

Example 3: Linear regression with empirical Bayes estimation for hyperparameters

Bayesian linear regression model

We have the model

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

where \mathbf{x}_i , $i = 1, \dots, n$ are known covariates, $\boldsymbol{\beta} \in \mathbb{R}^p$ includes the intercept, and is unknown, and \mathbf{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.

Exercise

Instead of fully variational approach, use VBEM to estimate the posterior by treating ϕ as a hyperparameter to update in the M-step. Implement the algorithm, and run on simulated data with one covariate $x_{i,1} \sim \mathcal{N}(0, 1)$, $n = 50$, $\beta_0 = -1$, $\beta_1 = 2$, $a_0 = b_0 = 0.001$. Play around with different initial values for ϕ - is this choice important?

Solution

Instead of fully variational approach, we use VBEM to estimate the posterior by treating ϕ as a hyperparameter 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

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

where

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

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

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

which gives the update.

$$\hat{\phi} = \frac{n}{\mathbf{y}^T \mathbf{y} + \text{tr}(\mathbf{X}^T \mathbf{X} (\mathbf{m}_n \mathbf{m}_n^T + \mathbf{S}_n)) - 2\mathbf{m}_n^T \mathbf{X}^T \mathbf{y}}.$$

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

R code is given in a separate file.