# 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.

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

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

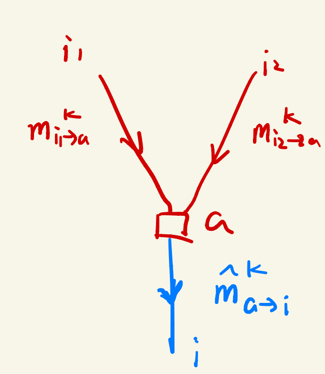
$$\rho^k_{i\to a}(x_i) = \frac{1}{\sqrt{2\pi v^k_{i\to a}}} \exp\left\{ -\frac{(x_i - m^k_{i\to a})^2}{2 v^k_{i\to a}} \right\}. \quad \text{density of } \mathcal{N}(m^k_{i\to a}, v^k_{i\to a})$$

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

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

Non-backtracking

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




$$(m^{k+1}_{i\to a}, v^{k+1}_{i\to a}) = (\text{mean}, \text{variance}) \text{ under } r^{k+1}_{i\to a}(x_i)$$

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

Extract marginals:

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

$\{ \rho^k_{i\to a}, \hat{\rho}^k_{a\to i} \}$ input beliefs, assumed to be Gaussian

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

Expectation propagation

② Example: LASSO with finite temperature $\beta$

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

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

LASSO MP

Update rule

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

⊗ $$\hat{r}_{a\to i}^k(x) \propto \int_{\mathbb{R}^{d-1}} \exp\left\{-\frac{\beta}{2}(y_a-\langle A_a,x\rangle)^2\right\}$$
$$\times \exp\left\{-\sum_{j\neq i}\frac{(x_j-m_{j\to a}^k)^2}{2v_{j\to a}^k}\right\}dx_{\setminus i}.$$

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

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

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

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

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

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

We are short of alphabets. Abuse notation:

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

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

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

$A_{ai} \sim \text{iid } N(0, \frac{1}{n})$

$\mathbb{E}[A_{ai}^2] = \frac{1}{n}$

Change of variable:

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

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

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

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

$i \in V \quad |V| = d$

$a \in F,\ |F| = n$

③ From message passing to approximate message passing.

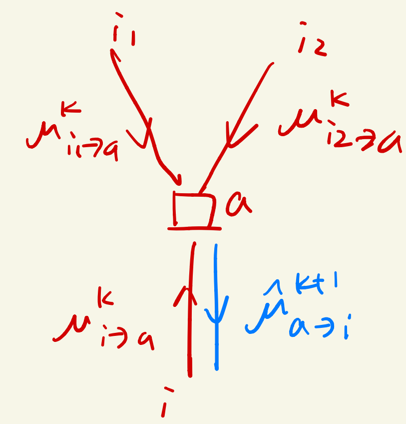
Goal: Simplify MP ( $2 \times n \times d$ # of messages)

after simplification, there are only $d$ messages.

A crude approximation: throw away non-backtracking property.

$\theta_i^k = \sum_b A_{bi} z_b^k$ $\Leftarrow$

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




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

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

$$s_i^k = \left(\frac{1}{n}\sum_b 1/\tau_b^k\right)^{-1} \Rightarrow \quad \zeta_i^k = \zeta^k = \tau^k.$$

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

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

Simplifying this gives

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

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

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

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

The equations for $\zeta$, $\nu$ and $\tau$ are fine.

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

# Derivation of AMP from MP.

⊗ Replace $\tau^k_{i\to a}$ by $\tau^k$, $\zeta^k_{i\to a}$ by $\zeta^k$.

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

⊗ Let

$$\theta^k_{i\to a} = \theta^k_i + \delta\theta^k_{i\to a} \qquad \delta\theta$$

$$m^k_{i\to a} = m^k_i + \delta m^k_{i\to a}$$

$$z^k_{a\to i} = z^k_a + \delta z^k_{a\to i}$$

(a) :

$$\theta^k_{i\to a} = \underbrace{\sum_b A_{bi}\, z^k_{b\to i}}_{\theta^k_i} - \underbrace{A_{ai}\, z^k_{a\to i}}_{\delta\theta^k_{i\to a}}$$

// $O(1)$

$$\begin{cases}
\theta^k_i = \sum_b A_{bi}\, z^k_{b\to i} = \sum_b A_{bi}\, z^k_b + \sum_b A_{bi}\, \delta z^k_{b\to i}\\
\delta\theta^k_{i\to a} = -A_{ai}\, z^k_{a\to i} = O\left(\frac{1}{\sqrt{n}}\right)
\end{cases}$$

(b) :

$$m^{k+1}_{i\to a} = \underbrace{\eta(\theta^k_i;\, \lambda\zeta^k)}_{m^{k+1}_i} + \underbrace{\eta'(\theta^k_i;\, \lambda\zeta^k)\cdot\delta\theta^k_{i\to a}}_{\delta m^{k+1}_i}$$

$$\begin{cases}
m^{k+1}_i = \eta(\theta^k_i;\, \lambda\zeta^k)\\
\delta m^{k+1}_{i\to a} = \eta'(\theta^k_i,\, \lambda\zeta^k)\cdot\delta\theta^k_{i\to a}
\end{cases}$$

(c) :

$$z^k_{a\to i} = \underbrace{y_a - \sum_b A_{bj}\, m^k_{j\to a}}_{z^k_a} + \underbrace{A_{ai}\, m^k_{i\to a}}_{\delta z^k_{a\to i}}$$

$$\begin{cases}
z^k_a = y_a - \sum_j A_{aj}\, m^k_{j\to a} = y_a - \sum_j A_{aj}\, m^k_j - \sum_j A_{aj}\,\delta m^k_{j\to a}\\
\delta z^k_{a\to i} = A_{ai}\, m^k_{i\to a} = A_{ai}(m^k_i + \delta m^k_{i\to a}) \approx A_{ai}\, m^k_i
\end{cases}$$

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

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

$$\theta_i^k = \sum_b A_{bi} z_b^k + \Big(\sum_b A_{bi}^2\Big) m_i^k$$

$$\approx m_i^k + \sum_b A_{bi} z_b^k .$$

$$z_a^k = y_a - \sum_j A_{aj} m_j^k + \sum_j A_{aj}^2 \, \partial\eta(\theta_j^k ; \lambda \zeta^k) z_a^{k-1}$$

$$\approx y_a - \sum_j A_{aj} m_j^k + \frac{1}{n} \sum_j \partial\eta(\theta_j^k ; \lambda \zeta^k) z_a^{k-1}$$

$$m^{k+1} = \eta(m^k + A^T z^k ; \lambda \zeta^k)$$

$$z^k = y - A m^k + \Big[\frac{1}{n} \sum_j \partial\eta(\theta_j^k ; \lambda \zeta^k)\Big] z^{k-1}$$

$$\zeta^{k+1} = \zeta^k \times \frac{1}{n} \sum_{i\in[n]} \eta'(m^k + A z^k ; \lambda \zeta^k)_i + 1$$