

Lab 2

Gustav Sternelöv

18 april 2016

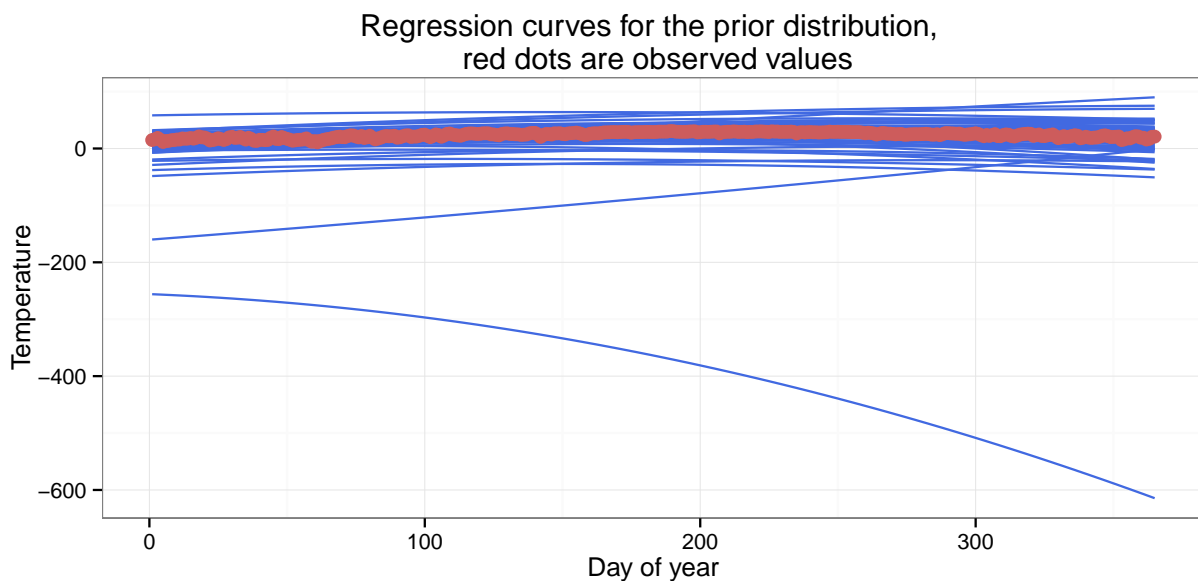
Assignment 2 - Linear and polynomial regression

a-b)

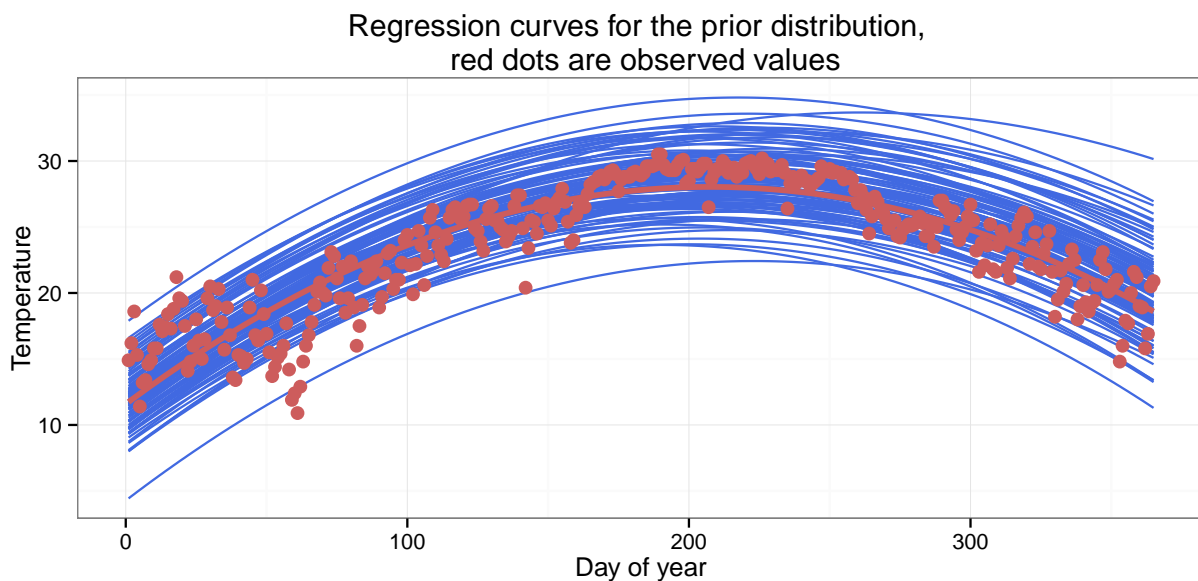
To get some information about what values that can be reasonable to use as priors is a quadratic model using plain least squares fitted. A summary of the results from this model is shown in the output below:

```
##
## Call:
## lm(formula = temp ~ time + I(time^2), data = JapanTemp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.9271 -1.2670  0.4481  1.4537  6.8899
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   11.5826     0.3445   33.62  <2e-16 ***
## time          57.8302     1.5864   36.45  <2e-16 ***
## I(time^2)    -50.8155     1.5321  -33.17  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.182 on 362 degrees of freedom
## Multiple R-squared:  0.7948, Adjusted R-squared:  0.7937
## F-statistic: 701.1 on 2 and 362 DF,  p-value: < 2.2e-16
```

By using the results from above the prior β vector is set to (11.58,58,-50) and the σ_0^2 prior is set to 122.82, the average sum of squares with 11.58 as the mean value. Regarding the degrees of freedom and the Ω_0 hyperparameter are more of a trial and error approach applied. First they are set to 1 and 1, which together with the values of the other hyperparameters gave the following results.



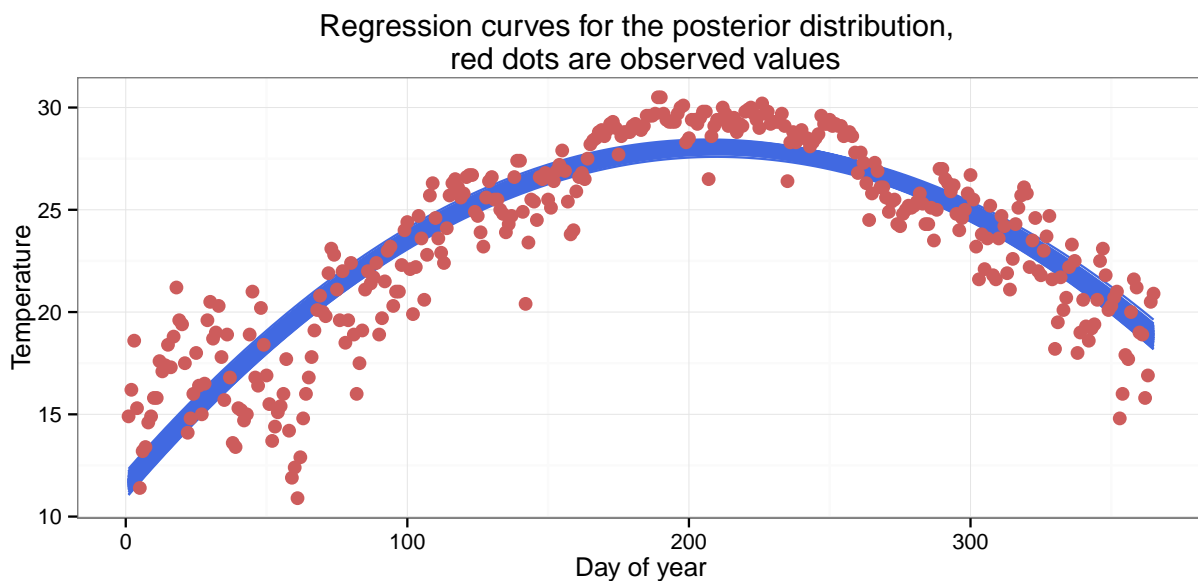
100 draws are made from the joint prior of all parameters and for each draw is a regression curve computed. As can be seen above are the regression curves given by the prior not so well fitted to data and in many cases quite far off the temperatures that can be expected. Hence, the prior hyperparameters v_0 and Ω_0 needs to be given other, more sensible, values. This is done by testing some different values for the respective parameters and it is concluded that v_0 equal to 10 and Ω_0 equal to 30 gives a reasonable prior distribution.



The regression curves obtained from the updated prior distribution are thought to be reasonable since they rather well agrees with our prior belief.

c)

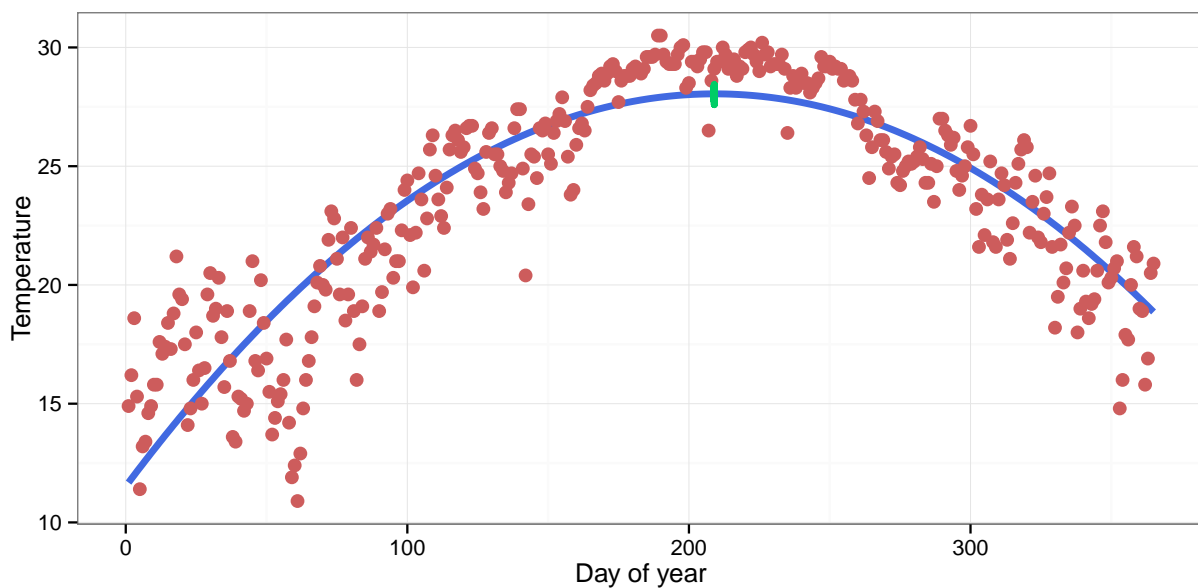
The joint posterior distribution for the β_0 , β_1 , β_2 and σ^2 is derived and then simulations are generated from it. 250 draws are simulated from the posterior distribution and the obtained regression curves are plotted in the graph below.



The regression curves obtained by the draws from the joint posterior distribution are thought to be reasonable. In general the curves follows the observed data rather well. The only exception is for the period with highest temperatures where the model consistently underestimates the temperature.

d)

If the mean temperature is calculated for every day for all the fitted values generated by the 250 regression curves obtained in *d*) it is concluded that day 209 is the day with the highest temperature.



e)

A suitable prior would be one that regulates so that the parameters not overfits data. That probably is a prior that regularizes the coefficients in the model quite heavily.

When the model has polynomials up to order 7 the aim is to regularize the parameters of higher grades. That since models with polynomials of high orders tends to be overfitted to data.

The regularization can be accomplished by setting the prior β parameters to zero and Ω_0 to a sufficiently high value. The effect of this is that each variable, or coefficient, needs to be important. Otherwise its parameter value will be close to zero and the importance of the parameter low.