setwd("C:/Users/Matrix/Desktop/New folder")

> filenameo <- "heart.csv"

> heartds <- read.csv(filenameo, header=TRUE)

> dim(heartds)

[1] 289 14

> heartds$output<- as.factor(heartds$output)

> sapply(heartds, class)

age sex cp trtbps chol fbs restecg thalachh exng oldpeak slp caa

"integer" "integer" "integer" "integer" "integer" "integer" "integer" "integer" "integer" "numeric" "integer" "integer"

thall output

"integer" "factor"

> head(heartds)

age sex cp trtbps chol fbs restecg thalachh exng oldpeak slp caa thall output

1 60 1 3 145 233 1 0 150 0 2.3 0 0 1 yes

2 35 1 2 130 250 0 1 187 0 3.5 0 0 2 yes

3 41 0 1 130 204 0 0 172 0 1.4 2 0 2 yes

4 55 1 1 120 236 0 1 178 0 0.8 2 0 2 yes

5 56 0 0 120 354 0 1 163 1 0.6 2 0 2 yes

6 55 1 0 140 192 0 1 148 0 0.4 1 0 1 yes

> summary(heartds)

age sex cp trtbps chol fbs restecg

Min. :29.00 Min. :0.0000 Min. :0.000 Min. : 94.0 Min. :126 Min. :0.0000 Min. :0.0000

1st Qu.:47.00 1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:120.0 1st Qu.:212 1st Qu.:0.0000 1st Qu.:0.0000

Median :54.00 Median :1.0000 Median :1.000 Median :130.0 Median :243 Median :0.0000 Median :1.0000

Mean :54.01 Mean :0.6782 Mean :1.021 Mean :131.4 Mean :248 Mean :0.1453 Mean :0.5156

3rd Qu.:60.00 3rd Qu.:1.0000 3rd Qu.:2.000 3rd Qu.:140.0 3rd Qu.:276 3rd Qu.:0.0000 3rd Qu.:1.0000

Max. :77.00 Max. :1.0000 Max. :3.000 Max. :200.0 Max. :564 Max. :1.0000 Max. :2.0000

thalachh exng oldpeak slp caa thall output

Min. : 71.0 Min. :0.0000 Min. :0.000 Min. :0.000 Min. :0.0000 Min. :0.000 no :124

1st Qu.:136.0 1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:2.000 yes:165

Median :154.0 Median :0.0000 Median :0.600 Median :1.000 Median :0.0000 Median :2.000

Mean :150.2 Mean :0.3183 Mean :1.008 Mean :1.419 Mean :0.7128 Mean :2.315

3rd Qu.:168.0 3rd Qu.:1.0000 3rd Qu.:1.600 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:3.000

Max. :202.0 Max. :1.0000 Max. :6.200 Max. :2.000 Max. :4.0000 Max. :3.000

> filenametr <- "train.csv"

> trainds <- read.csv(filenametr, header=TRUE)

> dim(trainds)

[1] 231 14

> trainds$output<- as.factor(trainds$output)

> sapply(trainds, class)

age sex cp trtbps chol fbs restecg thalachh exng oldpeak slp caa

"integer" "integer" "integer" "integer" "integer" "integer" "integer" "integer" "integer" "numeric" "integer" "integer"

thall output

"integer" "factor"

> head(trainds)

age sex cp trtbps chol fbs restecg thalachh exng oldpeak slp caa thall output

1 60 1 3 145 233 1 0 150 0 2.3 0 0 1 yes

2 35 1 2 130 250 0 1 187 0 3.5 0 0 2 yes

3 41 0 1 130 204 0 0 172 0 1.4 2 0 2 yes

4 56 0 0 120 354 0 1 163 1 0.6 2 0 2 yes

5 55 1 0 140 192 0 1 148 0 0.4 1 0 1 yes

6 56 0 1 140 294 0 0 153 0 1.3 1 0 2 yes

> summary(trainds)

age sex cp trtbps chol fbs restecg

Min. :29.00 Min. :0.0000 Min. :0 Min. : 94.0 Min. :126.0 Min. :0.0000 Min. :0.0000

1st Qu.:47.00 1st Qu.:0.0000 1st Qu.:0 1st Qu.:120.0 1st Qu.:212.5 1st Qu.:0.0000 1st Qu.:0.0000

Median :54.00 Median :1.0000 Median :1 Median :130.0 Median :245.0 Median :0.0000 Median :1.0000

Mean :54.28 Mean :0.6623 Mean :1 Mean :131.9 Mean :249.7 Mean :0.1472 Mean :0.5368

3rd Qu.:61.00 3rd Qu.:1.0000 3rd Qu.:2 3rd Qu.:140.0 3rd Qu.:279.5 3rd Qu.:0.0000 3rd Qu.:1.0000

Max. :77.00 Max. :1.0000 Max. :3 Max. :200.0 Max. :564.0 Max. :1.0000 Max. :2.0000

thalachh exng oldpeak slp caa thall output

Min. : 71.0 Min. :0.000 Min. :0.0000 Min. :0.000 Min. :0.0000 Min. :0.000 no : 99

1st Qu.:133.0 1st Qu.:0.000 1st Qu.:0.0000 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:2.000 yes:132

Median :155.0 Median :0.000 Median :0.6000 Median :1.000 Median :0.0000 Median :2.000

Mean :150.1 Mean :0.316 Mean :0.9874 Mean :1.429 Mean :0.7056 Mean :2.338

3rd Qu.:168.0 3rd Qu.:1.000 3rd Qu.:1.6000 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:3.000

Max. :202.0 Max. :1.000 Max. :5.6000 Max. :2.000 Max. :4.0000 Max. :3.000

> filenamete <- "test.csv"

> testds <- read.csv(filenamete, header=TRUE)

> dim(testds)

[1] 58 14

> testds$output<- as.factor(testds$output)

> sapply(testds, class)

age sex cp trtbps chol fbs restecg thalachh exng oldpeak slp caa

"integer" "integer" "integer" "integer" "integer" "integer" "integer" "integer" "integer" "numeric" "integer" "integer"

thall output

"integer" "factor"

> head(testds)

age sex cp trtbps chol fbs restecg thalachh exng oldpeak slp caa thall output

1 55 1 1 120 236 0 1 178 0 0.8 2 0 2 yes

2 44 1 1 120 263 0 1 173 0 0.0 2 0 3 yes

3 48 0 2 130 275 0 1 139 0 0.2 2 0 2 yes

4 59 1 2 150 212 1 1 157 0 1.6 2 0 2 yes

5 51 1 2 110 175 0 1 123 0 0.6 2 0 2 yes

6 44 1 1 130 219 0 0 188 0 0.0 2 0 2 yes

> summary(testds)

age sex cp trtbps chol fbs restecg

Min. :34.00 Min. :0.0000 Min. :0.000 Min. :100.0 Min. :157.0 Min. :0.0000 Min. :0.000

1st Qu.:45.75 1st Qu.:0.2500 1st Qu.:0.000 1st Qu.:113.5 1st Qu.:209.0 1st Qu.:0.0000 1st Qu.:0.000

Median :55.00 Median :1.0000 Median :1.000 Median :125.5 Median :234.5 Median :0.0000 Median :0.000

Mean :52.95 Mean :0.7414 Mean :1.103 Mean :129.1 Mean :240.9 Mean :0.1379 Mean :0.431

3rd Qu.:59.00 3rd Qu.:1.0000 3rd Qu.:2.000 3rd Qu.:138.0 3rd Qu.:264.5 3rd Qu.:0.0000 3rd Qu.:1.000

Max. :69.00 Max. :1.0000 Max. :3.000 Max. :178.0 Max. :409.0 Max. :1.0000 Max. :1.000

thalachh exng oldpeak slp caa thall output

Min. :103.0 Min. :0.0000 Min. :0.000 Min. :0.000 Min. :0.0000 Min. :0.000 no :25

1st Qu.:138.2 1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:2.000 yes:33

Median :150.5 Median :0.0000 Median :0.700 Median :1.000 Median :0.0000 Median :2.000

Mean :150.6 Mean :0.3276 Mean :1.088 Mean :1.379 Mean :0.7414 Mean :2.224

3rd Qu.:164.5 3rd Qu.:1.0000 3rd Qu.:1.800 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:3.000

Max. :188.0 Max. :1.0000 Max. :6.200 Max. :2.000 Max. :4.0000 Max. :3.000

>

> library(caret)

Loading required package: ggplot2

Loading required package: lattice

> library(MASS)

> set.seed(100)

> trctrl<-trainControl(method="repeatedcv",number=10,repeats=3)

> svm\_linear<-train(output~.,data=trainds,method="svmLinear",trControl=trctrl,preProcess=c("center","scale"),tuneLength=10)

> svm\_linear

Support Vector Machines with Linear Kernel

231 samples

13 predictor

2 classes: 'no', 'yes'

Pre-processing: centered (13), scaled (13)

Resampling: Cross-Validated (10 fold, repeated 3 times)

Summary of sample sizes: 208, 208, 208, 207, 207, 209, ...

Resampling results:

Accuracy Kappa

0.8501866 0.6888479

Tuning parameter 'C' was held constant at a value of 1

> test\_pred<-predict(svm\_linear,newdata=testds)

> test\_pred

[1] yes yes yes yes yes yes yes yes no yes yes yes yes yes yes yes yes yes yes no yes yes yes yes yes yes no yes yes no

[31] yes no no yes yes no no yes no no yes no no no no no no no no no no no no no yes yes yes no

Levels: no yes

> confusionMatrix(test\_pred,testds$output)

Confusion Matrix and Statistics

Reference

Prediction no yes

no 18 6

yes 7 27

Accuracy : 0.7759

95% CI : (0.6473, 0.8749)

No Information Rate : 0.569

P-Value [Acc > NIR] : 0.000826

Kappa : 0.5408

Mcnemar's Test P-Value : 1.000000

Sensitivity : 0.7200

Specificity : 0.8182

Pos Pred Value : 0.7500

Neg Pred Value : 0.7941

Prevalence : 0.4310

Detection Rate : 0.3103

Detection Prevalence : 0.4138

Balanced Accuracy : 0.7691

'Positive' Class : no

> grid<- expand.grid(C = seq(1,5,.25))

> svm\_Linear\_Grid<- train(output ~., data = trainds, method = "svmLinear",trControl=trctrl,preProcess = c("center", "scale"),tuneGrid = grid,tuneLength = 10)

> svm\_Linear\_Grid

Support Vector Machines with Linear Kernel

231 samples

13 predictor

2 classes: 'no', 'yes'

Pre-processing: centered (13), scaled (13)

Resampling: Cross-Validated (10 fold, repeated 3 times)

Summary of sample sizes: 208, 208, 208, 207, 208, 208, ...

Resampling results across tuning parameters:

C Accuracy Kappa

1.00 0.8442029 0.6776341

1.25 0.8457126 0.6807617

1.50 0.8513889 0.6928812

1.75 0.8513889 0.6928812

2.00 0.8484903 0.6868419

2.25 0.8471014 0.6841090

2.50 0.8471014 0.6841090

2.75 0.8471014 0.6841090

3.00 0.8485507 0.6868223

3.25 0.8513889 0.6923974

3.50 0.8513889 0.6923974

3.75 0.8513889 0.6923974

4.00 0.8500000 0.6894136

4.25 0.8514493 0.6926897

4.50 0.8500000 0.6898475

4.75 0.8528986 0.6958163

5.00 0.8528986 0.6958163

Accuracy was used to select the optimal model using the largest value.

The final value used for the model was C = 4.75.

> plot(svm\_Linear\_Grid)

> test\_pred\_grid<- predict(svm\_Linear\_Grid, newdata = testds)

> test\_pred\_grid

[1] yes yes yes yes yes yes yes yes no yes yes yes yes yes yes yes yes yes yes no yes yes yes yes yes yes no yes yes no

[31] yes no no yes yes no no yes no yes no no no no no no no no no no no no no no yes yes yes no

Levels: no yes

>

> confusionMatrix(table(test\_pred\_grid,testds$output))

Confusion Matrix and Statistics

test\_pred\_grid no yes

no 18 6

yes 7 27

Accuracy : 0.7759

95% CI : (0.6473, 0.8749)

No Information Rate : 0.569

P-Value [Acc > NIR] : 0.000826

Kappa : 0.5408

Mcnemar's Test P-Value : 1.000000

Sensitivity : 0.7200

Specificity : 0.8182

Pos Pred Value : 0.7500

Neg Pred Value : 0.7941

Prevalence : 0.4310

Detection Rate : 0.3103

Detection Prevalence : 0.4138

Balanced Accuracy : 0.7691

'Positive' Class : no