Model uncertainty and variable selection in Bayesian lasso regression

Chris Han 2009

Outline

Outline

1. Assessing model uncertainty using marginal likelihood
   - The purpose of this paper
   - Apply to Data set
2. Model uncertainty in higher dimensions

The purpose of this paper

- Introduce analytic and computation approaches for handling model uncertainty under the Bayesian lasso regression.
- How to find posterior marginal inclusion probabilities for the Bayesian lasso.
  - Exact ML vs. MCMC
  - Application to dataset

Bayesian lasso

- Park and Casella(2008) consider the Bayesian lasso regression model
$$y|\beta,\sigma^2 \sim N\left(X\beta,\sigma^2 I\right)$$
- The double-exponential prior for p-vector of regression coefficients β
$$\beta_j|\sigma^2,\tau \overset{iid}{\sim} DE\left(\tau/\sigma\right), \qquad j=1,\dots,p.$$
where y is an n-vector of observations and X is an $n\times p$ matrix of predictor variables
- $DE\left(\tau/\sigma\right)$ is the double-exponential distribution with density function
$$p\left(\beta_j|\tau,\sigma^2\right)=\frac{\tau}{2\sigma}e^{-\tau|\beta_j|/\sigma}.$$

The marginal likelihood for model γ

- For given model γ with k_γ predictor variables, the key to assessing model uncertainty is the ability to evaluate

$$
\begin{aligned}
m_\gamma\left(y|\sigma^2,\tau\right) &= \int N\left(y|X_\gamma\beta_\gamma,\sigma^2 I_n\right)\pi\left(\beta_\gamma|\sigma^2,\tau\right)d\beta_\gamma \\
&= \int \left(2\pi\sigma^2\right)^{-n/2} exp\left(-\frac{1}{2\sigma^2}\left(y-X_\gamma\beta_\gamma\right)^T\left(y-X_\gamma\beta_\gamma\right)\right) \\
&\times \left(\frac{\tau}{2\sigma}\right)^{k_\gamma} exp\left(-\tau\left|\left|\beta_\gamma\right|\right|_1\right)d\beta_\gamma
\end{aligned}
$$

where $||\beta||_1$ is the L_1-norm of β and we assume σ^2 and τ are known values for now.

Exact marginal likelihood calculation

- The integral in the previous slide can be re-expressed as follows:

$$m_\gamma\left(y|\sigma^2,\tau\right)=\omega_\gamma\left(\frac{\tau}{2\sigma}\right)^{k_\gamma}N\left(y|0,\sigma^2 I_n\right)$$

where $k_\gamma=\sum_{l=1}^{p}\gamma_l$ is the number of variables included in model γ and

$$\omega_\gamma=\sum_{z\in Z_{k_\gamma}}\frac{P\left(z,\left(X_\gamma^T X_\gamma\right)^{-1}\left(X_\gamma^T y-\tau\sigma z\right),\sigma^2\left(X_\gamma^T X_\gamma\right)^{-1}\right)}{N\left(0|\left(X_\gamma^T X_\gamma\right)^{-1}\left(X_\gamma^T y-\tau\sigma z\right),\sigma^2\left(X_\gamma^T X_\gamma\right)^{-1}\right)}.$$

when $k_\gamma=0$, then the weight defined to be 1.

Posterior probability of the regression model

- If we assume that the prior distribution on the model space is

$$\gamma_j \overset{iid}{\sim} Bernoulli\,(\rho = 0.5)$$

  all models are a priori equally likely.

- Thus, the posterior probability of the regression model are simply normalized marginal likelihood

$$\pi\,(\gamma|y) = \frac{m_\gamma\,(y)\,\pi\,(\gamma)}{\sum_{\gamma' \in \Gamma} m_{\gamma'}\,(y)\,\pi\,(\gamma')}$$
$$\underset{=}{aspt.}\ \frac{m_\gamma\,(y)}{\sum_{\gamma' \in \Gamma} m_{\gamma'}\,(y)}$$

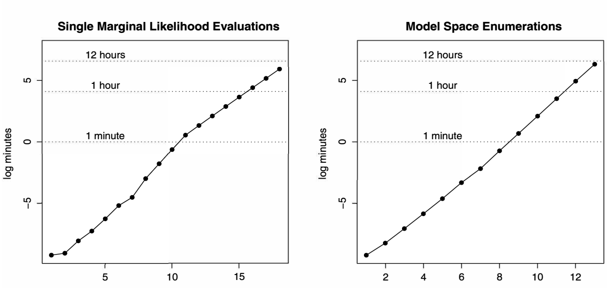
Outline

Computational limitations

- To calculate the marginal likelihood, we need 2^p marginal likelihood calculation and for given model size k_γ, 2^{k_γ} ratios must be computed. Thus,

$$\sum_{k=0}^{p} \binom{p}{k} 2^k = 3^p$$

ratios must be computed for the calculation.



Model uncertainty in higher dimensions

- Assuming $\gamma_j|\rho \overset{iid}{\sim} Bernoulli(\rho)$

$$\pi\left(\beta_j|\sigma^2,\tau,\rho\right) = (1-\rho)\,\delta_0(\beta_j) + \rho\,\frac{\tau}{2\sigma}e^{-\tau|\beta_j|/\sigma}$$

Equivalently,

$$\beta_j|\gamma_j \sim \begin{cases} point\ mass\ at\ 0 & \gamma_j = 0 \\ double-exponential & \gamma_j = 1 \end{cases}$$

Then model indicator γ is embedded in β.

Full conditional distribution of β_j

- An iteration of the Gibbs sampler cycles through the full conditional distributions $\beta_j|\beta_{-j},\sigma^2,\tau,y, \quad j=1,\ldots,p$

$$\pi\left(\beta_j|\beta_{-j},\sigma^2,\tau,y\right)=\phi_{0j}\delta_0\left(\beta_j\right)+\left(1-\phi_{0j}\right)\times \left\{\phi_j N^+\left(\beta_j|\mu_j^+,s_j^2\right)+\left(1-\phi_j\right)N^-\left(\beta_j|\mu_j^-,s_j^2\right)\right\}$$

where

$$\begin{cases}\phi_{0j}\equiv Pr\left(\beta_j=0|\beta_{-j},\sigma^2,\tau,\rho,y\right)\\ \phi_j\equiv\left\{\frac{\Phi\left(\mu_j^+/s_j\right)}{N\left(0|\mu_j^+,s_j^2\right)}\right\}/\left\{\frac{\Phi\left(\mu_j^+/s_j\right)}{N\left(0|\mu_j^+,s_j^2\right)}+\frac{\Phi\left(-\mu_j^-/s_j\right)}{N\left(0|\mu_j^-,s_j^2\right)}\right\}\end{cases}$$

- Full conditional distribution is a three component mixture: a point mass at 0, a positive normal distribution and a negative normal distribution with weights $\phi_{0j},\left(1-\phi_{0j}\right)\phi_j$ and $\left(1-\phi_{0j}\right)\left(1-\phi_j\right)$ respectively.

Full conditional distributions of σ^2, τ, ρ

- Prior:

$$\begin{aligned} \sigma^2 &\sim IG(a,b) \\ \tau &\sim Gamma(r,s) \\ \rho &\sim Beta(g,h) \end{aligned}$$

- Full conditionals

$$\begin{aligned} \pi\left(\sigma^2|\beta,\rho,y\right) &\propto \left(\sigma^2\right)^{-(a^*+1)} exp\left(-b^*/\sigma^2 - \tau\,||\beta||_1/\sigma\right) \\ \tau|\beta,\sigma^2,y &\sim Gamma\left(k_\gamma + r, \sigma^{-1}\,||\beta||_1 + s\right) \\ \rho|\beta,\sigma^2,\tau,y &\sim Beta\left(g + k_\gamma, h + p - k_\gamma\right) \end{aligned}$$

Computational results

Fixed Parameters	Method	AGE	SEX	BP	S1	S2	S3	S4	S6	Time
$\sigma^2 = 1, \tau = 4.25,$	ML	.192	.776	.983	.519	.372	.696	.402	.251	8.57
$\rho = 0.5$	MCMC	.192	.775	.983	.560	.372	.695	.401	.251	1.32
$\sigma^2 = .492, \tau = 4.25,$	ML	.191	.991	1.000	.658	.435	.797	.473	.307	8.19
$\rho = 0.5$	MCMC	.191	.991	1.000	.658	.436	.796	.473	.307	1.39
$\tau = 4.25, \rho = 0.5$	ML	.191	.987	1.000	.650	.432	.795	.470	.304	*
	MCMC	.191	.990	1.000	.660	.435	.793	.476	.307	1.51
$\rho = 0.5$	MCMC	.130	.989	1.000	.659	.429	.696	.414	.217	1.52
	MCMC	.381	.995	1.000	.816	.658	.781	.651	.503	1.72

Table: Posterior marginal inclusion probabilities

Additional approaches to model space MCMC

- Construct a Markov chain directly on the model indicator γ by marginalizing over β.

$$Pr\left(\gamma_j = 1 | \gamma_{-j}, \sigma^2, \tau, y\right) = \left(1 + \frac{1-\rho}{\rho} \times \frac{m_{\gamma 0}\left(y|\sigma^2, \tau\right)}{m_{\gamma 0}\left(y|\sigma^2, \tau\right)}\right)^{-1}$$

- Advantage: β is never sampled.
- Disadvantage: Computing of marginal likelihoods are expensive.

Comments

- This paper address the regression model uncertainty for the Bayesian LASSO.
- In the MCMC method, model indicator γ is embedded in β instead of use γ explicitly.
- By setting a non-zero probability of β_j being zero in prior distribution, it allows β_j to be exactly zero the sample of posterior distribution.
- A mixture of point mass and truncated normal distributions are used for posterior distribution of β.

Comments

- The paper didn't show the equivalence of $\beta_j = 0$ and $\gamma_j = 0$ and $\beta_j \neq 0$ and $\gamma_j = 1$.
- The problem of choosing the value of tuning parameter τ remains to be solved.
- Although the algorithm gives the marginal probabilities of each β_j being non-zero, what set of predictors should be in the model is still not clear.

Thank you!