More on Linear Models

PUBH 8442: Bayes Decision Theory and Data Analysis

Eric F. Lock
UMN Division of Biostatistics, SPH
elock@umn.edu

03/11/2024

Example: Body Fat

- Use iid normal prior for $\beta_i's$:

$$\beta \sim \text{Normal}(0, \sigma^2\tau^2 I)$$

- $IG(a, b)$ prior for σ^2
- Choose a and b so that

$$E(\sigma^2) = 10, \;\; \text{Var}(\sigma^2) = 100$$

$\rightarrow a = 3, b = 20.$

Choice of τ^2

- How to choose the hyper-parameter τ^2?
- Could use subjective intuition, or put a prior on τ^2
- Alternatively, choose τ^2 "empirically"

$$\hat{\tau}^2 = \underset{\tau^2}{\operatorname{argmax}}\ p(\mathbf{y} \mid \tau^2)$$

- Fix $\tau^2 = \hat{\tau}^2$ and proceed as if it were known.
- An "empirical Bayes" approach

Normal-inverse-gamma regression marginal

- For the normal-inverse-gamma regression model with $\beta_i's$ iid with variance τ^2,

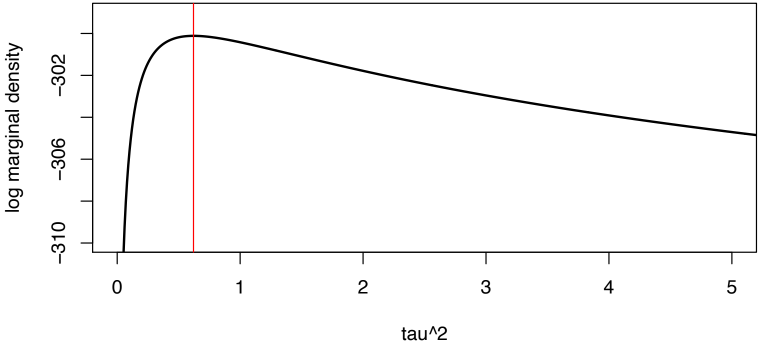
$$p(\mathbf{y} \mid \sigma^2, \tau^2) = N(\mathbf{0}, \sigma^2(I + \tau^2(XX^T))$$

- The marginal for $\mathbf{y}$ is a multivariate t-distribution with $2a$ df:

$$p(\mathbf{y} \mid \tau^2) = MVT_{2a}\left(\mathbf{0}, \frac{b}{a}(I - \tau^2 XX^T\right).$$

The multivariate t density is available in r under the 'mvtnorm' package

Example: Body Fat



http://www.ericfrazerlock.com/More_on_Linear_Models_Rcode1.r

- $p(\mathbf{y} \mid \tau^2)$ is maximized at $\hat{\tau}^2 = 0.62$

Example: Body Fat

- For hyperparameters $\tau^2 = 0.62$, $a = 3$ and $b = 20$, estimate

$$\tilde{\beta} = (X^T X + \frac{1}{0.62} I)^{-1} (X^T \mathbf{y})$$

- The marginal posterior for σ^2 is

$$p(\sigma^2 \mid \mathbf{y}) = IG\left(a_n, b_n\right)$$

where $a_n = 3 + \frac{100}{2}$ and
$b_n = 20 + \frac{1}{2}[\mathbf{y}^T \mathbf{y} - \tilde{\beta}^T (X^T X + \frac{1}{0.62} I)^{-1} \tilde{\beta}]$

- The marginal posterior for β is a multivariate t-distribution

$$\frac{\beta_i - \tilde{\beta}_i}{\sqrt{\frac{b_n}{a_n}((X^T X + \frac{1}{0.62} I)^{-1})_{ii}}} \sim t_{106}.$$

Example: Body Fat

- Estimates and 95% credible intervals for β_i's:

Variable	$\tilde{\beta}_i$	95% credible interval
Age	1.193	(0.157, 2.229)
Weight	-0.465	(-4.127, 3.197)
Height	-0.380	(-1.602, 0.842)
Neck	-0.143	(-1.749, 1.463)
Chest	-0.642	(-2.999, 1.714)
Abdomen	8.704	(6.286, 11.123)
Ankle	0.042	(-1.300, 1.384)
Biceps	0.207	(-1.104, 1.519)
Wrist	-2.176	(-3.683, -0.668)

http://www.ericfrazerlock.com/More_on_Linear_Models_Rcode1.r

Example: Body Fat (cont.)

- Recall: Predict centered $Y = BF\%$ from 9 standardized predictors X from 100 adult males:

$$\mathbf{y} = \boldsymbol{\beta} X + \boldsymbol{\epsilon}$$

where $\epsilon \sim \text{Normal}(\mathbf{0}, \sigma^2 I)$

- Considered uninformative model M_1:

$$p(\beta, \sigma^2) = \frac{1}{\sigma^2}$$

- And "shrinkage" model M_2:

$$\sigma^2 \sim IG(3, 20),$$

$$\beta \sim \text{Normal}(0, \sigma^2 \tau^2 I)$$

  - Fixed τ^2 empirically:

$$\hat{\tau}^2 = \underset{\tau^2}{\text{argmax}}\ p(\mathbf{y} \mid \tau^2) = 0.62.$$

Example: Body Fat (cont.)

- Would like to compare models M_1, M_2.
- Hard to use the Bayes factor
  - Marginal distribution not defined for M_1
  - Inherent bias because τ^2 chosen to maximize marginal under M_2.
- Recall instead the log psuedo marginal likelihood:

$$LPML = \log \prod_{i=1}^{n} p(y_i \mid \mathbf{y}_{(i)})$$

where $p(y_i \mid \mathbf{y}_{(i)})$ is the conditional predictive distribution for y_i.

Example: Body Fat (cont)

- Under M_1 the conditional predictive distribution is a non-central t

$$p(y_i \mid \mathbf{y}_{(i)}) = T_{90}\left(\mathbf{x}_i\hat{\beta}_{(i)}, s^2(1 + \mathbf{x}_i(X_{(i)}^T X_{(i)})^{-1}\mathbf{x}_i^T)\right)$$

- Under M_2 the conditional predictive distribution is a non-central t

$$p(y_i|\mathbf{y}_{(i)}) = T_{105}\left(\mathbf{x}_i\hat{\beta}_{(i)}, \frac{b_{n(i)}}{a_{n(i)}}(1 + \mathbf{x}_i(X_{(i)}^T X_{(i)} + (1/\tau^2)I)^{-1}\mathbf{x}_i^T)\right)$$

  - Compute for fixed $\tau^2 = 0.62$
  - Alternatively, recompute the MLE $\hat{\tau}^2$ based on $\mathbf{y}_{(i)}$ for $p(y_i \mid \mathbf{y}_{(i)})$ (adjusts for over-fitting).

Example: Body Fat (cont)

- Compute
  - *LPML* for $M1$ is -288.774
  - *LPML* for $M2$, with overall $\hat{\tau}^2$, is -288.6188
  - *LPML* for $M2$, re-estimating $\hat{\tau}^2$ for each sample, is -288.6155 http://www.ericfrazerlock.com/More_on_Linear_Models_Rcode1.r
- Suggests there (may) be a slight benefit for using the shrinkage model
- Suggests that overfitting τ^2 is not a concern.