

Bayesian Learning, 6 hp

Computer lab 2

You can use any programming language for the labs, but my hints, help and solutions will be in R.

You are supposed to work and submit your labs in pairs, but do make sure that both of you are contributing.

The **deadline** for this lab is **November 21**.

1. *Linear and polynomial regression*

The data set JapanTemp.dat contains daily temperatures (in Celcius degrees) at some japanese location over the course of a year. The response variable is *temp* and the covariate is

$$time = \frac{\text{the number of days since beginning of year}}{365}.$$

The task is to perform a Bayesian analysis of a quadratic regression

$$temp = \beta_0 + \beta_1 \cdot time + \beta_2 \cdot time^2 + \varepsilon, \varepsilon \stackrel{iid}{\sim} N(0, \sigma^2).$$

- (a) *Determining the prior distribution of the model parameter.* Use the conjugate prior for the linear regression model. Your task is to set the prior hyperparameters μ_0 , Ω_0 , ν_0 and σ_0^2 to sensible values. You may not be an expert in japanese temperatures, and I don't expect any deep expert knowledge, but do come up with something. You may simplify by assuming that Ω_0 is a diagonal matrix, if you want.

[Hint: it may be useful as a preliminary exploratory step to use the `lm()` command. The command `lm(y ~ x + I(x^2))` fits a quadratic model using plain least squares]

- (b) *Check if your prior from a) is sensible.* One way to check if a suggested prior is reasonable is to simulate draws from the joint prior of all parameters and for every draw compute the regression curve. This gives a collection of regression curves, one for each draw from the prior. Do the curves look reasonable? If not, change the prior hyperparameters until the collection of prior regression curves do agree with your prior beliefs about the regression curve.

[Hint: the R package `mvtnorm` will be handy.]

- (c) *Perform sensitivity analysis* with respect to some or all of the prior hyperparameters. That is, what is the effect on the posterior from changing the prior? You don't need to simulate from the posterior in this exercise, use the analytical formulas for the posterior distribution instead.

- (d) Write a program that *simulates from the joint posterior distribution* of $\beta_0, \beta_1, \beta_2$ and σ^2 . Try it out on the model in a).
[Hint: the R package `mvtnorm` will be handy.]
- (e) It is of interest to locate the day with the highest expected temperature (that is, the *time* where $E(temp|time)$ is maximal). Let's call this value \tilde{x} . Use the simulations in d) to simulate from the *posterior distribution of the day with highest temperature*, \tilde{x} .
- (f) Say now that you want to *estimate a polynomial model of order 7*, but you are worried that higher order terms may not be needed, and may ruin your pretty little quadratic model. Suggest a suitable prior that mitigates this potential problem.

MAY BAYES BE WITH YOU!