Bayesian linear Regression

Remy

1/9/2020

Introduction

Consider the following linear regression:
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \epsilon$, which can be written as:

$$y_i = \boldsymbol{x_i^T \beta} + \epsilon_i$$

or

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

$\epsilon_i \sim N(0, \sigma^2) \longleftrightarrow \boldsymbol{\epsilon} \sim N(0, \sigma^2 I_n)$

Normal likelihood

$y_i \sim N(x_i^T \beta, \sigma^2)$
$\boldsymbol{y} \sim N(\boldsymbol{X\beta, \sigma^2 I_n})$
Therefore, the distribution of the date is a multivariate normal with density

$$[y_i|\beta, \sigma^2] \propto (\sigma^2)^{-1/2} \exp\left\{-\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - x_i^T\beta)\right\}$$

Which is equivalent to

$$[\boldsymbol{y}|\boldsymbol{\beta, \sigma^2}] \propto (\sigma^2)^{-n/2} \exp\left\{\boldsymbol{(y - X\beta)^T \Sigma^{-1} (y - X\beta)}\right\}$$

with $\Sigma = \sigma^2 I_n$

Prior and posterior distributions

Given a normal likelihood, if we consider a prior for the normal variance $\sigma^2 \sim IG(a, b)$ to be an inverse gamma, and place a normal or a noninformative prior on the linear regression coefficients $\boldsymbol{\beta} = \beta_0, \beta_1, \ldots, \beta_p$, then the we obtain full conditional posteriors

- $\boldsymbol{\beta} \sim N(\boldsymbol{\mu_0, \Sigma_0})$
- Given the data: $\boldsymbol{y} \sim N(\boldsymbol{X\beta, \sigma^2 I_n})$
- $\boldsymbol{\beta|y, \sigma^2} \sim N(\boldsymbol{\mu_n, \Sigma_n})$
- $\sigma|\alpha, \gamma \sim IG(a, b = rate)$

Considering an noninformative prior on $\boldsymbol{\beta}$, we can use the ordinary least square results:

- $\boldsymbol{\hat{\beta} = (X^T X)^{-1} X^T y}$,

- $var(\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}) = \sigma^2(X^TX)^{-1}$.

It can be seen that $\hat{\boldsymbol{\beta}} = \underbrace{\boldsymbol{(X^TX)^{-1}X^T}}_{\boldsymbol{constant}}\boldsymbol{y}$ is a linear combination of multivariate normal random variables, which means that $\hat{\boldsymbol{\beta}}$ will also have a normal distribution. Therefore

$$\hat{\beta} \sim \boldsymbol{N(}\beta\boldsymbol{, \sigma^2(X^TX)^{-1})}$$

or

$$\beta \sim \boldsymbol{N(}\hat{\beta}\boldsymbol{, \sigma^2(X^TX)^{-1})}$$

Thus, $\hat{\boldsymbol{\beta}}$ is randomly distributed around its least square estimate with $\sigma^2(X^TX)^{-1}$ determining its randomness.

Posterior for σ^2

Given $\boldsymbol{\beta}$, the posterior for σ^2 can be calculated as follows:
$\sigma^2|\boldsymbol{y}, \boldsymbol{\beta} \sim N(\boldsymbol{X\beta, \sigma^2 I_n})IG(a, b)$, where

- $a_n = \frac{n}{2} + a$ and
- $b_n = \frac{1}{2\sigma^2}\boldsymbol{(y - X}\beta\boldsymbol{)^T(y - X\beta)} + b$

Posterior for $\boldsymbol{\beta}$ with a normal prior

$\beta|\boldsymbol{y, \sigma^2} \sim N(\mu_0, \Sigma_0)N(\boldsymbol{X\beta, \sigma^2 I_n}) \sim N(\boldsymbol{\mu_n, \Sigma_n})$, with

- $\Sigma_n = \left(\Sigma_0^{-1} + \sigma^2(X^TX)\right)^{-1}$
- $\mu_n = \Sigma_n\left(\Sigma_0^{-1}\mu_0 + \sigma^2(X^TX)\hat{\beta}\right)^{-1}$

Generally, the prior for $\boldsymbol{\beta}$ is $N(\boldsymbol{0, \sigma^2(X^TX)^{-1}})$