Mattias Villani and Måns Magnusson
Division of Statistics and Machine Learning
Dept. of Computer and Information Science
Linköping University
mattias.villani@liu.se

2017-04-20

Bayesian Learning, 6 hp

Computer lab 2

- You are strongly recommended to use R for solving the labs since the computer exam 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.
- It is OK to discuss the lab with other student pairs in general terms, but `it is not allowed to share exact solutions`.
- Submit your solutions via LISAM no later than **May** 2 at **midnight**.

1. **Linear and polynomial regression**
The dataset `TempLinkoping.txt` contains daily temperatures (in Celcius degrees) at Malmslätt, Linköping (close to our campus) over the course of the year 2016 (366 days since 2016 was a leap year). The response variable is *temp* and the covariate is

$$time = \frac{\text{the number of days since beginning of year}}{366}.$$

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

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

   (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. I don't expect any deep expert knowledge, just think about the problem and use what knowledge you have. You may simplify by assuming that Ω_0 is a diagonal matrix.
   [Hint: it may be useful as a preliminary exploratory step to use the `lm()` command. The command `lm(temp ~ time + I(time^2))` fits a quadratic model using ordinary 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. And use your Inv-χ^2 simulator from Lab 1.]

(c) Write a program that *simulates from the joint posterior distribution* of β_0, β_1,β_2 and σ^2. Produce a scatter plot of the temperature data and overlay a curve for the posterior mean of the regression function $f(time) = \beta_0 + \beta_1 \cdot time + \beta_2 \cdot time^2$. Also overlay a curve for the lower 5% and upper 95% posterior credible interval for $f(time)$. That is, compute the 95% equal tail posterior probability intervals for every value of *time* and then connect the lower and upper limits of the interval by curves.

(d) It is of interest to locate the *time* 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 c) to simulate from the *posterior distribution of the time with highest temperature*, $\tilde{x}$. [Hint: the regression curve is a quadratic. You can find a simple formula for $\tilde{x}$ given β_0, β_1 and β_2.]

(e) Say now that you want to *estimate a polynomial model of order* 7, but you suspect that higher order terms may not be needed, and you worry about overfitting. Suggest a suitable prior that mitigates this potential problem. You do not need to compute the posterior, just write down your prior. [Hint: the task is to specify μ_0 and Ω_0 in a smart way.]

2. **Posterior approximation for classification with logistic regression**

The dataset `WomenWork.dat` contains $n = 200$ observations (i.e. women) on the following nine variables:

Variable	Data type	Meaning	Role
Work	Binary	Whether or not the woman works	Response
Constant	1	Constant to the intercept	Feature
HusbandInc	Numeric	Husband's income	Feature
EducYears	Counts	Years of education	Feature
ExpYears	Counts	Years of experience	Feature
ExpYears2	Numeric	(Years of experience/10)^2	Feature
Age	Counts	Age	Feature
NSmallChild	Counts	Number of child ≤ 6 years in household	Feature
NBigChild	Counts	Number of child > 6 years in household	Feature

(a) Consider the logistic regression

$$\Pr(y = 1|\mathbf{x}) = \frac{\exp\left(\mathbf{x}^T\beta\right)}{1 + \exp\left(\mathbf{x}^T\beta\right)},$$

where y is the binary variable with $y = 1$ if the woman works and $y = 0$ if she does not. $\mathbf{x}$ is a 8-dimensional vector containing the eight features (including a one for the constant term that models the intercept). Fit the logistic regression using maximum likelihood estimation by the command: `glmModel <- glm(Work ~ 0 + ., data = WomenWork, family = binomial)`. Note how I added a zero in the model formula so that R doesn't add an extra intercept (we already have an intercept term from the `Constant` feature). Note also that a dot (.) in the model formula means to add all other variables in the dataset as features. `family = binomial` tells R that we want to fit a logistic regression (the glm function can fit many other models and data types, for example a regression where the response variable are counts following a Poisson distribution).

(b) Now the fun begins. Our goal is to approximate the posterior distribution of the 8-dim parameter vector β with a multivariate normal distribution

$$\beta|\mathbf{y}, \mathbf{X} \sim N\left(\tilde{\beta}, J_{\mathbf{y}}^{-1}(\tilde{\beta})\right),$$

where $\tilde{\beta}$ is the posterior mode and $J(\tilde{\beta}) = -\frac{\partial^2 \ln p(\beta|\mathbf{y})}{\partial\beta\partial\beta^T}|_{\beta=\tilde{\beta}}$ is the observed Hessian evaluated at the posterior mode. Note that $\frac{\partial^2 \ln p(\beta|\mathbf{y})}{\partial\beta\partial\beta^T}$ is an 8×8 matrix with second derivatives on the diagonal and cross-derivatives $\frac{\partial^2 \ln p(\beta|\mathbf{y})}{\partial\beta_i\partial\beta_j}$ on the off-diagonal. It is actually not hard to compute this derivative by hand, but don't worry, we will let the computer do it numerically for you. Now, both $\tilde{\beta}$ and $J(\tilde{\beta})$ are computed by the `optim` function in R. See my code `https://github.com/STIMALiU/BayesLearnCourse/raw/master/Code/MainOptimizeSpam.zip` where I have coded everything up for the spam prediction example (it also does probit regression, but that is not needed here). I want you to implement you own version of this. You can use my code as a template, but I want you to write your own file so that you understand every line of your code. Don't just copy my code.
Your report should include your code as well as numerical values for $\tilde{\beta}$ and $J_{\mathbf{y}}^{-1}(\tilde{\beta})$ for the `WomenWork` data. Compute an approximate 95% credible interval for the variable NSmallChild. Would you say that this feature is an important determinant of the probability that a women works?

(c) Write a function that simulates from the predictive distribution of the response variable in a logistic regression. Use your normal approximation from 2(b). Use that function to simulate and plot the predictive distribution for the `Work` variable for a 40 year old women, with two children (3 and 9 years old), 8 years of education, 10 years of experience. and a husband with an income of 10.
[Hint: the R package `mvtnorm` will again be handy. And remember my discussion on how Bayesian prediction can be done by simulation.]

HAVE FUN!