Variational Bayes EM & Methods Review

EM: reminder

$$\log p(X|\theta) \geq \mathcal{L}(\theta, q) = \mathbb{E}_{q(T)} \log \frac{p(X,T|\theta)}{q(T)} dT \to \max$$

Marginal likelihood

Variational lower bound

EM: reminder

$$\log p(X|\theta) \geq \mathcal{L}(\theta, q) = \mathbb{E}_{q(T)} \log \frac{p(X,T|\theta)}{q(T)} dT \to \max$$

Marginal likelihood

Variational lower bound

E-step

$$\mathcal{L}(\theta, q) \to \max_q$$

M-step

$$\mathcal{L}(\theta, q) \to \max_\theta$$

EM: reminder

$$\log p(X|\theta) \geq \mathcal{L}(\theta, q) = \mathbb{E}_{q(T)} \log \frac{p(X,T|\theta)}{q(T)} dT \to \max$$

Marginal likelihood

Variational lower bound

E-step

$$\mathcal{L}(\theta, q) \to \max_q \Leftrightarrow \mathcal{KL}[q(T) \,\|\, p(T|X,\theta)] \to \min_q$$

M-step

$$\mathcal{L}(\theta, q) \to \max_\theta \Leftrightarrow \mathbb{E}_{q(T)} \log p(X,T|\theta) \to \max_\theta$$

E-step

$$\mathcal{KL}[q(T) \,\|\, p(T|X,\theta)] \to \min_{q}$$

Full posterior

$$q(T) = p(T|X,\theta)$$

Variational inference

$$\mathcal{KL}[q(T) \,\|\, p(T|X,\theta)] \to \min_{q \in Q}$$

Model

Known: X data

Unknown: θ parameters

Unknown: T latent variables

Methods

Accurate

Inaccurate

- Full inference $p(T, \theta | X)$

Slow

Fast

Methods

Accurate

Inaccurate

Slow

Fast

- Full inference $p(T, \theta|X)$
- Mean field $q(T)q(\theta) \approx p(T, \theta|X)$

Methods

Accurate

Inaccurate

Slow

Fast

- Full inference $p(T, \theta|X)$
- Mean field $q(T)q(\theta) \approx p(T, \theta|X)$
- EM algorithm $q(T),\ \theta = \theta_{\mathrm{MP}}$

Methods

Accurate

Inaccurate

Slow

Fast

- Full inference $p(T, \theta|X)$
- Mean field $q(T)q(\theta) \approx p(T, \theta|X)$
- EM algorithm $q(T),\ \theta = \theta_{\mathrm{MP}}$
- Variational EM $q_1(T_1) \dots q_d(T_d),\ \theta = \theta_{\mathrm{MP}}$

Methods

Accurate

Inaccurate

Slow

Fast

- Full inference $p(T, \theta|X)$
- Mean field $q(T)q(\theta) \approx p(T, \theta|X)$
- EM algorithm $q(T),\ \theta = \theta_{\mathrm{MP}}$
- Variational EM $q_1(T_1) \ldots q_d(T_d),\ \theta = \theta_{\mathrm{MP}}$
- Crisp EM $T = T_{\mathrm{MP}},\ \theta = \theta_{\mathrm{MP}}$