Introduction to Machine Learning

Lab 3 Block 2

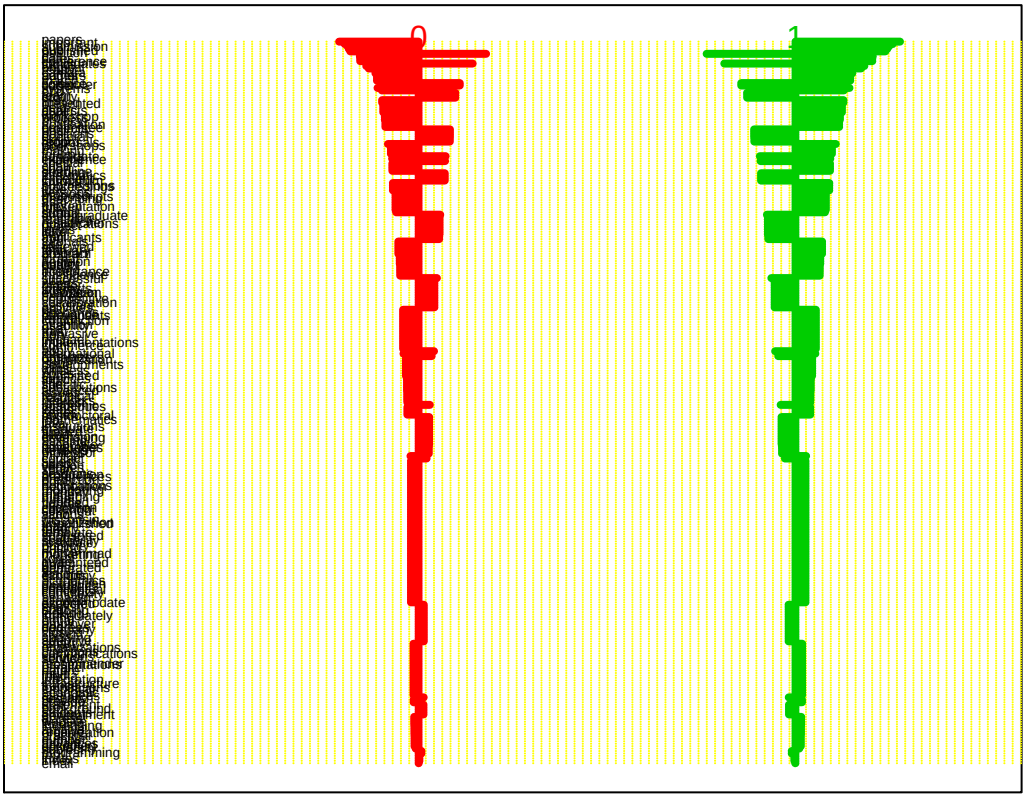
Rasmus Holm

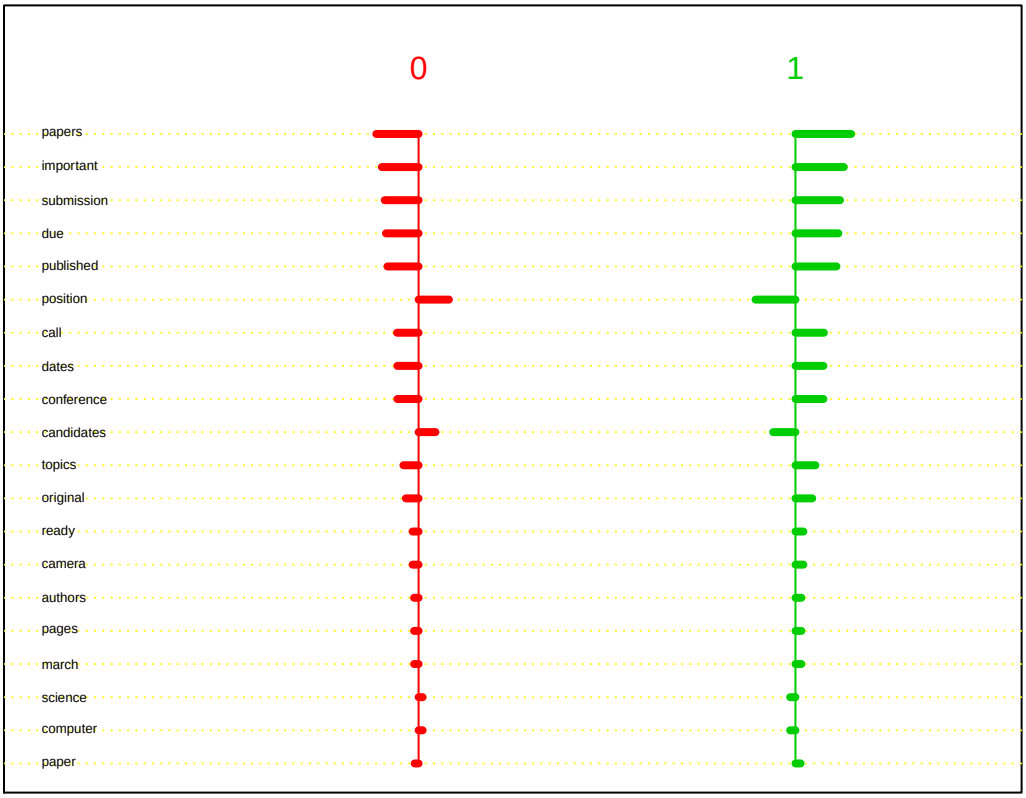
2016-12-14

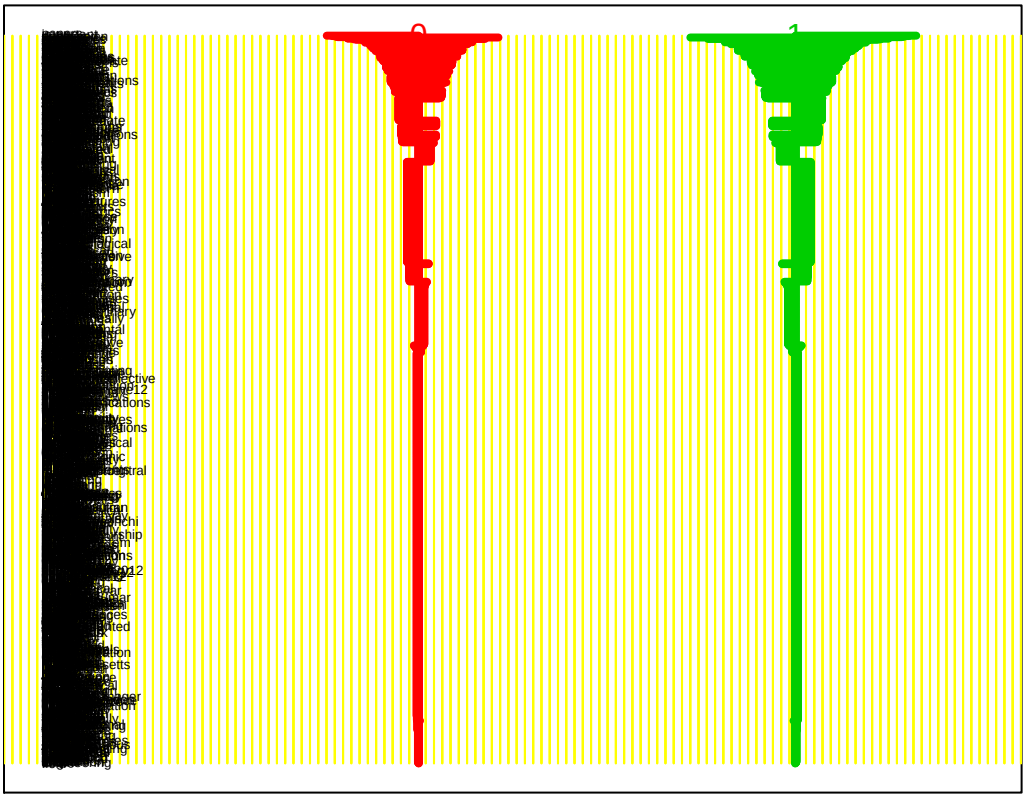
Contents

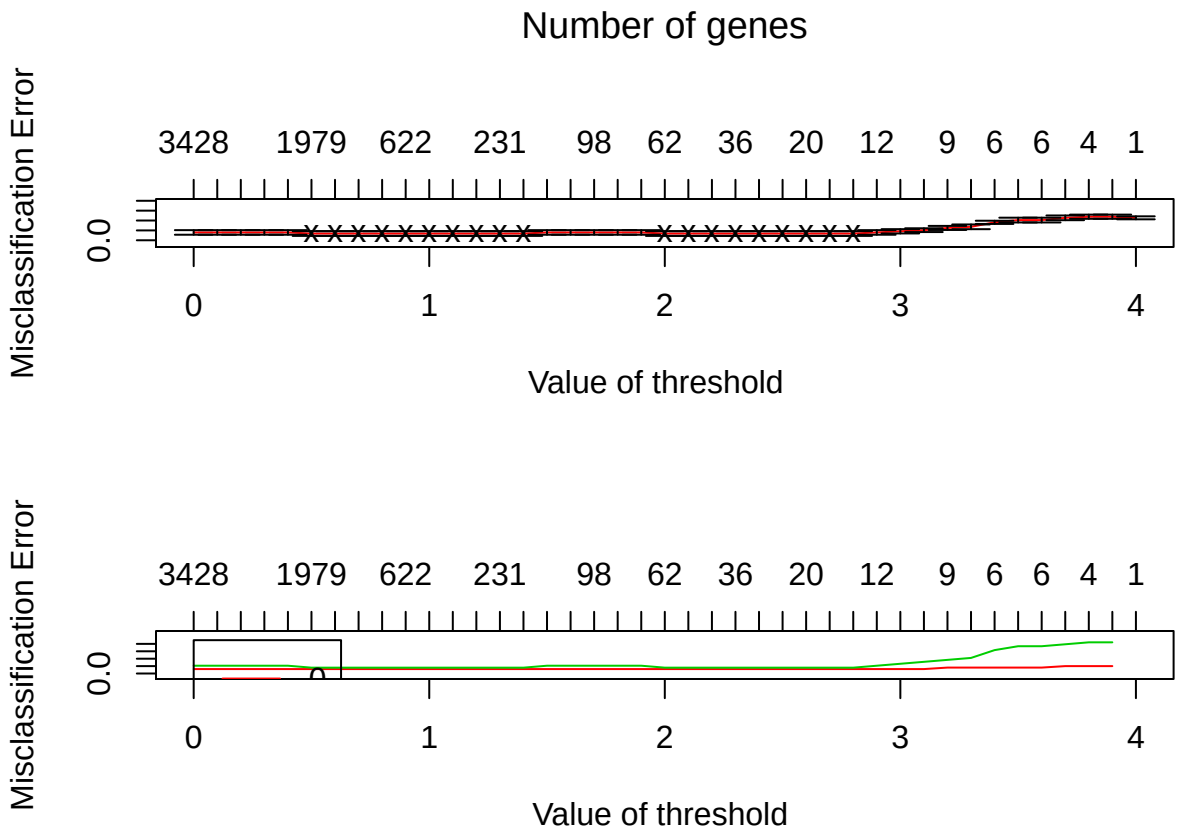
Assignment 1

1




0
1
papers
important
submission
due
published
position
call
dates
conference
candidates
topics
original
ready
camera
authors
pages
march
science
computer
paper


0
1



2

Elastic Net

```
#> [1] 0.1311628
#> s35
#>  38
#> [1] 0.15
```

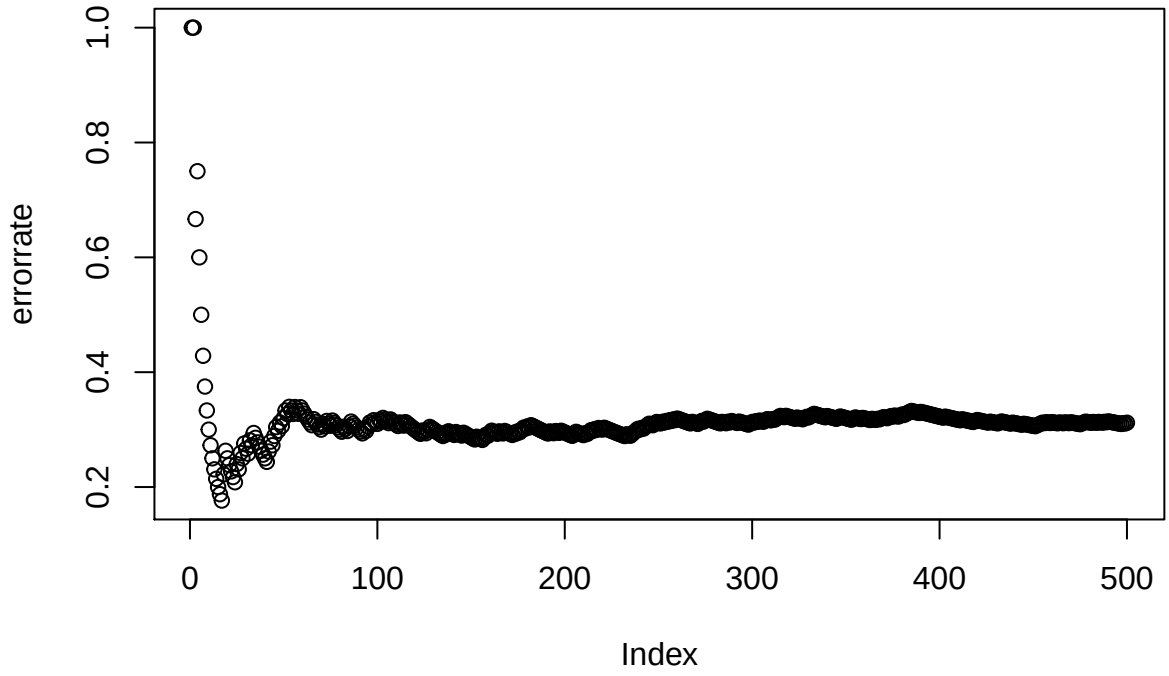
Support Vector Machine

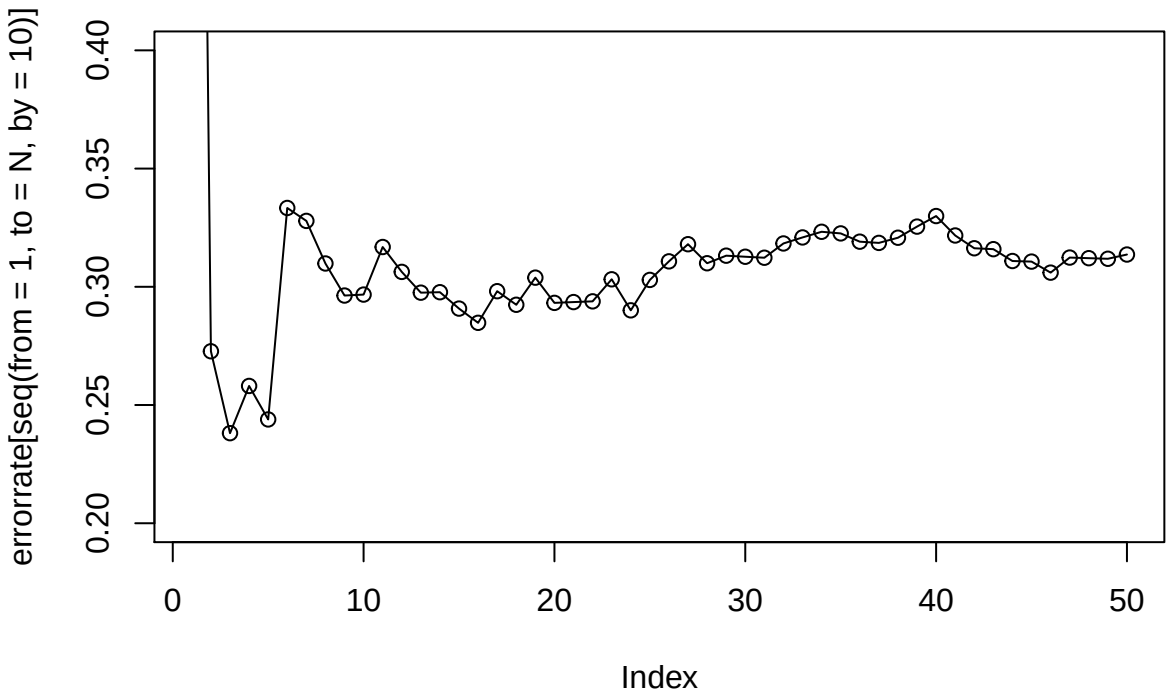
```
#>  Setting default kernel parameters
#> [1] 43
#> [1] 0.05
```

3

Assignment 2

```
#>   [1] 1.0000000 1.0000000 0.6666667 0.7500000 0.6000000 0.5000000 0.4285714 0.3750000 0.3333333 0.300
#>  [50] 0.3200000 0.3333333 0.3269231 0.3396226 0.3333333 0.3272727 0.3392857 0.3333333 0.3275862 0.338
#>  [99] 0.3131313 0.3100000 0.3168317 0.3137255 0.3203883 0.3173077 0.3142857 0.3113208 0.3177570 0.314
#> [148] 0.2905405 0.2885906 0.2866667 0.2847682 0.2828947 0.2875817 0.2857143 0.2838710 0.2820513 0.280
#> [197] 0.2944162 0.2979798 0.2964824 0.2950000 0.2935323 0.2920792 0.2906404 0.2892157 0.2926829 0.290
#> [246] 0.3089431 0.3076923 0.3104839 0.3132530 0.3120000 0.3107570 0.3134921 0.3122530 0.3149606 0.313
#> [295] 0.3118644 0.3108108 0.3097643 0.3087248 0.3110368 0.3133333 0.3122924 0.3145695 0.3135314 0.315
#> [344] 0.3197674 0.3188406 0.3208092 0.3227666 0.3218391 0.3209169 0.3200000 0.3190883 0.3181818 0.317
#> [393] 0.3282443 0.3274112 0.3265823 0.3257576 0.3249370 0.3241206 0.3233083 0.3225000 0.3216958 0.320
#> [442] 0.3099548 0.3092551 0.3085586 0.3101124 0.3094170 0.3087248 0.3080357 0.3073497 0.3066667 0.305
#> [491] 0.3136456 0.3130081 0.3123732 0.3117409 0.3111111 0.3104839 0.3098592 0.3112450 0.3106212 0.311
```




errorrate[seq(from = 1, to = N, by = 10)]
0.20
0.25
0.30
0.35
0.40
0
10
20
30
40
50
Index

```
#> [1] 20
#> [1] 0.312
```

Appendix

Code for Assignment 1

```
library(pamr)
library(glmnet)
library(kernlab)

data <- read.csv("../data/data.csv", sep=";", header=TRUE,
                 stringsAsFactors=FALSE, encoding="latin1")
rownames(data) <- 1:nrow(data)

set.seed(12345)
train_idx <- sample(nrow(data), size=floor(nrow(data) * 7 / 10))
train <- data[train_idx,]
test <- data[-train_idx,]

x <- t(train[, -ncol(data)])
y <- train[, ncol(data)]

x_test <- t(test[, -ncol(data)])
y_test <- test[, ncol(data)]
set.seed(12345)

nsc_data <- list(x=x, y=as.factor(y), geneid=as.character(1:nrow(x)), genenames=rownames(x))
model <- pamr.train(nsc_data, threshold=seq(0,4, 0.1))

cvmodel <- pamr.cv(model, nsc_data)

optimal_threshold <- cvmodel$threshold[which.min(cvmodel$error)]
optimal_size <- cvmodel$size[which.min(cvmodel$error)]

class_error <- 1 - (sum(pamr.predict(model, x_test,
                                     threshold=optimal_threshold) == y_test) /
                    length(y_test))

optimal_threshold
optimal_size
class_error

pamr.plotcen(model, nsc_data, threshold=1)
pamr.plotcen(model, nsc_data, threshold=2.5)
pamr.plotcen(model, nsc_data, threshold=optimal_threshold)


a <- pamr.listgenes(model, nsc_data, threshold=2.5)
cat(paste(colnames(data)[as.numeric(a[,1])], collapse='\n' ) )

a <- pamr.listgenes(model, nsc_data, threshold=optimal_threshold)
cat(paste(colnames(data)[as.numeric(a[,1])][1:10], collapse='\n' ) )

print(cvmodel)
pamr.plotcv(cvmodel)
```

```
set.seed(12345)

alpha <- 0.5
fit <- cv.glmnet(x=t(x), y=y, alpha=alpha, family="binomial")

optimal_lambda <- fit$lambda[which.min(fit$cvm)]
optimal_size <- fit$nzero[which.min(fit$cvm)]
class_error <- 1 - (sum(predict(fit, t(x_test), type="class") == y_test) / length(y_test))

optimal_lambda
optimal_size
class_error
set.seed(12345)

fit <- ksvm(x=t(x), y=y, kernel="vanilladot",
            type="C-svc", cross=10, scale=FALSE)

optimal_size <- fit@nSV
class_error <- 1 - (sum(predict(fit, t(x_test)) == y_test) / length(y_test))

optimal_size
class_error
```

Code for Assignment 2

```
library(pamr)
library(glmnet)
library(kernlab)

data <- read.csv("../data/data.csv", sep=";", header=TRUE,
                 stringsAsFactors=FALSE, encoding="latin1")
rownames(data) <- 1:nrow(data)

set.seed(12345)
train_idx <- sample(nrow(data), size=floor(nrow(data) * 7 / 10))
train <- data[train_idx,]
test <- data[-train_idx,]

x <- t(train[, -ncol(data)])
y <- train[, ncol(data)]

x_test <- t(test[, -ncol(data)])
y_test <- test[, ncol(data)]
```