Practical:Logistic regression

library(datasets)

> ir\_data<-(iris)

> head(ir\_data)

Sepal.Length Sepal.Width Petal.Length Petal.Width Species

1 5.1 3.5 1.4 0.2 setosa

2 4.9 3.0 1.4 0.2 setosa

3 4.7 3.2 1.3 0.2 setosa

4 4.6 3.1 1.5 0.2 setosa

5 5.0 3.6 1.4 0.2 setosa

6 5.4 3.9 1.7 0.4 setosa

> str(ir\_data)

'data.frame': 150 obs. of 5 variables:

$ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...

$ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...

$ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...

$ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...

$ Species : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...

> levels(ir\_data$Species)

[1] "setosa" "versicolor" "virginica"

> sum(is.na(ir\_data))

[1] 0

> ir\_data<-ir\_data[1:100,]

> set.seed(100)

> samp<-sample(1:100,80)

> ir\_test<-ir\_data[samp,]

> ir\_ctrl<-ir\_data[-samp,]

> install.packages("ggplot2")

Installing package into ‘C:/Users/PRAJAPATI GAURAV/Documents/R/win-library/4.1’

(as ‘lib’ is unspecified)

--- Please select a CRAN mirror for use in this session ---

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.1/ggplot2\_3.3.5.zip'

Content type 'application/zip' length 4130896 bytes (3.9 MB)

downloaded 3.9 MB

package ‘ggplot2’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\PRAJAPATI GAURAV\AppData\Local\Temp\RtmpimCVUC\downloaded\_packages

> library(ggplot2)

> install.packages("GGally")

Installing package into ‘C:/Users/PRAJAPATI GAURAV/Documents/R/win-library/4.1’

(as ‘lib’ is unspecified)

also installing the dependencies ‘generics’, ‘tidyselect’, ‘Rcpp’, ‘hms’, ‘prettyunits’, ‘purrr’, ‘cpp11’, ‘dplyr’, ‘forcats’, ‘plyr’, ‘progress’, ‘reshape’, ‘tidyr’

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.1/generics\_0.1.1.zip'

Content type 'application/zip' length 76719 bytes (74 KB)

downloaded 74 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.1/tidyselect\_1.1.1.zip'

Content type 'application/zip' length 204472 bytes (199 KB)

downloaded 199 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.1/Rcpp\_1.0.7.zip'

Content type 'application/zip' length 3263462 bytes (3.1 MB)

downloaded 3.1 MB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.1/hms\_1.1.1.zip'

Content type 'application/zip' length 104332 bytes (101 KB)

downloaded 101 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.1/prettyunits\_1.1.1.zip'

Content type 'application/zip' length 37957 bytes (37 KB)

downloaded 37 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.1/purrr\_0.3.4.zip'

Content type 'application/zip' length 429763 bytes (419 KB)

downloaded 419 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.1/cpp11\_0.4.2.zip'

Content type 'application/zip' length 327400 bytes (319 KB)

downloaded 319 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.1/dplyr\_1.0.7.zip'

Content type 'application/zip' length 1345382 bytes (1.3 MB)

downloaded 1.3 MB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.1/forcats\_0.5.1.zip'

Content type 'application/zip' length 358213 bytes (349 KB)

downloaded 349 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.1/plyr\_1.8.6.zip'

Content type 'application/zip' length 1500183 bytes (1.4 MB)

downloaded 1.4 MB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.1/progress\_1.2.2.zip'

Content type 'application/zip' length 85957 bytes (83 KB)

downloaded 83 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.1/reshape\_0.8.8.zip'

Content type 'application/zip' length 170776 bytes (166 KB)

downloaded 166 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.1/tidyr\_1.1.4.zip'

Content type 'application/zip' length 1075587 bytes (1.0 MB)

downloaded 1.0 MB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.1/GGally\_2.1.2.zip'

Content type 'application/zip' length 1637596 bytes (1.6 MB)

downloaded 1.6 MB

package ‘generics’ successfully unpacked and MD5 sums checked

package ‘tidyselect’ successfully unpacked and MD5 sums checked

package ‘Rcpp’ successfully unpacked and MD5 sums checked

package ‘hms’ successfully unpacked and MD5 sums checked

package ‘prettyunits’ successfully unpacked and MD5 sums checked

package ‘purrr’ successfully unpacked and MD5 sums checked

package ‘cpp11’ successfully unpacked and MD5 sums checked

package ‘dplyr’ successfully unpacked and MD5 sums checked

package ‘forcats’ successfully unpacked and MD5 sums checked

package ‘plyr’ successfully unpacked and MD5 sums checked

package ‘progress’ successfully unpacked and MD5 sums checked

package ‘reshape’ successfully unpacked and MD5 sums checked

package ‘tidyr’ successfully unpacked and MD5 sums checked

package ‘GGally’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\PRAJAPATI GAURAV\AppData\Local\Temp\RtmpimCVUC\downloaded\_packages

> library(GGally)

Registered S3 method overwritten by 'GGally':

method from

+.gg ggplot2

> ggpairs(ir\_test)

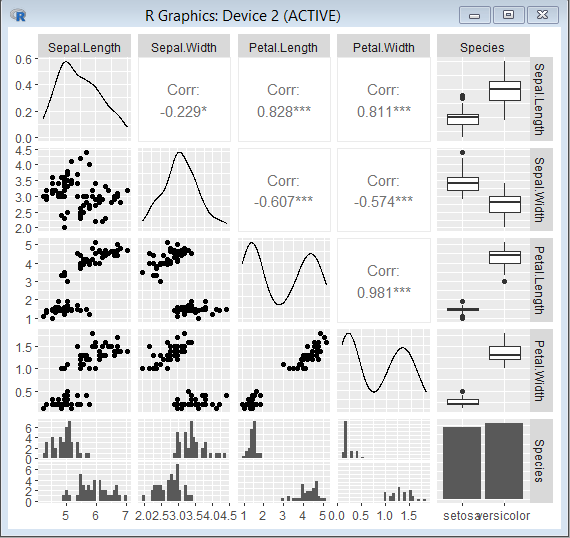
plot: [1,1] [=>------------------------------------------------] 4% est: 0s plot: [1,2] [===>----------------------------------------------] 8% est: 4s plot: [1,3] [=====>--------------------------------------------] 12% est: 4s plot: [1,4] [=======>------------------------------------------] 16% est: 3s plot: [1,5] [=========>----------------------------------------] 20% est: 3s plot: [2,1] [===========>--------------------------------------] 24% est: 3s plot: [2,2] [=============>------------------------------------] 28% est: 2s plot: [2,3] [===============>----------------------------------] 32% est: 2s plot: [2,4] [=================>--------------------------------] 36% est: 2s plot: [2,5] [===================>------------------------------] 40% est: 2s plot: [3,1] [=====================>----------------------------] 44% est: 2s plot: [3,2] [=======================>--------------------------] 48% est: 2s plot: [3,3] [=========================>------------------------] 52% est: 1s plot: [3,4] [===========================>----------------------] 56% est: 1s plot: [3,5] [=============================>--------------------] 60% est: 1s plot: [4,1] [===============================>------------------] 64% est: 1s plot: [4,2] [=================================>----------------] 68% est: 1s plot: [4,3] [===================================>--------------] 72% est: 1s plot: [4,4] [=====================================>------------] 76% est: 1s plot: [4,5] [=======================================>----------] 80% est: 1s plot: [5,1] [=========================================>--------] 84% est: 0s `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

plot: [5,2] [===========================================>------] 88% est: 0s `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

plot: [5,3] [=============================================>----] 92% est: 0s `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

plot: [5,4] [===============================================>--] 96% est: 0s `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

plot: [5,5] [==================================================]100% est: 0s



**y<-ir\_test$Species; x<-ir\_test$Sepal.Length**

**> glfit<-glm(y~x,family='binomial')**

**> summary(glfit)**

Call:

glm(formula = y ~ x, family = "binomial")

Deviance Residuals:

Min 1Q Median 3Q Max

-2.12681 -0.51865 0.02993 0.30652 2.25044

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -27.500 5.934 -4.634 3.59e-06 \*\*\*

x 5.112 1.109 4.611 4.01e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 110.854 on 79 degrees of freedom

Residual deviance: 48.818 on 78 degrees of freedom

AIC: 52.818

Number of Fisher Scoring iterations: 6

**> newdata<-data.frame(x=ir\_ctrl$Sepal.Length)**

**predicted\_val<-predict(glfit,newdata,type="response")**

**> prediction<-data.frame(ir\_ctrl$Sepal.Length,ir\_ctrl$Species,predicted\_val)**

**> prediction**

ir\_ctrl.Sepal.Length ir\_ctrl.Species predicted\_val

1 5.4 setosa 0.52665832

2 5.0 setosa 0.12584710

3 4.8 setosa 0.04923563

4 5.4 setosa 0.52665832

5 5.7 setosa 0.83759291

6 4.9 setosa 0.07948111

7 5.5 setosa 0.64975559

8 5.1 setosa 0.19357325

9 4.5 setosa 0.01104861

10 5.0 setosa 0.12584710

11 5.3 setosa 0.40023260

12 6.9 versicolor 0.99958015

13 5.7 versicolor 0.83759291

14 5.2 versicolor 0.28582944

15 5.6 versicolor 0.75569041

16 5.6 versicolor 0.75569041

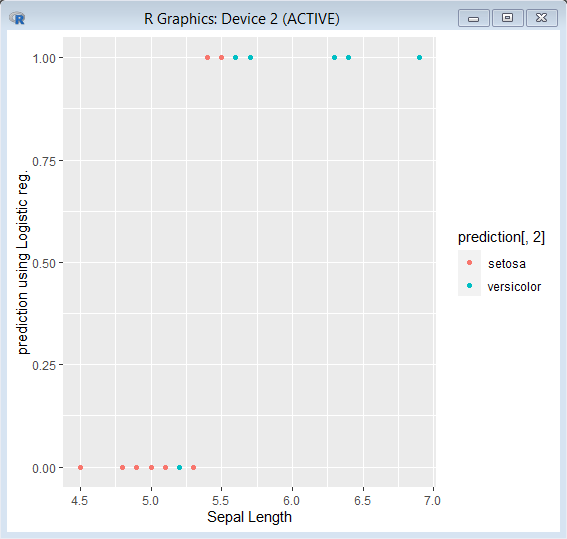
17 6.3 versicolor 0.99105619

18 6.4 versicolor 0.99461661

19 5.7 versicolor 0.83759291

20 5.7 versicolor 0.83759291

**qplot(prediction[,1],round(prediction[,3]),col=prediction[,2],xlab='Sepal Length', ylab='prediction using Logistic reg.')**

****

**library(datasets)**

**ir\_data<-(iris)**

**head(ir\_data)**

**str(ir\_data)**

**levels(ir\_data$Species)**

**sum(is.na(ir\_data))**

**ir\_data<-ir\_data[1:100,]**

**set.seed(100)**

**samp<-sample(1:100,80)**

**ir\_test<-ir\_data[samp,]**

**ir\_ctrl<-ir\_data[-samp,]**

**install.packages("ggplot2")**

**library(ggplot2)**

**install.packages("GGally")**

**library(GGally)**

**ggpairs(ir\_test)**

**y<-ir\_test$Species; x<-ir\_test$Sepal.Length**

**glfit<-glm(y~x,family='binomial')**

**summary(glfit)**

**newdata<-data.frame(x=ir\_ctrl$Sepal.Length)**

**predicted\_val<-predict(glfit,newdata,type="response")**

**prediction<-data.frame(ir\_ctrl$Sepal.Length,ir\_ctrl$Species,predicted\_val)**

**prediction**

**qplot(prediction[,1],round(prediction[,3]),col=prediction[,2],xlab='Sepal Length', ylab='prediction using Logistic reg.')**