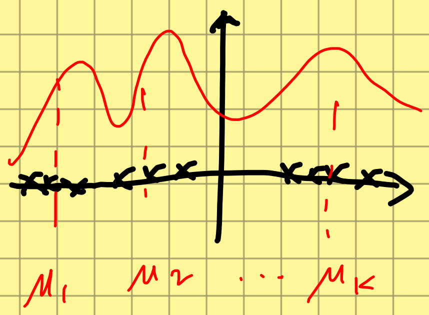
Lecture 16: Variational Inference Example.



Generative model:

Prior

1) $\mu_1 \cdots \mu_K \overset{iid}{\sim} \text{Normal}(\mu_0, \sigma_0^2)$

2) $Z_1 \cdots Z_n \overset{iid}{\sim} \text{Categorical}(\underset{\sim}{\pi})$

$\underset{\sim}{\pi} = (\pi_1, \cdots \pi_K)$

$X \sim \text{Multinomial}(\underset{\sim}{\pi})$

$K=5$

$(0\ 0\ 1\ 0\ 0)$

$X' \sim \text{Categorical}(\underset{\sim}{\pi})$

3

3) $X_i \sim \text{Normal}(\mu_{Z_i}, 1)$ independent.

assume known variance $=1$

Calculate the posterior of unknowns given data.

(*) $pr(Z_1, Z_2 \cdots Z_n, \mu_1, \cdots \mu_K \mid x_1, \cdots x_n, \mu_0, \sigma_0^2)$

approximate

Variational inference (mean-field)

Choose among a family of distn

$$q(\underset{\sim}{z}, \underset{\sim}{\mu}) = \prod_{k=1}^{K} q(\mu_k \mid \tilde{\mu}_k, \tilde{\sigma}_k^2) \times \prod_{i=1}^{n} q(Z_i \mid \phi_i)$$

1) Z_i : $i = 1, \cdots n$, (cluster membership indicator)

2) μ_k , $k = 1 \cdots K$: (component Gaussian means).

Update Z_i

$$q^*(Z_i) \propto \exp\left\{ E_{-i}\left[\underline{\log P(\underset{\sim}{\mu}, \underset{\sim}{Z}, \underset{\sim}{x})} \right] \right\}$$

$$\log P(\underset{\sim}{\mu}, \underset{\sim}{Z}, \underset{\sim}{x}) = \log P(\underset{\sim}{\mu}) + \underbrace{\log P(Z_i) + \log P(X_i \mid \mu_{Z_i}, Z_i)}_{\text{ith person}}$$

$$+ \underbrace{\sum_{i' \neq i} \log P(Z_{i'}) + \log P(X_{i'} \mid \mu_{Z_{i'}}, Z_{i'})}_{\text{other people}}$$

$$q^*(Z_i = k) \propto \exp\left(\log \pi_{Z_i} + \underbrace{E \log P(x_i \mid \mu_{Z_i})}_{E} \right)$$

$Z_i = k$; E w.r.t $q(\mu_{Z_i})$

$k = 1, 2, \cdots K$.

$$E = -\frac{1}{2}\log(2\pi) - \frac{1}{2}x_i^2 + x_i E(\mu_{Z_i}) - E[\mu_{Z_i}^2]/2)$$

- [ ] prior of μ_{Z_i}
- [x] variational distn of μ_{Z_i}

Mean-Field Approximation.

Update μ_k $k = 1, \ldots, K$.

Aside: Exponential family:

$$p(\underset{\sim}{x}) = h(\underset{\sim}{x}) \exp(\underset{\sim}{\eta}^t t(\underset{\sim}{x}) - a(\underset{\sim}{\eta}))$$

generic

sufficient statistics

natural parameters

normalizing constant

$$a(\underset{\sim}{\eta}) = \log\left(\int h(\underset{\sim}{x}) \exp(\eta^t t(\underset{\sim}{x})) \, d\underset{\sim}{x}\right)$$

1) $$\frac{\partial a(\underset{\sim}{\eta})}{\partial \underset{\sim}{\eta}} = E(t(\underset{\sim}{X})) \quad ; \quad \frac{\partial^2 a(\underset{\sim}{\eta})}{\partial \eta^2} = Var(t(\underset{\sim}{X}))$$

For Gaussian distn.

$$p(x) = \frac{1}{\sqrt{2\sigma^2 \pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(\frac{\mu}{\sigma^2} x - \frac{1}{2\sigma^2} x^2 - \frac{1}{2\sigma^2}\mu^2 - \log \sigma\right)$$

$$\underset{\sim}{\eta} = \left(\frac{\mu}{\sigma^2}, \frac{1}{\sigma^2}\right)$$

$$t(\underset{\sim}{x}) = \left(x, -\frac{1}{2}x^2\right)$$

Back to updating

$$q^*(\mu_k) \propto \exp(E_{-\{\mu_k\}} \log p(\underset{\sim}{\mu}, \underset{\sim}{z}, \underset{\sim}{x}))$$

we can simplify this update if we have assumed $q(\mu_k)$ comes from exponential family (Gaussian in this case)

$$\propto E_q(\eta)$$

$q(\mu_k \mid \tilde{\mu}_k, \tilde{\sigma}_k^2)$, $k = 1, \dots, K$

variational parameters

from a Gaussian family.

mean, variance parameters

$$\left(\eta_1 = \frac{\tilde{\mu}_k}{\tilde{\sigma}_k^2} \qquad \eta_2 = \frac{1}{\tilde{\sigma}_k^2} \right)$$

to find (η_1^*, η_2^*) that maximizes the $ELBO(\mu_k)$.

Natural Param. Updating

optimal

$$\begin{cases} \eta_1^* = \eta_{10} + \sum_{i=1}^n E\left(I(z_i = k)\right) \cdot x_i \\ \eta_2^* = \eta_{20} + \sum_{i=1}^n E\left(I(z_i = k)\right) \end{cases}$$

updating formula

(η_{10}, η_{20}) are the natural parameters of Prior Gaussian.

$$= \left(\frac{\mu_0}{\sigma_0^2}, \frac{1}{\sigma_0^2} \right).$$

equivalently

$$\mu_k^* = \frac{\eta_1^*}{\eta_2^*} = \frac{\frac{\mu_0}{\sigma_0^2} + \sum_{i=1}^n E\left(I(z_i = k)\right) x_i}{\frac{1}{\sigma_0^2} + \sum_{i=1}^n E\left(I(z_i = k)\right)}$$

$$\sigma_k^{2*} = \frac{1}{\eta_2^*} = \frac{1}{\frac{1}{\sigma_0^2} + \sum_{i=1}^n E\left(I(z_i = k)\right)}$$

mean-variance parametrization.