_{\theta_i^k} \underbrace{- A_{ai}\, z_{a\to i}^k}_{\delta\theta_{i\to a}^k}$$

$$\begin{cases}
\theta_i^k = \sum_b A_{bi}\, z_{b\to i}^k = \sum_b A_{bi}\, z_b^k + \sum_b A_{bi}\,\delta z_{b\to i}^k\\
\delta\theta_{i\to a}^k = - A_{ai}\, z_{a\to i}^k = - A_{ai}\,(z_a^k + \delta z_{a\to i}^k) \approx - A_{ai}\, z_a^k
\end{cases}$$

ⓑ :

$$m_{i\to a}^{k+1} = \eta(\theta_{i\to a}^k ; \lambda\varsigma^k) \approx \underbrace{\eta(\theta_i^k ; \lambda\varsigma^k)}_{m_i^{k+1}} + \underbrace{\partial\eta(\theta_i^k ; \lambda\varsigma^k)\,\delta\theta_{i\to a}^k}_{\delta m_{i\to a}^{k+1}}$$

$$\begin{cases}
m_i^{k+1} = \eta(\theta_i^k ; \lambda\varsigma^k)\\
\delta m_{i\to a}^{k+1} = \partial\eta(\theta_i^k ; \lambda\varsigma^k)\,\delta\theta_{i\to a}^k .
\end{cases}$$

ⓒ :

$$z_{a\to i}^k = \underbrace{y_a - \sum_b A_{bj}\, m_{j\to a}^k}_{z_a^k} \underbrace{+ A_{ai}\, m_{i\to a}^k}_{\delta z_{i\to a}^k}$$

$$\begin{cases}
z_a^k = y_a - \sum_j A_{bj}\, m_{j\to a}^k = y_a - \sum_j A_{bj}\, m_j^k - \sum_j A_{bj}\,\delta m_{j\to a}^k\\
\delta z_{i\to a}^k = A_{ai}\, m_{i\to a}^k = A_{ai}\,(m_i^k + \delta m_{i\to a}^k) \approx A_{ai}\, m_i^k .
\end{cases}$$

$$\Rightarrow \quad \boxed{\begin{aligned} m_i^{k+1} &= \eta(\theta_i^k ; \lambda \zeta^k) \\ \theta_i^k &= \sum_b A_{bi} z_b^k + \sum_b A_{bi}^2 m_i^k \\ &\approx m_i^k + \sum_b A_{bi} z_b^k \\ z_a^k &= y_a - \sum_j A_{bj} m_j^k + \sum_j A_{bj}^2 \partial \eta(\theta_j^k ; \lambda \zeta^k) z_{a-1}^k \\ &\approx y_a - \sum_j A_{bj} m_j^k + \frac{1}{n} \sum_j \partial \eta(\theta_j^k ; \lambda \zeta^k) z_{a-1}^k \end{aligned}}$$

$\mathbb{E} A_{bi}^2 = \frac{1}{n}.$

$$\Rightarrow \quad \boxed{\begin{aligned} m^{k+1} &= \eta(m^k + A^T z ; \lambda \zeta^k) \\ z^k &= y - A m^k + \underbrace{\left[\frac{1}{n} \sum_j \partial \eta(\theta_j^k ; \lambda \zeta^k)\right] z^k}_{\text{Onsager}} \\ \zeta^{k+1} &= \zeta^k \times \frac{1}{n} \sum_{i \in [n]} \eta'(\theta_j^k ; \lambda \zeta^k) + 1 \end{aligned}}$$

This is a version of the AMP for LASSO