Bayesian Linear Regression

Efim Abrikosov

January 1, 2024

1 Framework

1.1 Notations

$$y = x \cdot \boldsymbol{\beta} + \varepsilon,\ \varepsilon \sim \mathcal{N}(0, \sigma^2) \tag{1.1}$$

1. Individual observations $(x, y) \in \mathbb{R}^k \times \mathbb{R}$
2. Observed data $(X, Y) \in \mathbb{R}^{n \times k} \times \mathbb{R}^n$
3. Linear regression weights $\boldsymbol{\beta} \in \mathbb{R}^k$
4. Observation error variance σ^2

1.2 Model Assumptions

1. Observations x have full rank
2. Observation errors are independent, normally distributed with mean zero and variance σ^2
3. Relation 1.1 holds
4. Error variance σ^2 is either known, or its prior distribution is inverse gamma distribution with parameters $a_0,\ b_0$
5. $\boldsymbol{\beta}$ has a prior distribution $\mathcal{N}(\boldsymbol{\beta}_0, \sigma^2 \boldsymbol{\Sigma}_0)$

2 Known Observation Variance with Linear Weights Prior

2.1 Summary of Results

Posterior distribution of $\boldsymbol{\beta}$:

$$f(\boldsymbol{\beta} \mid Y, X) \sim \mathcal{N}(\boldsymbol{\beta}_1, \sigma^2 \boldsymbol{\Sigma}_1) \tag{2.1}$$

$$\boldsymbol{\Sigma}_1^{-1} = \boldsymbol{\Sigma}_0^{-1} + X^T X \tag{2.2}$$

$$\boldsymbol{\beta}_1 = \boldsymbol{\Sigma}_1 \left(X^T X \widehat{\boldsymbol{\beta}} + \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\beta}_0 \right) \tag{2.3}$$

$$\widehat{\boldsymbol{\beta}} = \left(X^T X \right)^{-1} X^T Y \tag{2.4}$$

Posterior prediction distribution $\widehat{y} \equiv x \cdot \boldsymbol{\beta}$:

$$f(\widehat{y} \mid Y, X, x) \sim \mathcal{N} \left(x\boldsymbol{\beta}_1, \sigma^2 x \left(X^T X \right)^{-1} x^T \right) \tag{2.5}$$

Posterior observation distribution $\widehat{y} \equiv x \cdot \boldsymbol{\beta} + e$:

$$f(y \mid Y, X, x) \sim \mathcal{N} \left(x\boldsymbol{\beta}_1, \sigma^2 \left(1 + x \left(X^T X \right)^{-1} x^T \right) \right) \tag{2.6}$$

Confidence interval, or t-test, for a fixed value $\underline{\boldsymbol{\beta}}$ and a linear constraint $\boldsymbol{c} \in \mathbb{R}^k$:

$$\frac{\boldsymbol{c}(\underline{\boldsymbol{\beta}} - \boldsymbol{\beta}_1)}{\sigma \sqrt{(\boldsymbol{c} \boldsymbol{\Sigma}_1 \boldsymbol{c}^T)}} \sim \mathcal{N}(0, 1) \tag{2.7}$$

Joint f-test for a set of linear constraints $\boldsymbol{C} \in \mathbb{R}^{l \times k}$

2.2 Uninformative Prior

3 Conjugate Priors For Observation Variance and Linear Weights