Derivation of a Variational-Bayes Linear Regression Algorithm using a Normal Inverse-Gamma Generative Model

Joaquín Rapela

June 12, 2017

Abstract

Here we give a brief introduction to variational inference, derive a variational-Bayes linear regression algorithm using a normal inverse-gamma generative model, and provide R sourced code implementing this algorithm.

1 Introduction

Section 2 summarizes key concepts of variational inference and Section 3 uses these concepts to derive a variational-Bayes linear regression algorithm using a Normal Inverse-Gamma generative model. This derivation builds on that of Bishop (2006, Chapter 10). For simplicity, the generative model in the latter derivation assumes the variance of the likelihood function is known and fixed to its true value. The generative model used below does not make such simplifying assumption and uses the more general generative model in Drugowitsch (2014). The derivations given here contain more details than those in Drugowitsch (2014).

2 Summary of Variational Inference

The variational inference framework is based on the following easy to check equations:

$$\begin{aligned}\ln p(\mathbf{x}) &= \mathcal{L}(q) + KL(q||p)\\ \mathcal{L}(q) &= \int q(\mathbf{z}) \ln\left\{\frac{p(\mathbf{x},\mathbf{z})}{q(\mathbf{z})}\right\} d\mathbf{z}\\ KL(q||p) &= -\int q(\mathbf{z}) \ln\left\{\frac{p(\mathbf{z}|\mathbf{x})}{q(\mathbf{z})}\right\} d\mathbf{z}\end{aligned}$$

We want to find values of the latent variable $\mathbf{z}$ that maximizes the posterior probability $p(\mathbf{z}|\mathbf{x})$ given an observable variable $\mathbf{x}$. Computing $p(\mathbf{z}|\mathbf{x})$ is challenging, so we are going to approximate it with a density $q(\mathbf{z})$. We seek the density $q^*(\mathbf{z})$ that minimizes the Kullback-Leibler divergence $KL(q||p)$ between $q(\mathbf{z})$ and $p(\mathbf{z}|\mathbf{x})$ over all possible q's.

However, we don't want to minimize the KL divergence directly because for this we need $p(\mathbf{z}|\mathbf{x})$, which is difficult to compute. Instead we maximize $\mathcal{L}(q)$ over all possible distribution q, which is easier since computing $p(\mathbf{x}, \mathbf{z})$ is simpler.

As shown in Bishop (2006, Eq. 10.9), if we take a factorization

$$q(\mathbf{z}) = \Pi_{i=1}^{M} q_i(\mathbf{z}_i) \tag{1}$$

and we keep $q_{i \neq j}$ fixed, then the q_j^* that maximizes $\mathcal{L}(q)$ is

$$\ln q_j^*(\mathbf{z_j}) = E_{i \neq j}\{\ln p(\mathbf{x}, \mathbf{z})\} + \text{const.} \tag{2}$$

3 Variational-Bayes Linear Regression

We use the following generative linear regression model:

$$p(\mathbf{y}, \mathbf{w}, \tau, \Phi, \alpha) = p(\mathbf{y}|\Phi, \mathbf{w}, \tau)\; p(\mathbf{w}, \tau|\alpha)\; p(\alpha) \tag{3}$$
$$p(\mathbf{y}|\Phi, \mathbf{w}, \tau) = N(\mathbf{y}|\Phi, \mathbf{w}, \tau^{-1} I) \tag{4}$$
$$p(\mathbf{w}, \tau|\alpha) = N(\mathbf{w}|\mathbf{0}, (\tau\alpha)^{-1} I)\; \text{Gam}(\tau|a_0, b_0) \tag{5}$$
$$p(\alpha) = \text{Gam}(\alpha|c_o, d_0) \tag{6}$$

where $\mathbf{y} \in \Re^N$ and $\mathbf{w} \in \Re^D$ are the dependent variable and weights of the linear regression model, respectively, τ is the precision of $\mathbf{y}$, Φ is the matrix of independent observations and α influences the precision of the prior on the weights. We assume that q factorizes as

$$q(\mathbf{w}, \tau, \alpha) = q(\mathbf{w}, \tau)\; q(\alpha) \tag{7}$$

and from Eq. 2 we obtain

$$\ln q^*(\mathbf{w}, \tau) = E_\alpha\{\ln p(\mathbf{y}, \mathbf{w}, \tau, \Phi, \alpha)\} + \text{const.} \tag{8}$$
$$\ln q^*(\alpha) = E_{\mathbf{w},\tau}\{\ln p(\mathbf{y}, \mathbf{w}, \tau, \Phi, \alpha)\} + \text{const.} \tag{9}$$

By calculating the expectation in the right-hand side of Eq. 8, Lemma 1 shows that

$$q^*(\mathbf{w}, \tau) = N(\mathbf{w}|\mathbf{m}_N, S_N)\; \text{Gam}(\tau|a_N, b_N) \tag{10}$$

with

$$\mathbf{m}_N = V_N \Phi^T \mathbf{y} \tag{11}$$
$$S_N^{-1} = \tau V_N^{-1} \tag{12}$$
$$V_N = (\Phi^T \Phi + E_\alpha\{\alpha\} I)^{-1}$$
$$a_N = a_0 + N/2 \tag{13}$$
$$b_N = b_0 + \frac{1}{2}(||\mathbf{y} - \Phi \mathbf{m}_N||^2 + E_\alpha\{\alpha\}||\mathbf{m}_N||^2) \tag{14}$$

and Corollary 1.1 proves

$$E_{\mathbf{w},\tau}\{\tau||\mathbf{w}||^2\} = \text{trace}\{V_N\} + ||\mathbf{m}_N||^2 \, a_N/b_N \tag{15}$$

By calculating the expectation in the right-hand side of Eq. 9, Lemma 2 shows that

$$q^*(\alpha) = \text{Gam}(\alpha|c_N, d_N) \tag{16}$$

with

$$c_N = c_0 + D/2 \tag{17}$$

$$d_N = d_0 + \frac{E_{\mathbf{w},\tau}\{\tau||\mathbf{w}||^2\}}{2} \tag{18}$$

and from Eq. 16 it follows

$$E_\alpha\{\alpha\} = \frac{c_N}{d_N}$$

It is remarkable that from only the generative model in Eq. 3-6 and from the factorization of q in Eq. 7 we can derive the parametrized close-form solution of the density $q(\mathbf{z})$ best approximating the posterior density $p(\mathbf{z}|\mathbf{x})$. To find the optimal parameters of q we proceed iteratively, as shown in Listing 1.

References

C.M. Bishop. *Pattern recognition and machine learning*. Springer, New York, NY, 2006.

J. Drugowitsch. Variational bayesian inference for linear and logistic regression. 2014.

A Proofs

Lemma 1. *For the generative model in Eqs. 3-6, the parametrized close-form expression of $q^*(\mathbf{w}, \tau)$ in Eq. 7 is given in Eq. 10.*

Proof.

$$\begin{aligned} \ln q^*(\mathbf{w}, \tau) &= E_\alpha\{\ln p(\mathbf{y}, \mathbf{w}, \tau, \Phi, \alpha)\} + \text{const.} \\ &= \ln p(\mathbf{y}|\Phi, \mathbf{w}, \tau) + E_\alpha\{\ln p(\mathbf{w}, \tau|\alpha)\} + \text{const.} \end{aligned} \tag{19}$$

The first equality is Eq. 8 and the second one follows from Eq. 3 by keeping only terms that depend on $\mathbf{w}$ and τ.
From Eq. 4

Listing 1 Variational-Bayes Linear Regression algorithm

Require: $\mathbf{y}, \Phi, a_0, b_0, c_0, d_0, \text{maxIter}$

$N \leftarrow \text{nrow}(\Phi)$
$D \leftarrow \text{ncol}(\Phi)$
converged $\leftarrow$ False
eAlpha $\leftarrow c_0/d_0$
$a_N \leftarrow a_0 + N/2$
$c_N \leftarrow c_0 + D/2$
$b_N \leftarrow b_0, d_N \leftarrow d_0$
lowerBound $\leftarrow -\text{largeNumber}$
iter $\leftarrow 1$
for iter $= 1 : \text{maxIter}$ **do**
  $V_N \leftarrow (\Phi^T\Phi + \text{eAlpha } I_D)^{-1}$
  $m_N \leftarrow V_N\Phi^T\mathbf{y}$
  oldLowerBound $\leftarrow$ lowerBound
  lowerBound $\leftarrow$ computeLowerBound()
  if lowerBound $-$ oldLowerBound $< \epsilon$ **then**
    converged $\leftarrow$ True
    break
  end if
  $b_N \leftarrow b_0 + \frac{1}{2}(||\mathbf{y} - \Phi\mathbf{m}_N||^2 + \text{eAlpha}||\mathbf{m}_N||^2)$
  eTauWTW2 $\leftarrow \text{trace}\{V_N\} + ||\mathbf{m}_N||^2 a_N/b_N$
  $d_N \leftarrow d_0 + \frac{\text{eTauWTW2}}{2}$
  eAlpha $\leftarrow c_N/d_N$
end for
return $\mathbf{m}_N, V_N, a_N, b_N, c_N, d_N, \text{converged}$

$$p(\mathbf{y}|\Phi,\mathbf{w},\tau) = \frac{\tau^{N/2}}{(2\pi)^{N/2}} \exp\left(-\frac{\tau}{2}||\mathbf{y}-\Phi\mathbf{w}||^2\right)$$

then

$$\ln p(\mathbf{y}|\Phi,\mathbf{w},\tau) = \frac{N}{2}\ln\tau - \frac{\tau}{2}||\mathbf{y}-\Phi\mathbf{w}||^2 + \text{const.} \tag{20}$$

where const. groups all terms that do not depend on $\mathbf{w}$ or τ.
From Eq. 5

$$p(\mathbf{w},\tau|\alpha) = \frac{\tau^{D/2}\alpha^{D/2}}{(2\pi)^{D/2}} \exp\left(-\frac{1}{2}\tau\alpha||\mathbf{w}||^2\right) \frac{1}{\Gamma(a_0)} b_0^{a_0}\tau^{a_0-1}\exp(-b_0\tau) \tag{21}$$

then

$$E_\alpha\{\ln(p(\mathbf{w},\tau|\alpha)\} = \left(\frac{D}{2}+(a_0-1)\right)\ln\tau - \left(\frac{||\mathbf{w}||^2}{2}E_\alpha\{\alpha\}+b_0\right)\tau + \text{const.} \tag{22}$$

Replacing Eqs. 20 and 22 into Eq. 19 we obtain

$$\begin{aligned}\ln q^*(\mathbf{w},\tau) = &\left(\frac{N}{2}+(a_0-1)+\frac{D}{2}\right)\ln\tau \\ &- \frac{1}{2}\tau\left(||\mathbf{y}-\Phi\mathbf{w}||^2 + E_\alpha\{\alpha\}||\mathbf{w}||^2\right) \\ &- b_0\tau + \text{const.}\end{aligned} \tag{23}$$

Completing squares on the second term of Eq. 23 and re-arranging we obtain

$$\begin{aligned}\ln q^*(\mathbf{w},\tau) = &\left((a_0-1)+\frac{N}{2}\right)\ln\tau \\ &- \frac{\tau}{2}\left(||\mathbf{y}-\Phi\mathbf{m}_N||^2 + E_\alpha\{\alpha\}||m_N||^2 + 2b_0\right) \\ &- \frac{1}{2}(\mathbf{w}-\mathbf{m}_N)^T S_N^{-1}(\mathbf{w}-\mathbf{m}_N) \\ &+ \frac{D}{2}\ln\tau + \text{const.}\end{aligned} \tag{24}$$

with $\mathbf{m}_N$ and S_N given in Eqs. 11 and 12, respectively. Defining a_N and b_N as in Eqs. 13 and 14, respectively, from Eq. 24 we obtain

$$\ln q^*(\mathbf{w},\tau)) = \ln N(\mathbf{w}|\mathbf{m}_N, S_N)\ \text{Gam}(\tau|a_N,b_N)$$

□

Corollary 1.1. *Given $q^*(\mathbf{w}, \tau)$ in Eq. 10, $E_{\mathbf{w},\tau}\left\{\tau||\mathbf{w}||^2\right\}$ is given in Eq. 15*

Proof.

$$\begin{aligned} E_{\mathbf{w},\tau}\left\{\tau||\mathbf{w}||^2\right\} &= \int \tau \int \ldots \int ||\mathbf{w}||^2 q^*(\mathbf{w}, \tau)\, d\mathbf{w} d\tau \\ &= \int \tau\, \mathrm{Gam}(\tau|a_N, b_N) \left(\int \ldots \int ||\mathbf{w}||^2 N(\mathbf{w}|\mathbf{m}_N, S_N)\, d\mathbf{m}_N \right) d\tau \\ &= \int \tau\, \mathrm{Gam}(\tau|a_N, b_N) E_{\mathbf{w}}\left\{||\mathbf{w}||^2\right\}\, d\tau \end{aligned} \tag{25}$$

Next we derive $E_{\mathbf{w}}\left\{||\mathbf{w}||^2\right\}$

$$\begin{aligned} E_{\mathbf{w}}\left\{||\mathbf{w}||^2\right\} &= \mathrm{trace}\left\{\mathrm{cor}(\mathbf{w})\right\} \\ &= \mathrm{trace}\left\{cov(\mathbf{w}) + E\{\mathbf{w}\}E\{\mathbf{w}\}^T\right\} \\ &= \mathrm{trace}\left\{cov(\mathbf{w})\right\} + E\{\mathbf{w}\}^T E\{\mathbf{w}\} \\ &= \mathrm{trace}\left\{S_N\right\} + \mathbf{m}_n^T \mathbf{m}_N \\ &= \tau^{-1}\mathrm{trace}\left\{V_N\right\} + ||\mathbf{m}_N||^2 \end{aligned} \tag{26}$$

Inserting Eq. 26 into Eq. 25 and integrating gives Eq. 15. □

Lemma 2. *For the generative model in Eqs. 3-6, the parametrized close-form expression of $q^*(\alpha)$ in Eq. 7 is given in Eq. 16.*

Proof.

$$\begin{aligned} \ln q^*(\alpha) &= E_{\mathbf{w},\tau}\{\ln p(\mathbf{y}, \mathbf{w}, \tau, \Phi, \alpha)\} + \text{const.} \\ &= E_{\mathbf{w},\tau}\{\ln p(\mathbf{w}, \tau|\alpha)\} + \ln p(\alpha) + \text{const.} \end{aligned} \tag{27}$$

The first equality is Eq. 9 and the second one follows from Eq. 3 by keeping only terms that depend on α.
From Eq. 21

$$E_{\mathbf{w},\tau}\{\ln p(\mathbf{w}, \tau|\alpha)\} = \frac{D}{2}\ln\alpha - \frac{\alpha}{2}E_{\mathbf{w},\tau}\{\tau||\mathbf{w}||^2\} + \text{const.} \tag{28}$$

From Eq. 6

$$p(\alpha) = \frac{1}{\Gamma(c_0)} d_0^{c_0} \alpha^{c_0 - 1} \exp(-d_0\alpha)$$

then

$$\ln p(\alpha) = (c_0 - 1)\ln\alpha - d_0\alpha + \text{const.} \tag{29}$$

Replacing Eqs. 28 and 29 into Eq. 27 we obtain

$$\begin{aligned} \ln q^*(\alpha)) &= \left(c_0 + \frac{D}{2} - 1\right)\ln\alpha - \left(\frac{E_{\mathbf{w},\tau}\{\tau||\mathbf{w}||^2\}}{2}\} + d_0\right)\alpha + \text{const.} \\ &= \mathrm{Gam}(\alpha|c_N, d_N) \end{aligned}$$

with c_N and d_N given in Eqs. 17 and 18, respectively. □