

SVM

```
set.seed(123)

library(e1071)
library(caret)

## Loading required package: ggplot2

## Loading required package: lattice

source("../cleandata.R")

## 
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
## 
##     filter, lag

## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union

source("../partition.R")

## -- Attaching packages ----- tidymodels 1.4.1 --

## v broom      1.0.9   v rsample      1.3.1
## v dials      1.4.2   v tailor       0.1.0
## v infer      1.0.9   v tidyverse    1.3.1
## v modeldata   1.5.1   v tune         2.0.1
## v parsnip     1.4.0   v workflows    1.3.0
## v purrr       1.1.0   v workflowsets 1.1.1
## v recipes     1.3.1   v yardstick    1.3.2

## Warning: package 'parsnip' was built under R version 4.5.2

## -- Conflicts ----- tidymodels_conflicts() --
## x rsample::calibration() masks caret::calibration()
## x purrr::discard()     masks scales::discard()
## x dplyr::filter()      masks stats::filter()
## x dplyr::lag()         masks stats::lag()
## x purrr::lift()        masks caret::lift()
```

```

## x rsample::permutations()  masks e1071::permutations()
## x yardstick::precision()   masks caret::precision()
## x yardstick::recall()     masks caret::recall()
## x yardstick::sensitivity() masks caret::sensitivity()
## x yardstick::specificity() masks caret::specificity()
## x recipes::step()         masks stats::step()
## x tune::tune()            masks parsnip::tune(), e1071::tune()

source("../confusionmatrix.R")
source("../eval.R")

diabetes_raw <- read.csv("../data/diabetes_012_health_indicators_BRFSS2015.csv")
diabetes <- cleandata(diabetes_raw)
head(diabetes)

##   Diabetes_012 HighBP HighChol CholCheck BMI Smoker Stroke HeartDiseaseorAttack
## 1   NoDiabetes      1       1       1   40      1       0                      0
## 2   NoDiabetes      0       0       0   25      1       0                      0
## 3   NoDiabetes      1       1       1   28      0       0                      0
## 4   NoDiabetes      1       0       1   27      0       0                      0
## 5   NoDiabetes      1       1       1   24      0       0                      0
## 6   NoDiabetes      1       1       1   25      1       0                      0
##   PhysActivity Fruits Veggies HvyAlcoholConsump AnyHealthcare NoDocbcCost
## 1             0     0       1                 0           1              0
## 2             1     0       0                 0           0              1
## 3             0     1       0                 0           1              1
## 4             1     1       1                 0           1              0
## 5             1     1       1                 0           1              0
## 6             1     1       1                 0           1              0
##   GenHlth MentHlth PhysHlth DiffWalk Sex Age Education Income
## 1      5      18      15       1   0    9        4      3
## 2      3       0       0       0   0    7        6      1
## 3      5      30      30       1   0    9        4      8
## 4      2       0       0       0   0   11        3      6
## 5      2       3       0       0   0   11        5      4
## 6      2       0       2       0   1   10        6      8

diabetes |> count(Diabetes_012)

##   Diabetes_012      n
## 1   NoDiabetes 213703
## 2  PreDiabetes   4631
## 3    Diabetes   35346

split <- partition(diabetes, 0.8)
train <- split$train
test <- split$test

train |> count(Diabetes_012)

## # A tibble: 3 x 2
##   Diabetes_012      n

```

```

##      <ord>      <int>
## 1 NoDiabetes    3705
## 2 PreDiabetes   3705
## 3 Diabetes      3705

svm.model <- svm(
  Diabetes_012 ~ .,
  data = train,
  kernel = "radial",
  cost = 100,
  gamma = 0.001
)

#tune.out <- e1071::tune(
#  svm,
#  Diabetes_012 ~ .,
#  data = train,
#  kernel = "radial",
#  ranges = list(
#    cost = c(0.1, 1, 10, 100),
#    gamma = c(0.001, 0.01, 0.1, 1)
#  )
#)

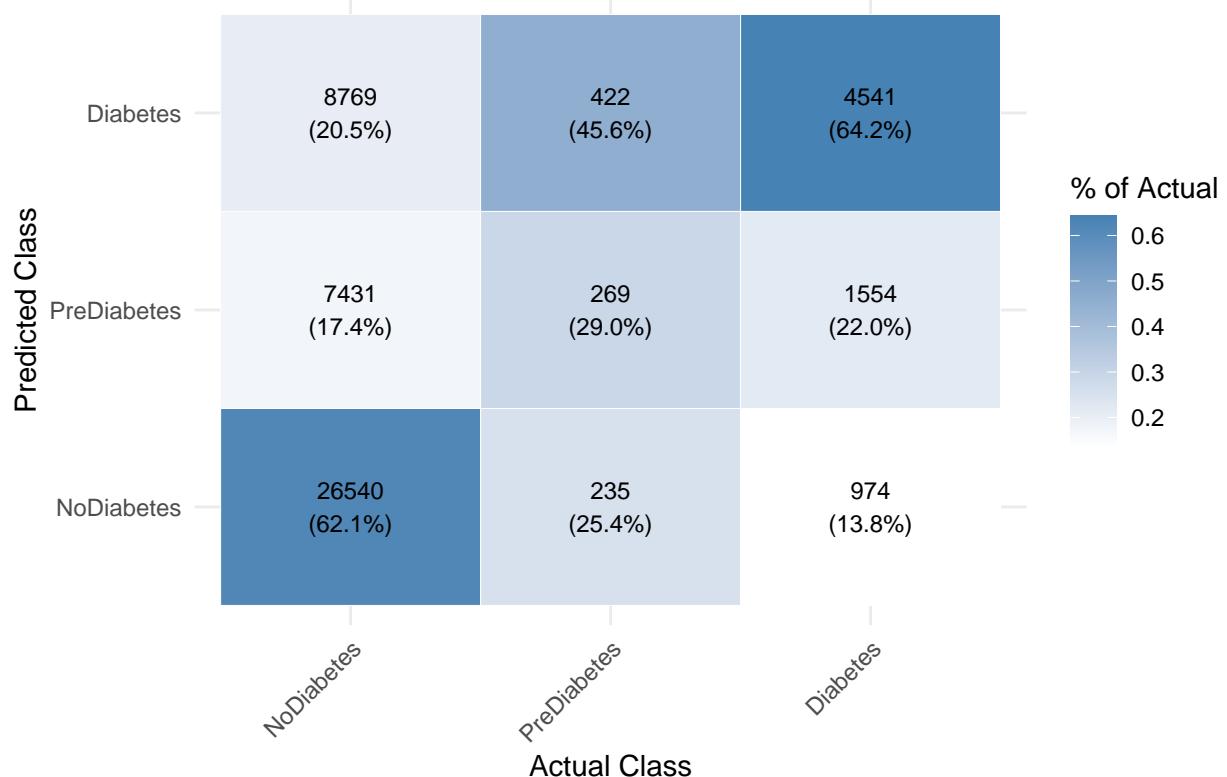
summary(svm.model)

##
## Call:
## svm(formula = Diabetes_012 ~ ., data = train, kernel = "radial",
##       cost = 100, gamma = 0.001)
##
## Parameters:
##   SVM-Type: C-classification
##   SVM-Kernel: radial
##   cost: 100
##
## Number of Support Vectors:  9537
##
## ( 2653 3688 3196 )
##
## Number of Classes: 3
##
## Levels:
## NoDiabetes PreDiabetes Diabetes

pred <- predict(svm.model, newdata = test)
cm <- confusionMatrix(pred, test$Diabetes_012)
plot_cm(cm)

```

Confusion Matrix (Shaded by % of Actual Class)



```
qwk_score(pred, test$Diabetes_012)
```

```
## [1] "trying..."
```

```
## [1] 0.3208412
```

```
mae_ord(pred, test$Diabetes_012)
```

```
## [1] 0.5741204
```

model	QWK	MAE
CumLogit	0.35	0.96
SVM	0.32	0.57
RandForest	0.32	0.55
OrdForest	0.31	0.56