Homework - Statistical modelling and inference

October 19, 2015

Students	**Student IDs**
Max van Esso	73539
Marco Fayet	125593
Felix Gutmann	125604

1 Homework Lecture 3

1.1 Exercise 1

To show: $(\boldsymbol{D} + q\boldsymbol{\Phi}^T\boldsymbol{\Phi})\boldsymbol{w}_{Bayes} = q\boldsymbol{\Phi}^T t + \boldsymbol{D}\boldsymbol{\mu}$

Starting from the (-2 times) log likelihood (l) function. Expanding and simplifying it leads to:

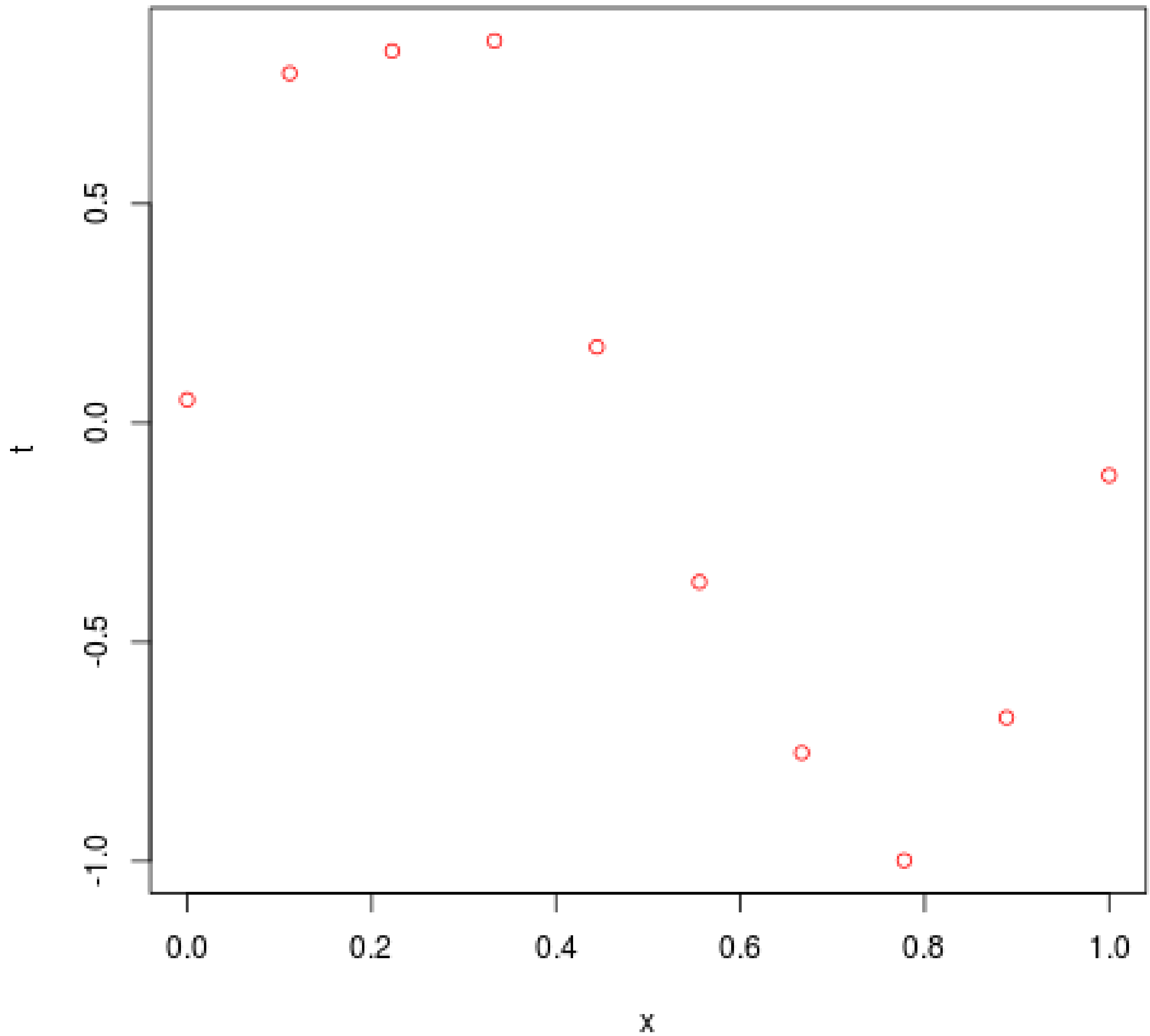
$$
\begin{aligned}
l &= (\boldsymbol{t} - \boldsymbol{\Phi w})^T q\boldsymbol{I}(\boldsymbol{t} - \boldsymbol{\Phi w}) + (\boldsymbol{w} - \boldsymbol{\mu})^T\boldsymbol{D}(\boldsymbol{w} - \boldsymbol{\mu}) + Const \\
&= \boldsymbol{t}^T q\boldsymbol{t} - \boldsymbol{t}^T q\boldsymbol{\Phi w} - \boldsymbol{w}^T\boldsymbol{\Phi}^T q\boldsymbol{t} + \boldsymbol{w}^T\boldsymbol{\Phi}^T q\boldsymbol{\Phi w} + \boldsymbol{w}^T\boldsymbol{D w} - \boldsymbol{w}^T\boldsymbol{D\mu} - \boldsymbol{\mu}^T\boldsymbol{D w} + \boldsymbol{\mu}^T\boldsymbol{D\mu} + Const \\
&= \boldsymbol{t}^T q\boldsymbol{t} - 2q\boldsymbol{t}^T\boldsymbol{\Phi w} + q\boldsymbol{w}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi w} + \boldsymbol{w}^T\boldsymbol{D w} - 2\boldsymbol{\mu}^T\boldsymbol{D w} + \boldsymbol{\mu}^T\boldsymbol{D\mu} + Const
\end{aligned}
$$

Take the derivative and set it to zero:

$$
\begin{aligned}
\frac{\partial l}{\partial \boldsymbol{w}} &\overset{!}{=} 0 \\
&= -2q\boldsymbol{t}^T\boldsymbol{\Phi} + 2q\boldsymbol{w}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi} + 2\boldsymbol{w}^T\boldsymbol{D} - 2\boldsymbol{\mu}^T\boldsymbol{D} = 0 \\
2q\boldsymbol{w}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi} + 2\boldsymbol{w}^T\boldsymbol{D} &= 2q\boldsymbol{t}^T\boldsymbol{\Phi} + 2\boldsymbol{\mu}^T\boldsymbol{D} \\
q\boldsymbol{w}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \boldsymbol{w}^T\boldsymbol{D} &= q\boldsymbol{t}^T\boldsymbol{\Phi} + \boldsymbol{\mu}^T\boldsymbol{D} \\
(q\boldsymbol{w}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \boldsymbol{w}^T\boldsymbol{D})^T &= (q\boldsymbol{t}^T\boldsymbol{\Phi} + \boldsymbol{\mu}^T\boldsymbol{D})^T \\
q\boldsymbol{\Phi}^T\boldsymbol{\Phi w} + \boldsymbol{D w} &= q\boldsymbol{\Phi}^T\boldsymbol{t} + \boldsymbol{D\mu} \\
(q\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \boldsymbol{D})\boldsymbol{w} &= q\boldsymbol{\Phi}^T\boldsymbol{t} + \boldsymbol{D\mu}
\end{aligned}
$$

1.2 Exercise in R

1.2.1 Plot the data from the data set



1.2.2 phix function

Listing 1: "R-Code for the phix function"

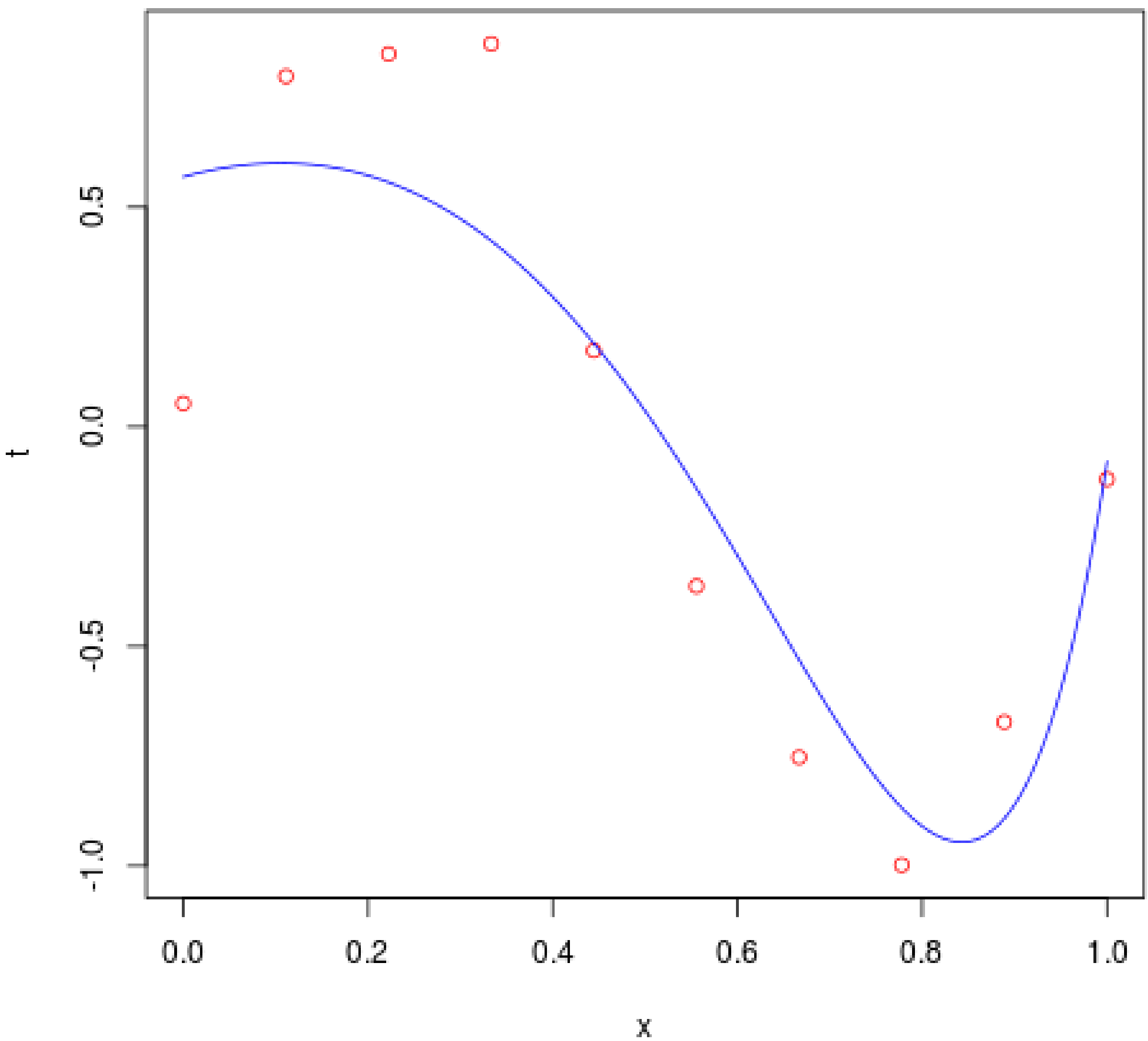
```
phix <- function(x,M,categorial){
  #Argument categorial - poly
  if(categorial=="poly"){
      num.basis.poly <- 0:M
      poly <- sapply(num.basis.poly,function(b){x^b})
      return(poly)
  #Argument categorial - gauss
    } else if (categorial=="Gauss"){
      num.basis.gauss <- 0:(M-1)
      gauss <- sapply(num.basis.gauss, function(i){ exp( -( x - i / (M-1) )^2 /.1 ) } )
      #Add intercept
      gauss <- cbind(1,gauss)
      return(gauss)
    }
  }
```

1.2.3 post.param function

Listing 2: "R-Code for the post.param function"

```
post.params <- function(dependent,independent,M,cat,delta,q){
    phi        <<- phix(independent,M,cat)
    lambda      <- delta/q
    lambda.mat  <- diag( ncol( t(phi) %*% phi ) ) * lambda
    wbayes      <- solve( ( lambda.mat + t(phi) %*% phi ) ) %*% t(phi) %*% dependent
    Q           <- q * (lambda.mat + t(phi) %*% phi)
    output      <- list(wbayes,Q)
    return(output)
   }
```

1.2.4 Estimated linear predictor vs. training data



2 Homework Lecture 4

2.1 Exercise 1

To show: $\phi(\boldsymbol{x})^T \boldsymbol{w}_{Bayes} = \sum_{n=1}^{N} q\phi(\boldsymbol{x})^T \boldsymbol{Q}^{-1}\phi(\boldsymbol{x}_n)t_n$

Prerequisite information:

- $\boldsymbol{w}_{Bayes} = (\lambda \boldsymbol{I} + \boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}^T \boldsymbol{t}$
- $\boldsymbol{Q} = q(\lambda \boldsymbol{I} + \boldsymbol{\Phi}^T \boldsymbol{\Phi})$

Proof:

$$\phi(\boldsymbol{x})^T \boldsymbol{w}_{Bayes} = \phi(\boldsymbol{x})^T (\lambda \boldsymbol{I} + \boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}^T \boldsymbol{t}$$

Using the equation for $\boldsymbol{Q}$ we can rephrase $(\lambda \boldsymbol{I} + \boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1}$ as follows:

$$\begin{aligned}
\boldsymbol{Q} &= q(\lambda \boldsymbol{I} + \boldsymbol{\Phi}^T \boldsymbol{\Phi}) \\
\frac{1}{q}\boldsymbol{Q} &= (\lambda \boldsymbol{I} \boldsymbol{\Phi}^T \boldsymbol{\Phi}) \\
(\frac{1}{q}\boldsymbol{Q})^{-1} &= (\lambda \boldsymbol{I} \boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1} \\
q\boldsymbol{Q}^{-1} &= (\lambda \boldsymbol{I} \boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1}
\end{aligned}$$

Substituting that in the former equation and rewriting it leads to:

$$\begin{aligned}
&= q\phi(\boldsymbol{x})^T \boldsymbol{Q}^{-1} \boldsymbol{\Phi}^T \boldsymbol{t} \\
&= q\phi(\boldsymbol{x})^T \boldsymbol{Q}^{-1} \sum_{n=1}^{N} \phi(\boldsymbol{x_n}) * t_n \\
&= \sum_{n=1}^{N} q\phi(\boldsymbol{x})^T \boldsymbol{Q}^{-1} \phi(\boldsymbol{x_n}) * t_n
\end{aligned}$$

2.2 Exercise 2

To show: The weight of t_n is $k(\boldsymbol{x}, \boldsymbol{x}_n)$

Prerequisite information:

- $k := q\phi(\boldsymbol{x})^T \boldsymbol{Q}^{-1} \phi(\boldsymbol{y})$

Starting with the result from the last exercise:

$$\phi(\boldsymbol{x})^T \boldsymbol{w}_{Bayes} = \sum_{n=1}^{N} q\phi(\boldsymbol{x})^T \boldsymbol{Q}^{-1} \boldsymbol{\Phi}(\boldsymbol{x}_n) t_n$$

Notice that $q\phi(\boldsymbol{x})^T\boldsymbol{Q}^{-1}\boldsymbol{\Phi}(x_n)$ fits the former definition of k. Substituting it into the last equation leads to:

$$\phi(\boldsymbol{x})^T\boldsymbol{w}_{Bayes} = \sum_{n=1}^{N} k(\boldsymbol{x}, \boldsymbol{x}_n)t_n$$

2.3 Exercise 3

To show: $\boldsymbol{K} = q\boldsymbol{\Phi}\boldsymbol{Q}^{-1}\boldsymbol{\Phi}^T$

Starting with the weight of t_n from the last exercise:

$$k(\boldsymbol{x}, \boldsymbol{x}_n) = q\phi(\boldsymbol{x})^T\boldsymbol{Q}^{-1}\phi(\boldsymbol{x_n})$$

Extend this to the first variable $\boldsymbol{x}_k$ to get an entry of $\boldsymbol{K}$

$$k(\boldsymbol{x}_k, \boldsymbol{x}_n) = q\phi(\boldsymbol{x_k})^T\boldsymbol{Q}^{-1}\phi(\boldsymbol{x_n})$$

All entries of matrix $\boldsymbol{K}$ can be expressed by the following matrix product and thus:

$$\boldsymbol{K} = q\boldsymbol{\Phi}\boldsymbol{Q}^{-1}\boldsymbol{\Phi}^T$$

2.4 Exercise 4

To show: $\boldsymbol{K}$ is equal to the hat matrix of the linear regression for $\lambda = 0$:

Prerequisite information:

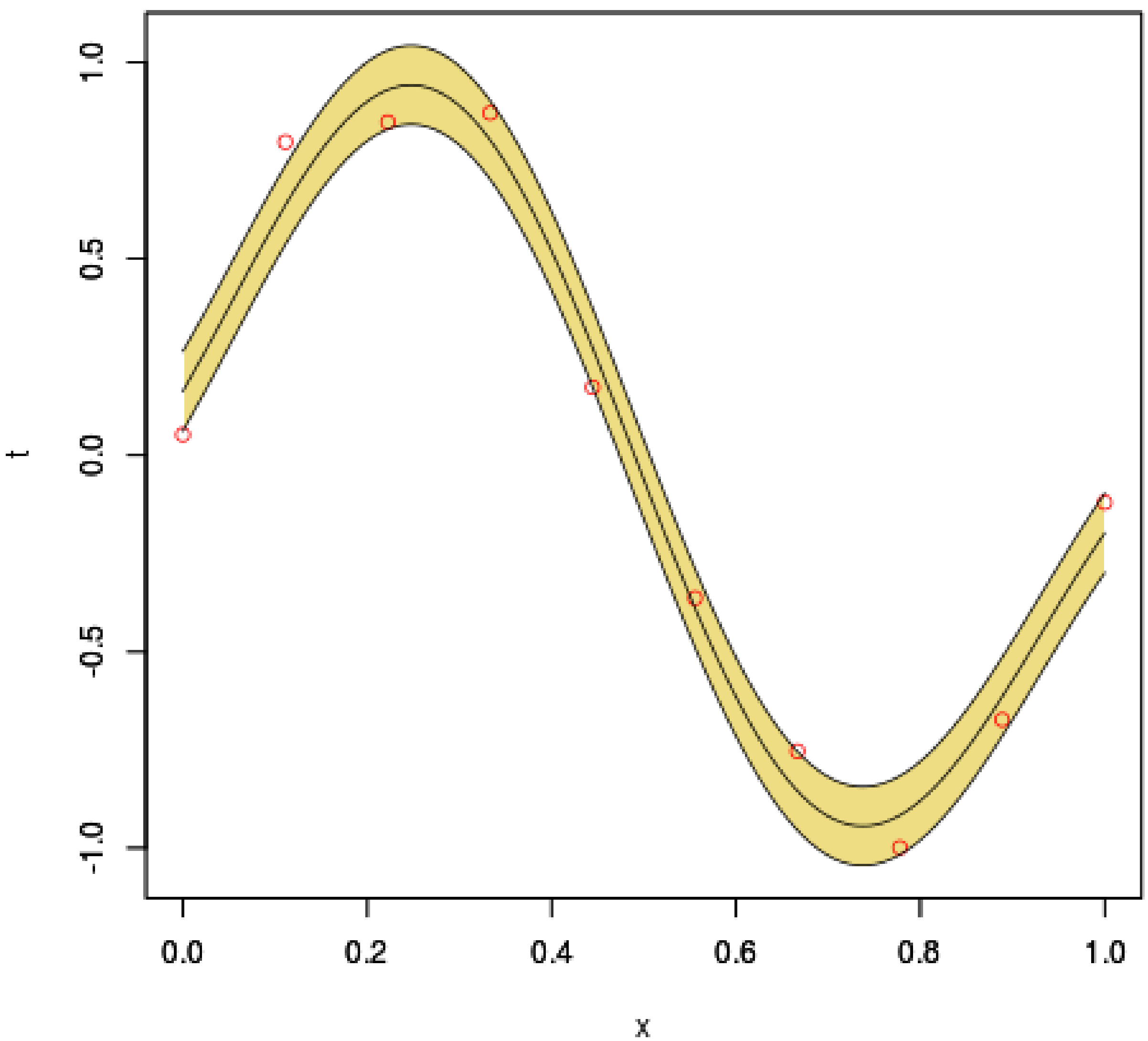
- $\lambda = \frac{\delta}{q}$

Proof:

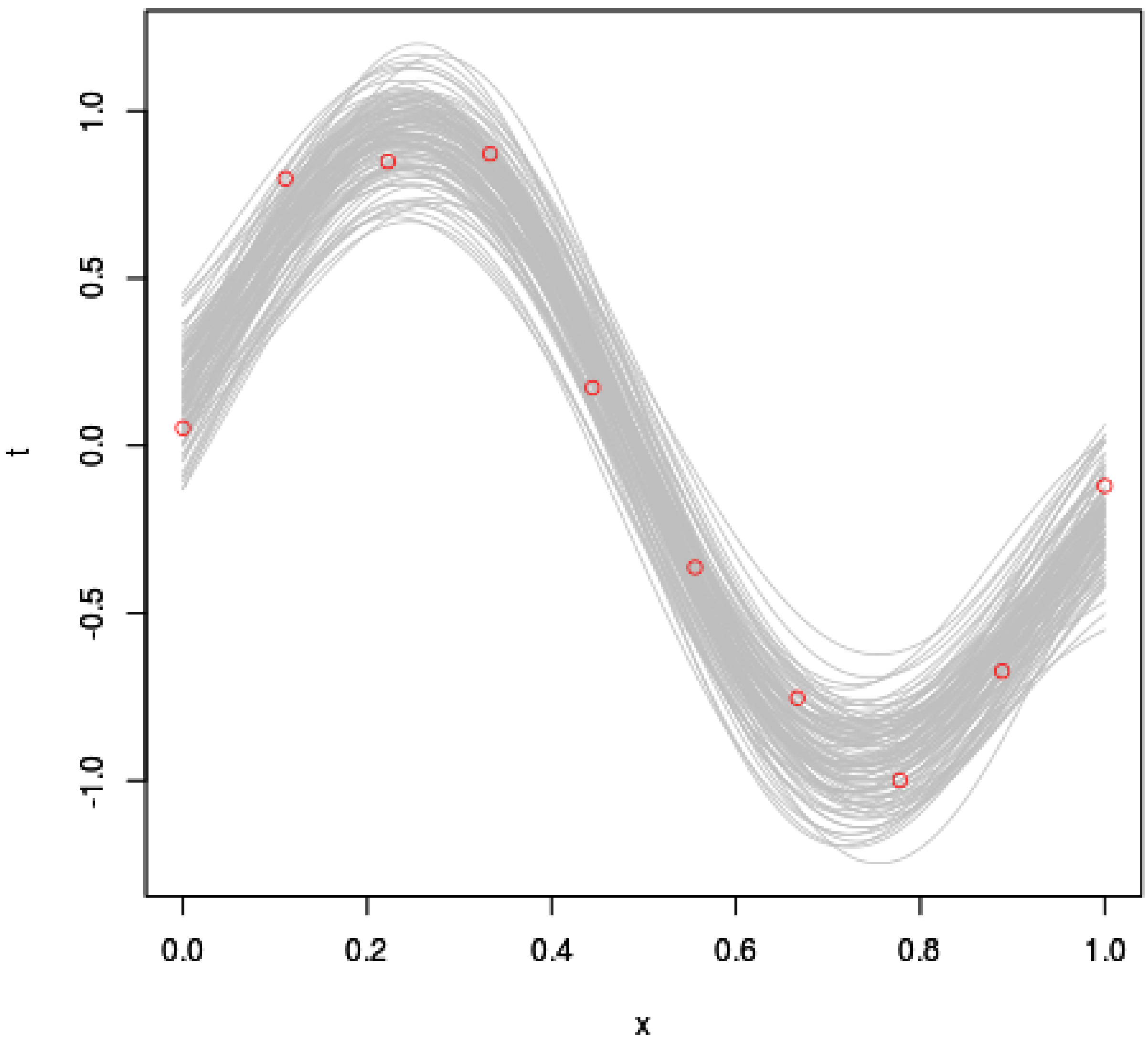
$$\begin{aligned}
\boldsymbol{K} &= q\boldsymbol{\Phi}(\delta\boldsymbol{I} + q\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T \\
&= q * \frac{1}{q}\boldsymbol{\Phi}(\delta\frac{1}{q}\boldsymbol{I} + \boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T \\
&= \boldsymbol{\Phi}(\lambda\boldsymbol{I} + \boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T \\
&= \boldsymbol{\Phi}(0 + \boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T \\
&= \boldsymbol{\Phi}(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T = \boldsymbol{H}
\end{aligned}$$

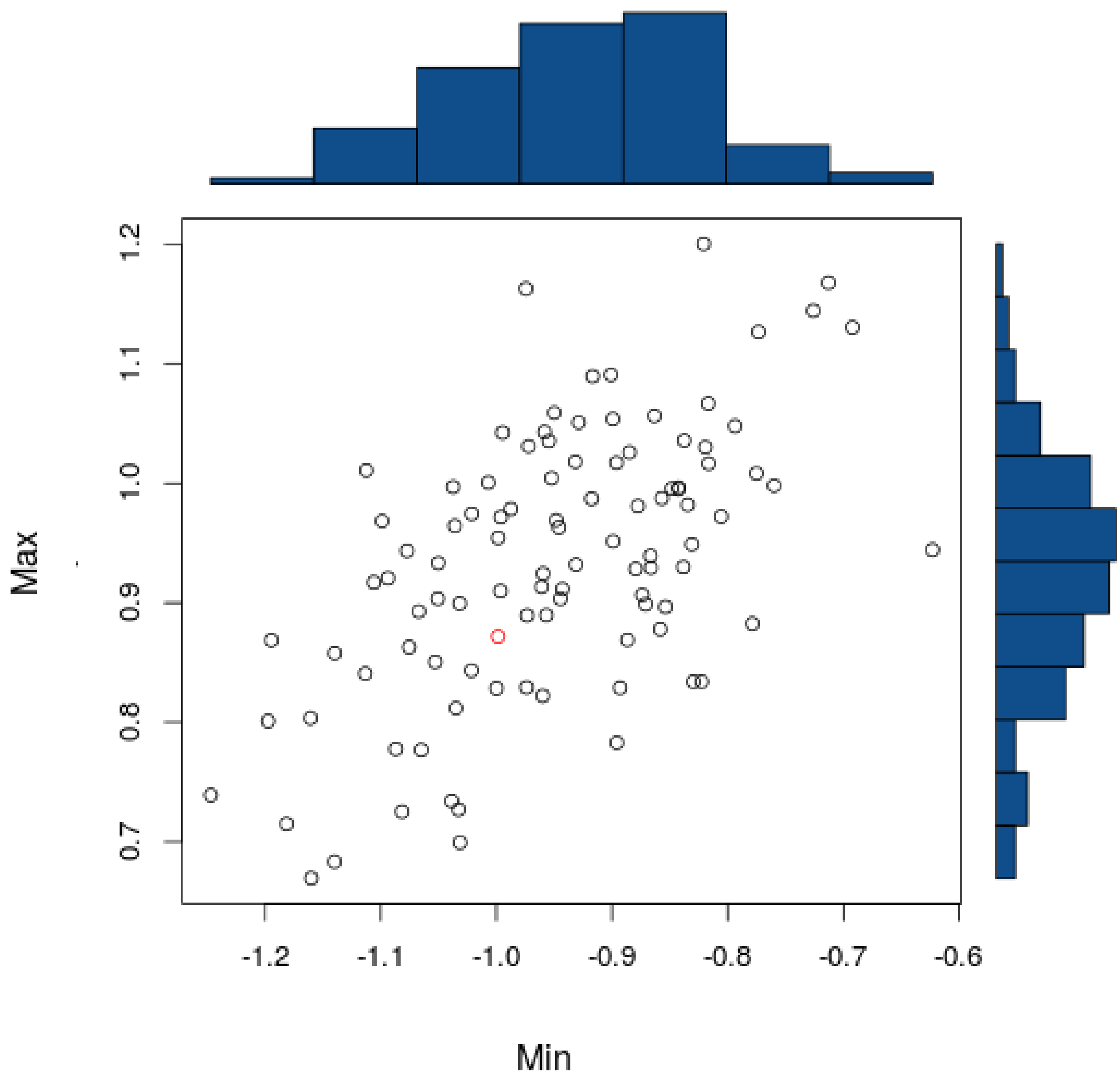
2.5 Exercise in R

2.5.1 Training data vs. predicted mean and one standard deviation



2.5.2 Simulations




Max
Min
1.2
1.1
1.0
0.9
0.8
0.7
-1.2
-1.1
-1.0
-0.9
-0.8
-0.7
-0.6