Advanced Model Predictive Control
Prof. M. Zeilinger
Dr. A. Carron
Dr. L. Hewing
Dr. J. Köhler

Fall 2022

Exercise 8

Bayesian Linear Regression

Alexandre Didier and Jérôme Sieber

1 Exercise

Your colleague told you that a correlation exists between the amount of chocolate consumed per year per capita and the number of Nobel prize winners per 10 million population[1]. While researching the topic, you notice that the number of Nobel laureates seems to be missing for a number of countries, including Switzerland. When asking your colleague about it, he tells you that Bayesian Linear regression can be used to predict the number of Swiss Nobel prize winners given its chocolate consumption. Use the following model to determine the number of Swiss Nobel prize winners per 10 million population.

$$f(x_i) = x_i\theta, \tag{1}$$
$$y_i = f(x_i) + w_i, \tag{2}$$

where x_i is the amount of chocolate consumed per year per capita, y_i is the number of Nobel laureates per 10 million population and the noise $w_i \sim \mathcal{N}(0, \sigma_n^2) = \mathcal{N}(0, 2)$.

The posterior distribution for a prior with non-zero mean $\mathcal{N}(\bar{\theta}_p, \Sigma_p)$ is given by

$$p(\theta|X, y) = \mathcal{N}(\bar{\theta}, A^{-1}), \tag{3}$$

where

$$\bar{\theta} = A^{-1}(\Sigma_p^{-1}\bar{\theta}_p + \frac{1}{\sigma_n^2}Xy), \tag{4}$$
$$A = \frac{1}{\sigma_n^2}XX^T + \Sigma_p^{-1}. \tag{5}$$

1. The United States of America has $y_{\text{USA}} = 11.7$ laureates per 10 million people[2] and consumes $x_{\text{USA}} = 4.4$kg of chocolate per capita per year[3]. Assuming a prior knowledge $p(\theta) = \mathcal{N}(5, 4)$, compute the posterior distribution $p(\theta|x_{\text{USA}}, y_{\text{USA}})$.
2. In the UK, $x_{\text{UK}} = 7.6$kg of chocolate is consumed per capita per year and there are $y_{\text{UK}} = 19.4$ laureates per 10 million people. Recursively update the posterior $p(\theta|X, y)$ with the new data.
3. What is the most likely number of Swiss Nobel prize winners per 10 million population given Switzerland consumes $x_{\text{CH}} = 8.8$kg of chocolate per capita per year? What is the variance of the predictive distribution $p(f_*|x_*, X, y)$?

[1] Franz H. Messerli. Chocolate consumption, cognitive function, and Nobel laureates. *N ENGL J MED*. 2012;367:1562-4.
[2] https://en.wikipedia.org/wiki/List_of_countries_by_Nobel_laureates_per_capita
[3] https://www.statista.com/statistics/819288/worldwide-chocolate-consumption-by-country/

4. You show your colleague your estimate of the number of Swiss Nobel laureates, but he believes it is too low. He suggests using a non-linear model $f(x) = \Phi(x)^T\theta$, with

$$\Phi(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}. \tag{6}$$

and a prior

$$p(\theta) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \frac{1}{2}\mathbb{I}\right). \tag{7}$$

He also gives you four more data points:

Sweden	$x_{\text{SW}} = 6.6$	$y_{\text{SW}} = 30.0$
Austria	$x_{\text{AT}} = 8.1$	$y_{\text{AT}} = 25.1$
France	$x_{\text{FR}} = 4.3$	$y_{\text{FR}} = 10.7$
Japan	$x_{\text{JP}} = 1.2$	$y_{\text{JP}} = 2.2$

Use all the given data to predict the number of Swiss laureates per 10 million population with the provided MATLAB code.