P8451 Machine Learning in Public Health - Assignment 10

2023-4-4

In preparation for all the analyses below, we will load the following libraries:

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.4.1 ✔ purrr 1.0.1  
## ✔ tibble 3.1.8 ✔ dplyr 1.1.0  
## ✔ tidyr 1.3.0 ✔ stringr 1.5.0  
## ✔ readr 2.1.4 ✔ forcats 1.0.0  
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(caret)

## Loading required package: lattice  
##   
## Attaching package: 'caret'  
##   
## The following object is masked from 'package:purrr':  
##   
## lift

library(rpart.plot)

## Loading required package: rpart

# Part 1: Import Data & Exploratory Data Analysis

## 1.1 Load .Rdata & Merge

Below we merge all data into a single data frame.

#Load data using path of where file is stored  
load("./exposome.RData")  
  
#Merge all data frames into a single data frame.  
studydata <- merge(exposome,phenotype,by="ID") %>% merge(covariates, by="ID")  
  
#Strip off ID Variable  
studydata$ID <- NULL

## 1.2 Exploratory Data Analysis

Next, we will generate some descriptive measures for the following features in the newly merged data frame:

* hs\_asthma: Doctor diagnosed asthma (ever) at 6-11 years of age (*phenotype*)
* hs\_popdens\_h\_Sqrt: Population density at home (*exposome*)
* hs\_popdens\_s\_Sqrt: Population density at school (*exposome*)
* hs\_dif\_hours\_total\_None: Total hours of sleep; mean weekdays and night (*exposome*)
* h\_NO2\_Log: Concentration of indoor NO2 at home (*exposome*)
* h\_PM\_Log: Concentration of particulate matter at home (*exposome*)

To do so, we will first generate a data set with only 6 variables mentioned above and convert the hs\_asthma variable from numeric to a 2-level factor variable. Finally, we will apply the summary function to generate descriptive statistics.

helix\_data = studydata %>%   
 select(hs\_asthma, hs\_popdens\_h\_Sqrt, hs\_popdens\_s\_Sqrt, hs\_dif\_hours\_total\_None, h\_NO2\_Log, h\_PM\_Log) %>%   
 mutate(hs\_asthma = factor(hs\_asthma,   
 labels = c("Asthma", "No\_Asthma")))  
  
skimr::skim(helix\_data)

Data summary

|  |  |
| --- | --- |
| Name | helix\_data |
| Number of rows | 1301 |
| Number of columns | 6 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| factor | 1 |
| numeric | 5 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: factor**

| skim\_variable | n\_missing | complete\_rate | ordered | n\_unique | top\_counts |
| --- | --- | --- | --- | --- | --- |
| hs\_asthma | 0 | 1 | FALSE | 2 | Ast: 1159, No\_: 142 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| hs\_popdens\_h\_Sqrt | 0 | 1 | 67.65 | 45.76 | 1.73 | 30.04 | 67.41 | 84.99 | 261.50 | ▅▇▁▁▁ |
| hs\_popdens\_s\_Sqrt | 0 | 1 | 68.10 | 38.11 | 0.00 | 38.56 | 69.26 | 84.99 | 210.95 | ▆▇▇▁▁ |
| hs\_dif\_hours\_total\_None | 0 | 1 | 10.30 | 0.72 | 7.90 | 9.79 | 10.33 | 10.74 | 12.85 | ▁▅▇▂▁ |
| h\_NO2\_Log | 0 | 1 | 3.83 | 1.05 | 1.57 | 2.98 | 3.62 | 4.58 | 7.09 | ▃▇▆▃▁ |
| h\_PM\_Log | 0 | 1 | 2.44 | 0.52 | 1.55 | 2.07 | 2.30 | 2.70 | 5.24 | ▇▆▂▁▁ |

The newly generated helix\_data data set contains **1301 rows** and **6 columns** (i.e., features). There are **no missing values** for any of the 6 features. There are 5 continuous features and 1 binary feature in this data set.

The 1 binary feature in the helix\_data data set is the hs\_asthma variable. A total of **1,159** individuals reported an asthma diagnosis, and **142** individuals reported no asthma diagnosis.

Descriptive statistics (i.e., mean, median and range) for the 5 continuous variables in the helix\_data data set are shown in the table below:

| Variable | Mean | Median | Range |
| --- | --- | --- | --- |
| hs\_popdens\_h\_Sqrt | 67.652 | 67.405 | (1.732, 261.500) |
| hs\_popdens\_s\_Sqrt | 68.10 | 69.26 | (0.00, 210.95) |
| hs\_dif\_hours\_total\_None | 10.296 | 10.330 | (7.901, 12.852) |
| h\_NO2\_Log | 3.833 | 3.617 | (1.573, 7.093) |
| h\_PM\_Log | 2.443 | 2.304 | (1.549, 5.236) |

# Part II: Developing Research Question

The following is the research question of interest:

Can whether a child is doctor-diagnosed with asthma at 6-11 years of age be predicted using population density at home and at school, total hours of sleep (mean weekdays and night), and concentrations of indoor NO2 and particulate matter at home?

# Part III: Implement Pipeline to Address Research Question

## 3.1 Data Preprocessing: Centering and Scaling & Partitioning Data

Below, we center and scale these data. In general, it is always good practice to do so!

helix\_numeric = helix\_data %>%   
 select(where(is.numeric))   
  
preprocess\_setup <- preProcess(helix\_numeric, method = c("center", "scale"))  
transformed.vals = predict(preprocess\_setup, helix\_numeric)

For the purposes of this analysis, we will partition the data into training and testing using a 70/30 split. This process involves applying the createDataPartition function to generate a set of training and testing data with equal proportion of individual with the outcome of interest, i.e., Diabetes. The new object train\_index contains all the indexes of the rows in the original data set contained in the 70% split. The rows indexed to be in the 70% is assigned to a new training data set, and the remaining 30% is assigned to a new testing data set.

train\_index = createDataPartition(helix\_data$hs\_asthma, p = 0.7, list = FALSE)  
  
helix\_train <- helix\_data[train\_index,]  
helix\_test <- helix\_data [-train\_index,]

## 3.2 Developing the Support Vector Classifier Model

In the code chunk below, we will use the trainControl function to set our validation method. For the purposes of this analysis, we will use the 10-fold cross validation method and will generate predicted probabilities.

train\_control\_svm = trainControl(method = "cv", number = 10, classProbs = T)

Next, we will incorporate different values for cost (C) into the model. We will also show information about the final model, and generate the metrics of accuracy from training using the confusionMatrix function.

set.seed(123)  
  
svm\_helix = train(hs\_asthma ~ .,   
 data = helix\_train,   
 method = "svmLinear",   
 trControl = train\_control\_svm,   
 preProcess = c("center", "scale"),   
 tuneGrid = expand.grid(C = seq(0.001, 10, length = 30)))

## maximum number of iterations reached -0.0002310126 -0.0002311194maximum number of iterations reached -0.0001219233 -0.000121952maximum number of iterations reached 0.000355389 0.0003552654maximum number of iterations reached 0.0001444532 0.0001444416maximum number of iterations reached -0.0001460444 -0.0001460521maximum number of iterations reached -0.0004396265 -0.000439754maximum number of iterations reached 0.0007914193 0.0007907321maximum number of iterations reached 0.0006250988 0.0006247398maximum number of iterations reached -0.0003703784 -0.0003706161maximum number of iterations reached 0.0004169221 0.000416839maximum number of iterations reached -0.0002838524 -0.0002839171maximum number of iterations reached -0.0004185816 -0.0004186644maximum number of iterations reached 0.0002353105 0.0002353023maximum number of iterations reached 5.945949e-05 5.94554e-05maximum number of iterations reached 4.111747e-05 4.111469e-05maximum number of iterations reached 0.0002075094 0.0002074232maximum number of iterations reached 0.0004248738 0.0004247021maximum number of iterations reached 0.0003389704 0.0003388887maximum number of iterations reached 0.00116177 0.001159854maximum number of iterations reached 0.0001087249 0.0001087201maximum number of iterations reached -0.0004198922 -0.0004201325maximum number of iterations reached -0.000101802 -0.0001018092maximum number of iterations reached 0.0003012497 0.0003012239maximum number of iterations reached -4.196064e-05 -4.196304e-05maximum number of iterations reached -0.0001404469 -0.0001404539maximum number of iterations reached -0.0006062972 -0.0006064358maximum number of iterations reached 0.0005905007 0.0005902041maximum number of iterations reached -0.0003929531 -0.0003929843maximum number of iterations reached -0.0001812008 -0.0001812219maximum number of iterations reached -4.188827e-05 -4.189088e-05maximum number of iterations reached 0.000328725 0.000328644maximum number of iterations reached 0.0005062991 0.0005061849maximum number of iterations reached -0.0004014622 -0.0004015775maximum number of iterations reached -0.0004372967 -0.0004374258maximum number of iterations reached -0.0004099922 -0.0004100175maximum number of iterations reached 0.0001353081 0.0001353083maximum number of iterations reached 0.000154012 0.0001540037maximum number of iterations reached 0.0009712681 0.0009704424maximum number of iterations reached -0.0005214823 -0.0005216867maximum number of iterations reached 0.0001162482 0.0001162368maximum number of iterations reached 0.0002453353 0.0002453137maximum number of iterations reached -0.0002226634 -0.0002226704maximum number of iterations reached -0.0005409328 -0.000541094maximum number of iterations reached -0.0002071862 -0.0002072057maximum number of iterations reached -6.428096e-05 -6.428394e-05maximum number of iterations reached -6.761521e-05 -6.762028e-05maximum number of iterations reached -0.0004184257 -0.000418471maximum number of iterations reached 9.463953e-05 9.463585e-05maximum number of iterations reached -0.0001081241 -0.0001081293maximum number of iterations reached 0.0007512933 0.0007507062maximum number of iterations reached -0.0004922268 -0.0004925591maximum number of iterations reached -0.0001715095 -0.0001715189maximum number of iterations reached -0.0001904372 -0.0001904572maximum number of iterations reached 9.456525e-05 9.456127e-05maximum number of iterations reached -0.0001668739 -0.0001669079maximum number of iterations reached -0.001166742 -0.001167755maximum number of iterations reached -0.0003218305 -0.000321888maximum number of iterations reached 0.0002466637 0.0002466302maximum number of iterations reached 0.0001955973 0.0001955824maximum number of iterations reached 0.0002337885 0.0002337417maximum number of iterations reached -0.0001251283 -0.0001251475maximum number of iterations reached -0.0004967773 -0.000496852maximum number of iterations reached -0.0002600164 -0.000260056maximum number of iterations reached -0.0001090359 -0.0001090611maximum number of iterations reached 1.199728e-05 1.199716e-05maximum number of iterations reached -0.0003485568 -0.000348622maximum number of iterations reached 0.0002883837 0.0002883161maximum number of iterations reached -0.0007595669 -0.0007600409maximum number of iterations reached -0.0008016399 -0.0008020378maximum number of iterations reached 0.0006538555 0.0006534164maximum number of iterations reached -1.944177e-05 -1.944243e-05maximum number of iterations reached 0.0004491646 0.0004490262maximum number of iterations reached 2.684862e-05 2.684844e-05maximum number of iterations reached 7.136322e-05 7.136083e-05maximum number of iterations reached 0.0001351528 0.0001351413maximum number of iterations reached -0.001114641 -0.001115452maximum number of iterations reached -0.0003136485 -0.0003137648maximum number of iterations reached 8.399962e-05 8.39988e-05maximum number of iterations reached -0.0003562302 -0.0003563752maximum number of iterations reached 0.0001279807 0.0001279663maximum number of iterations reached 0.0006865715 0.0006863636maximum number of iterations reached -0.0002797572 -0.0002798074maximum number of iterations reached -0.0002440709 -0.0002441072maximum number of iterations reached 0.0007240999 0.000723765maximum number of iterations reached -0.0003388336 -0.0003388681maximum number of iterations reached -0.0008328286 -0.0008338007maximum number of iterations reached 0.0009289961 0.0009278582maximum number of iterations reached -0.0001627468 -0.0001627691maximum number of iterations reached -0.0006995618 -0.0007001747maximum number of iterations reached 0.0001969209 0.0001969001maximum number of iterations reached 9.026191e-05 9.026012e-05maximum number of iterations reached 0.000452925 0.0004528568maximum number of iterations reached 0.0008687067 0.0008675476maximum number of iterations reached -0.0005015066 -0.0005016498maximum number of iterations reached -3.200998e-05 -3.201009e-05maximum number of iterations reached -0.0001498673 -0.0001498884maximum number of iterations reached 0.0001178773 0.0001178745maximum number of iterations reached -0.0007093474 -0.0007097792maximum number of iterations reached 0.0003612883 0.0003611784maximum number of iterations reached 0.0005657344 0.0005656135maximum number of iterations reached -0.000642309 -0.0006427724maximum number of iterations reached -2.853509e-05 -2.85357e-05maximum number of iterations reached -0.0004084049 -0.0004085886maximum number of iterations reached -0.001068141 -0.001068627maximum number of iterations reached 0.0006686806 0.0006684234maximum number of iterations reached 0.000573337 0.000573106maximum number of iterations reached -0.0003079328 -0.0003080219maximum number of iterations reached -0.000810341 -0.0008118446maximum number of iterations reached 0.0003565167 0.0003564479maximum number of iterations reached 0.0001873041 0.0001872823maximum number of iterations reached 0.0002435541 0.0002434867maximum number of iterations reached -3.506759e-05 -3.506851e-05maximum number of iterations reached -0.0003886875 -0.0003887983maximum number of iterations reached 0.0001619906 0.0001619772maximum number of iterations reached 0.0005997824 0.0005996268maximum number of iterations reached -0.0001932873 -0.000193295maximum number of iterations reached 0.000318422 0.0003183667maximum number of iterations reached -2.819708e-05 -2.81974e-05maximum number of iterations reached -0.0001355597 -0.0001355682maximum number of iterations reached -0.0003520038 -0.0003520552maximum number of iterations reached -0.001289835 -0.001291231maximum number of iterations reached -0.0007059132 -0.0007062815maximum number of iterations reached 0.000261333 0.0002612667maximum number of iterations reached 0.0003630729 0.0003629472maximum number of iterations reached 0.000323762 0.0003236695maximum number of iterations reached 0.00141028 0.001408058maximum number of iterations reached 0.0001972127 0.0001971962maximum number of iterations reached -0.0004439569 -0.0004442049maximum number of iterations reached -0.0004720878 -0.000472298maximum number of iterations reached -0.0003203256 -0.0003203506maximum number of iterations reached -0.0009385568 -0.0009391119maximum number of iterations reached -0.0002626201 -0.000262653maximum number of iterations reached 0.001046601 0.001045169maximum number of iterations reached 0.0005203228 0.0005202154maximum number of iterations reached 0.0004219572 0.0004218451maximum number of iterations reached 0.001975044 0.00197007maximum number of iterations reached 6.189803e-05 6.189895e-05maximum number of iterations reached -0.000578106 -0.0005784336maximum number of iterations reached 0.001137636 0.001136608maximum number of iterations reached 0.000219779 0.0002196913maximum number of iterations reached 0.0002679436 0.0002679166maximum number of iterations reached 0.0009692709 0.0009687461maximum number of iterations reached -6.127347e-05 -6.127635e-05maximum number of iterations reached -0.0001359457 -0.0001359597maximum number of iterations reached 0.0004015402 0.0004014936maximum number of iterations reached -0.0005058886 -0.0005060289maximum number of iterations reached -1.333206e-05 -1.333276e-05maximum number of iterations reached -0.0004497392 -0.0004499158maximum number of iterations reached 0.0002089656 0.000208957maximum number of iterations reached -3.431368e-05 -3.431509e-05maximum number of iterations reached 4.422636e-05 4.422324e-05maximum number of iterations reached 0.0003381885 0.0003380432maximum number of iterations reached 0.0002040421 0.0002040334maximum number of iterations reached -0.0009660157 -0.0009664514maximum number of iterations reached -0.0001473451 -0.0001473673maximum number of iterations reached 0.000127225 0.0001272233maximum number of iterations reached -0.0001069661 -0.0001069766maximum number of iterations reached 0.0008652984 0.0008643686maximum number of iterations reached 0.0008214618 0.0008211751maximum number of iterations reached 5.34147e-05 5.341272e-05maximum number of iterations reached 0.0005810367 0.0005808932maximum number of iterations reached 9.485832e-05 9.48562e-05maximum number of iterations reached 0.001002722 0.001001203maximum number of iterations reached -0.0002479662 -0.0002480113maximum number of iterations reached 0.0002743142 0.0002742997maximum number of iterations reached -0.0003493737 -0.0003494552maximum number of iterations reached 9.715126e-05 9.714474e-05maximum number of iterations reached -0.0008012195 -0.0008021957maximum number of iterations reached -0.0004069163 -0.000406956maximum number of iterations reached 0.0005556406 0.0005554745maximum number of iterations reached 0.0001231933 0.0001231902maximum number of iterations reached 0.0001734431 0.000173433maximum number of iterations reached 0.0001201866 0.0001201609maximum number of iterations reached -0.0004428259 -0.0004429215maximum number of iterations reached -0.0004518966 -0.0004520444maximum number of iterations reached 0.0001958284 0.000195815maximum number of iterations reached -0.0003437201 -0.0003438217maximum number of iterations reached 0.0001706614 0.0001706574maximum number of iterations reached 0.0004618786 0.00046163maximum number of iterations reached -2.361897e-05 -2.361886e-05maximum number of iterations reached 0.0003779941 0.0003779273maximum number of iterations reached 0.000207207 0.0002071655maximum number of iterations reached 0.0006404186 0.0006399934maximum number of iterations reached -0.001027982 -0.001029741maximum number of iterations reached 7.930726e-05 7.93064e-05maximum number of iterations reached 0.0005543328 0.0005537525maximum number of iterations reached -3.049284e-05 -3.04938e-05maximum number of iterations reached -3.912139e-05 -3.912209e-05maximum number of iterations reached 0.0008369623 0.0008363882maximum number of iterations reached 0.0005652463 0.0005649436maximum number of iterations reached -0.0003675197 -0.0003675998maximum number of iterations reached -0.0008497963 -0.0008504603maximum number of iterations reached -0.0004009738 -0.0004010607maximum number of iterations reached 0.0007117678 0.0007115714maximum number of iterations reached -0.0003110155 -0.0003110324maximum number of iterations reached -0.0009632315 -0.0009637692maximum number of iterations reached 0.0004304471 0.0004302063maximum number of iterations reached -0.0003919592 -0.0003920608maximum number of iterations reached 0.0001020729 0.0001020696maximum number of iterations reached -0.0001362403 -0.0001362521maximum number of iterations reached 0.0001241711 0.0001241672maximum number of iterations reached 0.0002599822 0.0002599633maximum number of iterations reached 4.377558e-05 4.377538e-05maximum number of iterations reached 0.0002826862 0.0002826248maximum number of iterations reached 0.0001136874 0.0001136796maximum number of iterations reached 0.0001903284 0.0001902924maximum number of iterations reached 0.000510765 0.0005105516maximum number of iterations reached 0.0003590727 0.0003590322maximum number of iterations reached -0.0002165317 -0.0002165472maximum number of iterations reached -0.000936663 -0.0009371263maximum number of iterations reached -0.0004126947 -0.0004127409maximum number of iterations reached -7.261795e-05 -7.262221e-05maximum number of iterations reached -0.000444459 -0.0004446792maximum number of iterations reached -0.000260587 -0.0002606667maximum number of iterations reached -0.0001649623 -0.0001649817maximum number of iterations reached -5.191135e-05 -5.191248e-05maximum number of iterations reached 3.546504e-05 3.546357e-05maximum number of iterations reached -0.0004687719 -0.0004688659maximum number of iterations reached 0.0001209879 0.0001209826maximum number of iterations reached -0.0002087047 -0.000208741maximum number of iterations reached -2.080438e-05 -2.080422e-05maximum number of iterations reached 0.001011933 0.001010932maximum number of iterations reached -3.285675e-05 -3.285716e-05maximum number of iterations reached 5.428e-05 5.427902e-05maximum number of iterations reached 0.0005438574 0.0005437401maximum number of iterations reached 8.549219e-05 8.548894e-05maximum number of iterations reached 0.0001890177 0.0001890018maximum number of iterations reached -0.0003120187 -0.0003120824maximum number of iterations reached -0.0002292833 -0.0002292924maximum number of iterations reached 8.7191e-05 8.718833e-05maximum number of iterations reached 1.225532e-05 1.22556e-05maximum number of iterations reached -0.0006045939 -0.0006048893maximum number of iterations reached -0.0001062379 -0.0001062469maximum number of iterations reached 7.833777e-05 7.833171e-05maximum number of iterations reached -0.0004328664 -0.0004329378maximum number of iterations reached -0.0004735991 -0.0004738156maximum number of iterations reached 0.0009519882 0.000951167maximum number of iterations reached 0.000198237 0.0001981829maximum number of iterations reached -0.00024612 -0.0002461654maximum number of iterations reached -0.0002300367 -0.0002300787maximum number of iterations reached 0.0007316703 0.0007311965maximum number of iterations reached -0.0004450027 -0.0004451953maximum number of iterations reached 0.000442785 0.0004426657maximum number of iterations reached 0.00131699 0.001314616maximum number of iterations reached 2.829668e-05 2.82955e-05maximum number of iterations reached -0.0009569415 -0.00095759maximum number of iterations reached 0.000735455 0.0007349823maximum number of iterations reached -0.0005015867 -0.0005016745maximum number of iterations reached 8.208278e-05 8.207988e-05maximum number of iterations reached -0.0002189153 -0.0002189311maximum number of iterations reached 0.0008075852 0.0008069195maximum number of iterations reached 0.0003832777 0.0003832451maximum number of iterations reached -0.0003953213 -0.0003954154maximum number of iterations reached 0.0006609931 0.0006603438maximum number of iterations reached -0.00045937 -0.0004595431maximum number of iterations reached 6.431958e-05 6.431922e-05maximum number of iterations reached 9.784205e-05 9.783992e-05maximum number of iterations reached -0.0008246996 -0.0008251723maximum number of iterations reached 8.278194e-05 8.2779e-05maximum number of iterations reached 0.0002982677 0.0002982065maximum number of iterations reached -9.697856e-05 -9.698583e-05maximum number of iterations reached 0.0008033938 0.0008026018maximum number of iterations reached 0.0005273959 0.0005270837maximum number of iterations reached 0.0005811345 0.0005808704maximum number of iterations reached 0.0001866331 0.0001866216maximum number of iterations reached -0.0003333845 -0.0003335588maximum number of iterations reached 0.00146408 0.001461063maximum number of iterations reached 4.110166e-05 4.110223e-05maximum number of iterations reached -0.0002038628 -0.000203881maximum number of iterations reached 0.0003074617 0.0003073478maximum number of iterations reached 0.00141691 0.001411569maximum number of iterations reached 6.577809e-05 6.577739e-05maximum number of iterations reached 0.0003303417 0.0003302937maximum number of iterations reached -0.000532065 -0.0005321762maximum number of iterations reached 0.0006065143 0.0006062957maximum number of iterations reached -0.0002200187 -0.0002200412maximum number of iterations reached -9.216201e-05 -9.216681e-05maximum number of iterations reached 0.0002508019 0.0002507837maximum number of iterations reached 0.000496783 0.0004966713maximum number of iterations reached -0.00032503 -0.000325193maximum number of iterations reached 0.0002609178 0.0002608907maximum number of iterations reached 0.0005313748 0.0005311889maximum number of iterations reached 0.0002950186 0.0002949377maximum number of iterations reached 0.0001171306 0.0001171262maximum number of iterations reached -5.903396e-05 -5.903983e-05maximum number of iterations reached 0.0006830365 0.0006825975maximum number of iterations reached 0.0004865075 0.0004863934maximum number of iterations reached -3.19382e-05 -3.193844e-05maximum number of iterations reached 0.0001530188 0.0001529972maximum number of iterations reached 0.0002603078 0.0002602781maximum number of iterations reached -0.0008397858 -0.0008404657

svm\_helix$finalModel

## Support Vector Machine object of class "ksvm"   
##   
## SV type: C-svc (classification)   
## parameter : cost C = 0.001   
##   
## Linear (vanilla) kernel function.   
##   
## Number of Support Vectors : 201   
##   
## Objective Function Value : -0.2   
## Training error : 0.109649   
## Probability model included.

confusionMatrix(svm\_helix)

## Cross-Validated (10 fold) Confusion Matrix   
##   
## (entries are percentual average cell counts across resamples)  
##   
## Reference  
## Prediction Asthma No\_Asthma  
## Asthma 89 11  
## No\_Asthma 0 0  
##   
## Accuracy (average) : 0.8904

Based on the output above, the accuracy of the SVC model is **0.8904**, and the cost value is **0.001**.

## 3.3 Calculate Final Evaluation Metrics in Test Set with the Optimal Model

We will now apply the SVC model to the testing data set, and generate evaluation metrics using the confusionMatrix function.

set.seed(123)  
  
svm\_pred\_helix\_test = predict(svm\_helix, helix\_test)  
  
confusionMatrix(svm\_pred\_helix\_test, helix\_test$hs\_asthma)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction Asthma No\_Asthma  
## Asthma 347 42  
## No\_Asthma 0 0  
##   
## Accuracy : 0.892   
## 95% CI : (0.8569, 0.9211)  
## No Information Rate : 0.892   
## P-Value [Acc > NIR] : 0.541   
##   
## Kappa : 0   
##   
## Mcnemar's Test P-Value : 2.509e-10   
##   
## Sensitivity : 1.000   
## Specificity : 0.000   
## Pos Pred Value : 0.892   
## Neg Pred Value : NaN   
## Prevalence : 0.892   
## Detection Rate : 0.892   
## Detection Prevalence : 1.000   
## Balanced Accuracy : 0.500   
##   
## 'Positive' Class : Asthma   
##

The kappa value is **0**, and the Mcnemar’s Test p-value is **2.509e-10**. The accuracy level of the SVC model is **0.892**, with a 95% confidence interval of **0.8569 to 0.9211**. The sensitivity of this model is **1.000** and the specificity of this model is **0.000**. The reported prevalence is **0.892**.