Assessment_2

2023-03-15

Kernel Methods For Regression

Part 1

Question 1

The Gaussian kernel is an example of an kernel for which the model is identifiable. The model is unidentifiable for the kernel: $k(x, y) = 1$ for $x, y \in \mathbb{R}^p$. This kernel is positive semi-definite and by the Moore-Aronszajn theorem there exists a unique RKHS for which k is the reproducing kernel. In this RKHS all the functions are constant, let's pick two functions $f_1, f_2 \in H_k$ where $f_1(x) = c$, $f_2(x) = d$ for all $x \in \mathcal{X}$. Then if we let the α we use with f_2, $\alpha_2 = \alpha_1 - d + c$, where α_1 is the α we use with f_1, then these are the same model. Hence our model is unidentifiable.

Question 2

We have for some $f \in H_k$: $f = f_1 + f_2$, for some $f_1 \in \tilde{H}, f_2 \in \tilde{H}^{\perp}$. By orthogonality:

$$||f_1 + f_2||^2 = ||f_1||^2 + ||f_2||^2$$

And by the reproducing property:

$$\frac{1}{2n}\sum_{i=1}^{n} \log f(y_i; g^{-1}(\alpha + (f_1 + f_2)(x_i^0)), \phi)$$

$$= \frac{1}{2n}\sum_{i=1}^{n} \log f(y_i; g^{-1}(\alpha + f_1(x_i^0)), \phi)$$

Combining the above,

$$\frac{1}{2n}\sum_{i=1}^{n} \log f(y_i; g^{-1}(\alpha + (f_1 + f_2)(x_i^0)), \phi) - \lambda||f_1 + f_2||^2$$

$$= \frac{1}{2n}\sum_{i=1}^{n} \log f(y_i; g^{-1}(\alpha + f_1(x_i^0)), \phi) - \lambda||f_1||^2 + \lambda||f_2||^2$$

$$\geq \frac{1}{2n}\sum_{i=1}^{n} \log f(y_i; g^{-1}(\alpha + f_1(x_i^0)), \phi) - \lambda||f_1||^2$$

Hence, we have that $\hat{f}_\lambda \in \tilde{H}$ and therefore can write $\hat{f}_\lambda$ as a linear combination of $k(x_1^0, \cdot), ..., k(x_n^0, \cdot)$ therefore we can write:

$$\hat{f}_\lambda = \sum_{i=1}^{n} \hat{\beta}_{\lambda,i} k(x_i^0, \cdot)$$

Question 3

We have that,

$$\begin{aligned}
-\lambda||f||_{H_k}^2 &= -\lambda||\sum_{i=1}^{n}\beta_{\lambda,i}k(x_i^0,\cdot)||_{H_k}^2 \\
&= -\lambda < \sum_{i-1}^{n}\beta_{\lambda,i}k(x_i^0,\cdot), \sum_{j=1}^{n}\beta_{\lambda,j}k(x_j^0,\cdot) >_k \\
&= -\lambda \sum_{i-1}^{n} < \beta_{\lambda,i}k(x_i^0,\cdot), \sum_{j=1}^{n}\beta_{\lambda,j}k(x_j^0,\cdot) >_k \\
&= -\lambda \sum_{i-1}^{n}\sum_{j=1}^{n} < \beta_{\lambda,i}\beta_{\lambda,j}k(x_i^0,\cdot), k(x_j^0,\cdot) >_k \\
&= -\lambda \sum_{i-1}^{n}\sum_{j=1}^{n} \beta_{\lambda,i}\beta_{\lambda,j}k(x_i^0,x_j^0) \\
&= -\lambda \beta_\lambda^T K \beta_\lambda
\end{aligned}$$

Where $K = [k(x_i^0, x_j^0)]_{i,j}$. Hence we can rewrite (2) as:

$$\frac{1}{2n}\sum_{i=1}^{n}\log f(y_i; g^{-1}(\alpha + \beta_\lambda^T \cdot K^{(i)}), \phi) - \lambda\beta_\lambda^T K \beta_\lambda$$

Question 4

Let $m < n + 2$, consider the optimization problem posed in the previous question, the Nyonstrom method involves reducing the dimension of the gram matrix, K, hence reducing the dimensionality of the optimization problem. We will approximate the kernel, k, by $\tilde{k}^{(m)}$ such that the matrix $\tilde{K}^{(m)}$ obtained by replacing k with $\tilde{k}^{(m)}$ has rank $\leq m$. We will let,

$$\tilde{k}^{(m)}(x, x') = k_m(x)^T (K_m)^{-1} k_m(x')$$

where K_m is the first m rows and columns of K and

$$k_m(x) = (k(x_1^0, x), ..., k(x_m^0, x))$$

Let's now rewrite f using our new kernel:

$$\begin{aligned}
f_\lambda(x) &\approx \beta_\lambda^T \tilde{k}^{(m)}(x) \\
&= \beta_\lambda^T K(X_{1:m}, X)^T (K_m^0)^{-1} K(X_{1:m}, x) \\
&= \gamma^T (K_m^0)^{-1} K(X_{1:m}, x)
\end{aligned}$$

where $K(A, B) = [k(a_i, b_j)]_{i,j}$, X is our data matrix where $X_{1:m}$ means the data matrix including only the first m datapoints and we let $\gamma^T = \beta_\lambda^T K(X_{1:m}, X)^T$. Let's now rewrite the penalty:

$$\begin{aligned}
-\lambda\beta_\lambda^T K \beta_\lambda &\approx -\lambda\beta_\lambda^T \tilde{K}^{(m)} \beta_\lambda \\
&= -\lambda\beta_\lambda^T K(X_{1:m}, X)^T (K_m^0)^{-1} K(X_{1:m}, X)\beta_\lambda \\
&= -\lambda\gamma^T (K_m^0)^{-1}\gamma
\end{aligned}$$

This leaves us with the following optimization problem,

$$\arg \max_{\alpha\in\mathbb{R}, \phi\in(0,\inf), \gamma\in\mathbb{R}^m} \frac{1}{2n}\sum_{i=1}^{n}\log f(y_i; g^{-1}(\alpha + \gamma^T (K_m^0)^{-1} K(X_{1:m}, x_i^0)), \phi) - \lambda\gamma^T (K_m^0)^{-1}\gamma$$

Question 5

Let's first find the spectral decomposition of $(K_m^0)^{-1}$,

$$(K_m^0)^{-1} = S\Lambda S^{-1}$$

Then we can write,

$$\gamma^T (K_m^0)^{-1}\gamma = \gamma^T S^{-T}\Lambda S^{-1}\gamma \quad = ||\Lambda^{\frac{1}{2}} S^{-1}\gamma||_2^2$$

Glmnet estimates parameters that minimize the following (if using the ridge penalty):

$$-\frac{1}{n}\sum_{i=1}^{n} \log f(y_i; g^{-1}(\alpha + \gamma^T X_i), \phi) + \frac{\lambda}{2}||\gamma_{glm}||_2^2$$

which is the same as maximizing optimization problem as ours, except for if we instead try to find the minimum of the negative of our optimization problem, a factor of two and if we substitute for γ_{glm}. We have that,

$$\gamma_{glm} = \Lambda^{\frac{1}{2}} S\gamma$$

$$\Rightarrow \gamma = S^{-1}\Lambda^{-\frac{1}{2}}\gamma_{glm}$$

Therefore we can find our value for γ using the estimate we get from glmnet where for X_i we use $(K_m^0)^{-1}K(X_{1:m}, x_i^0)$.

Part 2

We are going to use the wesdr dataset:

```
library(gss)
data(wesdr)
head(wesdr)
```

```
##    dur  gly  bmi ret
## 1 10.3 13.7 23.8   0
## 2  9.9 13.5 23.5   0
## 3 15.6 13.8 24.8   0
## 4 26.0 13.0 21.6   1
## 5 13.8 11.1 24.6   1
## 6 31.1 11.3 24.6   1
```

Let's now split it into a testing and training set:

```
n.test <- round(0.15 * nrow(wesdr))
test_ind <- sample(seq_len(nrow(wesdr)), size = n.test)

train <- wesdr[-test_ind, ]
test <- wesdr[test_ind, ]
```

Question 6

We are going to use a binomial distribution as we are modelling a response variable that is either 0 or 1, we are going to set $\alpha = 0$ so that we are using the ridge penalty and we will use the radial basis kernel function for which the model is identifiable.

```
library(glmnet)
```

```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.1-6
```

```r
library(kernlab)

gaussian_kernel <- function(x, y, sigma) {
  exp(-sum((x - y)^2) / (2*sigma^2))
}

fit_model <- function(X, y, lambda, sigma, m){
  n <- nrow(X)
  rbf <- rbfdot(sigma = sigma)
  K <- kernelMatrix(rbf, X)
  K_m_inverse <- solve(K[1:m, 1:m])
  K_mn <- matrix(0, m, n)
  X_m <- X[1:m,]
  for (i in 1:m) {
    for (j in 1:n) {
    K_mn[i,j] <- gaussian_kernel(X_m[i,], X[j,], sigma)
    }
  }
  input <- t(K_m_inverse %*% K_mn)
  results <- glmnet(input, y, family = "binomial" , alpha = 0, lambda = lambda)
  gamma_glm <- results$beta

  eig <- eigen(K_m_inverse)
  S <- eig$vectors
  L <- diag(eig$values)

  gamma <- solve(S) %*% sqrt(solve(L)) %*% gamma_glm
  list(gamma = gamma, results= results)
}

X <- as.matrix(train[,-4])
y <- train[,4]
head(fit_model(X, y, 0.1, 1, 50))
```

```
## $gamma
## 50 x 1 Matrix of class "dgeMatrix"
##                s0
##  [1,] -1.60440236
##  [2,]  1.91049004
##  [3,]  1.20842762
##  [4,]  0.06052861
##  [5,]  0.66606875
##  [6,] -0.46944090
##  [7,]  2.01952518
##  [8,] -0.31374185
##  [9,] -2.74610904
## [10,]  1.08667732
## [11,]  0.87790162
## [12,] -0.33901408
## [13,] -0.26918396
## [14,]  0.37501646
## [15,]  0.32290458
## [16,]  0.03496745
## [17,] -1.80486805
```

```
## [18,]  0.63346531
## [19,]  2.02958252
## [20,]  0.77003666
## [21,]  1.07216753
## [22,] -1.36466366
## [23,]  1.75448879
## [24,] -2.80198283
## [25,]  3.11027165
## [26,]  0.53932317
## [27,]  0.86212862
## [28,]  2.60839961
## [29,] -1.29322286
## [30,] -0.25718187
## [31,] -2.01565360
## [32,] -1.64046870
## [33,]  1.61240689
## [34,] -2.80782100
## [35,] -1.69175051
## [36,] -0.08286890
## [37,]  0.61514286
## [38,] -1.07512878
## [39,]  0.18112783
## [40,] -0.19490723
## [41,]  2.56773697
## [42,] -0.41486552
## [43,]  0.56424404
## [44,]  1.31427989
## [45,]  2.72301774
## [46,]  0.48488163
## [47,]  1.20385634
## [48,] -0.45302684
## [49,]  1.57948177
## [50,] -3.12751456
##
## $results
##
## Call:  glmnet(x = input, y = y, family = "binomial", alpha = 0, lambda = lambda)
##
##   Df  %Dev Lambda
## 1 50 12.37    0.1
```

Question 7

Let's fit our model using a grid of parameters to see for which parameters it works and doesnt't and to identify which (if any) parameters lead to errors.

```
# Define the parameter values to search over
lambda_range <- c(0.001, 0.01, 0.1, 0.5, 1.0, 2, 5, 10, 100, 1000, 10000)
sigma_range <- c(0.01, 0.1, 1.0)
m_range <- c(2, 5, 10, 20, 100, 400)

# Generate all combinations of parameters
param_combinations <- expand.grid(lambda_range, sigma_range, m_range)

# Loop through each parameter combination and calculate the score
```

```r
for (i in 1:nrow(param_combinations)) {
  lambda_val <- param_combinations[i, 1]
  sigma_val <- param_combinations[i, 2]
  m_val <- param_combinations[i, 3]
  cat("Current parameters: lambda=", lambda_val, ", sigma=", sigma_val, ", m=", m_val, "\n")
  # Call function to calculate score using the current parameters
  gamma <- fit_model(X, y, lambda_val, sigma_val, m_val)
}
```

```
## Current parameters: lambda= 0.001 , sigma= 0.01 , m= 2
## Current parameters: lambda= 0.01 , sigma= 0.01 , m= 2
## Current parameters: lambda= 0.1 , sigma= 0.01 , m= 2
## Current parameters: lambda= 0.5 , sigma= 0.01 , m= 2
## Current parameters: lambda= 1 , sigma= 0.01 , m= 2
## Current parameters: lambda= 2 , sigma= 0.01 , m= 2
## Current parameters: lambda= 5 , sigma= 0.01 , m= 2
## Current parameters: lambda= 10 , sigma= 0.01 , m= 2
## Current parameters: lambda= 100 , sigma= 0.01 , m= 2
## Current parameters: lambda= 1000 , sigma= 0.01 , m= 2
## Current parameters: lambda= 10000 , sigma= 0.01 , m= 2
## Current parameters: lambda= 0.001 , sigma= 0.1 , m= 2
## Current parameters: lambda= 0.01 , sigma= 0.1 , m= 2
## Current parameters: lambda= 0.1 , sigma= 0.1 , m= 2
## Current parameters: lambda= 0.5 , sigma= 0.1 , m= 2
## Current parameters: lambda= 1 , sigma= 0.1 , m= 2
## Current parameters: lambda= 2 , sigma= 0.1 , m= 2
## Current parameters: lambda= 5 , sigma= 0.1 , m= 2
## Current parameters: lambda= 10 , sigma= 0.1 , m= 2
## Current parameters: lambda= 100 , sigma= 0.1 , m= 2
## Current parameters: lambda= 1000 , sigma= 0.1 , m= 2
## Current parameters: lambda= 10000 , sigma= 0.1 , m= 2
## Current parameters: lambda= 0.001 , sigma= 1 , m= 2
## Current parameters: lambda= 0.01 , sigma= 1 , m= 2
## Current parameters: lambda= 0.1 , sigma= 1 , m= 2
## Current parameters: lambda= 0.5 , sigma= 1 , m= 2
## Current parameters: lambda= 1 , sigma= 1 , m= 2
## Current parameters: lambda= 2 , sigma= 1 , m= 2
## Current parameters: lambda= 5 , sigma= 1 , m= 2
## Current parameters: lambda= 10 , sigma= 1 , m= 2
## Current parameters: lambda= 100 , sigma= 1 , m= 2
## Current parameters: lambda= 1000 , sigma= 1 , m= 2
## Current parameters: lambda= 10000 , sigma= 1 , m= 2
## Current parameters: lambda= 0.001 , sigma= 0.01 , m= 5
## Current parameters: lambda= 0.01 , sigma= 0.01 , m= 5
## Current parameters: lambda= 0.1 , sigma= 0.01 , m= 5
## Current parameters: lambda= 0.5 , sigma= 0.01 , m= 5
## Current parameters: lambda= 1 , sigma= 0.01 , m= 5
## Current parameters: lambda= 2 , sigma= 0.01 , m= 5
## Current parameters: lambda= 5 , sigma= 0.01 , m= 5
## Current parameters: lambda= 10 , sigma= 0.01 , m= 5
## Current parameters: lambda= 100 , sigma= 0.01 , m= 5
## Current parameters: lambda= 1000 , sigma= 0.01 , m= 5
## Current parameters: lambda= 10000 , sigma= 0.01 , m= 5
## Current parameters: lambda= 0.001 , sigma= 0.1 , m= 5
```

```
## Current parameters: lambda= 0.01 , sigma= 0.1 , m= 5
## Current parameters: lambda= 0.1 , sigma= 0.1 , m= 5
## Current parameters: lambda= 0.5 , sigma= 0.1 , m= 5
## Current parameters: lambda= 1 , sigma= 0.1 , m= 5
## Current parameters: lambda= 2 , sigma= 0.1 , m= 5
## Current parameters: lambda= 5 , sigma= 0.1 , m= 5
## Current parameters: lambda= 10 , sigma= 0.1 , m= 5
## Current parameters: lambda= 100 , sigma= 0.1 , m= 5
## Current parameters: lambda= 1000 , sigma= 0.1 , m= 5
## Current parameters: lambda= 10000 , sigma= 0.1 , m= 5
## Current parameters: lambda= 0.001 , sigma= 1 , m= 5
## Current parameters: lambda= 0.01 , sigma= 1 , m= 5
## Current parameters: lambda= 0.1 , sigma= 1 , m= 5
## Current parameters: lambda= 0.5 , sigma= 1 , m= 5
## Current parameters: lambda= 1 , sigma= 1 , m= 5
## Current parameters: lambda= 2 , sigma= 1 , m= 5
## Current parameters: lambda= 5 , sigma= 1 , m= 5
## Current parameters: lambda= 10 , sigma= 1 , m= 5
## Current parameters: lambda= 100 , sigma= 1 , m= 5
## Current parameters: lambda= 1000 , sigma= 1 , m= 5
## Current parameters: lambda= 10000 , sigma= 1 , m= 5
## Current parameters: lambda= 0.001 , sigma= 0.01 , m= 10
## Current parameters: lambda= 0.01 , sigma= 0.01 , m= 10
## Current parameters: lambda= 0.1 , sigma= 0.01 , m= 10
## Current parameters: lambda= 0.5 , sigma= 0.01 , m= 10
## Current parameters: lambda= 1 , sigma= 0.01 , m= 10
## Current parameters: lambda= 2 , sigma= 0.01 , m= 10
## Current parameters: lambda= 5 , sigma= 0.01 , m= 10
## Current parameters: lambda= 10 , sigma= 0.01 , m= 10
## Current parameters: lambda= 100 , sigma= 0.01 , m= 10
## Current parameters: lambda= 1000 , sigma= 0.01 , m= 10
## Current parameters: lambda= 10000 , sigma= 0.01 , m= 10
## Current parameters: lambda= 0.001 , sigma= 0.1 , m= 10
## Current parameters: lambda= 0.01 , sigma= 0.1 , m= 10
## Current parameters: lambda= 0.1 , sigma= 0.1 , m= 10
## Current parameters: lambda= 0.5 , sigma= 0.1 , m= 10
## Current parameters: lambda= 1 , sigma= 0.1 , m= 10
## Current parameters: lambda= 2 , sigma= 0.1 , m= 10
## Current parameters: lambda= 5 , sigma= 0.1 , m= 10
## Current parameters: lambda= 10 , sigma= 0.1 , m= 10
## Current parameters: lambda= 100 , sigma= 0.1 , m= 10
## Current parameters: lambda= 1000 , sigma= 0.1 , m= 10
## Current parameters: lambda= 10000 , sigma= 0.1 , m= 10
## Current parameters: lambda= 0.001 , sigma= 1 , m= 10
## Current parameters: lambda= 0.01 , sigma= 1 , m= 10
## Current parameters: lambda= 0.1 , sigma= 1 , m= 10
## Current parameters: lambda= 0.5 , sigma= 1 , m= 10
## Current parameters: lambda= 1 , sigma= 1 , m= 10
## Current parameters: lambda= 2 , sigma= 1 , m= 10
## Current parameters: lambda= 5 , sigma= 1 , m= 10
## Current parameters: lambda= 10 , sigma= 1 , m= 10
## Current parameters: lambda= 100 , sigma= 1 , m= 10
## Current parameters: lambda= 1000 , sigma= 1 , m= 10
## Current parameters: lambda= 10000 , sigma= 1 , m= 10
```

```
## Current parameters: lambda= 0.001 , sigma= 0.01 , m= 20
## Current parameters: lambda= 0.01 , sigma= 0.01 , m= 20
## Current parameters: lambda= 0.1 , sigma= 0.01 , m= 20
## Current parameters: lambda= 0.5 , sigma= 0.01 , m= 20
## Current parameters: lambda= 1 , sigma= 0.01 , m= 20
## Current parameters: lambda= 2 , sigma= 0.01 , m= 20
## Current parameters: lambda= 5 , sigma= 0.01 , m= 20
## Current parameters: lambda= 10 , sigma= 0.01 , m= 20
## Current parameters: lambda= 100 , sigma= 0.01 , m= 20
## Current parameters: lambda= 1000 , sigma= 0.01 , m= 20
## Current parameters: lambda= 10000 , sigma= 0.01 , m= 20
## Current parameters: lambda= 0.001 , sigma= 0.1 , m= 20
## Current parameters: lambda= 0.01 , sigma= 0.1 , m= 20
## Current parameters: lambda= 0.1 , sigma= 0.1 , m= 20
## Current parameters: lambda= 0.5 , sigma= 0.1 , m= 20
## Current parameters: lambda= 1 , sigma= 0.1 , m= 20
## Current parameters: lambda= 2 , sigma= 0.1 , m= 20
## Current parameters: lambda= 5 , sigma= 0.1 , m= 20
## Current parameters: lambda= 10 , sigma= 0.1 , m= 20
## Current parameters: lambda= 100 , sigma= 0.1 , m= 20
## Current parameters: lambda= 1000 , sigma= 0.1 , m= 20
## Current parameters: lambda= 10000 , sigma= 0.1 , m= 20
## Current parameters: lambda= 0.001 , sigma= 1 , m= 20
## Current parameters: lambda= 0.01 , sigma= 1 , m= 20
## Current parameters: lambda= 0.1 , sigma= 1 , m= 20
## Current parameters: lambda= 0.5 , sigma= 1 , m= 20
## Current parameters: lambda= 1 , sigma= 1 , m= 20
## Current parameters: lambda= 2 , sigma= 1 , m= 20
## Current parameters: lambda= 5 , sigma= 1 , m= 20
## Current parameters: lambda= 10 , sigma= 1 , m= 20
## Current parameters: lambda= 100 , sigma= 1 , m= 20
## Current parameters: lambda= 1000 , sigma= 1 , m= 20
## Current parameters: lambda= 10000 , sigma= 1 , m= 20
## Current parameters: lambda= 0.001 , sigma= 0.01 , m= 100
## Current parameters: lambda= 0.01 , sigma= 0.01 , m= 100
## Current parameters: lambda= 0.1 , sigma= 0.01 , m= 100
## Current parameters: lambda= 0.5 , sigma= 0.01 , m= 100
## Current parameters: lambda= 1 , sigma= 0.01 , m= 100
## Current parameters: lambda= 2 , sigma= 0.01 , m= 100
## Current parameters: lambda= 5 , sigma= 0.01 , m= 100
## Current parameters: lambda= 10 , sigma= 0.01 , m= 100
## Current parameters: lambda= 100 , sigma= 0.01 , m= 100
## Current parameters: lambda= 1000 , sigma= 0.01 , m= 100
## Current parameters: lambda= 10000 , sigma= 0.01 , m= 100
## Current parameters: lambda= 0.001 , sigma= 0.1 , m= 100
## Current parameters: lambda= 0.01 , sigma= 0.1 , m= 100
## Current parameters: lambda= 0.1 , sigma= 0.1 , m= 100
## Current parameters: lambda= 0.5 , sigma= 0.1 , m= 100
## Current parameters: lambda= 1 , sigma= 0.1 , m= 100
## Current parameters: lambda= 2 , sigma= 0.1 , m= 100
## Current parameters: lambda= 5 , sigma= 0.1 , m= 100
## Current parameters: lambda= 10 , sigma= 0.1 , m= 100
## Current parameters: lambda= 100 , sigma= 0.1 , m= 100
## Current parameters: lambda= 1000 , sigma= 0.1 , m= 100
```

```
## Current parameters: lambda= 10000 , sigma= 0.1 , m= 100
## Current parameters: lambda= 0.001 , sigma= 1 , m= 100
## Current parameters: lambda= 0.01 , sigma= 1 , m= 100
## Current parameters: lambda= 0.1 , sigma= 1 , m= 100
## Current parameters: lambda= 0.5 , sigma= 1 , m= 100
## Current parameters: lambda= 1 , sigma= 1 , m= 100
## Current parameters: lambda= 2 , sigma= 1 , m= 100
## Current parameters: lambda= 5 , sigma= 1 , m= 100
## Current parameters: lambda= 10 , sigma= 1 , m= 100
## Current parameters: lambda= 100 , sigma= 1 , m= 100
## Current parameters: lambda= 1000 , sigma= 1 , m= 100
## Current parameters: lambda= 10000 , sigma= 1 , m= 100
## Current parameters: lambda= 0.001 , sigma= 0.01 , m= 400
```

```
## Error in solve.default(K[1:m, 1:m]): system is computationally singular: reciprocal condition number
```

From the above we see that it works for a large range of values for lambda and sigma, however, when m get's too large we are getting an error. This is happening when we try and invert our matrix K_m, we can fix this by adding some small ϵ to the diagonal of this matrix, let's modify our function to include this:

```
fit_model <- function(X, y, lambda, sigma, m, eps=0.0001){
  n <- nrow(X)
  rbf <- rbfdot(sigma = sigma)
  K <- kernelMatrix(rbf, X)
  K_m_inverse <- solve(K[1:m, 1:m] + diag(m) * eps)
  K_mn <- matrix(0, m, n)
  X_m <- X[1:m,]
  for (i in 1:m) {
    for (j in 1:n) {
    K_mn[i,j] <- gaussian_kernel(X_m[i,], X[j,], sigma)
    }
  }
  input <- t(K_m_inverse %*% K_mn)
  results <- glmnet(input, y, family = "binomial" , alpha = 0, lambda = lambda)
  gamma_glm <- results$beta

  eig <- eigen(K_m_inverse)
  S <- eig$vectors
  L <- diag(eig$values)

  gamma <- solve(S) %*% sqrt(solve(L)) %*% gamma_glm
  list(gamma = gamma, results= results)
}
```

Let's now try our grid search to see if our function now works for large values of m:

```
# Define the parameter values to search over
lambda_range <- c(0.001, 0.01, 0.1, 0.5, 1.0, 2, 5, 10, 100, 1000, 10000)
sigma_range <- c(0.01, 0.1, 1.0)
m_range <- c(2, 5, 10, 20, 100, 400)

# Generate all combinations of parameters
param_combinations <- expand.grid(lambda_range, sigma_range, m_range)

# Loop through each parameter combination and calculate the score
for (i in 1:nrow(param_combinations)) {
```

```
  lambda_val <- param_combinations[i, 1]
  sigma_val <- param_combinations[i, 2]
  m_val <- param_combinations[i, 3]
  cat("Current parameters: lambda=", lambda_val, ", sigma=", sigma_val, ", m=", m_val, "\n")
  # Call function to calculate score using the current parameters
  gamma <- fit_model(X, y, lambda_val, sigma_val, m_val)
}
```

```
## Current parameters: lambda= 0.001 , sigma= 0.01 , m= 2
## Current parameters: lambda= 0.01 , sigma= 0.01 , m= 2
## Current parameters: lambda= 0.1 , sigma= 0.01 , m= 2
## Current parameters: lambda= 0.5 , sigma= 0.01 , m= 2
## Current parameters: lambda= 1 , sigma= 0.01 , m= 2
## Current parameters: lambda= 2 , sigma= 0.01 , m= 2
## Current parameters: lambda= 5 , sigma= 0.01 , m= 2
## Current parameters: lambda= 10 , sigma= 0.01 , m= 2
## Current parameters: lambda= 100 , sigma= 0.01 , m= 2
## Current parameters: lambda= 1000 , sigma= 0.01 , m= 2
## Current parameters: lambda= 10000 , sigma= 0.01 , m= 2
## Current parameters: lambda= 0.001 , sigma= 0.1 , m= 2
## Current parameters: lambda= 0.01 , sigma= 0.1 , m= 2
## Current parameters: lambda= 0.1 , sigma= 0.1 , m= 2
## Current parameters: lambda= 0.5 , sigma= 0.1 , m= 2
## Current parameters: lambda= 1 , sigma= 0.1 , m= 2
## Current parameters: lambda= 2 , sigma= 0.1 , m= 2
## Current parameters: lambda= 5 , sigma= 0.1 , m= 2
## Current parameters: lambda= 10 , sigma= 0.1 , m= 2
## Current parameters: lambda= 100 , sigma= 0.1 , m= 2
## Current parameters: lambda= 1000 , sigma= 0.1 , m= 2
## Current parameters: lambda= 10000 , sigma= 0.1 , m= 2
## Current parameters: lambda= 0.001 , sigma= 1 , m= 2
## Current parameters: lambda= 0.01 , sigma= 1 , m= 2
## Current parameters: lambda= 0.1 , sigma= 1 , m= 2
## Current parameters: lambda= 0.5 , sigma= 1 , m= 2
## Current parameters: lambda= 1 , sigma= 1 , m= 2
## Current parameters: lambda= 2 , sigma= 1 , m= 2
## Current parameters: lambda= 5 , sigma= 1 , m= 2
## Current parameters: lambda= 10 , sigma= 1 , m= 2
## Current parameters: lambda= 100 , sigma= 1 , m= 2
## Current parameters: lambda= 1000 , sigma= 1 , m= 2
## Current parameters: lambda= 10000 , sigma= 1 , m= 2
## Current parameters: lambda= 0.001 , sigma= 0.01 , m= 5
## Current parameters: lambda= 0.01 , sigma= 0.01 , m= 5
## Current parameters: lambda= 0.1 , sigma= 0.01 , m= 5
## Current parameters: lambda= 0.5 , sigma= 0.01 , m= 5
## Current parameters: lambda= 1 , sigma= 0.01 , m= 5
## Current parameters: lambda= 2 , sigma= 0.01 , m= 5
## Current parameters: lambda= 5 , sigma= 0.01 , m= 5
## Current parameters: lambda= 10 , sigma= 0.01 , m= 5
## Current parameters: lambda= 100 , sigma= 0.01 , m= 5
## Current parameters: lambda= 1000 , sigma= 0.01 , m= 5
## Current parameters: lambda= 10000 , sigma= 0.01 , m= 5
## Current parameters: lambda= 0.001 , sigma= 0.1 , m= 5
## Current parameters: lambda= 0.01 , sigma= 0.1 , m= 5
```

```
## Current parameters: lambda= 0.1 , sigma= 0.1 , m= 5
## Current parameters: lambda= 0.5 , sigma= 0.1 , m= 5
## Current parameters: lambda= 1 , sigma= 0.1 , m= 5
## Current parameters: lambda= 2 , sigma= 0.1 , m= 5
## Current parameters: lambda= 5 , sigma= 0.1 , m= 5
## Current parameters: lambda= 10 , sigma= 0.1 , m= 5
## Current parameters: lambda= 100 , sigma= 0.1 , m= 5
## Current parameters: lambda= 1000 , sigma= 0.1 , m= 5
## Current parameters: lambda= 10000 , sigma= 0.1 , m= 5
## Current parameters: lambda= 0.001 , sigma= 1 , m= 5
## Current parameters: lambda= 0.01 , sigma= 1 , m= 5
## Current parameters: lambda= 0.1 , sigma= 1 , m= 5
## Current parameters: lambda= 0.5 , sigma= 1 , m= 5
## Current parameters: lambda= 1 , sigma= 1 , m= 5
## Current parameters: lambda= 2 , sigma= 1 , m= 5
## Current parameters: lambda= 5 , sigma= 1 , m= 5
## Current parameters: lambda= 10 , sigma= 1 , m= 5
## Current parameters: lambda= 100 , sigma= 1 , m= 5
## Current parameters: lambda= 1000 , sigma= 1 , m= 5
## Current parameters: lambda= 10000 , sigma= 1 , m= 5
## Current parameters: lambda= 0.001 , sigma= 0.01 , m= 10
## Current parameters: lambda= 0.01 , sigma= 0.01 , m= 10
## Current parameters: lambda= 0.1 , sigma= 0.01 , m= 10
## Current parameters: lambda= 0.5 , sigma= 0.01 , m= 10
## Current parameters: lambda= 1 , sigma= 0.01 , m= 10
## Current parameters: lambda= 2 , sigma= 0.01 , m= 10
## Current parameters: lambda= 5 , sigma= 0.01 , m= 10
## Current parameters: lambda= 10 , sigma= 0.01 , m= 10
## Current parameters: lambda= 100 , sigma= 0.01 , m= 10
## Current parameters: lambda= 1000 , sigma= 0.01 , m= 10
## Current parameters: lambda= 10000 , sigma= 0.01 , m= 10
## Current parameters: lambda= 0.001 , sigma= 0.1 , m= 10
## Current parameters: lambda= 0.01 , sigma= 0.1 , m= 10
## Current parameters: lambda= 0.1 , sigma= 0.1 , m= 10
## Current parameters: lambda= 0.5 , sigma= 0.1 , m= 10
## Current parameters: lambda= 1 , sigma= 0.1 , m= 10
## Current parameters: lambda= 2 , sigma= 0.1 , m= 10
## Current parameters: lambda= 5 , sigma= 0.1 , m= 10
## Current parameters: lambda= 10 , sigma= 0.1 , m= 10
## Current parameters: lambda= 100 , sigma= 0.1 , m= 10
## Current parameters: lambda= 1000 , sigma= 0.1 , m= 10
## Current parameters: lambda= 10000 , sigma= 0.1 , m= 10
## Current parameters: lambda= 0.001 , sigma= 1 , m= 10
## Current parameters: lambda= 0.01 , sigma= 1 , m= 10
## Current parameters: lambda= 0.1 , sigma= 1 , m= 10
## Current parameters: lambda= 0.5 , sigma= 1 , m= 10
## Current parameters: lambda= 1 , sigma= 1 , m= 10
## Current parameters: lambda= 2 , sigma= 1 , m= 10
## Current parameters: lambda= 5 , sigma= 1 , m= 10
## Current parameters: lambda= 10 , sigma= 1 , m= 10
## Current parameters: lambda= 100 , sigma= 1 , m= 10
## Current parameters: lambda= 1000 , sigma= 1 , m= 10
## Current parameters: lambda= 10000 , sigma= 1 , m= 10
## Current parameters: lambda= 0.001 , sigma= 0.01 , m= 20
```

```
## Current parameters: lambda= 0.01 , sigma= 0.01 , m= 20
## Current parameters: lambda= 0.1 , sigma= 0.01 , m= 20
## Current parameters: lambda= 0.5 , sigma= 0.01 , m= 20
## Current parameters: lambda= 1 , sigma= 0.01 , m= 20
## Current parameters: lambda= 2 , sigma= 0.01 , m= 20
## Current parameters: lambda= 5 , sigma= 0.01 , m= 20
## Current parameters: lambda= 10 , sigma= 0.01 , m= 20
## Current parameters: lambda= 100 , sigma= 0.01 , m= 20
## Current parameters: lambda= 1000 , sigma= 0.01 , m= 20
## Current parameters: lambda= 10000 , sigma= 0.01 , m= 20
## Current parameters: lambda= 0.001 , sigma= 0.1 , m= 20
## Current parameters: lambda= 0.01 , sigma= 0.1 , m= 20
## Current parameters: lambda= 0.1 , sigma= 0.1 , m= 20
## Current parameters: lambda= 0.5 , sigma= 0.1 , m= 20
## Current parameters: lambda= 1 , sigma= 0.1 , m= 20
## Current parameters: lambda= 2 , sigma= 0.1 , m= 20
## Current parameters: lambda= 5 , sigma= 0.1 , m= 20
## Current parameters: lambda= 10 , sigma= 0.1 , m= 20
## Current parameters: lambda= 100 , sigma= 0.1 , m= 20
## Current parameters: lambda= 1000 , sigma= 0.1 , m= 20
## Current parameters: lambda= 10000 , sigma= 0.1 , m= 20
## Current parameters: lambda= 0.001 , sigma= 1 , m= 20
## Current parameters: lambda= 0.01 , sigma= 1 , m= 20
## Current parameters: lambda= 0.1 , sigma= 1 , m= 20
## Current parameters: lambda= 0.5 , sigma= 1 , m= 20
## Current parameters: lambda= 1 , sigma= 1 , m= 20
## Current parameters: lambda= 2 , sigma= 1 , m= 20
## Current parameters: lambda= 5 , sigma= 1 , m= 20
## Current parameters: lambda= 10 , sigma= 1 , m= 20
## Current parameters: lambda= 100 , sigma= 1 , m= 20
## Current parameters: lambda= 1000 , sigma= 1 , m= 20
## Current parameters: lambda= 10000 , sigma= 1 , m= 20
## Current parameters: lambda= 0.001 , sigma= 0.01 , m= 100
## Current parameters: lambda= 0.01 , sigma= 0.01 , m= 100
## Current parameters: lambda= 0.1 , sigma= 0.01 , m= 100
## Current parameters: lambda= 0.5 , sigma= 0.01 , m= 100
## Current parameters: lambda= 1 , sigma= 0.01 , m= 100
## Current parameters: lambda= 2 , sigma= 0.01 , m= 100
## Current parameters: lambda= 5 , sigma= 0.01 , m= 100
## Current parameters: lambda= 10 , sigma= 0.01 , m= 100
## Current parameters: lambda= 100 , sigma= 0.01 , m= 100
## Current parameters: lambda= 1000 , sigma= 0.01 , m= 100
## Current parameters: lambda= 10000 , sigma= 0.01 , m= 100
## Current parameters: lambda= 0.001 , sigma= 0.1 , m= 100
## Current parameters: lambda= 0.01 , sigma= 0.1 , m= 100
## Current parameters: lambda= 0.1 , sigma= 0.1 , m= 100
## Current parameters: lambda= 0.5 , sigma= 0.1 , m= 100
## Current parameters: lambda= 1 , sigma= 0.1 , m= 100
## Current parameters: lambda= 2 , sigma= 0.1 , m= 100
## Current parameters: lambda= 5 , sigma= 0.1 , m= 100
## Current parameters: lambda= 10 , sigma= 0.1 , m= 100
## Current parameters: lambda= 100 , sigma= 0.1 , m= 100
## Current parameters: lambda= 1000 , sigma= 0.1 , m= 100
## Current parameters: lambda= 10000 , sigma= 0.1 , m= 100
```

```
## Current parameters: lambda= 0.001 , sigma= 1 , m= 100
## Current parameters: lambda= 0.01 , sigma= 1 , m= 100
## Current parameters: lambda= 0.1 , sigma= 1 , m= 100
## Current parameters: lambda= 0.5 , sigma= 1 , m= 100
## Current parameters: lambda= 1 , sigma= 1 , m= 100
## Current parameters: lambda= 2 , sigma= 1 , m= 100
## Current parameters: lambda= 5 , sigma= 1 , m= 100
## Current parameters: lambda= 10 , sigma= 1 , m= 100
## Current parameters: lambda= 100 , sigma= 1 , m= 100
## Current parameters: lambda= 1000 , sigma= 1 , m= 100
## Current parameters: lambda= 10000 , sigma= 1 , m= 100
## Current parameters: lambda= 0.001 , sigma= 0.01 , m= 400
## Current parameters: lambda= 0.01 , sigma= 0.01 , m= 400
## Current parameters: lambda= 0.1 , sigma= 0.01 , m= 400
## Current parameters: lambda= 0.5 , sigma= 0.01 , m= 400
## Current parameters: lambda= 1 , sigma= 0.01 , m= 400
## Current parameters: lambda= 2 , sigma= 0.01 , m= 400
## Current parameters: lambda= 5 , sigma= 0.01 , m= 400
## Current parameters: lambda= 10 , sigma= 0.01 , m= 400
## Current parameters: lambda= 100 , sigma= 0.01 , m= 400
## Current parameters: lambda= 1000 , sigma= 0.01 , m= 400
## Current parameters: lambda= 10000 , sigma= 0.01 , m= 400
## Current parameters: lambda= 0.001 , sigma= 0.1 , m= 400
## Current parameters: lambda= 0.01 , sigma= 0.1 , m= 400
## Current parameters: lambda= 0.1 , sigma= 0.1 , m= 400
## Current parameters: lambda= 0.5 , sigma= 0.1 , m= 400
## Current parameters: lambda= 1 , sigma= 0.1 , m= 400
## Current parameters: lambda= 2 , sigma= 0.1 , m= 400
## Current parameters: lambda= 5 , sigma= 0.1 , m= 400
## Current parameters: lambda= 10 , sigma= 0.1 , m= 400
## Current parameters: lambda= 100 , sigma= 0.1 , m= 400
## Current parameters: lambda= 1000 , sigma= 0.1 , m= 400
## Current parameters: lambda= 10000 , sigma= 0.1 , m= 400
## Current parameters: lambda= 0.001 , sigma= 1 , m= 400
## Current parameters: lambda= 0.01 , sigma= 1 , m= 400
## Current parameters: lambda= 0.1 , sigma= 1 , m= 400
## Current parameters: lambda= 0.5 , sigma= 1 , m= 400
## Current parameters: lambda= 1 , sigma= 1 , m= 400
## Current parameters: lambda= 2 , sigma= 1 , m= 400
## Current parameters: lambda= 5 , sigma= 1 , m= 400
## Current parameters: lambda= 10 , sigma= 1 , m= 400
## Current parameters: lambda= 100 , sigma= 1 , m= 400
## Current parameters: lambda= 1000 , sigma= 1 , m= 400
## Current parameters: lambda= 10000 , sigma= 1 , m= 400
```

It does! we have solved our problem.

Question 8

This function will compute the approximate solution to (2) for any $c \in C$ and integer $m \leq n + 2$ where lambda is chosen using 10-fold cross validation using the missclassification error, here is the function and it ran on an example so we can check it works:

```
fit_model_cv_l <- function(X, y, sigma, m, eps=0.0001){
  n <- nrow(X)
```

```
  rbf <- rbfdot(sigma = sigma)
  K <- kernelMatrix(rbf, X)
  K_m_inverse <- solve(K[1:m, 1:m] + diag(m) * eps)
  K_mn <- matrix(0, m, n)
  X_m <- X[1:m,]
  for (i in 1:m) {
    for (j in 1:n) {
    K_mn[i,j] <- gaussian_kernel(X_m[i,], X[j,], sigma)
    }
  }
  input <- t(K_m_inverse %*% K_mn)
  cv <- cv.glmnet(input, y, family = "binomial" , alpha = 0)
  lambda <- cv$lambda.min
  fit_model(X, y, lambda, sigma, m)
}

X <- as.matrix(train[,-4])
y <- train[,4]
head(fit_model_cv_l(X, y, 1, 50))
```

```
## $gamma
## 50 x 1 Matrix of class "dgeMatrix"
##                 s0
##  [1,] -0.462820836
##  [2,]  0.544547657
##  [3,]  0.425338129
##  [4,] -0.028036385
##  [5,]  0.157677908
##  [6,] -0.016263649
##  [7,]  0.587411472
##  [8,] -0.104811440
##  [9,] -0.759762262
## [10,]  0.310796763
## [11,]  0.305467560
## [12,] -0.082385046
## [13,]  0.073586168
## [14,]  0.192698613
## [15,]  0.152727165
## [16,]  0.096630815
## [17,] -0.560821716
## [18,]  0.227131256
## [19,]  0.652098766
## [20,]  0.266057458
## [21,]  0.375983270
## [22,] -0.464892968
## [23,] -0.520344105
## [24,] -0.888896221
## [25,]  0.795010099
## [26,] -0.319793781
## [27,]  0.346037199
## [28,] -0.094992874
## [29,]  1.020163420
## [30,] -0.103809474
## [31,] -0.687727390
```

```
## [32,] -0.531075087
## [33,]  0.559736015
## [34,] -0.876601438
## [35,] -0.617213273
## [36,] -0.002724833
## [37,]  0.150899788
## [38,] -0.384154937
## [39,] -0.089217641
## [40,]  0.079449029
## [41,] -0.812630340
## [42,]  0.134488197
## [43,]  0.204006779
## [44,]  0.527285532
## [45,] -1.030757128
## [46,]  0.219829510
## [47,]  0.490800236
## [48,] -0.114128234
## [49,]  0.614800959
## [50,] -1.107081754
##
## $results
##
## Call:  glmnet(x = input, y = y, family = "binomial", alpha = 0, lambda = lambda)
##
##   Df %Dev Lambda
## 1 50 5.71 0.7098
```

Question 9

Let's now make it so that sigma is also chosen by cross-validation, here is the function:

```
predict_our_model <- function(fit, train_X, test_X, sigma, m, eps=0.0001){
  n <- nrow(test_X)
  rbf <- rbfdot(sigma = sigma)
  K <- kernelMatrix(rbf, train_X)
  K_m_inverse <- solve(K[1:m, 1:m] + diag(m) * eps)
  K_mn <- matrix(0, m, n)
  X_m <- train_X[1:m,]
  for (i in 1:m) {
    for (j in 1:n) {
    K_mn[i,j] <- gaussian_kernel(X_m[i,], test_X[j,], sigma)
    }
  }
  input <- t(K_m_inverse %*% K_mn)

  out <- predict(fit$results, newx=input)
  ifelse(out>0.5, 1, 0)
}

fit_model_cv_l_sigma <- function(X, y, m, sigma_list=c(0.01, 0.1, 1.0)) {

  # initialize variables to store the best parameter and misclassification error
  best_param <- NULL
  best_error <- Inf
  best_fit <- NULL
```

```
  # loop over the parameter values
  for (i in 1:length(sigma_list)) {

    # set the parameter value
    sigma <- sigma_list[i]

    # carry out cross-validation using the misclassification error
    folds <- cut(seq(1,nrow(X)), breaks=10, labels=FALSE)
    misclass_error <- rep(NA, 10)
    for (j in 1:10) {
      test_idx <- which(folds == j)
      train_idx <- which(folds != j)
      train_data <- X[train_idx, ]
      test_data <- X[test_idx, ]
      train_y <- y[train_idx]
      test_y <- y[test_idx]
      fit <- fit_model_cv_l(train_data, train_y, sigma, m)
      pred <- predict_our_model(fit, X, test_data, sigma, m)
      misclass_error[j] <- mean(pred != test_y)
    }
    mean_misclass_error <- mean(misclass_error)

    # check if the current parameter gives a lower mean misclassification error than the previous best
    if (mean_misclass_error < best_error) {
      best_param <- sigma_list[i]
      best_error <- mean_misclass_error
      best_fit <- fit
    }
  }

  # return the best parameter value and misclassification error as a list
  best_fit[["sigma"]] = best_param
  best_fit[["MisclassError"]] = best_error
  return(best_fit)
}

X <- as.matrix(train[,-4])
y <- train[,4]
head(fit_model_cv_l_sigma(X, y, 50))
```

```
## $gamma
## 50 x 1 Matrix of class "dgeMatrix"
##                   s0
##  [1,] -2.251361e-06
##  [2,]  3.976049e-06
##  [3,]  5.603991e-06
##  [4,]  1.684639e-06
##  [5,]  1.397109e-06
##  [6,]  7.483336e-06
##  [7,]  1.310969e-06
##  [8,]  3.186262e-06
##  [9,]  2.609344e-06
## [10,]  1.877686e-06
```

```
## [11,]  1.113746e-06
## [12,]  5.556009e-06
## [13,]  1.510892e-05
## [14,]  1.193626e-06
## [15,]  1.586852e-05
## [16,] -8.665329e-07
## [17,]  7.255815e-06
## [18,]  5.815453e-06
## [19,]  2.039161e-05
## [20,]  1.783108e-05
## [21,] -3.533980e-05
## [22,] -1.473606e-06
## [23,] -1.163293e-05
## [24,]  5.975308e-06
## [25,] -1.547631e-05
## [26,] -5.089571e-08
## [27,]  3.753316e-05
## [28,] -2.036002e-05
## [29,] -3.605203e-05
## [30,]  2.130693e-05
## [31,] -4.685337e-06
## [32,] -1.437389e-05
## [33,]  5.401744e-05
## [34,] -7.215700e-05
## [35,]  2.640436e-06
## [36,] -1.614215e-05
## [37,] -7.187167e-05
## [38,]  5.119470e-05
## [39,] -1.545914e-05
## [40,]  1.258507e-05
## [41,]  2.557927e-05
## [42,] -2.951897e-05
## [43,] -7.769601e-06
## [44,]  3.702747e-06
## [45,]  3.261114e-05
## [46,]  1.881216e-05
## [47,]  1.731062e-05
## [48,]  1.938809e-06
## [49,] -6.834239e-06
## [50,]  1.545074e-06
##
## $results
##
## Call:  glmnet(x = input, y = y, family = "binomial", alpha = 0, lambda = lambda)
##
##   Df %Dev Lambda
## 1 50 0.04  40.12
##
## $sigma
## [1] 0.01
##
## $MisclassError
## [1] 0.4078947
```

Again we run it on an example to check it works.

Question 10

Let's now fit our model for different values of m and calculate the error on the test set:

```
X_test <- as.matrix(test[,-4])
y_test <- test[,4]

m_list <- c(10, 25, 50, 100, 200, 400)
misclass_error <- rep(NA, length(m_list))
for(i in 1:length(m_list)){
  m <- m_list[i]
  results <- fit_model_cv_l_sigma(X, y, m)
  pred <- predict_our_model(results, X, X_test, results$sigma, m)
  misclass_error[i] <- mean(pred != y_test)
}
```

```
## Error in solve(S) %*% sqrt(solve(L)) %*% gamma_glm: invalid class 'NA' to dup_mMatrix_as_dgeMatrix
```

```
m_list
```

```
## [1]  10  25  50 100 200 400
```

```
misclass_error
```

```
## [1] 0.46 0.46 0.46 0.46 0.46   NA
```

We see that for different values of m we are getting similar values for the missclassification error (basically the same), also note that this misclassification error is not particularly small. This suggests that if lambda and sigma are chosen well, then the choice of m is not particularly important. Which means we should set m small when trying to fit our model as we will achieve similar performance as if we use a large m but with less computation needed.

Question 11

Let's now find the GAM estimate of model (1), where we will choose the penalty parameters, $\{\lambda_j\}_{j=1}^p$, using Generalized Cross Validation (GCV) which is the method gam uses by default:

```
library(mgcv)
```

```
## Loading required package: nlme
```

```
## This is mgcv 1.8-42. For overview type 'help("mgcv-package")'.
```

```
fit <- gam(ret~s(dur)+s(gly)+s(bmi), data = train)
preds <- predict(fit, newdata = test[,-4])
preds <- ifelse(preds >0.5, 1, 0)
misclass_error <- mean(pred != test[,4])
misclass_error
```

```
## [1] 0.46
```

This missclassification error is the same as that achieved by our model, with much less computation involved partly down to the smaller number of hyper-parameters, which are chosen by cross validation (which is costly).