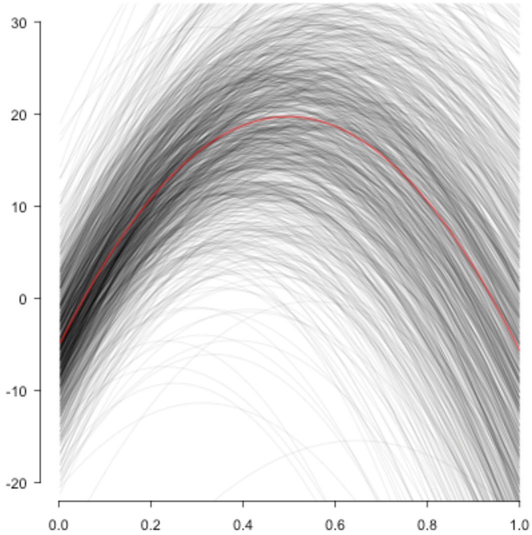
```
######################### Lab 2 #########################
## Meta-info
# Linear regression
# Polynomial regression
# Logistic regressions
# Credible interval
# Maximum likelihood
# Optim
# Hessian
# Mode of beta
# Predictive distributions

# Task 1: Linear and polynomial regression
# a) Set the prior hyperparameters μ0, Ω0, ν0 and σ2 to sensible values
# b) Check if your prior from a) is sensible
# Simulate draws from joint prior and compute regression curve of each draw
# c) Simulates from the joint posterior distribution of β0, β1,β2 and σ2
# Plot:
# • Posterior mean of simulations
# • Curve of lower and upper credible interval of f(time)
# d) Simulate highest expected temperatures
# Simulate from posterior distribution of time with highest expected temperatures
# What to do to mitigate risk of overfitting high order polynomial regression?

# Task 2: Posterior approximation for classification with logistic regression
# a) Fit logistic regression using maximum likelihood estimations
# b) Approximate the posterior distribution of the 8-dim parameter vector β with a multivariate normal distribution
# c) Simulates from the predictive distribution of the response variable in a logistic regression

####################################################################################################

# Functions
scal_inv_schsq <- function(v, σ_2, nDraws) {
  X <- rchisq(n = nDraws, df = v)
  return (v*σ_2/X)
}

############# Task 1 #############
# Respons variable: temp = β0 + β1*time + β2 * time^2 + ε, ε ~ N(0, σ2)
# Covariate: time = No. of days since beginning of year/366

# a)
# Conjugate priors
# β | μ ~ N(μ0, σ2*Ω0_(-1))
# σ2 ~ Inv-X(ν0, σ0_2)
#
# Set prior hyperparameters:
# μ0: The expected value of the betas [vector]
# Ω0: How sure we are about the betas (scales the variance) [matrix]
# ν0: How sure we are about our prior knowledge of the sigmas [scalar]
# σ0_2: The variance of the betas [vector]

dataset <- read.table("TempLinkoping.txt", header = TRUE)

prior.mu0 <- c(-5, 100, -100)
prior.ν0 <- 10
prior.σ0_2 <- (7/1.96)^2 # 10 = 1.95 * σ -> 10 degrees are in the confidence interval 95% of the times
prior.Ω0 <- matrix(c(0.5, 0, 0, 0, 0.1, 0, 0, 0, 0.1),
                   nrow = 3,
                   ncol = 3)
prior.inv_Ω0 <- solve(prior.Ω0)

# b)
nDraws = 1000
beta_draws <- matrix(nrow = nDraws, ncol = 3)

plot.new()
plot.window(xlim=c(0,1), ylim=c(-20, 30))
axis(side=1)
axis(side=2)

# Sample betas & plot regression curves
for (i in 1:nDraws) {
  sigma2 <- scal_inv_schsq(prior.ν0, prior.σ0_2, 1) # Sample sigma2 from scaled inverse chi squared
  beta_draws[i,] <- rmvnorm(n = 1, mean = prior.mu0, sigma = sigma2*prior.inv_Ω0) # Sample betas from Multinormal

  # Generates regression point for each dataset$time. All of them are then plotted as a line
  lines(dataset$time,
        beta_draws[i,1] + beta_draws[i,2]*dataset$time + beta_draws[i,3]*dataset$time^2,
        col=rgb(0, 0, 0, 0.1))
}

lines(dataset$time,
      mean(beta_draws[,1]) + mean(beta_draws[,2])*dataset$time + mean(beta_draws[,3])*dataset$time^2,
      col=rgb(1, 0, 0, 1))
```




```
# Priors not sensible. New priors
prior.mu0 <- c(-5, 100, -100)
prior.v0 <- 10
prior.σ0_2 <- (7/1.96)^2 # 7 = 1.95 * σ -> 7 degrees are in the confidence interval 95% of the times
prior.Ω0 <- matrix(c(0.5, 0, 0, 0, 0.1, 0, 0, 0, 0.1),
                   nrow = 3,
                   ncol = 3)
prior.inv_Ω0 <- solve(prior.Ω0)
n <- dim(dataset)[1]

# c)

# Setup
X <- cbind(1, dataset$time, dataset$time^2) # Create X with constant row for beta_0
Y <- dataset$temp
n <- dim(dataset)[1]
CI <- 0.90
CI_lower <- (1-CI)/2 # 0.05
CI_upper <- 1-(1-CI)/2 # 0.95

# Posterior
beta_hat <- solve(t(X)%*%X)%*%t(X)%*%Y # Beta_hat by classic by OLS (Ordinary Least Square)
posterior.mun <- solve(t(X) %*% X + prior.Ω0) %*% (t(X) %*% X %*% beta_hat + prior.Ω0 %*% prior.mu0) # Mu_n
posterior.Ωn <- t(X)%*%X + prior.Ω0 # Omega_n
posterior.inv_Ωn <- solve(posterior.Ωn) # Inverse Omega_n
posterior.vn <- prior.v0 + n # v_n
posterior.sigman_2 <- (t(Y) %*% Y + t(prior.mu0) %*% prior.Ω0 %*% prior.mu0 - t(posterior.mun) %*% posterior.Ωn %*% posterior.mun)/post

nDraws <- 10000
beta_post_draws <- matrix(nrow = nDraws,
                          ncol = 3)
posterior_y <- matrix(nrow = nDraws,
                      ncol = n)

# Draws of sigma2 and beta posteriors
for (i in 1:nDraws) {
  sigma2 <- as.vector(scal_inv_schsq(posterior.vn, posterior.sigman_2, 1))
  beta_post_draws[i,] <- rmvnorm(n = 1,
                             mean = posterior.mun,
                             sigma = sigma2*posterior.inv_Ωn)
}

# Generates a large matrix (10000 x 366), For each time/column (366) -> 10000 predicted temps/rows, one for each beta-draw
temp_pred_each_time <-  t(X %*% t(beta_post_draws))
temp_pred_each_time_sorted = apply(X = temp_pred_each_time,
                                   MARGIN = 2,
                                   sort) # Sort each time/row

# Extract lower and upper
temp_pred_CI90 <- rbind(temp_pred_each_time_sorted[round(nDraws*CI_lower),],
                        temp_pred_each_time_sorted[round(nDraws*CI_upper),])

# Plot data and mean regression line
plot(dataset)
lines(dataset$time, mean(beta_post_draws[,1]) + mean(beta_post_draws[,2]) * dataset$time + mean(beta_post_draws[,3]) * dataset$time^2,
      col=rgb(1, 0, 0, 1))
lines(dataset$time, temp_pred_CI90[1,], col=rgb(0, 1, 0, 1))
lines(dataset$time, temp_pred_CI90[2,], col=rgb(0, 1, 0, 1))

# d)
# Time with highest expected temperature: time = -B1/2B2
# Calculated from the derivation of f(time) set to 0.

posterior_max_time <- -beta_post_draws[,2]/(2*beta_post_draws[,3])
abline(v = mean(posterior_max_time), col=rgb(0, 0, 1, 1))

# e)
# To mitigate the risk of overfitting due to higher order polynomials we want to have a small
# mu_n for larger betas. This will lead to a smaller coefficient for the higher polynomials,
# thus making them affect the end result less.
# 1) Set my_0 corresponding to higher polynomials low
# 2) Set the diagonal indices of Ometa_0 corresponding higher polynomial high

############# Task 2 #############
```

```r
# Posterior approximation for classification with logistic regression
# Dataset:
# Response variable: Work
# Covariates: Constant  HustbandInc EducYear  ExpYear   ExpYear2  Age   NSmallChild   NBigChild

dataset <- read.table("WomenWork.dat", header=TRUE)

# a)

logRegFit <- glm(formula = Work ~.-Constant, data = dataset, family = "binomial")

# b)
# Approx posterior distribution of 8-dim parameter vector β
# Posterior: β | y, X ~ N(Beta_mode, Inv_Hessian_At_Beta_bode)
# Likelihood: y | β, X = exp(x_i*β)^(y_i)/[ 1+exp(x_i*β) ]
# Prior: β ~ N(0, τ_2*I), τ = 10

# Functions
# !!!!! Important to remember !!!!!
## 1) Always use log posterior as it's more stable and avoids problems with to small or large numbers
## 2) Don't forget that in log -> posterior = log.likelihood + log.prior
## 3) Don't forget to handle Infinity
postLogReg <- function(β, mu_0, X, Y, tau) {
  no_of_betas <- length(β) # Number of covariates
  sigma <- tau^2 * diag(no_of_betas) # Calculate sigma tau^2 * I

  # Likelihood
  # Logarithm of prod[ exp(x*β)^Y / (1 + exp(x*β))]
  log.likelihood <- sum(t(Y) %*% X %*% β) - log(prod(1 + exp(X %*% β)))

  if (abs(log.likelihood) == Inf) log.likelihood = -20000;

  # Prior
  log.prior <- dmvnorm(β, mean = mu_0, sigma = sigma, log = TRUE)

  return (log.likelihood + log.prior)
}

# Setup
n_parameters <- dim(dataset[,-1])[2] # No. of covariates
X <- as.matrix(dataset[, -1])
Y <- as.matrix(dataset[, 1])
nDraws = 10000
CI_interval = c(0.025, 0.975)

# Prior
β0 <- as.matrix(rep(0, n_parameters)) # Initial beta-values
beta.prior.mu_0 <- rep(0, n_parameters) # Mu_0
beta.prior.tau <- 10 # Tau: Given in the task

# Optim
# par: Initial values for parameter to me optimized
# fn: Function to be minimized/maximized
# Variables: Pass all variables except the one being maximized
optim.res <- optim(par = β0,
                   fn = postLogReg,
                   gr = NULL,
                   mu_0 = beta.prior.mu_0,
                   X = X,
                   Y = Y,
                   tau = beta.prior.tau,
                   method = "BFGS",
                   control = list(fnscale=-1),
                   hessian = TRUE
                   )

β.mode <- optim.res$par # Mode of beta
β.hessian.neg.inv <- -solve(optim.res$hessian) # Negative inverse hessian of beta

# Beta draws
β.draws <- matrix(nrow = nDraws,
                  ncol = n_parameters)

for (i in 1:nDraws) {
  β.draws[i, ] <- mvrnorm(n = 1, mu = β.mode, Sigma = β.hessian.inv)
}
```

```
## Error in mvrnorm(n = 1, mu = β.mode, Sigma = β.hessian.inv): could not find function "mvrnorm"
```

```r
# Calculate Credible Interval of NSmallChild
NSmallChild.draws <- sort(β.draws[, 7]) # Sort all draws in ascending order
NSmallChild.CI <- c(NSmallChild.draws[round(nDraws*CI_interval[1])],
                    NSmallChild.draws[round(nDraws*CI_interval[2])]) # Extract Credible intervals

# Hist of draws of NSmallChild
breaks <- 200
h <- hist(β.draws[,7], breaks = breaks, plot = FALSE)
```

```
## Error in hist.default(β.draws[, 7], breaks = breaks, plot = FALSE): 'x' must be numeric
```

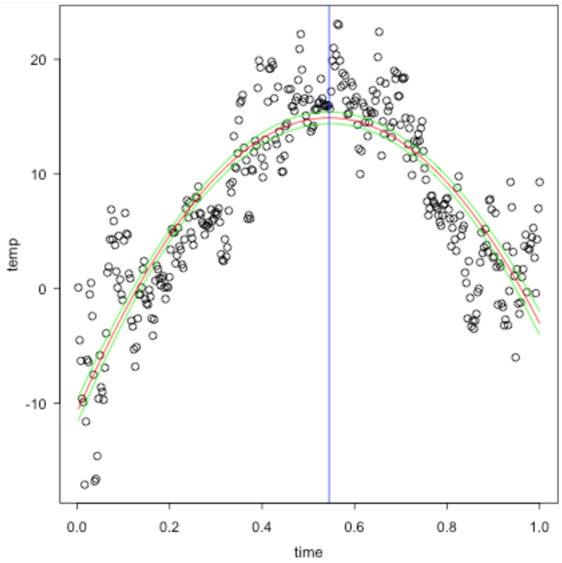
```r
cut <- cut(h$breaks, c(NSmallChild.CI[1], NSmallChild.CI[2]))
```

```
## Error in cut(h$breaks, c(NSmallChild.CI[1], NSmallChild.CI[2])): object 'h' not found
```

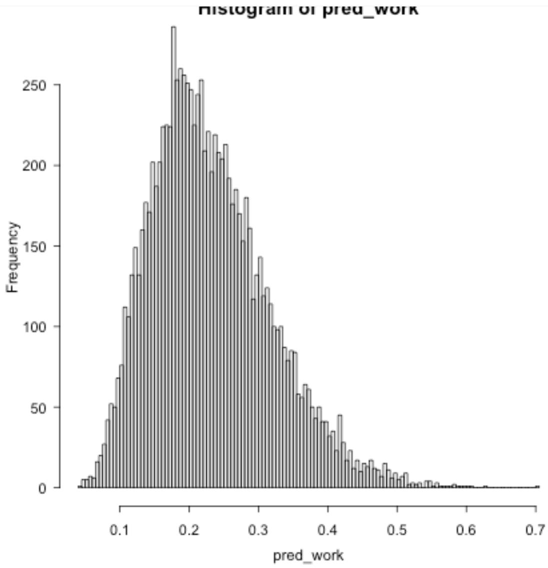
```r
plot(h,
     col = cut,
     main = "Draws of NSmallChild",
     xlab = "Value")
```

```
## Error in plot(h, col = cut, main = "Draws of NSmallChild", xlab = "Value"): object 'h' not found
```

```
abline(v = NSmallChild.CI[1], col = rgb(1, 0, 0, 1))
abline(v = NSmallChild.CI[2], col = rgb(1, 0, 0, 1))
```




```
# c)

# Functions
sigmoid <- function(x) {
  return (exp(x) / (1 + exp(x)))
}

drawPredDist <- function(βmode, negInvHess, y, nDraws) {
  beta_draws <- rmvnorm(n = nDraws,
                        mean = βmode,
                        sigma = negInvHess) # Draw from posterior beta distribution

  return (sigmoid(beta_draws %*% y))
}

# Setup
target <- c(1, 10, 8, 10, (10/10)^2, 40, 1, 1) # Target woman covariates
pred <- numeric() # Vector to collect results
nDraws <- 10000 # No. of draws

# Distribution of logistic regression of target
pred_work <- drawPredDist(βmode = β.mode,
                          negInvHess = β.hessian.neg.inv,
                          y = target,
                          nDraws = nDraws)

# Histogram of logistic regression
hist(pred_work, breaks=100)
```




```
# ~99.5% of the values are below 0.5.
# The data is indicating that the target woman doesn't work.
percent_below_05 <- sum(ifelse(pred_work < 0.5, 1, 0))/length(pred_work)
```