Example 3: Linear regression with empirical Bayes estimation for hyperparameters

Bayesian linear regression model

We have the model

$$\begin{aligned} y_i \mid \boldsymbol{\beta} &\sim \mathcal{N}(\boldsymbol{x}_i^T \boldsymbol{\beta}, \phi^{-1}), \quad i = 1, \ldots, n; \\ \boldsymbol{\beta} \mid \kappa &\sim \mathcal{N}(0, \kappa^{-1}\mathrm{I}); \\ \kappa &\sim \mathrm{Gamma}(a_0, b_0); \end{aligned}$$

where $\boldsymbol{x}_i, \ i = 1, \ldots, n$ are known covariates, $\boldsymbol{\beta} \in \mathbb{R}^p$ includes the intercept, and is unknown, and I is the identity matrix. Assume a_0 and b_0 are known. We now treat the precision parameter $\phi = 1/\sigma^2$ as unknown.

Solution

Instead of fully variational approach, we use VBEM to estimate the posterior by treating ϕ as a hyper-parameter to update in the M-step. As the marginal likelihood is not available, we use the ELBO as a proxy.

The ELBO is given by

$$\begin{aligned} \mathrm{ELBO} &= \mathbb{E}_q \left[\log p(\boldsymbol{\beta}, \kappa, \boldsymbol{y})\right] - \mathbb{E}_q \left[\log q(\boldsymbol{\beta}, \kappa)\right] \\ &= \mathbb{E}_{q_{\boldsymbol{\beta}}} \left[\log p(\boldsymbol{y} \mid \boldsymbol{\beta})\right] + \mathbb{E}_{q_{\boldsymbol{\beta},\kappa}} \left[\log p(\boldsymbol{\beta} \mid \kappa)\right] + \mathbb{E}_{q_\kappa} \left[\log p(\kappa)\right] - \mathbb{E}_{q_{\boldsymbol{\beta}}} \left[\log q(\boldsymbol{\beta})\right] - \mathbb{E}_{q_\kappa} \left[\log q(\kappa)\right]; \end{aligned}$$

where

$$\begin{aligned} \mathbb{E}_{q_{\boldsymbol{\beta}}} \left[\log p(\boldsymbol{y} \mid \boldsymbol{\beta})\right] &= \frac{n}{2} \log(\frac{\phi}{2\pi}) - \frac{\phi}{2} \boldsymbol{y}^T \boldsymbol{y} + \phi \boldsymbol{m}_n^T \boldsymbol{X}^T \boldsymbol{y} - \frac{\phi}{2} \mathrm{tr}(\boldsymbol{X}^T \boldsymbol{X}(\boldsymbol{m}_n \boldsymbol{m}_n^T + \boldsymbol{S}_n)) \\ \mathbb{E}_{q_{\boldsymbol{\beta},\kappa}} \left[\log p(\boldsymbol{\beta} \mid \kappa)\right] &= -\frac{p}{2} \log(2\pi) + \frac{p}{2}(\Psi(a_n) - \log b_n) - \frac{a_n}{2b_n}(\boldsymbol{m}_n^T \boldsymbol{m}_n + \mathrm{tr}(\boldsymbol{S}_n)) \\ \mathbb{E}_{q_\kappa} \left[\log p(\kappa)\right] &= a_0 \log b_0 + (a_0 - 1)(\Psi(a_n) - \log b_n) - b_0 \frac{a_n}{b_n} - \log \Gamma(a_0) \\ \mathbb{E}_{q_{\boldsymbol{\beta}}} \left[\log q(\boldsymbol{\beta})\right] &= -\frac{1}{2} \log |\boldsymbol{S}_n| - \frac{p}{2}(1 + \log(2\pi)) \\ \mathbb{E}_{q_\kappa} \left[\log q(\kappa)\right] &= -\log \Gamma(a_n) + (a_n - 1)\Psi(a_n) + \log b_n - a_n \end{aligned}$$

and where $\Gamma(\cdot)$ and $\Psi(\cdot)$ denote the Gamma and digamma function, respectively.

Taking the derivative of the ELBO w.r.t. ϕ, the VBEM update for ϕ must satisfy

$$\frac{\partial \mathrm{ELBO}}{\partial \phi} = \frac{n}{2\phi} - \frac{1}{2} \boldsymbol{y}^T \boldsymbol{y} + \boldsymbol{m}_n^T \boldsymbol{X}^T \boldsymbol{y} - \frac{1}{2} \mathrm{tr}(\boldsymbol{X}^T \boldsymbol{X}(\boldsymbol{m}_n \boldsymbol{m}_n^T + \boldsymbol{S}_n)) = 0;$$

which gives the update.

$$\widehat{\phi} = \frac{n}{\boldsymbol{y}^T\boldsymbol{y} + \mathrm{tr}(\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{m}_n\boldsymbol{m}_n^T + \boldsymbol{S}_n)) - 2\boldsymbol{m}_n^T\boldsymbol{X}^T\boldsymbol{y}}.$$

Notably, as this uses the ELBO as a proxy for the marginal likelihood, we should ensure the ELBO has converged before updating $\widehat{\phi}$. Thus, each time we update $\widehat{\phi}$ we should first run the VI updating scheme for $\boldsymbol{\beta}$ and κ fully (until convergence).