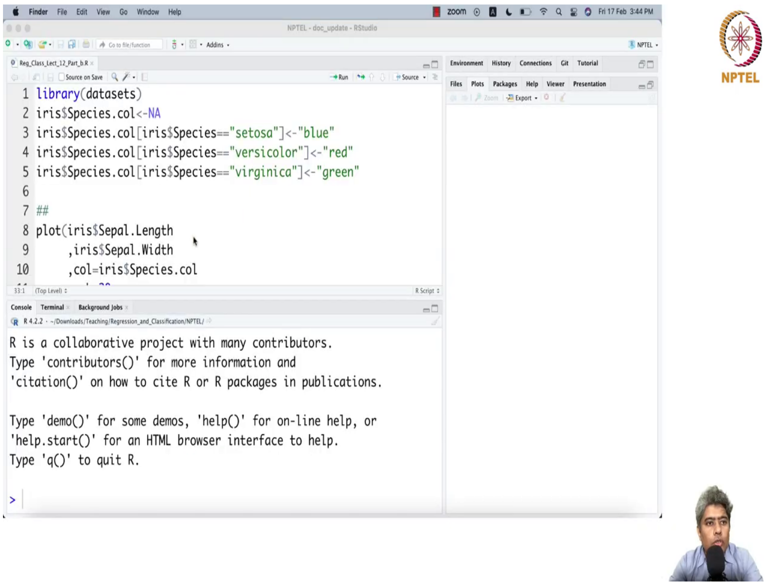
Predictive Analytics - Regression and Classification
Prof. Sourish Das
Department of Mathematics
Chennai Mathematical Institute

Lecture - 41
Hands on with R with Iris dataset

Welcome to the part b of lecture 12. In this video, we are going to do some Hands on. So, in this hands on we are going to load the dataset library which has the; which has the iris dataset.

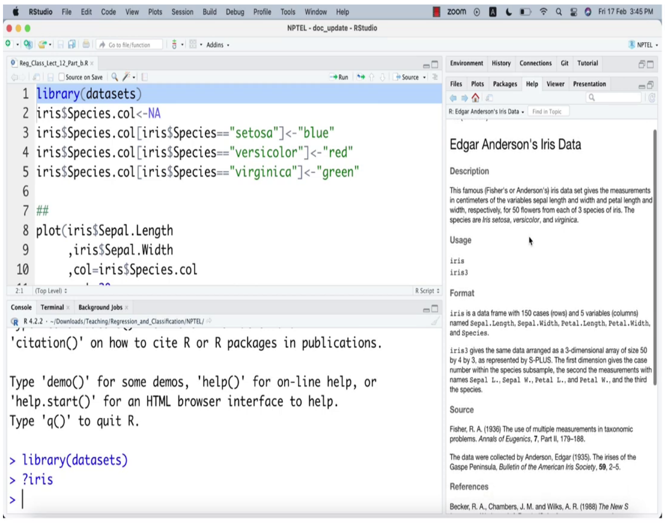
(Refer Slide Time: 00:23)


Finder File Edit View Go Window Help
Fri 17 Feb 3:44 PM
NPTEL - doc_update - RStudio
Go to file/function
Addins
NPTEL
Reg_Class_Lect_12_Part_b.R
Source on Save
Run
Source
Environment History Connections Git Tutorial
Files Plots Packages Help Viewer Presentation
Zoom
Export
33:1
(Top Level)
R Script
Console Terminal Background Jobs
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
NPTEL

```
library(datasets)
iris$Species.col<-NA
iris$Species.col[iris$Species=="setosa"]<-"blue"
iris$Species.col[iris$Species=="versicolor"]<-"red"
iris$Species.col[iris$Species=="virginica"]<-"green"

##
plot(iris$Sepal.Length
     ,iris$Sepal.Width
     ,col=iris$Species.col
```

```
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
```

(Refer Slide Time: 00:32)


RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
NPTEL - doc_update - RStudio
Reg_Class_Lect_12_Part_b.R
Source on Save
Run
Source
R Script
Console Terminal Background Jobs
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
Environment History Connections Git Tutorial
Files Plots Packages Help Viewer Presentation
R: Edgar Anderson's Iris Data
Find in Topic

```r
library(datasets)
iris$Species.col<-NA
iris$Species.col[iris$Species=="setosa"]<-"blue"
iris$Species.col[iris$Species=="versicolor"]<-"red"
iris$Species.col[iris$Species=="virginica"]<-"green"

##
plot(iris$Sepal.Length
     ,iris$Sepal.Width
     ,col=iris$Species.col
```

```
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> library(datasets)
> ?iris
> 
```

Edgar Anderson's Iris Data

Description

This famous (Fisher's or Anderson's) iris data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are *Iris setosa*, *versicolor*, and *virginica*.

Usage

```
iris
iris3
```

Format

`iris` is a data frame with 150 cases (rows) and 5 variables (columns) named `Sepal.Length`, `Sepal.Width`, `Petal.Length`, `Petal.Width`, and `Species`.

`iris3` gives the same data arranged as a 3-dimensional array of size 50 by 4 by 3, as represented by S-PLUS. The first dimension gives the case number within the species subsample, the second the measurements with names `Sepal L.`, `Sepal W.`, `Petal L.`, and `Petal W.`, and the third the species.

Source

Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, Part II, 179–188.

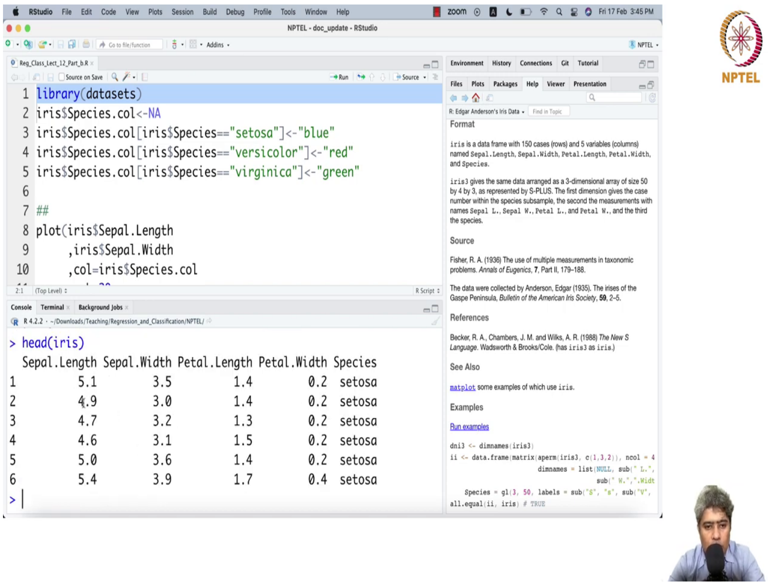
The data were collected by Anderson, Edgar (1935). The irises of the Gaspe Peninsula, *Bulletin of the American Iris Society*, **59**, 2–5.

References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S*

So, if you go to the load the, you write iris. So, Edgar Anderson’s iris datasets, it is also known as Fisher’s iris dataset. So, it has iris is a English flower, which has three subspecies, one is called Setosa, one is called versicolor, and one is called virginica.

(Refer Slide Time: 01:02)



RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
zoom Fri 17 Feb 3:45 PM
NPTEL - doc_update - RStudio
Reg_Class_Lect_12_Part_b.R
Source on Save
Run
Source
Environment History Connections Git Tutorial
Files Plots Packages Help Viewer Presentation
R: Edgar Anderson's Iris Data
Find in Topic
Console Terminal Background Jobs
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
R Script
NPTEL

```r
library(datasets)
iris$Species.col<-NA
iris$Species.col[iris$Species=="setosa"]<-"blue"
iris$Species.col[iris$Species=="versicolor"]<-"red"
iris$Species.col[iris$Species=="virginica"]<-"green"

##
plot(iris$Sepal.Length
     ,iris$Sepal.Width
     ,col=iris$Species.col
```

```
> head(iris)
  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
>
```

Format

iris is a data frame with 150 cases (rows) and 5 variables (columns) named Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, and Species.

iris3 gives the same data arranged as a 3-dimensional array of size 50 by 4 by 3, as represented by S-PLUS. The first dimension gives the case number within the species subsample, the second the measurements with names Sepal L., Sepal W., Petal L., and Petal W., and the third the species.

Source

Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, Part II, 179–188.

The data were collected by Anderson, Edgar (1935). The irises of the Gaspe Peninsula, *Bulletin of the American Iris Society*, **59**, 2–5.

References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole. (has iris3 as iris.)

See Also

matplot some examples of which use iris.

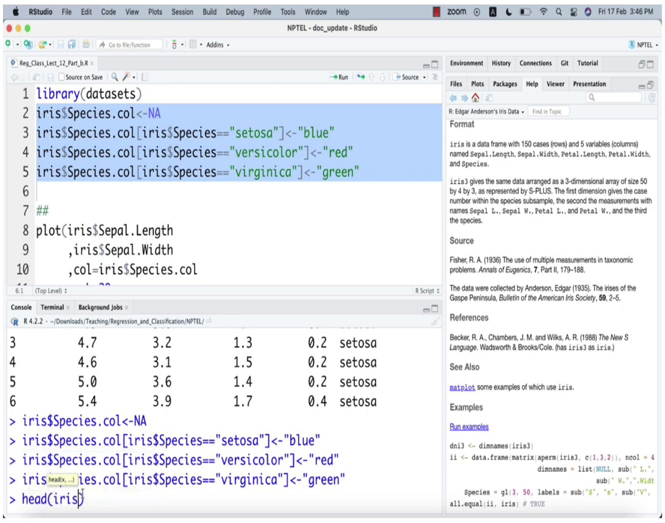
Examples

Run examples

```r
dni3 <- dimnames(iris3)
ii <- data.frame(matrix(aperm(iris3, c(1,3,2)), ncol = 4
                        dimnames = list(NULL, sub(" L.",
                                             sub(" W.",".Widt
    Species = gl(3, 50, labels = sub("S", "s", sub("V",
all.equal(ii, iris) # TRUE
```

In this dataset, what happens is if you just say, let us say, iris, head iris. So, it has four predictors, sepal length, sepal width, petal length, and petal width. This four based on these four phenotype can you say which species the flower belongs to? Ok. So, what we I have done here, I for different colors setosa, versicolor, and virginica.

(Refer Slide Time: 01:48)



```r
library(datasets)
iris$Species.col<-NA
iris$Species.col[iris$Species=="setosa"]<-"blue"
iris$Species.col[iris$Species=="versicolor"]<-"red"
iris$Species.col[iris$Species=="virginica"]<-"green"

##
plot(iris$Sepal.Length
     ,iris$Sepal.Width
     ,col=iris$Species.col
```

```
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
> iris$Species.col<-NA
> iris$Species.col[iris$Species=="setosa"]<-"blue"
> iris$Species.col[iris$Species=="versicolor"]<-"red"
> iris        ies.col[iris$Species=="virginica"]<-"green"
> head(iris)
```

Format

iris is a data frame with 150 cases (rows) and 5 variables (columns) named Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, and Species.

iris3 gives the same data arranged as a 3-dimensional array of size 50 by 4 by 3, as represented by S-PLUS. The first dimension gives the case number within the species subsample, the second the measurements with names Sepal L., Sepal W., Petal L., and Petal W., and the third the species.

Source

Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, Part II, 179–188.

The data were collected by Anderson, Edgar (1935). The irises of the Gaspe Peninsula, *Bulletin of the American Iris Society*, **59**, 2–5.

References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole. (has iris3 as iris.)

See Also

matplot some examples of which use iris.

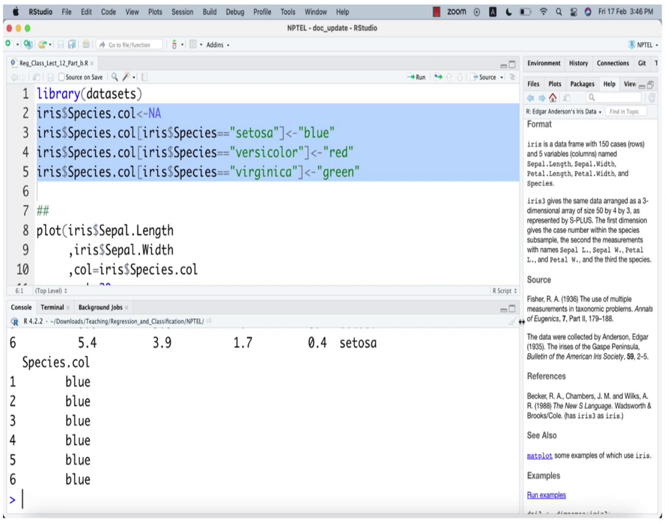
Examples

Run examples

```r
dni3 <- dimnames(iris3)
ii <- data.frame(matrix(aperm(iris3, c(1,3,2)), ncol = 4
                        dimnames = list(NULL, sub(" L.",
                                        sub(" W.",".Widt
    Species = gl(3, 50, labels = sub("S", "s", sub("V",
all.equal(ii, iris) # TRUE
```

I have given different color, for different species, I have now had different color. So, now, if you go to head iris.

(Refer Slide Time: 01:55)



RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
Fri 17 Feb 3:46 PM
NPTEL - doc_update - RStudio
Go to file/function
Addins
NPTEL
Reg_Class_Lect_12_Part_b.R
Source on Save
Run
Source
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
Environment
History
Connections
Git
Files
Plots
Packages
Help
View
R: Edgar Anderson's Iris Data
Find in Topic

```
library(datasets)
iris$Species.col<-NA
iris$Species.col[iris$Species=="setosa"]<-"blue"
iris$Species.col[iris$Species=="versicolor"]<-"red"
iris$Species.col[iris$Species=="virginica"]<-"green"

##
plot(iris$Sepal.Length
     ,iris$Sepal.Width
     ,col=iris$Species.col
```

```
6          5.4         3.9          1.7         0.4  setosa
  Species.col
1        blue
2        blue
3        blue
4        blue
5        blue
6        blue
>
```

Format

iris is a data frame with 150 cases (rows) and 5 variables (columns) named Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, and Species.

iris3 gives the same data arranged as a 3-dimensional array of size 50 by 4 by 3, as represented by S-PLUS. The first dimension gives the case number within the species subsample, the second the measurements with names Sepal L., Sepal W., Petal L., and Petal W., and the third the species.

Source

Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, Part II, 179–188.

The data were collected by Anderson, Edgar (1935). The irises of the Gaspe Peninsula, *Bulletin of the American Iris Society*, **59**, 2–5.

References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole. (has iris3 as iris.)

See Also

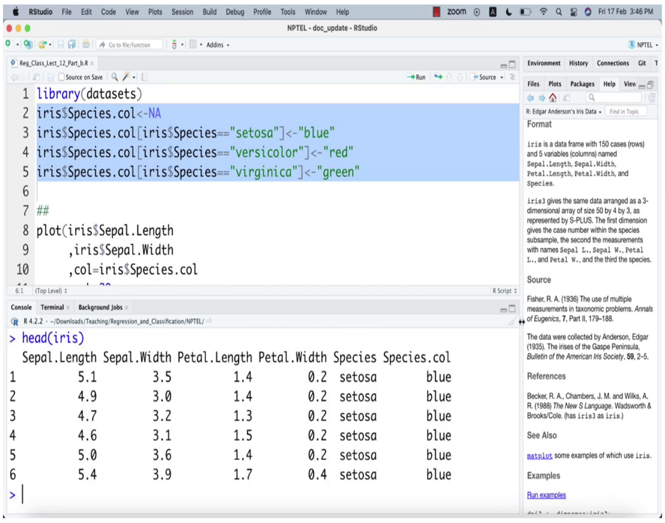
matplot some examples of which use iris.

Examples

Run examples

So, you will see that in along with that, let me just.

(Refer Slide Time: 01:59)


NPTEL - doc_update - RStudio
Reg_Class_Lect_12_Part_b.R
Run
Source
Console
Terminal
Background Jobs
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
Environment
History
Connections
Git
Files
Plots
Packages
Help
View
R: Edgar Anderson's Iris Data
Find in Topic
Fri 17 Feb 3:46 PM

```
library(datasets)
iris$Species.col<-NA
iris$Species.col[iris$Species=="setosa"]<-"blue"
iris$Species.col[iris$Species=="versicolor"]<-"red"
iris$Species.col[iris$Species=="virginica"]<-"green"

##
plot(iris$Sepal.Length
     ,iris$Sepal.Width
     ,col=iris$Species.col
```

```
> head(iris)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species Species.col
1          5.1         3.5          1.4         0.2  setosa        blue
2          4.9         3.0          1.4         0.2  setosa        blue
3          4.7         3.2          1.3         0.2  setosa        blue
4          4.6         3.1          1.5         0.2  setosa        blue
5          5.0         3.6          1.4         0.2  setosa        blue
6          5.4         3.9          1.7         0.4  setosa        blue
> 
```

Format

iris is a data frame with 150 cases (rows) and 5 variables (columns) named Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, and Species.

iris3 gives the same data arranged as a 3-dimensional array of size 50 by 4 by 3, as represented by S-PLUS. The first dimension gives the case number within the species subsample, the second the measurements with names Sepal L., Sepal W., Petal L., and Petal W., and the third the species.

Source

Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, Part II, 179–188.

The data were collected by Anderson, Edgar (1935). The irises of the Gaspe Peninsula, *Bulletin of the American Iris Society*, **59**, 2–5.

References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole. (has iris3 as iris.)

See Also

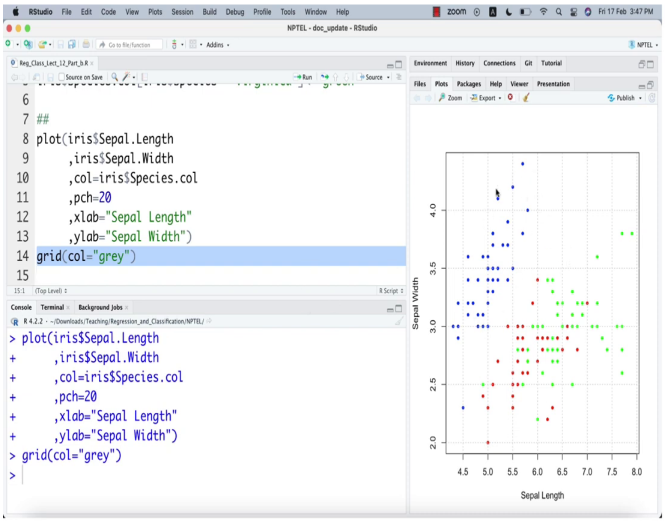
matplot some examples of which use iris.

Examples

Run examples

Yeah. So, now, this is the new column, that last column is a new column. For setosa, I had blue, versicolor I have given red, and virginica I given green.

(Refer Slide Time: 02:18)



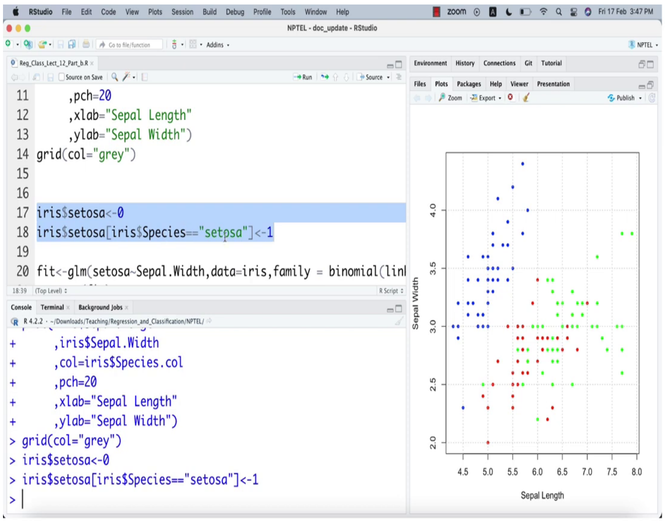
```
##
plot(iris$Sepal.Length
     ,iris$Sepal.Width
     ,col=iris$Species.col
     ,pch=20
     ,xlab="Sepal Length"
     ,ylab="Sepal Width")
grid(col="grey")
```

```
> plot(iris$Sepal.Length
+      ,iris$Sepal.Width
+      ,col=iris$Species.col
+      ,pch=20
+      ,xlab="Sepal Length"
+      ,ylab="Sepal Width")
> grid(col="grey")
>
```

Now, if I now if you run this piece of code, I just plotting from iris dataset, I am just taking the sepal length and sepal width. I am just extracting it and giving it as a x value and y values. And if you run this plot, then you got this plot, you want, you can give a grid also.

So, you can see these are like bluer setosa, we have given bluer setosa, reds are red points are all versicolor, and the green points are virginica.

(Refer Slide Time: 02:58)

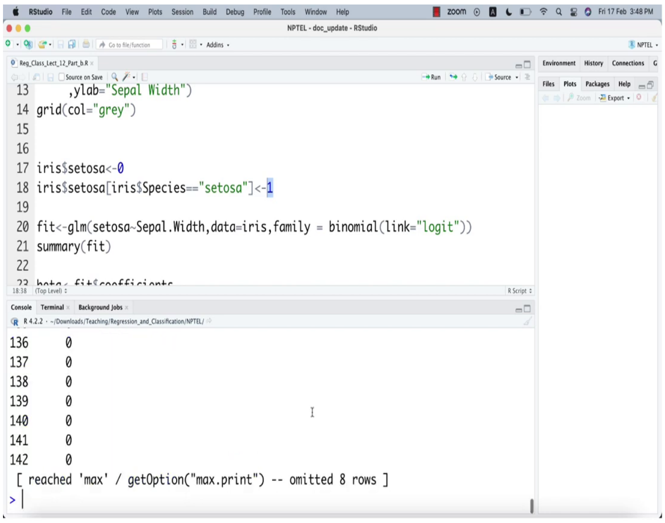

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
NPTEL - doc_update - RStudio
Reg_Class_Lect_12_Part_b.R
Source on Save
Run
Source
Environment History Connections Git Tutorial
Files Plots Packages Help Viewer Presentation
Zoom
Export
Publish
18:39
(Top Level)
R Script
Console Terminal Background Jobs
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
Sepal Width
Sepal Length
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0

```
      ,pch=20
      ,xlab="Sepal Length"
      ,ylab="Sepal Width")
grid(col="grey")


iris$setosa<-0
iris$setosa[iris$Species=="setosa"]<-1

fit<-glm(setosa~Sepal.Width,data=iris,family = binomial(linl
```

```
+       ,iris$Sepal.Width
+       ,col=iris$Species.col
+       ,pch=20
+       ,xlab="Sepal Length"
+       ,ylab="Sepal Width")
> grid(col="grey")
> iris$setosa<-0
> iris$setosa[iris$Species=="setosa"]<-1
> |
```

Now, what I have done in that dataset, I have created a one-hot encoding for setosa. If it is setosa, then it will have a 0 or it is 1.

(Refer Slide Time: 03:21)

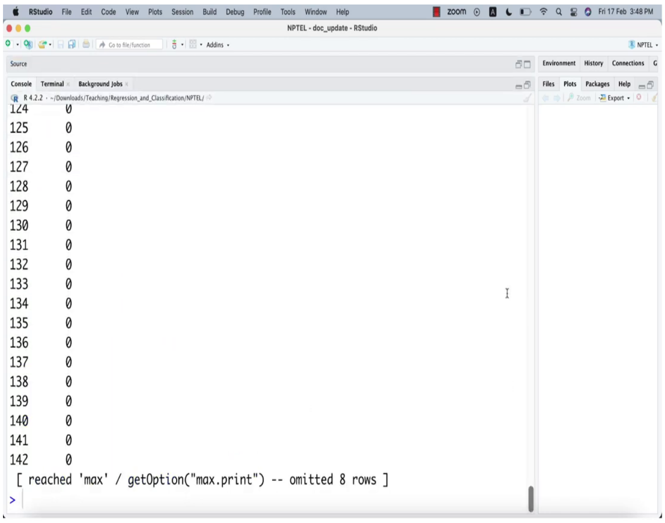

NPTEL - doc_update - RStudio
Reg_Class_Lect_12_Part_b.R
Source on Save
Run
Source
Environment
History
Connections
Files
Plots
Packages
Help
Export
Console
Terminal
Background Jobs
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/

```
      ,ylab="Sepal Width")
grid(col="grey")


iris$setosa<-0
iris$setosa[iris$Species=="setosa"]<-1

fit<-glm(setosa~Sepal.Width,data=iris,family = binomial(link="logit"))
summary(fit)

beta<-fit$coefficients
```

```
136      0
137      0
138      0
139      0
140      0
141      0
142      0
 [ reached 'max' / getOption("max.print") -- omitted 8 rows ]
>
```

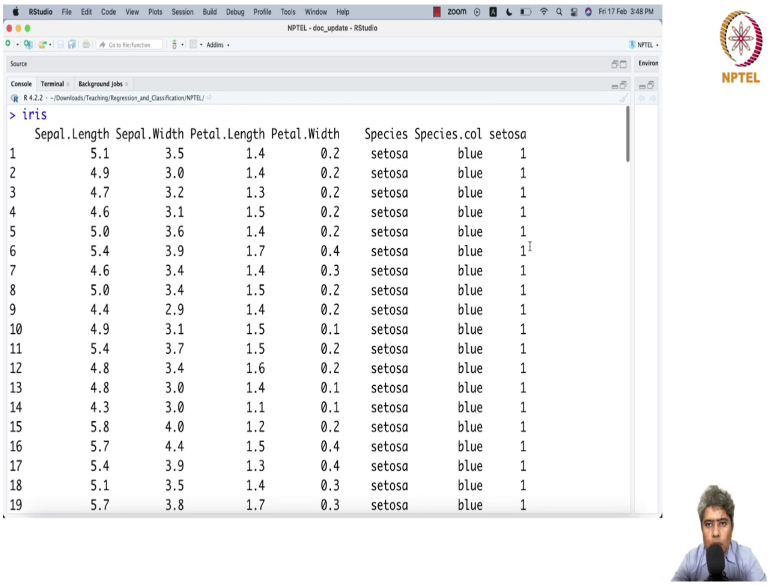
So, for all it is 0, but whenever it will find setosa, it will get a 1 otherwise, it will be 0. So, now the dataset, let me just let me just show you how the dataset look likes now, iris, ok.

(Refer Slide Time: 03:36)


RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
Fri 17 Feb 3:48 PM
NPTEL - doc_update - RStudio
Addins
NPTEL
Source
Console Terminal Background Jobs
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
Environment History Connections
Files Plots Packages Help
Zoom Export

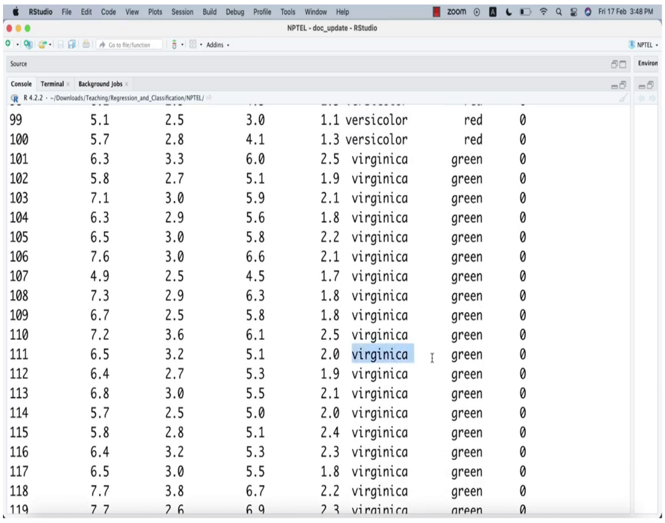
```
124     0
125     0
126     0
127     0
128     0
129     0
130     0
131     0
132     0
133     0
134     0
135     0
136     0
137     0
138     0
139     0
140     0
141     0
142     0
 [ reached 'max' / getOption("max.print") -- omitted 8 rows ]
>
```

(Refer Slide Time: 03:40)


RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
Fri 17 Feb 3:48 PM
NPTEL - doc_update - RStudio
Go to file/function
Addins
NPTEL
Source
Console
Terminal
Background Jobs
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
NPTEL

```
> iris
    Sepal.Length Sepal.Width Petal.Length Petal.Width    Species Species.col setosa
1            5.1         3.5          1.4         0.2     setosa        blue      1
2            4.9         3.0          1.4         0.2     setosa        blue      1
3            4.7         3.2          1.3         0.2     setosa        blue      1
4            4.6         3.1          1.5         0.2     setosa        blue      1
5            5.0         3.6          1.4         0.2     setosa        blue      1
6            5.4         3.9          1.7         0.4     setosa        blue      1
7            4.6         3.4          1.4         0.3     setosa        blue      1
8            5.0         3.4          1.5         0.2     setosa        blue      1
9            4.4         2.9          1.4         0.2     setosa        blue      1
10           4.9         3.1          1.5         0.1     setosa        blue      1
11           5.4         3.7          1.5         0.2     setosa        blue      1
12           4.8         3.4          1.6         0.2     setosa        blue      1
13           4.8         3.0          1.4         0.1     setosa        blue      1
14           4.3         3.0          1.1         0.1     setosa        blue      1
15           5.8         4.0          1.2         0.2     setosa        blue      1
16           5.7         4.4          1.5         0.4     setosa        blue      1
17           5.4         3.9          1.3         0.4     setosa        blue      1
18           5.1         3.5          1.4         0.3     setosa        blue      1
19           5.7         3.8          1.7         0.3     setosa        blue      1
```

(Refer Slide Time: 03:42)

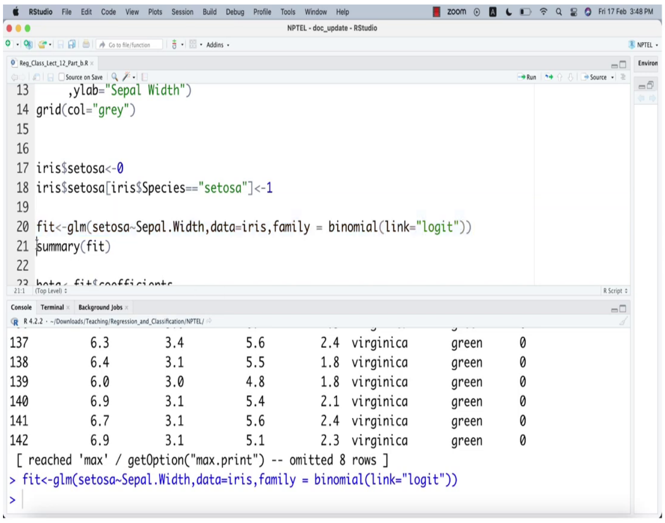


RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
Fri 17 Feb 3:48 PM
NPTEL - doc_update - RStudio
Go to file/function
Addins
NPTEL
Source
Environ
Console
Terminal
Background Jobs
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/

```
99          5.1        2.5         3.0         1.1 versicolor      red     0
100         5.7        2.8         4.1         1.3 versicolor      red     0
101         6.3        3.3         6.0         2.5  virginica    green     0
102         5.8        2.7         5.1         1.9  virginica    green     0
103         7.1        3.0         5.9         2.1  virginica    green     0
104         6.3        2.9         5.6         1.8  virginica    green     0
105         6.5        3.0         5.8         2.2  virginica    green     0
106         7.6        3.0         6.6         2.1  virginica    green     0
107         4.9        2.5         4.5         1.7  virginica    green     0
108         7.3        2.9         6.3         1.8  virginica    green     0
109         6.7        2.5         5.8         1.8  virginica    green     0
110         7.2        3.6         6.1         2.5  virginica    green     0
111         6.5        3.2         5.1         2.0  virginica    green     0
112         6.4        2.7         5.3         1.9  virginica    green     0
113         6.8        3.0         5.5         2.1  virginica    green     0
114         5.7        2.5         5.0         2.0  virginica    green     0
115         5.8        2.8         5.1         2.4  virginica    green     0
116         6.4        3.2         5.3         2.3  virginica    green     0
117         6.5        3.0         5.5         1.8  virginica    green     0
118         7.7        3.8         6.7         2.2  virginica    green     0
119         7.7        2.6         6.9         2.3  virginica    green     0
```

So, let me just, yeah. So, first few values are setosa and then when it is versicolor, it is coded as 0, when it is virginica, this is coded as 0, ok.

(Refer Slide Time: 04:25)

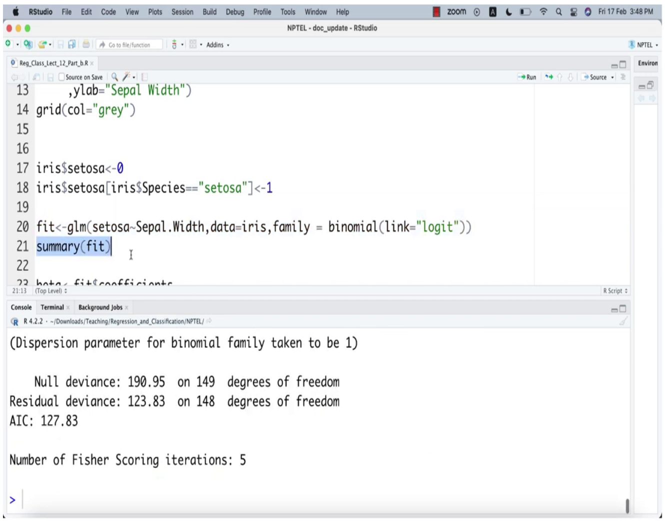

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
zoom
Fri 17 Feb 3:48 PM
NPTEL - doc_update - RStudio
Go to file/function
Addins
NPTEL
Reg_Class_Lect_12_Part_b.R
Source on Save
Run
Source
Environ
21:1
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/

```
      ,ylab="Sepal Width")
grid(col="grey")


iris$setosa<-0
iris$setosa[iris$Species=="setosa"]<-1

fit<-glm(setosa~Sepal.Width,data=iris,family = binomial(link="logit"))
summary(fit)

beta<-fit$coefficients
```

```
137          6.3         3.4          5.6         2.4  virginica     green    0
138          6.4         3.1          5.5         1.8  virginica     green    0
139          6.0         3.0          4.8         1.8  virginica     green    0
140          6.9         3.1          5.4         2.1  virginica     green    0
141          6.7         3.1          5.6         2.4  virginica     green    0
142          6.9         3.1          5.1         2.3  virginica     green    0
 [ reached 'max' / getOption("max.print") -- omitted 8 rows ]
> fit<-glm(setosa~Sepal.Width,data=iris,family = binomial(link="logit"))
> 
```

And the column name is setosa. So, that is how I created a one-hot encoding or binary class variable, sometime it is called indicator variable. So, let me run the model, glm, setosa as a function of Sepal Width, data equal to iris, you give family equal to binomial link equal to logit.

(Refer Slide Time: 04:29)

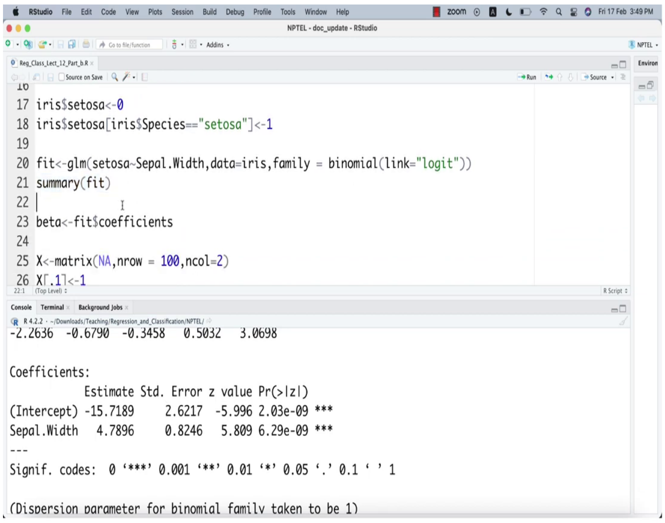


RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
NPTEL - doc_update - RStudio
Reg_Class_Lect_12_Part_b.R
Source on Save
Run
Source
R Script
Console
Terminal
Background Jobs
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/

```
      ,ylab="Sepal Width")
grid(col="grey")


iris$setosa<-0
iris$setosa[iris$Species=="setosa"]<-1

fit<-glm(setosa~Sepal.Width,data=iris,family = binomial(link="logit"))
summary(fit)

beta<-fit$coefficients
```

```
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 190.95  on 149  degrees of freedom
Residual deviance: 123.83  on 148  degrees of freedom
AIC: 127.83

Number of Fisher Scoring iterations: 5

>
```

(Refer Slide Time: 04:32)

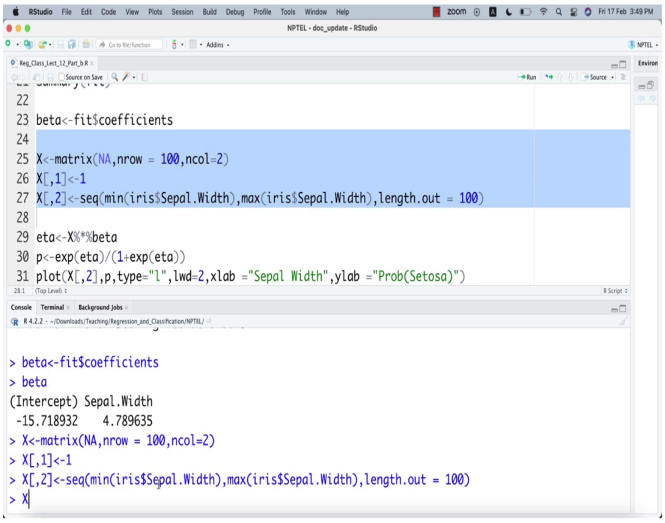

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
zoom Fri 17 Feb 3:49 PM
NPTEL - doc_update - RStudio
Go to file/function
Addins
NPTEL
Reg_Class_Lect_12_Part_b.R
Source on Save
Run
Source
Environ
22:1
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/

```
iris$setosa<-0
iris$setosa[iris$Species=="setosa"]<-1

fit<-glm(setosa~Sepal.Width,data=iris,family = binomial(link="logit"))
summary(fit)


beta<-fit$coefficients

X<-matrix(NA,nrow = 100,ncol=2)
X[,1]<-1
```

```
-2.2636  -0.6790  -0.3458   0.5032   3.0698

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -15.7189     2.6217  -5.996 2.03e-09 ***
Sepal.Width   4.7896     0.8246   5.809 6.29e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)
```

And if you run summary fit, then this is the fit that you will get.

(Refer Slide Time: 04:35)

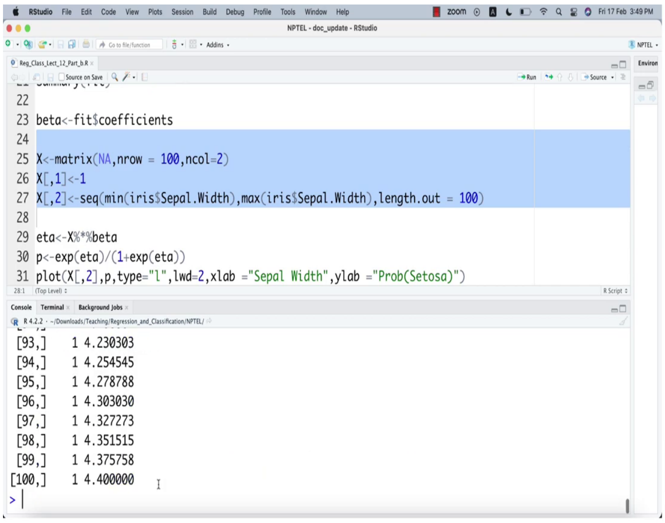

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
zoom Fri 17 Feb 3:49 PM
NPTEL - doc_update - RStudio
Go to file/function
Addins
NPTEL
Reg_Class_Lect_12_Part_b.R
Source on Save
Run
Source
Environ
28:1
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/

```
beta<-fit$coefficients

X<-matrix(NA,nrow = 100,ncol=2)
X[,1]<-1
X[,2]<-seq(min(iris$Sepal.Width),max(iris$Sepal.Width),length.out = 100)

eta<-X%*%beta
p<-exp(eta)/(1+exp(eta))
plot(X[,2],p,type="l",lwd=2,xlab ="Sepal Width",ylab ="Prob(Setosa)")
```

```
> beta<-fit$coefficients
> beta
(Intercept) Sepal.Width
 -15.718932    4.789635
> X<-matrix(NA,nrow = 100,ncol=2)
> X[,1]<-1
> X[,2]<-seq(min(iris$Sepal.Width),max(iris$Sepal.Width),length.out = 100)
> X
```

If you run the beta, you can from the fit, you can extract the coefficients. Now, in the beta, I have the coefficient. Now, what I am going to do, I am going to calculate eta X matrix with 100 values with which X 2 takes values with minimum value of Sepal Width to maximum value of Sepal Width.

(Refer Slide Time: 05:04)



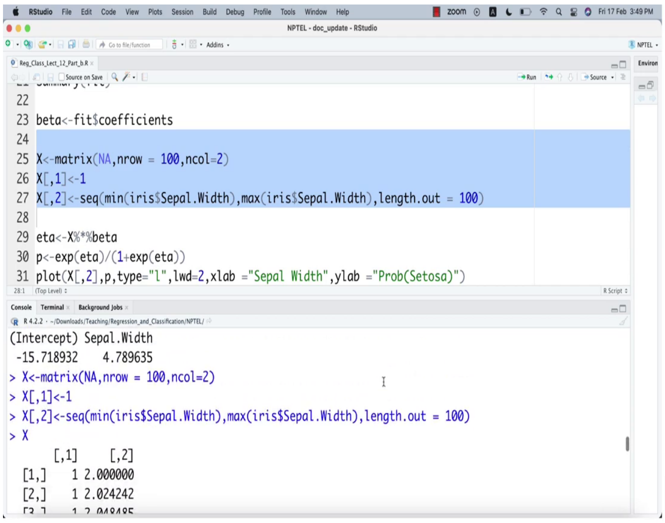
RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
Fri 17 Feb 3:49 PM
NPTEL - doc_update - RStudio
Go to file/function
Addins
NPTEL
Reg_Class_Lect_12_Part_b.R
Source on Save
Run
Source
Environ
28:1
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/

```
beta<-fit$coefficients

X<-matrix(NA,nrow = 100,ncol=2)
X[,1]<-1
X[,2]<-seq(min(iris$Sepal.Width),max(iris$Sepal.Width),length.out = 100)

eta<-X%*%beta
p<-exp(eta)/(1+exp(eta))
plot(X[,2],p,type="l",lwd=2,xlab ="Sepal Width",ylab ="Prob(Setosa)")
```

```
 [93,]    1 4.230303
 [94,]    1 4.254545
 [95,]    1 4.278788
 [96,]    1 4.303030
 [97,]    1 4.327273
 [98,]    1 4.351515
 [99,]    1 4.375758
[100,]    1 4.400000
>
```

So, if you just run it. So, minimum is I think 2.2 is the minimum.

(Refer Slide Time: 05:08)



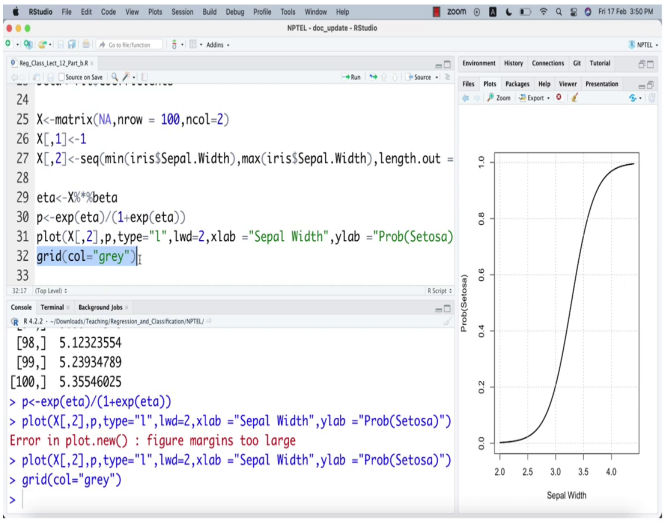
RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
Fri 17 Feb 3:49 PM
NPTEL - doc_update - RStudio
Reg_Class_Lect_12_Part_b.R
Source on Save
Run
Source
Console
Terminal
Background Jobs
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/

```r
beta<-fit$coefficients

X<-matrix(NA,nrow = 100,ncol=2)
X[,1]<-1
X[,2]<-seq(min(iris$Sepal.Width),max(iris$Sepal.Width),length.out = 100)

eta<-X%*%beta
p<-exp(eta)/(1+exp(eta))
plot(X[,2],p,type="l",lwd=2,xlab ="Sepal Width",ylab ="Prob(Setosa)")
```

```
(Intercept) Sepal.Width
 -15.718932    4.789635
> X<-matrix(NA,nrow = 100,ncol=2)
> X[,1]<-1
> X[,2]<-seq(min(iris$Sepal.Width),max(iris$Sepal.Width),length.out = 100)
> X
       [,1]     [,2]
  [1,]    1 2.000000
  [2,]    1 2.024242
  [3,]    1 2.048485
```

And max is the 4.4. And in between, it just fill up with some values in a equal width. And then I calculate the eta, these are the eta values or z values, latent variable values. And then I calculate the probability.

(Refer Slide Time: 05:29)



RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
NPTEL - doc_update - RStudio
Reg_Class_Lect_12_Part_b.R
Source on Save
Run
Source
Environment History Connections Git Tutorial
Files Plots Packages Help Viewer Presentation
Zoom
Export
32:17
(Top Level)
R Script
Console Terminal Background Jobs
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
Prob(Setosa)
0.0
0.2
0.4
0.6
0.8
1.0
2.0
2.5
3.0
3.5
4.0
Sepal Width

```
X<-matrix(NA,nrow = 100,ncol=2)
X[,1]<-1
X[,2]<-seq(min(iris$Sepal.Width),max(iris$Sepal.Width),length.out =

eta<-X%*%beta
p<-exp(eta)/(1+exp(eta))
plot(X[,2],p,type="l",lwd=2,xlab ="Sepal Width",ylab ="Prob(Setosa)
grid(col="grey")
```

```
 [98,]  5.12323554
 [99,]  5.23934789
[100,]  5.35546025
> p<-exp(eta)/(1+exp(eta))
> plot(X[,2],p,type="l",lwd=2,xlab ="Sepal Width",ylab ="Prob(Setosa)")
Error in plot.new() : figure margins too large
> plot(X[,2],p,type="l",lwd=2,xlab ="Sepal Width",ylab ="Prob(Setosa)")
> grid(col="grey")
>
```

And then I plot the X 2 versus Sepal Width versus sorry, but that sepal width versus the probability of setosa. So, if the Sepal Width increases, clearly probability that the flower is setosa increases. So, you can put a grid also in this. So, that is how we, this is that is how we can, you can draw probability of p, you can plot p against some predictor values.

So, I will stop here, but this iris data set shows you that your target variable could be not necessarily has to be binary class. It could be multi class variable. And then in that case, you have to you, you have a multi class classification. Because you have three class here. Remember that you have to do three class, not binary class.

You have a setosa, versicolor and virginica, three sub species are there. When you have more than two class, basically it is a k class problem or multi class problem. The most popular and the old and tested method is linear discriminant analysis.

So, in the next video, we are going to start linear discriminant analysis or Fisher's linear discriminant analysis. So, for now, thank you very much for your attention. See you in the next video.