

Variational Approach for Bayesian Density Regression

Eric Chuu

Texas A&M University

ericchuu@tamu.com

December 5, 2018

- 1 Introduction and Setup
- 2 Notation and Prior Specification
- 3 Variational Approach
 - CAVI Updates
 - Example Calculation for $q(\mathbf{Z})$
 - Bouchard Bound for Problematic Quantity
 - Variational Algorithm
 - Monitoring Convergence

Problem Setup

- Given data $(y_n, x_n)_{n=1}^N$ and we want to estimate the density of $y \mid x$
- Gaussian mixture models are a common choice due to flexibility

$$f(y \mid x) = \sum_k \pi_k \mathcal{N}(y \mid \mu_k(x), \tau_k^{-1})$$

- Covariate-independent weights have limited flexibility in practice

Covariate Dependent Weights

- Allow for covariate dependence in the mixing weights

$$f(y | x) = \sum_k \pi_k(x) \mathcal{N}(y | \mu_k(x), \tau_k^{-1})$$

- Existing methods that use kernel stick breaking process (Dunson and Park, 2008), logit stick breaking prior (Rigon and Durante, 2017)
- We allow covariates to enter the weights via a logit link function

$$\pi_k(x) \propto \exp\{x^\top \gamma_k\}$$

Notation

- Observed Data: $\mathbf{y} = \{y_1, \dots, y_N\}$, $\mathbf{X} = \{x_1, \dots, x_N\} \subseteq \mathbb{R}^D$
- Gaussian Components: $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_K\}$, $\boldsymbol{\tau} = \{\tau_1, \dots, \tau_k\}$

$$f(y | x) = \sum_k \pi_k(x) \mathcal{N}(y | x^\top \beta_k, \tau_k^{-1})$$

- coefficient vectors in the weights: $\boldsymbol{\gamma} = \{\gamma_1, \dots, \gamma_k\}$
- Latent variables: $\mathbf{Z} = \{z_1, \dots, z_N\}$
 - $z_n \in \mathbb{R}^K$ and $z_{nk} = 1$ iff y_n belongs to the k -th cluster

Prior Specification

With the introduction of latent variables, we have the simplified conditional distribution,

$$p(\mathbf{y} | \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\tau}, \mathbf{Z}) = \prod_n \prod_k \mathcal{N}(y_n | \mathbf{x}_n^\top \boldsymbol{\beta}_k, \tau_k^{-1})^{z_{nk}} \quad (1)$$

The conditional distribution of \mathbf{Z} given $\mathbf{X}, \boldsymbol{\gamma}$

$$p(\mathbf{Z} | \mathbf{X}, \boldsymbol{\gamma}) = \prod_n \prod_k \pi_k(x_n)^{z_{nk}} = \prod_n \prod_k \left(\frac{\exp\{\mathbf{x}_n^\top \boldsymbol{\gamma}_k\}}{\sum_{j=1}^K \exp\{\mathbf{x}_n^\top \boldsymbol{\gamma}_j\}} \right)^{z_{nk}} \quad (2)$$

Priors over $\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\gamma}$

- $p(\boldsymbol{\gamma}) = \prod_k p(\boldsymbol{\gamma}_k) = \mathcal{N}(\boldsymbol{\gamma}_k | 0, \mathbf{I}_D)$
- $p(\boldsymbol{\beta}, \boldsymbol{\tau}) = p(\boldsymbol{\beta} | \boldsymbol{\tau}) p(\boldsymbol{\tau}) = \prod_k \mathcal{N}(\boldsymbol{\beta}_k | m_0, (\tau_k \boldsymbol{\Lambda}_0)^{-1}) \text{Ga}(\tau_k | a_0, b_0)$

General Approach

Full joint distribution of all random variables

$$p(\mathbf{y}, \mathbf{X}, \beta, \tau, \mathbf{Z}) = p(\mathbf{y} \mid \mathbf{X}, \beta, \tau, \mathbf{Z}) p(\mathbf{Z} \mid \mathbf{X}, \gamma) p(\gamma) p(\beta, \tau)$$

Formulate an approximating distribution

$$q(\mathbf{Z}, \beta, \tau, \gamma) = q(\mathbf{Z})q(\beta, \tau, \gamma)$$

Derive update equations for the variational distributions

$$\ln q^*(\mathbf{Z}) = \mathbb{E}_{-q(\mathbf{Z})} \left[\ln \{ p(\mathbf{y}, \mathbf{X}, \beta, \tau, \mathbf{Z}, \gamma) \} \right]$$

$$\ln q^*(\beta, \tau, \gamma) = \mathbb{E}_{-q(\beta, \tau, \gamma)} \left[\ln \{ p(\mathbf{y}, \mathbf{X}, \beta, \tau, \mathbf{Z}, \gamma) \} \right]$$

Variational Approximation

Joint Distribution of all random variables

$$\begin{aligned}\ln p(\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\tau}, \mathbf{Z}) = & \sum_n \sum_k z_{nk} \left\{ -\frac{1}{2} \ln(2\pi) + \frac{1}{2} \ln \tau_k - \frac{\tau_k}{2} (y_n - \mathbf{x}_n^\top \boldsymbol{\beta}_k)^2 \right\} \\ & + \sum_n \sum_k z_{nk} \left\{ \mathbf{x}_n^\top \boldsymbol{\gamma}_k - \ln \left(\sum_{j=1}^K \exp\{\mathbf{x}_n^\top \boldsymbol{\gamma}_j\} \right) \right\} \\ & + \sum_k \left\{ -\frac{D}{2} \ln(2\pi) + \frac{D}{2} \ln \tau_k + \ln |\boldsymbol{\Lambda}_0| - \frac{\tau_k}{2} (\boldsymbol{\beta}_k - \mathbf{m}_0)^\top \boldsymbol{\Lambda}_0 (\boldsymbol{\beta}_k - \mathbf{m}_0) \right\} \\ & + \sum_k \left\{ (a_0 - 1) \ln \tau_k - b_0 \tau_k \right\}\end{aligned}$$

Approximating Distribution

$$q(\mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\gamma}) = q(\mathbf{Z})q(\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\gamma}) \quad (3)$$

Example calculation for $q(\mathbf{Z})$

$$\begin{aligned}\ln q^*(\mathbf{Z}) &= \sum_n \sum_k z_{nk} \left\{ -\frac{1}{2} \ln(2\pi) + \frac{1}{2} E_{q(\tau)}[\ln \tau_k] \right. \\ &\quad \left. - \frac{1}{2} E_{q(\beta, \tau)}[\tau_k (y_n - \mathbf{x}_n^\top \beta_k)^2] \right. \\ &\quad \left. + \mathbf{x}_n^\top E_{q(\gamma)}[\gamma_k] - E_{q(\gamma)} \left[\ln \left(\sum_j \exp\{\mathbf{x}_n^\top \gamma_j\} \right) \right] \right\} \\ &= \sum_n \sum_k z_{nk} \ln \rho_{nk}\end{aligned}$$

Exponentiating and normalizing, we arrive at the optimal variational distribution. Then, $\mathbb{E}[z_{nk}] = r_{nk}$.

$$q^*(\mathbf{Z}) = \prod_n \prod_k r_{nk}^{z_{nk}}, \quad r_{nk} = \frac{\rho_{nk}}{\sum_j \rho_{nj}} \quad (4)$$

Example calculation for $q(\mathbf{Z})$

Consider the quantity $\ln \rho_{nk}$:

- $E_{q(\tau)}[\ln \tau_k]$
- $E_{q(\beta, \tau)}[\tau_k (y_n - x_n^\top \beta_k)^2]$
- $E_{q(\gamma)}[\gamma_k]$
- $E_{q(\gamma)} \left[\ln \left(\sum_j \exp \{ x_n^\top \gamma_j \} \right) \right]$

Variational Updates

Our choice of conjugate families tells us that we will have the following form of the variational distributions, for $k = 1, \dots, K$,

$$q^*(\gamma_k) = \mathcal{N}(\gamma_k \mid \mu_k, Q_k^{-1}) \quad (5)$$

$$q^*(\beta_k \mid \tau_k) = \mathcal{N}(\beta_k \mid m_k, (\tau_k V_k)^{-1}) \quad (6)$$

$$q^*(\tau_k) = \text{Ga}(\tau_k \mid a_k, b_k) \quad (7)$$

Example calculation for $q(\mathbf{Z})$

Consider the quantity $\ln \rho_{nk}$:

- $E_{q(\tau)}[\ln \tau_k] = \psi(a_k) - \psi(b_k)$
- $E_{q(\beta, \tau)}[\tau_k (y_n - x_n^\top \beta_k)^2] = \frac{a_k}{b_k} (y_n + m_k^\top x_n)^2 + x_n^\top V_k^{-1} x_n$
- $E_{q(\gamma)}[\gamma_k] = \mu_k$
- $E_{q(\gamma)} \left[\ln \left(\sum_j \exp\{x_n^\top \gamma_j\} \right) \right] = \dots \text{ oh no!}$

Bouchard's Bound for Problematic Quantity

Upper bound on the sum of exponentials

$$\sum_{j=1}^K e^{t_j} \leq \prod_{j=1}^K (1 + e^{t_j})$$

Tangential Bound (Jaakkola and Jordan, 1996)

For $x \in \mathbb{R}, \alpha \in \mathbb{R}, \xi \in [0, \infty)$

$$\log(1 + e^x) \leq \lambda(\xi)(x^2 - \xi^2) + \frac{x - \xi}{2} + \log(1 + e^\xi)$$

Taking log the first bound and setting $t_j = x_n^\top \gamma_j - \alpha_n$, we have

$$\begin{aligned} \log \sum_{j=1}^K \exp\{x_n^\top \gamma_j\} &\leq \alpha_n + \sum_{j=1}^K \frac{x_n^\top \gamma_j - \alpha_n + \xi_{nj}}{2} + \lambda(\xi_{nj}) ((x_n^\top \gamma_j - \alpha_n)^2 - \xi_{nj}^2) \\ &\quad + \log(1 + e^{\xi_{nj}}) \end{aligned}$$

Bouchard's Bound for Problematic Quantity

Upper Bound for the problematic expectation

$$\begin{aligned} & \mathbb{E}_{q(\gamma)} \left[\ln \left(\sum_j^K \exp\{x_n^\top \gamma_j\} \right) \right] \\ & \leq \alpha_n + \sum_j^K \lambda(\xi_{nj}) \left((x_n^\top \mu_j - \alpha_k)^2 - \xi_{nj}^2 + x_n^\top Q_k^{-1} x_n \right) + \log(1 + e^{\xi_{nj}}) \\ & \quad + \frac{1}{2} (x_n^\top \mu_j - \alpha_n + \xi_{nj}) \end{aligned}$$

Two new variational parameters that need to be updated

The additional parameters introduced in the two upper bounds can be updated using the following equations

$$\xi_{nk} = \sqrt{(\mu_k^\top x_n - \alpha_n)^2 + x_n^\top Q_k^{-1} x_n} \quad \forall k$$

$$\alpha_n = \frac{\frac{1}{2} \left(\frac{K}{2} - 1 \right) + \sum_{j=1}^K \lambda(\xi_{nj}) \mu_j^\top x_n}{\sum_{j=1}^K \lambda(\xi_{nj})}$$

Variational Algorithm

Input

- 1 Data $(y_n, x_n)_{n=1}^N$
- 2 Number of components, K
- 3 Prior mean, precision for coefficients vectors, $\beta_{1:K}$
- 4 Prior shape, rate parameters for precision parameters, $\tau_{1:K}$

Output

- 1 A variational density,

$$q(\mathbf{Z}, \beta, \tau, \gamma) = q(\mathbf{Z})q(\beta, \tau, \gamma) = q(\mathbf{Z}) \prod_k q(\beta_k, \tau_k)q(\gamma_k)$$

- 2 Fully specified by the variational parameters

Variational Algorithm

Algorithm. CAVI for Conditional Density Estimation

while *the ELBO has not converged* **do**

for $n \in \{1, \dots, N\}$ **do****for** $k \in \{1, \dots, K\}$ **do**

$$\text{Set } r_{nk} \propto \exp \left\{ -\frac{1}{2} \ln(2\pi) + \frac{1}{2} \mathbb{E}[\ln \tau_k] + x_n^T \mathbb{E}[\gamma_k] \right. \\ \left. - \frac{1}{2} \mathbb{E}[\tau_k (y_n - x_n^T \beta_k)^2] - \mathbb{E} \left[\ln \left(\sum_{j=1}^K \exp \{ x_n^T \gamma_j \} \right) \right] \right\}$$

end

end

for $n \in \{1, \dots, N\}$ **do****for** $k \in \{1, \dots, K\}$ **do**

$$\text{Set } \xi_{nk} \leftarrow \sqrt{(x_n^\top \mu_k - \alpha_n)^2 + x_n^\top Q_k^{-1} x_n}$$

end

$$\text{Set } \alpha_n \leftarrow \left[\frac{1}{2} \left(\frac{K}{2} - 1 \right) + \sum_k \lambda(\xi_{nk}) \mu_k^T x_n \right] / \left[\sum_k \lambda(\xi_{nk}) \right]$$

end

```
/** Remaining Variational Updates on Next Slide **/
```

end

Variational Algorithm (cont.)

Algorithm. CAVI for Conditional Density Estimation

while *the ELBO has not converged* **do**

```
⋮
for  $k \in \{1, \dots, K\}$  do
     $Q_k \leftarrow I_D + 2 \sum_n r_{nk} \lambda(\xi_{nk}) x_n x_n^T$            /* gamma_k cov */
     $\eta_k \leftarrow \sum_n r_{nk} \left[ \frac{1}{2} + 2\lambda(\xi_{nk}) \alpha_n \right] x_n$ 
     $\mu_k \leftarrow Q_k^{-1} \eta_k$                                    /* gamma_k mean */
     $V_k \leftarrow \sum_n r_{nk} x_n x_n^T + \Lambda_0$              /* beta_k cov */
     $\zeta_k \leftarrow \sum_n r_{nk} y_n x_n + \Lambda_0 m_0$ 
     $m_k \leftarrow V_k^{-1} \zeta_k$                                /* beta_k mean */
     $a_k \leftarrow a_0 + N_k$                                    /* tau_k shape */
     $b_k \leftarrow b_0 + \frac{1}{2} [\sum_n r_{nk} y_n^2 + m_0^T \Lambda_0 m_0 - \zeta_k^T V_k^{-1} \zeta_k]$  /* tau_k rate */
```

end

Compute ELBO using updated parameters

end

return $q(\mathbf{Z}, \beta, \tau, \gamma)$

At the end of each iteration, we compute the ELBO:

$$\mathcal{L}(q) = \sum_{\mathbf{z}} \int \int \int q(\beta, \tau, \gamma, \mathbf{Z}) \ln \left\{ \frac{p(\mathbf{y}, \mathbf{X}, \beta, \tau, \gamma, \mathbf{Z})}{q(\beta, \tau, \gamma, \mathbf{Z})} \right\} d\beta d\tau d\gamma$$

7 quantities to calculate:

- $\mathbb{E}[\ln p(\mathbf{y} \mid \mathbf{X}, \beta, \tau, \mathbf{Z})]$
- $\mathbb{E}[\ln p(\mathbf{Z} \mid \mathbf{X}, \gamma)]$
- $\mathbb{E}[\ln p(\gamma)]$
- $\mathbb{E}[\ln p(\beta, \tau)]$
- $\mathbb{E}[\ln q(\mathbf{Z})]$
- $\mathbb{E}[\ln q(\beta, \tau)]$
- $\mathbb{E}[\ln q(\gamma)]$

References

Please see report submission for updated set of references.

Thank you!