Empirical Bayes Estimate of Covariance for Multivariate Normal Distribution

Emt
CS,UBC

January 29, 2009 (Modified on July 31, 2009)

Abstract

In this note, I derive an empirical Bayes estimate for covariance matrix of multivariate normal distribution. I assume an inverse-Wishart prior on covariance and integrate it out to obtain the marginal likelihood. I estimate the hyperparameters of inverse-Wishart distribution by maximizing the marginal likelihood. Estimate of inverse scale matrix can be found in closed form, while we can obtain the estimate of degree of freedom by iteratively minimizing a convex function. Finally posterior distribution of covariance matrix is obtain by plugging the estimated value of hyperparameters.

Let us say that we have n i.i.d. observations $y_{1:n}$ from a multivariate normal distribution $\mathcal{N}_d(0,\Sigma)$, where d is the number of variables. We assume an inverse-Wishart (IW) prior on Σ,

$$p(\Sigma|\delta,\Phi) = h(\delta,\Phi)|\Sigma|^{-\frac{1}{2}(\delta+d+1)}\exp\left[-\frac{1}{2}\mathrm{tr}(\Phi\Sigma^{-1})\right] \tag{1}$$

where $\delta > 0$, Φ is a $d\times d$ positive definite matrix, and h is normalizing constant for inverse-Wishart distribution, given as follows,

$$h(\delta,\Phi) = \frac{|\Phi|^{\delta/2}}{2^{d\delta/2}\Gamma_d(\delta/2)} \tag{2}$$

where $\Gamma_d(a)$ is the multivariate Gamma function,

$$\Gamma_d(a) = \pi^{d(d-1)/4}\prod_{i=1}^{d}\Gamma\left(a-\frac{i-1}{2}\right) \tag{3}$$

We now derive an estimator for Σ using an empirical Bayes procedure.

The marginal likelihood is given as follows,

$$\begin{aligned}
p(y|\delta,\Phi) &= \int p(y|\Sigma)p(\Sigma|\delta,\Phi)d\Sigma && (4)\\
&= \int h(\delta,\Phi)|\Sigma|^{-\frac{1}{2}(\delta+d+1)}\exp\left[-\frac{1}{2}\mathrm{tr}(\Phi\Sigma^{-1})\right](2\pi)^{-nd/2}|\Sigma|^{-n/2}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\sum_{i=1}^{n}y_iy_i^T\Sigma^{-1}\right)\right] && (5)\\
&= (2\pi)^{-nd/2}h(\delta,\Phi)\int|\Sigma|^{-(\delta+n+d+1)/2}\exp\left[-\frac{1}{2}\mathrm{tr}\left\{\left(\Phi+\sum_{i=1}^{n}y_iy_i^T\right)\Sigma^{-1}\right\}\right]d\Sigma && (6)\\
&= (2\pi)^{-nd/2}\frac{h(\delta,\Phi)}{h(\delta^*,\Phi^*)}\int\mathcal{IW}(\Sigma|\delta^*,\Phi^*)d\Sigma && (7)\\
&= (2\pi)^{-nd/2}\frac{h(\delta,\Phi)}{h(\delta^*,\Phi^*)} && (8)
\end{aligned}$$

where $\delta^* = \delta + n$ and $\Phi^* = \Phi + S_y$, and $S_y = \sum_i y_i y_i^T$. Taking logarithm and negating it, we get the following cost function with respect to Φ and δ,

$$\begin{aligned}
\mathcal{L}(\Phi,\delta) &\equiv -\log p(y|\delta,\Phi) && (9)\\
&= \frac{nd}{2}\log(2\pi) - \log h(\delta,\Phi) + \log h(\delta^*,\Phi^*) && (10)\\
&= -\frac{\delta}{2}\log|\Phi| + \frac{\delta^*}{2}\log|\Phi^*| + \frac{d\delta}{2}\log 2 - \frac{d\delta^*}{2}\log 2 + \log\frac{\Gamma_d(\delta/2)}{\Gamma_d(\delta^*/2)} + \text{const} && (11)\\
&= -\frac{\delta}{2}\log|\Phi| + \frac{\delta^*}{2}\log|\Phi^*| + \log\frac{\Gamma_d(\delta/2)}{\Gamma_d(\delta^*/2)} + \text{const} && (12)
\end{aligned}$$

Differentiating and setting to zero,

$$-\frac{\delta}{2}\Phi^{-1} + \frac{\delta^*}{2}(\Phi+S_y)^{-1} \Rightarrow \Phi_{opt} = \frac{\delta}{n}S_y \qquad (13)$$

We can now substitute this in the cost function and optimize over δ,

$$\begin{aligned}
\tilde{\mathcal{L}}(\delta) &\equiv -\frac{\delta}{2}\log|\frac{\delta}{n}S_y| + \frac{\delta^*}{2}\log|(1+\frac{\delta}{n})S_y| + \log\frac{\Gamma_d(\delta/2)}{\Gamma_d(\delta^*/2)} + \text{const} && (14)\\
&= -\frac{\delta}{2}\log\left[\left(\frac{\delta}{n}\right)^d |S_y|\right] + \frac{\delta^*}{2}\log\left[\left(\frac{\delta+n}{n}\right)^d |S_y|\right] + \log\frac{\Gamma_d(\delta/2)}{\Gamma_d(\delta^*/2)} + \text{const} && (15)\\
&= -\frac{\delta d}{2}\log\frac{\delta}{n} + \frac{\delta+n}{2}d\log\frac{\delta+n}{n} + \log\frac{\Gamma_d(\delta/2)}{\Gamma_d(\delta^*/2)} + \text{const} && (16)\\
&= \frac{\delta d}{2}\log\frac{\delta+n}{\delta} + \frac{nd}{2}\log\frac{\delta+n}{n} + \log\frac{\Gamma_d(\delta/2)}{\Gamma_d(\delta^*/2)} + \text{const} && (17)
\end{aligned}$$

Differentiate and simplify,

$$\begin{aligned}
\frac{\partial}{\partial\delta}\tilde{\mathcal{L}}(\delta) &= \frac{d}{2}\log\frac{\delta+n}{\delta} + \frac{\delta d}{2}\frac{\delta}{\delta+n}\left(-\frac{n}{\delta^2}\right) + \frac{nd}{2}\frac{n}{\delta+n}\frac{1}{n} + \frac{\partial}{\partial\delta}\log\frac{\Gamma_d(\delta/2)}{\Gamma_d(\delta^*/2)} && (18)\\
&= \frac{d}{2}\log\frac{\delta+n}{\delta} + \frac{\partial}{\partial\delta}\log\frac{\Gamma_d(\delta/2)}{\Gamma_d(\delta^*/2)} && (19)\\
& && (20)
\end{aligned}$$

Expanding the multivariate gamma expression in the last term we get the following,

$$\begin{aligned}
\log\frac{\Gamma_d(\delta/2)}{\Gamma_d(\delta^*/2)} &= \frac{d(d-1)}{4}\log\pi + \sum_{i=1}^{n}\log\Gamma\left(\frac{\delta-i-1}{2}\right) - \frac{d(d-1)}{4}\log\pi - \sum_{i=1}^{n}\log\Gamma\left(\frac{\delta+n-i-1}{2}\right) && (21)\\
&= \sum_{i=1}^{n}\log\frac{\Gamma\left(\frac{\delta-i-1}{2}\right)}{\Gamma\left(\frac{\delta+n-i-1}{2}\right)} && (22)\\
&= -\sum_{i=1}^{n}\sum_{j=1}^{n}\log\left(\frac{\delta+n-i-1-j}{2}\right) && (23)
\end{aligned}$$

Differentiating we get the following,

$$\frac{\partial}{\partial\delta}\log\frac{\Gamma_d(\delta/2)}{\Gamma_d(\delta^*/2)} = -\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{\delta+n-i-1-j} \qquad (24)$$

This gives us the final gradient which we set to zero,

$$\frac{\partial}{\partial\delta}\tilde{\mathcal{L}}(\delta) = \frac{d}{2}\log\frac{\delta+n}{\delta} - \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{\delta+n-i-1-j} = 0 \qquad (25)$$

There is no closed form expression for the solution of above equation, but we can use an iterative optimization algorithm. Note that the objective function in Eq. 17 is convex (can be verified by taking second derivative which is positive) and standard solvers can be used to optimize this function.

Substituting the values into the posterior, we get the distribution over Σ,

$$\begin{aligned} p(\Sigma|y, \delta_{opt}, \Phi_{opt}) &= \mathcal{IW}\left(\delta_{opt} + n, \Phi_{opt} + S_y\right) && (26) \\ &= \mathcal{IW}\left(\Sigma|\delta_{opt} + n, \frac{\delta_{opt} + n}{n} S_y\right) && (27) \end{aligned}$$