Homework 2: Classification Methods

Wen Li Teng

February 2 2020

The Bayes Classifier

```
# Load packages
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(rsample)
```

```
## Warning: package 'rsample' was built under R version 3.6.2
```

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 3.6.2

## Loading required package: lattice

##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
##     lift
```

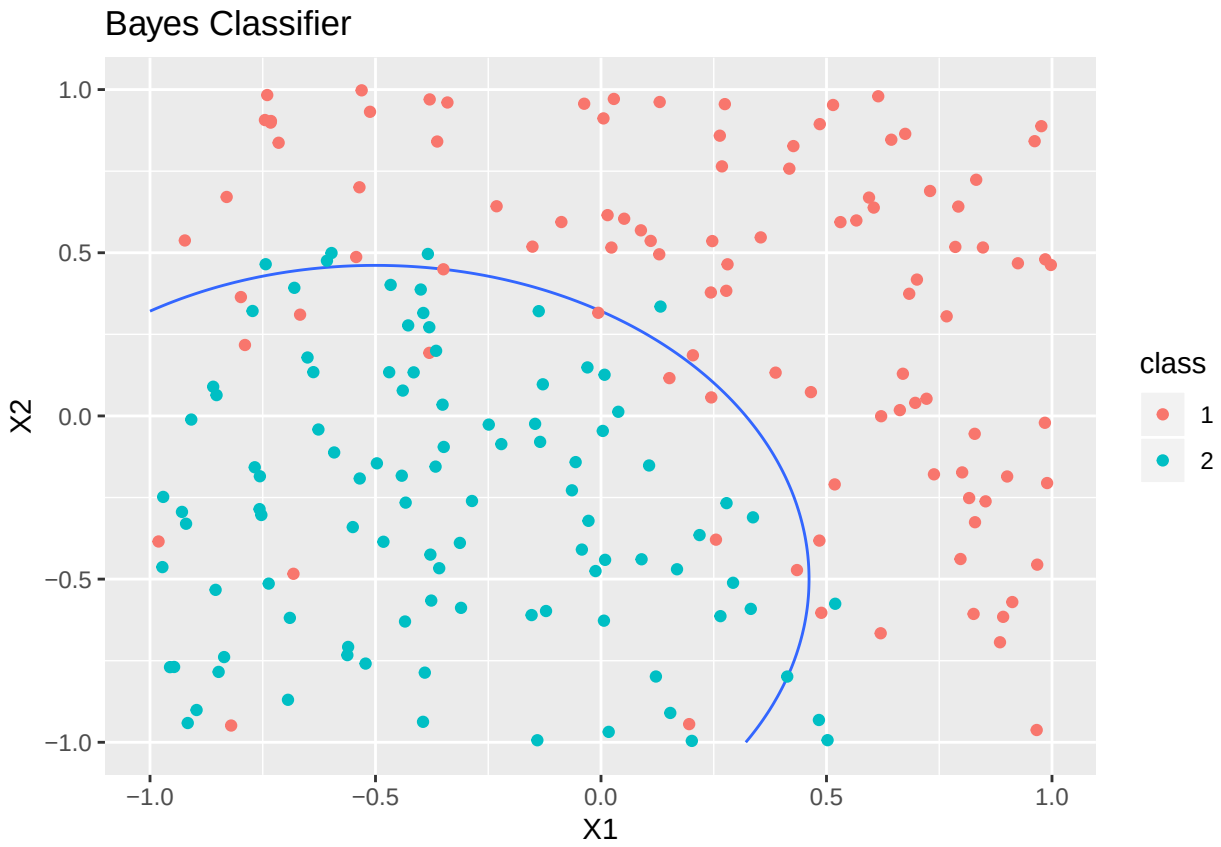
```
library(ggplot2)
```

```
# 1(a) Set random generator seed
set.seed(1234)
```

```
# 1(b) Simulate a dataset
n_obs <- 200
min <- -1
max <- 1
q1_data <- tibble(X1 = runif(n_obs, min, max), X2 = runif(n_obs, min, max))
```

```r
# 1(c) Calculate Y
# 1(d) Calculate the probability of success

logit2prob <- function(x){
  exp(x) / (1 + exp(x))
}

codeY <- function(Y) {
  if(Y >= 0.5)
    output <- 1
  if (Y < 0.5)
    output <- 2
  return(output)
}

q1_data <- q1_data %>%
  mutate (E = rnorm(200, 0, sqrt(0.25))) %>%
  mutate(Y = X1 + X1^2 + X2 + X2^2 + E) %>%
  mutate (prob = logit2prob(Y)) %>%
  rowwise() %>%
  mutate(class = codeY(Y))

q1_data$class <- as.factor(q1_data$class)
```

```r
# 1(e) Plot data points on graph
# 1(f) Overlay plot with Bayes decision boundary
# 1(g) Give plot title and axis labels

x1 <- seq(from = -1, to = 1, by = 0.01)
x2 <- seq(from = -1, to = 1, by = 0.01)
grid <- expand.grid(x1, x2)

grid_new <- grid %>%
  rename(X1 = Var1) %>%
  rename(X2 = Var2) %>%
  mutate(Y = X1 + X1^2 + X2 + X2^2) %>%
  mutate (prob = logit2prob(Y)) %>%
  rowwise() %>%
  mutate(class = codeY(Y))

ggplot(grid_new, aes(x=X1, y=X2)) +
  geom_contour(aes(z = prob), bins = 1) +
  geom_point(data = q1_data, aes(color = class)) +
  labs(title = "Bayes Classifier",
       x = "X1",
       y = "X2")
```


Bayes Classifier
1.0
0.5
0.0
−0.5
−1.0
X2
−1.0
−0.5
0.0
0.5
1.0
X1
class
1
2