

Logistic Regression

Load Data

```
path <- "~/Users/pulkitbatra/Desktop/CACSC19/Unit-2 R Programming/Learning R/Assignment/Placement_Data_Full_Class.csv"

library(dplyr)
library(ggplot2)
location <- "~/input/factors-affecting-campus-placement/Placement_Data_Full_Class.csv"
placement.df <- read.csv(path)
# select only relevant columns
placement.lr <- placement.df %>% select(ends_with("p"), -etest_p, status)
table(placement.lr$status)
```

```
Not Placed    Placed
        67         148
```

```
placement.lr$status <- ifelse(placement.lr$status == "Not Placed", 1, 0)
table(placement.lr$status)
```

```
0    1
148  67
```

```
library(caTools)
```

```
# Train and Test data
library(caTools) # to split data into train and test
set.seed(101)
sample <- sample.split(placement.lr$status, SplitRatio = 0.80)
train.lr = subset(placement.lr, sample == TRUE)
test.lr = subset(placement.lr, sample == FALSE)
#check the splits
prop.table(table(train.lr$status))
```

```
0    1
0.6860465 0.3139535
```

```
prop.table(table(test.lr$status))
```

```
0    1
0.6976744 0.3023256
```

```
# Train the model
model.lr <- glm(status ~ degree_p, family = binomial, data = train.lr)
summary(model.lr)
```

```
Call:
glm(formula = status ~ degree_p, family = binomial, data = train.lr)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  11.43688    2.24817   5.087 3.63e-07 ***
degree_p     -0.18851    0.03509  -5.372 7.79e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 214.05  on 171  degrees of freedom
Residual deviance: 173.35  on 170  degrees of freedom
AIC: 177.35

Number of Fisher Scoring iterations: 5
```

```
# prediction
lr.pred <- predict(model.lr, newdata = test.lr, type = "response")
head(lr.pred)
```

```
15    17    22    25    33    35
0.88198047 0.28303502 0.01008494 0.03139780 0.25345579 0.83675747
```

```
# The probabilities always refer to the class dummy-coded as "1"
head(test.lr$status)
```

```
[1] 1 0 0 0 0 1
```

```
# Classification Table
# categorize into groups based on the predicted probability
lr.pred.class <- ifelse(lr.pred>=0.5, 1, 0)
head(lr.pred.class)
```

```
15 17 22 25 33 35
1  0  0  0  0  1
```

```
table(lr.pred.class)
```

```
lr.pred.class
0    1
34    9
```

```
table(test.lr$status)
```

```
0    1
30   13
```

```
conf.matrix <- table(test.lr$status, lr.pred.class)
conf.matrix
```

```
lr.pred.class
0    1
0 30  0
1  4  9
```

```
rownames(conf.matrix) <- c("Placed", "Not Placed")
colnames(conf.matrix) <- c("Placed", "Not Placed")
addmargins(conf.matrix)
```

```
lr.pred.class
      Placed Not Placed Sum
Placed      30      0    30
Not Placed   4      9    13
Sum          34      9    43
```

```
# model accuracy
mean((test.lr$status == lr.pred.class))
```

```
[1] 0.9069767
```

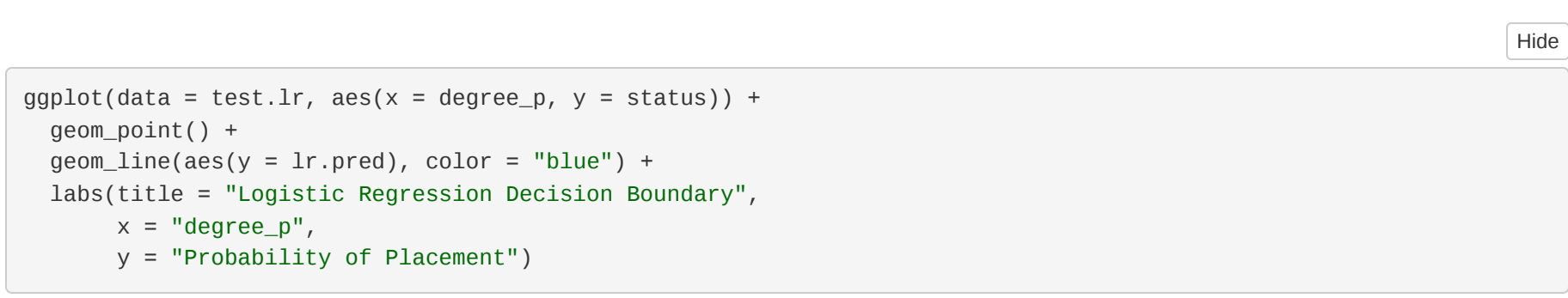
```
# different cut-off
lr.pred.class1 <- ifelse(lr.pred>=0.35, 1, 0)
conf.matrix1 <- table(test.lr$status, lr.pred.class1)
conf.matrix1
```

```
lr.pred.class1
0    1
0 27  3
1  2 11
```

```
Plots
```

```
ggplot(data = test.lr, aes(x = degree_p, y = status)) +
  geom_point() +
  geom_line(aes(y = lr.pred), color = "blue") +
  labs(title = "Logistic Regression Decision Boundary",
       x = "degree_p",
       y = "Probability of Placement")
```

```
Logistic Regression Decision Boundary
```



```
install.packages("pROC")
```

```
Installing package into '/opt/homebrew/lib/R/4.3/site-library'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/src/contrib/pROC_1.18.5.tar.gz'
Content type 'application/x-gzip' length 696162 bytes (679 KB)
=====
downloaded 679 KB

* installing 'source' package 'pROC' ...
** package 'pROC' successfully unpacked and MD5 sums checked
** using staged installation
** libs
using C++ compiler: 'Apple clang version 15.0.0 (clang-1500.0.40.1)'
using SDK: 'MacOSX14.2.sdk'
```

```
clang++ -std=gnu++17 -I"/opt/homebrew/Cellar/r/4.3.2/lib/R/include" -DNDEBUG -I"/opt/homebrew/lib/R/4.3/site-library/Rcpp/include" -I"/opt/homebrew/opt/gettext/include" -I"/opt/homebrew/opt/readline/include" -I"/opt/homebrew/opt/xz/include" -I"/opt/homebrew/include" -fPIC -g -O2 -c RcppExports.cpp -o RcppExports.o
clang++ -std=gnu++17 -I"/opt/homebrew/Cellar/r/4.3.2/lib/R/include" -DNDEBUG -I"/opt/homebrew/lib/R/4.3/site-library/Rcpp/include" -I"/opt/homebrew/opt/gettext/include" -I"/opt/homebrew/opt/readline/include" -I"/opt/homebrew/opt/xz/include" -I"/opt/homebrew/include" -fPIC -g -O2 -c RcppVersion.cpp -o RcppVersion.o
clang++ -std=gnu++17 -I"/opt/homebrew/Cellar/r/4.3.2/lib/R/include" -DNDEBUG -I"/opt/homebrew/lib/R/4.3/site-library/Rcpp/include" -I"/opt/homebrew/opt/gettext/include" -I"/opt/homebrew/opt/readline/include" -I"/opt/homebrew/opt/xz/include" -I"/opt/homebrew/include" -fPIC -g -O2 -c delong.cpp -o delong.o
clang++ -std=gnu++17 -I"/opt/homebrew/Cellar/r/4.3.2/lib/R/include" -DNDEBUG -I"/opt/homebrew/lib/R/4.3/site-library/Rcpp/include" -I"/opt/homebrew/opt/gettext/include" -I"/opt/homebrew/opt/readline/include" -I"/opt/homebrew/opt/xz/include" -I"/opt/homebrew/include" -fPIC -g -O2 -c perfAll.cpp -o perfAll.o
clang++ -std=gnu++17 -dynamiclib -Wl,-headerpad_max_install_names -undefined dynamic_lookup -L/opt/homebrew/Cellar/r/4.3.2/lib/R/lib -L/opt/homebrew/opt/gettext/lib -L/opt/homebrew/opt/readline/lib -L/opt/homebrew/opt/xz/lib -L/opt/homebrew/lib -o pROC.so RcppExports.o RcppVersion.o delong.o perfAll.o -L/opt/homebrew/Cellar/r/4.3.2/lib/R/lib -L -lR -lintl -Wl,-framework -Wl,CoreFoundation
```

```
installing to /opt/homebrew/lib/R/4.3/site-library/00LOCK-pROC/00new/pROC/libs
** R
** data
*** moving datasets to lazyload DB
** inst
** byte-compile and prepare package for lazy loading
** help
*** installing help indices
** building package indices
** testing if installed package can be loaded from temporary location
** checking absolute paths in shared objects and dynamic libraries
** testing if installed package can be loaded from final location
** testing if installed package keeps a record of temporary installation path
* DONE (pROC)

The downloaded source packages are in
'/private/var/folders/gs/jr7fq_pj3kdbf9s3vfs7680000gn/T/RtmpLbjxs0/downloaded_packages'
```

```
library(pROC)
```

```
Type 'citation("pROC")' for a citation.
```

```
Attaching package: 'pROC'
```

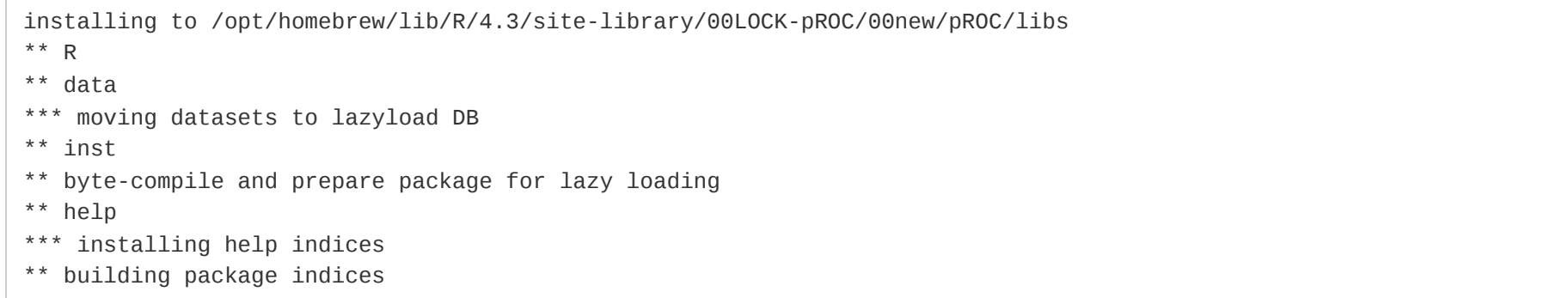
```
The following objects are masked from 'package:stats':
```

```
cov, smooth, var
```

```
roc_curve <- roc(test.lr$status, lr.pred)
```

```
Setting levels: control = 0, case = 1
Setting direction: controls < cases
```

```
plot(roc_curve, main = "ROC Curve", col = "blue", lwd = 2)
```



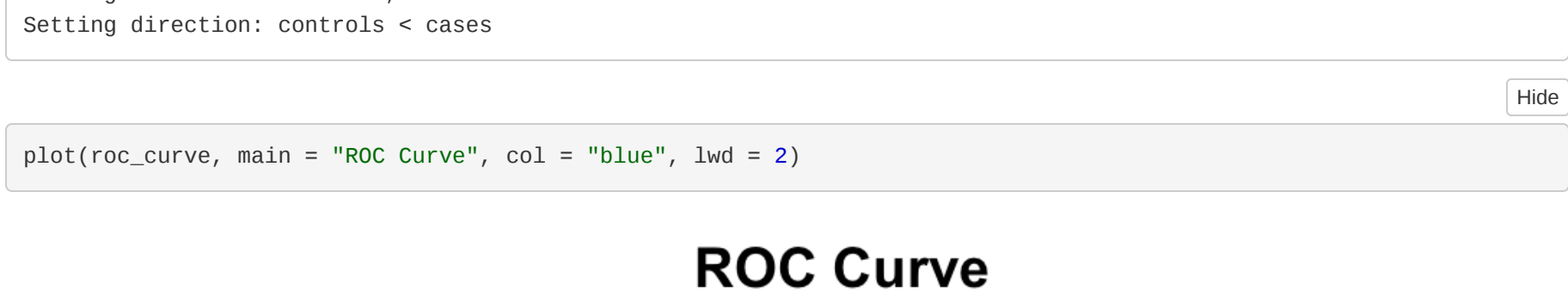
```
library(ggplot2)
```

```
# Convert confusion matrix to a data frame
conf_matrix_df <- as.data.frame.matrix(conf.matrix)
conf_matrix_df <- cbind(Actual = rownames(conf_matrix_df), conf_matrix_df)
```

```
# Reshape data for ggplot
conf_matrix_long <- tidyr::gather(conf_matrix_df, key = "Predicted", value = "Frequency", -Actual)
```

```
# Create heatmap using ggplot2
ggplot(data = conf_matrix_long, aes(x = Predicted, y = Actual, fill = Frequency)) +
  geom_tile() +
  labs(title = "Confusion Matrix", x = "Predicted", y = "Actual") +
  scale_fill_gradient(low = "white", high = "blue") +
  theme_minimal()
```

```
Confusion Matrix
```



```
NA
NA
```