Lab 10: Bayesian linear regression

Olivier Binette

Friday October 23, 2020

Agenda

- Announcements
- Review of Bayesian linear regression
- Lab 10

Announcements

Announcements

- Very stressful week and we're thinking about you.
- We're trying to give the best learning experience despite everything.
  - Feedback is appreciated (box is open until November 18): https://app.suggestionox.com/r/OlivierSuggestionBox
- Let your TAs know if you are facing any issue.
  - I do not grade and I can bring things up anonymously.

Review of Bayesian linear regression

Review of Bayesian linear regression

Model:

$$Y \mid X, \beta, \sigma^2 \sim MVN(X\beta, \sigma^2 I_n)$$

Prior (known variance):

$$\beta \sim MVN(\beta_0, \Sigma_0)$$

Review of Bayesian linear regression

Posterior distribution:

$$\beta \mid Y, X, \sigma^2 \sim MVN(\beta_n, \Sigma_n)$$

where

$$\beta_n = \left(\Sigma_0^{-1} + X^T X/\sigma^2\right)^{-1} \left(\Sigma_0^{-1}\beta_0 + X^T Y/\sigma^2\right),$$

$$\Sigma_n = \left(\Sigma_0^{-1} + X^T X/\sigma^2\right)^{-1}.$$

Lab 10 (linear regression)

Lab 10

Exercice 9.1 in Hoff's book.

https://github.com/resteorts/modern-bayes/blob/master/labs/10-linear-regression/11-linear-regression_v2.pdf

Lab 10

The problem:

- Data on four swimmers describing the evolution of their lap time up until now.
- We want to (1) predict their lap time two weeks from now and (2) predict who will be the fastest swimmer.

Lab 10

More precisely:

For each swimmer, we have lap times $Y_i = (Y_{i,1}, Y_{i,2}, \ldots, Y_{i,6})$ associated with a week $x_i \in \{1, 3, 5, 7, 9, 11\}$.

The approach:

We'll independently fit linear models for each swimmer.

Lab 10

Let's first load the data.

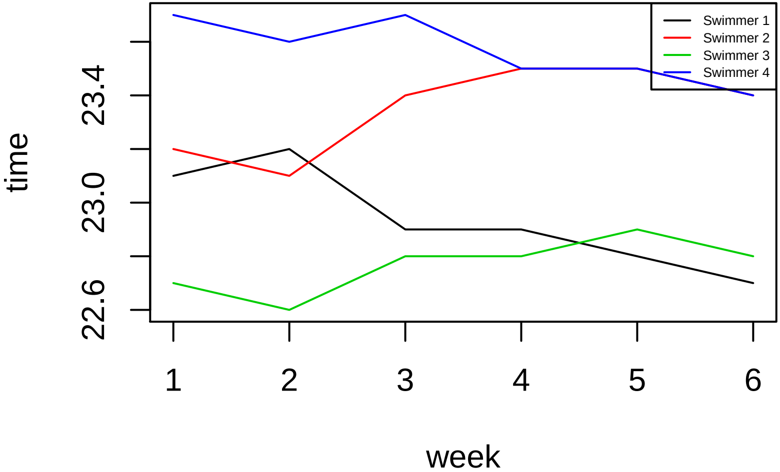
```
data = read.table(url("https://raw.githubusercontent.com/resteor
weeks = c(1,3,5,7,9,11)
```

Lab 10

And let's take a look.

```
plot(weeks, data[1,], ylim=range(data),
     type="l", col=1, ylab="time")
for (i in 2:4) {
  lines(weeks, data[i,], col=i)
}
legend("topright", legend=paste("Swimmer", 1:4),
       lty=1, col=1:4, cex=0.45)
```

Lab 10



Swimmer 1
Swimmer 2
Swimmer 3
Swimmer 4
time
22.6
23.0
23.4
1
2
3
4
5
6
week

Lab 10

Task 1

We will fit a separate linear regression model for each swimmer, with swimming time as the response and week as the explanatory variable. Let $Y_i \in \mathbb{R}^6$ be the 6 recorded times for swimmer $i = 1, 2, 3, 4$. Let

$$X_i = \begin{pmatrix} 1 & 1 \\ 1 & 3 \\ ... & \\ 1 & 9 \\ 1 & 11 \end{pmatrix}$$

be the design matrix for swimmer $i = 1, 2, 3, 4$.

Lab 10

Task 1

Then we use the following linear regression model:

$$\begin{aligned} Y_i \mid \beta_i, \tau_i &\sim \mathcal{N}_6\left(X\beta_i, \tau_i^{-1}\mathcal{I}_6\right) \\ \beta_i &\sim \mathcal{N}_2\left(\beta_0, \Sigma_0\right) \\ \tau_i &\sim \text{Gamma}(a, b). \end{aligned}$$

Derive full conditionals for β_i and τ_i. Assume that β_0, Σ_0, a, b are known.

Lab 10

Solution

1. Using the the result on slide 7, derive the full conditional $\beta_i \mid Y_i, X, \sigma^2$.
2. Derive the full conditional $\tau_i \mid Y_i, X, \beta_i$:

- First write down the likelihood.
- Next write the joint posterior (up to a normalizing constant).
- Finally derive the full conditional.

Lab 10

Task 2

Complete the prior specification by choosing a, b, β_0, and Σ_0. Let your choices be informed by the fact that times for this age group tend to be between 22 and 24 seconds.

Lab 10

Task 3

Code a Gibbs sampler to fit each of the models. For each swimmer i, obtain draws from the posterior predictive distribution for y_i^*, the time of swimmer i if they were to swim two weeks from the last recorded time.

Lab 10

Task 4

The coach has to decide which swimmer should compete in a meet two weeks from the last recorded time. Using the posterior predictive distributions, compute $\Pr\left\{y_i^* = \max\left(y_1^*, y_2^*, y_3^*, y_4^*\right)\right\}$ for each swimmer i and use these probabilities to make a recommendation to the coach.