STAT847 Lab 4

Anh Pham - ewa589

September 28, 2023

Contents

Question 1 **1**

Question 2 **1**

Question 3 **2**

Question 4 **3**

Question 1

Load the `mtcars.csv` file into an object names "mtcars.data". You can use `?mtcars` to get more information on the data set

```
mtcars.data <-read.csv('D:mtcars.csv')
head(mtcars.data)
```

```
##               model  mpg cyl disp  hp drat    wt  qsec vs am gear carb
## 1         Mazda RX4 21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## 2     Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## 3        Datsun 710 22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## 4    Hornet 4 Drive 21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## 5 Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## 6           Valiant 18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
?mtcars
```

```
## starting httpd help server ... done
```

Question 2

Fit a logistic regression model to predict `am`(transmission) using `mpg`, `disp`, `hp`, and `wt`.

```
myfit <- glm(am ~ mpg + disp + hp + wt, data = mtcars.data,
family = "binomial")
summary(myfit)
```

```
##
## Call:
## glm(formula = am ~ mpg + disp + hp + wt, family = "binomial",
##     data = mtcars.data)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -18.48207   40.90451  -0.452    0.651
## mpg           1.13503    1.55720   0.729    0.466
## disp         -0.02588    0.04087  -0.633    0.527
## hp            0.10871    0.09837   1.105    0.269
## wt           -4.80560    3.97978  -1.208    0.227
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 43.230  on 31  degrees of freedom
## Residual deviance:  8.162  on 27  degrees of freedom
## AIC: 18.162
##
## Number of Fisher Scoring iterations: 9
```

Print out the model coefficients

```
coef(myfit)
```

```
## (Intercept)         mpg        disp          hp          wt
## -18.4820714   1.1350288  -0.0258781   0.1087068  -4.8056004
```

Question 3

Use the `predict()` function to get the fitted probabilities. Then convert the probabilities to 1 or 0 using the threshold of 0.5. For example, if the fitted probability is greater than 0.5, it would equate to a 1 values, if the fitted probability is less than or equal to 0.5 it equates to a 0 value.

First, we get the fitted probability

```
glm.probs=predict(myfit, type="response")
head(glm.probs)
```

```
##           1           2           3           4           5           6
## 0.6410939258 0.3440465534 0.9724145105 0.0126014813 0.0166145986 0.0001266719
```

Now we convert the probilities to 1 or 0 using the threshold of 0.5

```
glm.pred<- ifelse(glm.probs > 0.5,1,0)
head(glm.pred)
```

```
## 1 2 3 4 5 6
## 1 0 1 0 0 0
```

Question 4

Create a confusion matrix to assess the accuracy of the fitted logistic model

Create a confusion matrix

```
table(glm.pred, mtcars.data$am)
```

```
##
## glm.pred  0  1
##        0 18  1
##        1  1 12
```

True classification rate

```
mean(glm.pred==mtcars.data$am)
```

```
## [1] 0.9375
```

mis-classification rate

```
mean(glm.pred!=mtcars.data$am)
```

```
## [1] 0.0625
```