Bayesian learning Lab 2

Mohamed Ali

2023-05-10

Linear and polynomial regression

To answer this question a conjugate prior for the linear regression model will be used in which: The joint prior for $\beta \; and \; \sigma^2$:

$$\beta|\sigma^2 \sim \mathcal{N}(\mu_0, \sigma^2{\Omega_0}^{-1})$$
$$\sigma^2 \sim Inv - \chi^2(v_0, \sigma^2)$$

While the posterior:

$$\begin{aligned}
\beta|\sigma^2, y &\sim \mathcal{N}(\mu_n, \sigma^2{\Omega_n}^{-1}) \\
\sigma^2 &\sim Inv - \chi^2(v_n, {\sigma^2}_n) \\
where: & \\
\mu_n &= (X'X\Omega_0)^{-1}\;(X'X\hat{\beta} + \Omega_0\mu_0) \\
\Omega_n &= X'X + \Omega_0 \\
v_n &= v_0 + n \\
v_n\sigma^2 &= v_n\sigma^2 + (y'y + \mu_0'\Omega_0\mu_0 - \mu_n'\Omega_n\mu_n)
\end{aligned}$$

First we start by reading the files and then we Create the covariate_time vaiable as (the number of days sinvce the beginning of the year / 365)

```r
#reading the files
df<-read_xlsx("Linkoping2022.xlsx")
#Creating the covariate_time vaiable as
#(the number of days sinvce the beginning of the year / 365)
a<- df$datetime
#begunning of the year
b<- '2022-01-01'
a<-format.POSIXlt(strptime(a,'%Y-%m-%d'))
b<-format.POSIXlt(strptime(b,'%Y-%m-%d'))
#time diff from the
x<-as.vector(difftime(a,b,units='days'))
df$cov_tm<-x/365
```

A

We assume to use a conjugate prior from the linear regression in Lec 5 , we have been given the prior hyperparamteres as follow:

```r
#We assume to use a conjugate prior from
#the linear regression in Lec 5 ,
#we have been given the prior hyperparamteres as follow:
mu_0= as.matrix(c(0,100,-100),ncol=3)
omega_0=0.01*diag(3)
v_0=1
segma2_0=1
n= length(df$temp)
ndraws=10
```

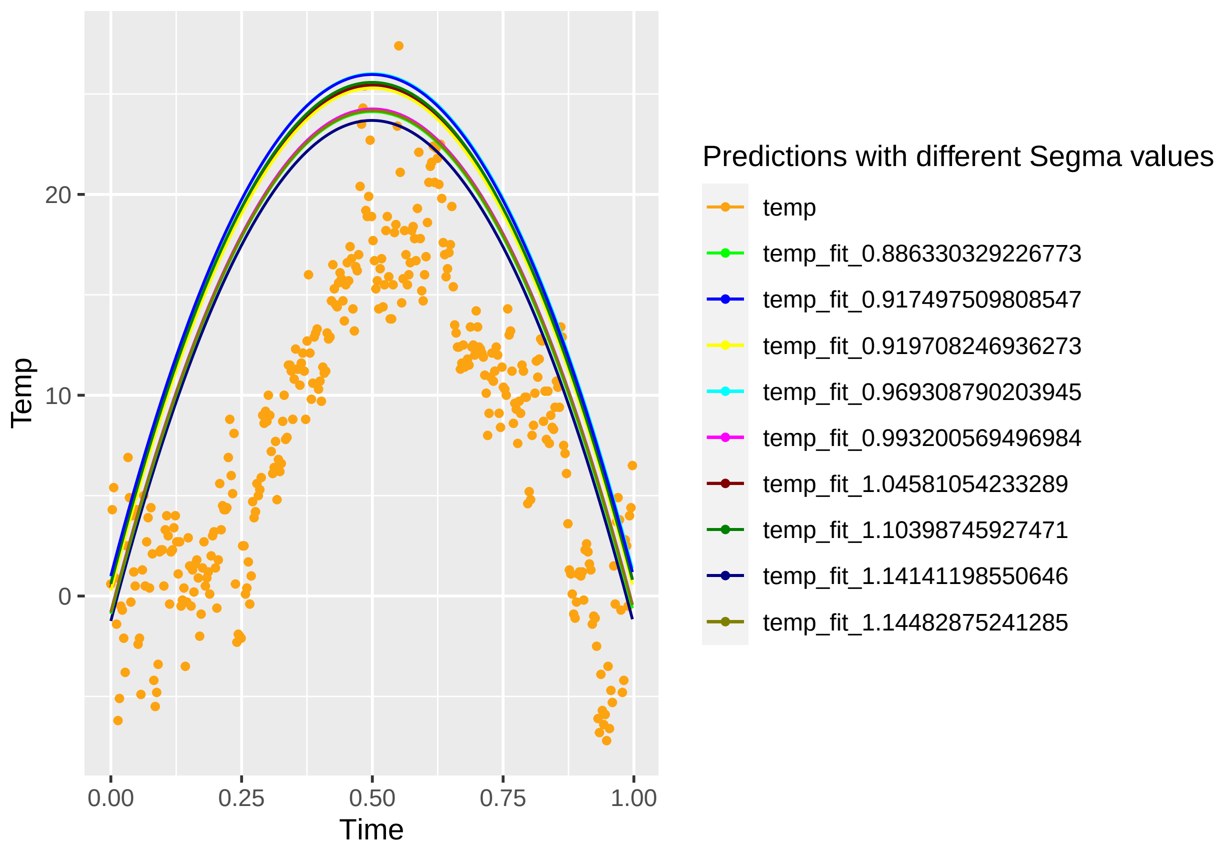
We have the joint prior for beta and segma2 defined as $\beta|\sigma^2 \sim \mathcal{N}(\mu_0, \sigma^2\Omega_0^{-1})$ and $\sigma^2 \sim Inv - \chi^2(v_o, \sigma_0^2)$ follows Inv-Chi(v0,sigma2_0)

First we draw our random samaple from inv-chi2 using the below defined function from Lec 3 slide 5

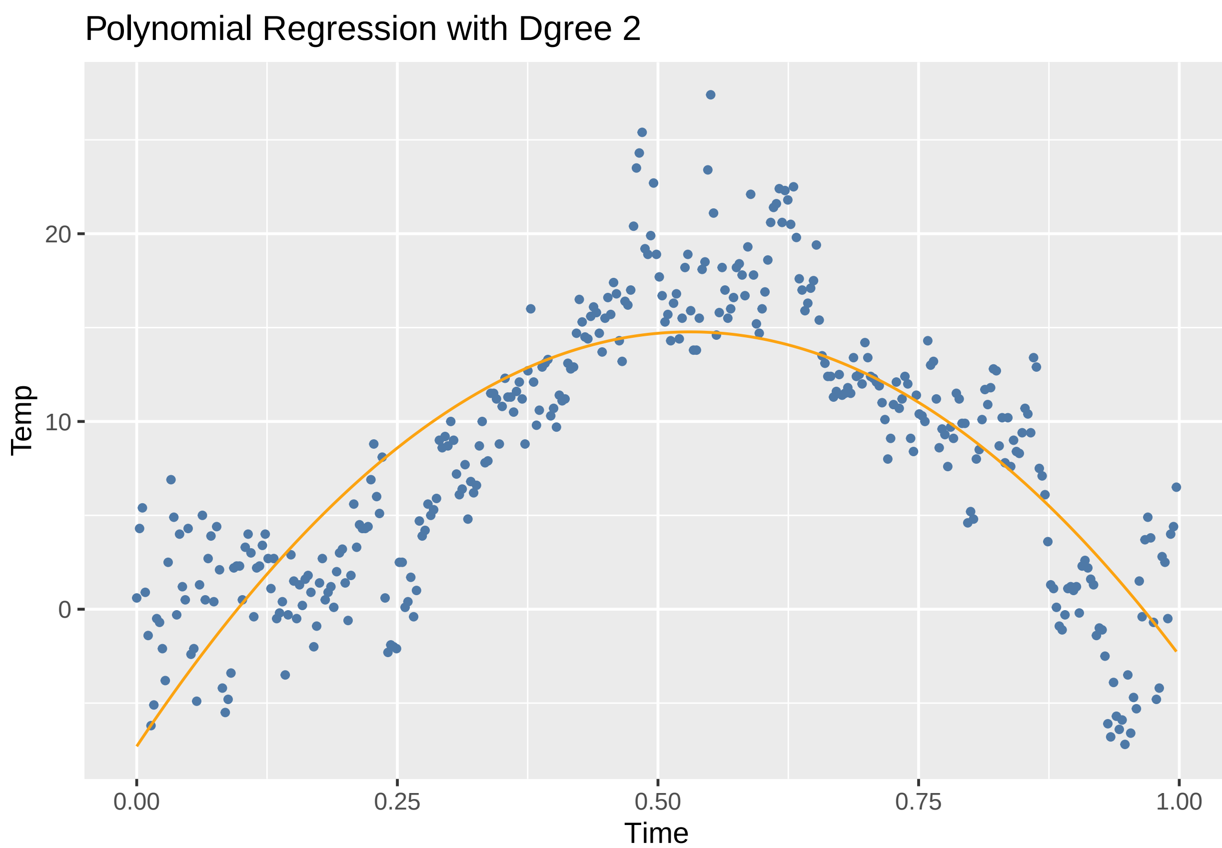
Now we estimate the betas values using the formula $\beta|\sigma^2 \sim \mathcal{N}(\mu_0, \sigma^2\Omega_0^{-1})$ and we fit the regression based ontemp = beta0 + beta1 time + beta 2 time^2 + erorr

*Note we have our error follows the normal distrbution by 0 and σ^2

After we done with the draws now for every draw compute the regression curve

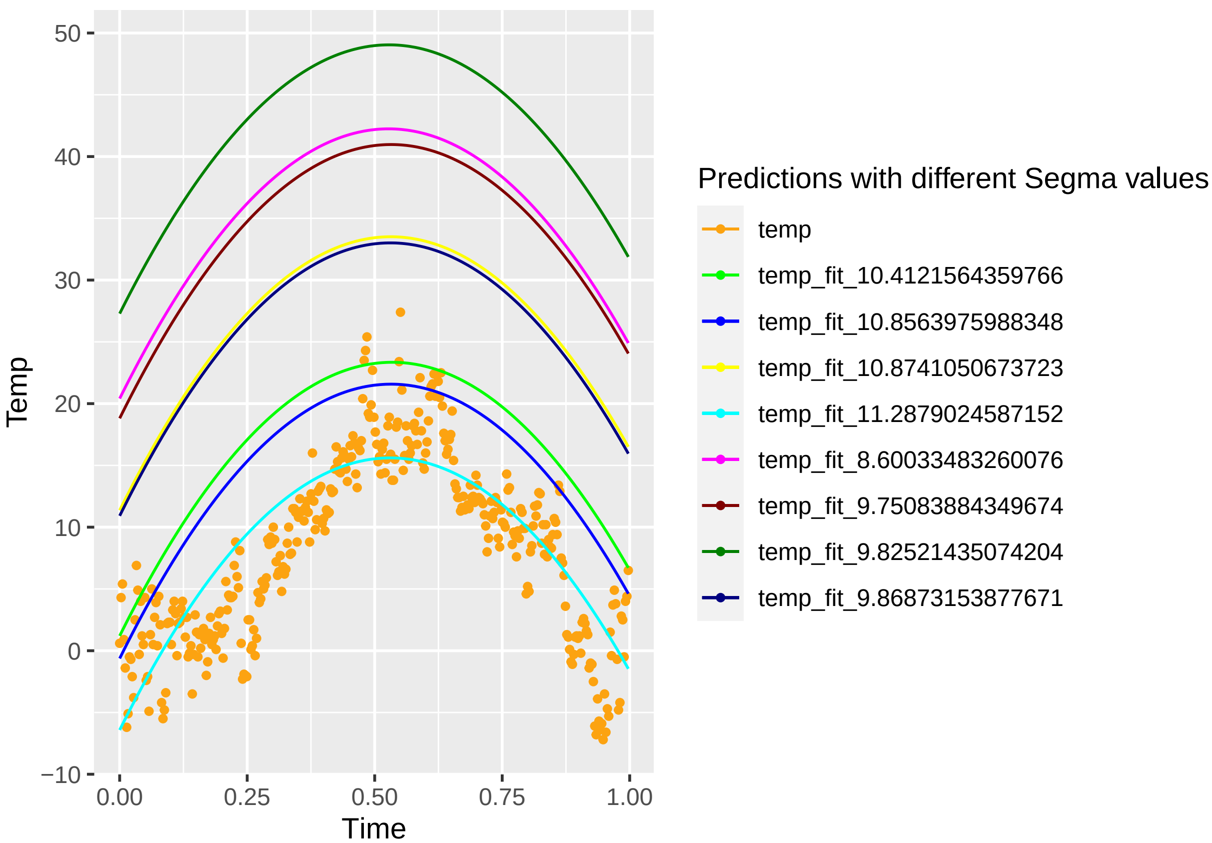


Comparing the above chart with our prior beliefs which we can see by running the polynomial model using the *lm* function in R with df = 2 we can see the fitted regression line with to some extend follow our regression line when σ^2 is equal to *1.037*.




```
##
## Call:
## lm(formula = y ~ poly(x, degree, raw = TRUE))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.6557  -2.8525  -0.1874   2.5052  12.6580
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -7.3039     0.6582  -11.10   <2e-16 ***
## poly(x, degree, raw = TRUE)1  83.1568     3.0492   27.27   <2e-16 ***
## poly(x, degree, raw = TRUE)2 -78.3093     2.9600  -26.46   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.215 on 362 degrees of freedom
## Multiple R-squared:  0.6726, Adjusted R-squared:  0.6708
## F-statistic: 371.9 on 2 and 362 DF,  p-value: < 2.2e-16
```

Now we draw simulations using the modified μ_0 from the model results we have



B

As we want to estimate the uncertainty in the model parameters. Simulating from the joint posterior allows us to obtain a set of plausible values for all the parameters in the model, taking into account the observed data and prior information. to do so: we have our non-informative prior:

$$p(\beta, \sigma^2) \propto \sigma^-1$$

Our joint posterior of β and σ^2:

$$\begin{aligned}
\beta|\sigma^2, y &\sim \mathcal{N}(\hat{\beta}, \sigma^2(X\mathfrak{t}X)^{-1}) \\
\sigma^2|y &\sim Inv - \chi^{\in}(n-k, s^2) \\
where &: \\
k &= number\ of\ \beta s\ in\ our\ case\ 3 \\
\hat{\beta} &= (X\mathfrak{t}X)^{-1}X\mathfrak{t}y \\
s^2 &= \frac{1}{n-k}(y - X\hat{\beta})\mathfrak{t}(y - X\hat{\beta})
\end{aligned}$$

Thus to simulate from the joint posterior we need to simulate from: 1- $p(\sigma|y$ 2- $p(\beta|\sigma^2, y)$

And then we find the marginal posterior of β:

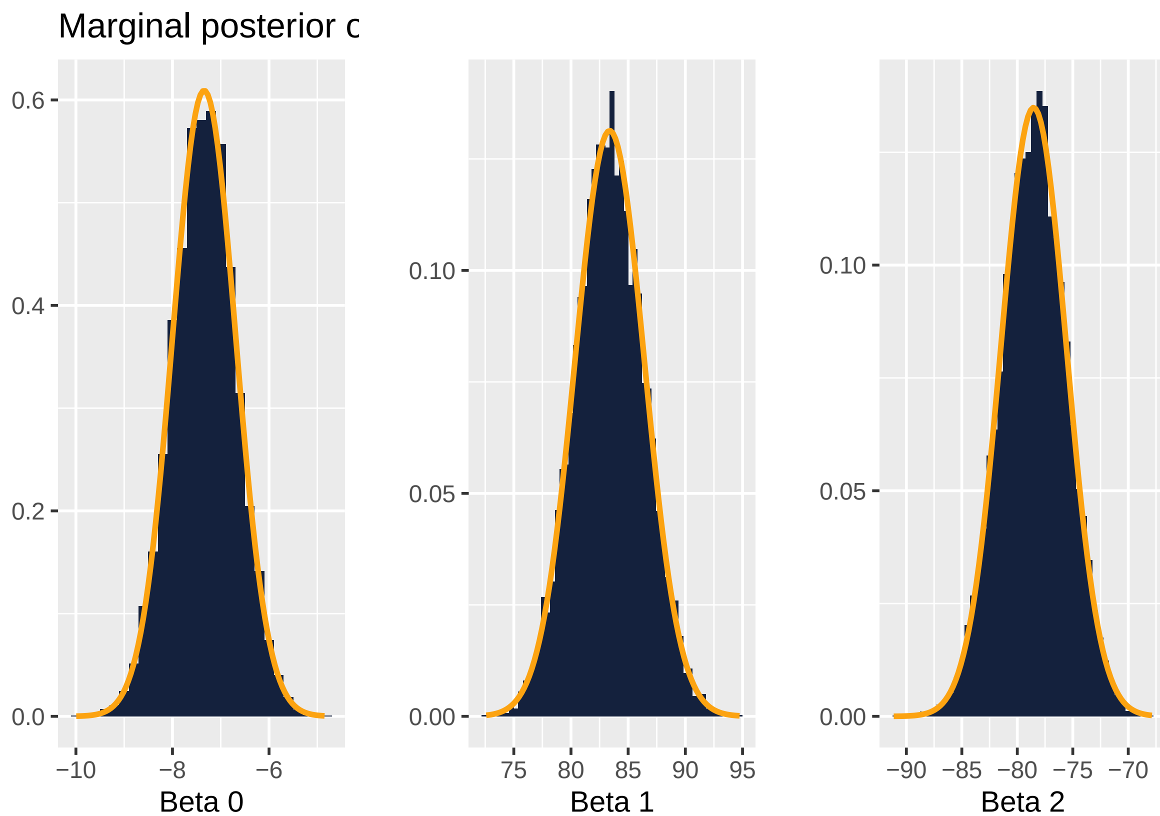
$$\beta|y \sim t_n - k(\hat{\beta}, s^2(X\mathfrak{t}X)^-1)$$

There for to draw a sample from the joint posterior distribution of $\beta_0, \beta_2, \beta_3\ and\ \sigma^2$: we need first to calculate the value of $\hat{\beta} = (X\mathfrak{t}X)^-1X\mathfrak{t}y$

Then we calculate the values of Ω_n , v_n , μ_n *and* $\sigma^2{}_n$ we use the same format as in eq 1

Now we generate the valuse of β from $t_{n-k}(\hat{\beta},\ s^2(X\text{t}X)^{-1})$ using μ_n as our delta and σ_n^2 as sigma. in R we use function *mvtnorm::rmvt*

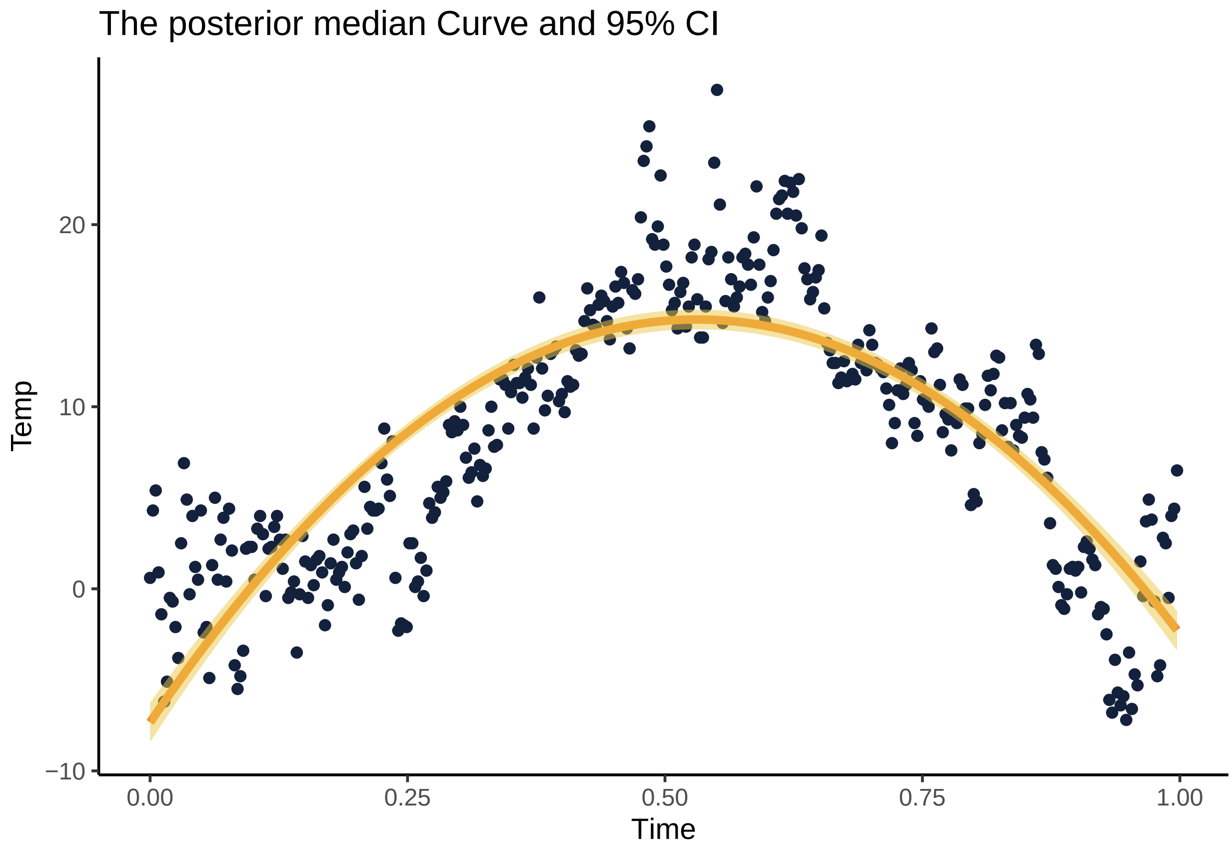
i/ Now we Plot a histogram for each marginal posterior of the parameters $\beta's$



ii/ Make a scatter plot of the temperature data and overlay a curve for the posterior median of the regression function $f(time) = E[temp|time] = \beta_0 + \beta_1.time + \beta_2.time^2$. First we need to calculate the $f(time)$ by using the results from i we can define:

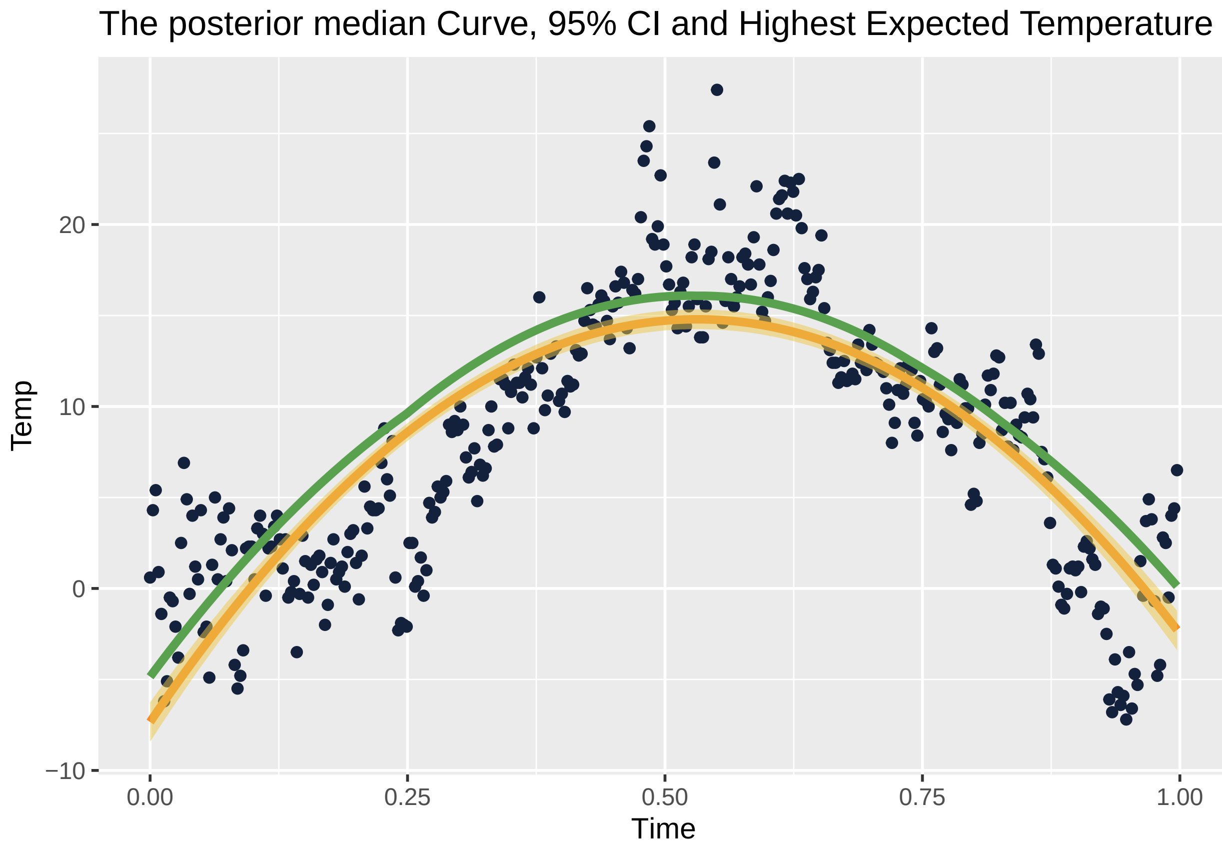
b

by using the results from above we can fit the curve line for the posterior median and 95% CI on the scatter plot of the temperature data as below:



c

To simulate from the posterior distribution of the highest expected temperature $\bar{X}$ we can use the all possible prediction values that we generated in the task before, then by taking the max value we get the point point wise highest expected temperature over time:



D

To mitigate the problem of over fitting when estimating a polynomial resgression of order 10 without being worry of the over fitting, one can use Smoothness/Shrinkage/Regularization of the prior by introducing λ a penalization parameter (See lec 5 Slide 9 Lasso) in this method we have:

$$\beta_j|\sigma^2 \sim N(o, \frac{\sigma^2}{\lambda})$$

Here we have a large values of λ gives smoother fit. More shrinkage. where:

$$\mu_0 = 0$$
$$\Omega_0 = \lambda I$$

Which equivalent to *Penalized Likelihood*:

$$-2log\ p(\beta|\sigma^2, y, x) \propto (y - X\beta)'(y - X\beta) + \lambda\beta'\beta$$

Thus, the Posterior mean/mode gives ridge regressoin estimator:

$$\tilde{\beta} = (X\grave{t}X + \lambda I)^{-1}X\grave{t}y$$
$$if\ X\grave{t}X = I$$
$$Then$$
$$\tilde{\beta} = \frac{1}{(1+\lambda)}\hat{\beta}$$

We might also be interested to determine the lambda, which could be by performing a cross validation on the test data pf using the Bayesian inference, where to us a prior for λ. we have this hierarchical setup:

$$
\begin{aligned}
y|\beta, \sigma^2, x &\sim N(X\beta, \sigma^2 I_n) \\
\beta|\sigma^2, \lambda &\sim N(0, \sigma^2 \lambda^{-1} I_m) \\
\sigma^2 &\sim Inv - \chi^2(v_o, \sigma_o^2) \\
\lambda &\sim Inv - \chi^2(\eta_0, \lambda_0) \\
so,\ \mu_o &= 0,\ \Omega = \lambda I_m
\end{aligned}
$$

and we have the joint posterior of β, σ^2, and λ is:

$$
\begin{aligned}
\beta|\sigma^2, \lambda, y &\sim N(\mu_n, \sigma^2 \Omega_n^{-1}) \\
\sigma^2|\lambda, y &\sim Inv - \chi^2(v_n, \sigma_n^2) \\
p(\lambda|y) &\propto \sqrt{\frac{|\Omega_0|}{|X^T X + \Omega_0|}} (\frac{v_n \sigma_n^2}{2})^{\frac{-v_n}{2}} .p(\lambda) \\
where,\ \Omega_0 &= \lambda I_m \ and\ p(\lambda)\ is\ the\ prior\ for\ \lambda\ and: \\
\mu_n &= (X^T X + \Omega_0)^{-1} X^T y \\
\Omega_n &= X^T X + \Omega_0 \\
v_n &= v_0 + n \\
v_n \sigma_n^2 &= v_0 \sigma_0^2 + y^T y - \mu_0^T \Omega_n \mu_n
\end{aligned}
$$

Posterior approximation for classifcation with logistic regression

a

Firstly we want to calculate the value of $\tilde{\beta}$ the posterior mode and the negative of the observed 7x7 hessian evaluated at the posterior mode $J(\tilde{\beta}) = -\frac{\partial^2\ lnp(\beta|y)}{\partial\beta\partial\beta^T}|_{\beta=\tilde{\beta}}$. Using code snippets from my demo of logistic regression in Lecture 6. First we want to calculate the vale of $\tilde{\beta}$ and $J(\tilde{\beta}) = -\frac{\partial^2\ lnp(\beta|y)}{\partial\beta\partial\beta^T}|_{\beta=\tilde{\beta}}$ by using the *optim*, Note that we have $\tau = 2$ and $\beta \sim N(0, \frac{1}{\lambda} I)$.

```
## [1] "The posterior mode is:"
```

```
##              [,1]
## [1,]  0.226662508
## [2,] -0.210975839
## [3,]  0.244068539
## [4,]  0.521031519
## [5,] -0.193480984
## [6,] -0.411117301
## [7,] -0.000204361
## attr(,"names")
## [1] "Constant"    "HusbandInc"  "EducYears"   "ExpYears"    "Age"
## [6] "NSmallChild" "NBigChild"
```

```
## [1] "The Hessian Matrix:"
```

```
##             [,1]      [,2]      [,3]      [,4]      [,5]       [,6]
## [1,] -75.1606178  -3.199556  2.176572  9.004046  1.564092  0.5417135
```

```
## [2,]  -3.1995561 -77.178127 -29.655683  -2.259583 -11.944041   5.9946343
## [3,]   2.1765716 -29.655683 -72.614886  -0.665412   8.091300 -16.2311250
## [4,]   9.0040458  -2.259583  -0.665412 -58.908053 -21.945032   6.0479660
## [5,]   1.5640924 -11.944041   8.091300 -21.945032 -70.630463  30.8279319
## [6,]   0.5417135   5.994634 -16.231125   6.047966  30.827932 -72.0625824
## [7,] -95.7438620   6.140173  16.222540  33.849372  34.696528   2.2833795
##             [,7]
## [1,]  -95.743862
## [2,]    6.140173
## [3,]   16.222540
## [4,]   33.849372
## [5,]   34.696528
## [6,]    2.283379
## [7,] -255.200443
```

```
## [1] "The approximate posterior standard deviation is:"
```

```
##    Constant  HusbandInc   EducYears    ExpYears         Age NSmallChild
##  0.16591941  0.12836587  0.13537952  0.14181584  0.15053667  0.13571113
##   NBigChild
##  0.09593617
```

Now we compare with the results from the regression model: *glmModel<- glm(Work ~ 0 + ., data = WomenAtWork, family = binomial)*

```
##
## Call:
## glm(formula = Work ~ 0 + ., family = binomial, data = wat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3210  -0.9799   0.4423   0.9707   1.9131
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## Constant     0.02263    1.93083   0.012 0.990649
## HusbandInc  -0.03796    0.02229  -1.703 0.088573 .
## EducYears    0.18447    0.10007   1.844 0.065253 .
## ExpYears     0.12132    0.03353   3.618 0.000297 ***
## Age         -0.04858    0.03323  -1.462 0.143686
## NSmallChild -1.56485    0.51078  -3.064 0.002187 **
## NBigChild   -0.02526    0.17716  -0.143 0.886618
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 182.99  on 132  degrees of freedom
## Residual deviance: 146.73  on 125  degrees of freedom
## AIC: 160.73
##
## Number of Fisher Scoring iterations: 4
```

The Coefficients of model results about shows relationship between each predictor variable and the log-odds of the outcome variable. variables like *HusbandInc,Age,NSmallChild and NBigChild* have a negative impact on the y variable in which the direction of the relationship, while the rest have a positive impact, our results from the teammate tell the same direction of the model estimates (i.e variables with negative impact *HusbandInc,Age,NSmallChild and NBigChild*). However the magnitude of the coefficient which indicates the strength of the relationship is a bit different.

Now we Compute an approximate 95% equal tail posterior probability interval for the regression coefficient to the variable NSmallChild

```
## [1] "An approximate 95% equal tail posterior probability interval for the regression coeffcient to th
```

```
##       2.5%      97.5%
## -0.6730329 -0.1455234
```

The results of the CI tell us that the NSmallChild regression coefficient have a significant impact on the log-odds of the outcome variable as we can see the coefficient fall between the lower and the upper bound of the CI, the direction of this impact in negative and have the highest magnitude in the model coefficients.