Labs: Trees, Hierarchical Clustering, Project Dataset One-on-One with the instructor

Thilanka Munasinghe

Data Analytics

ITWS-4600/ITWS-6600/MATP-4450/CSCI-4960

Group 3 - Lab 4, 7th October 2022

Remaining Labs: Group1 & 2

- Continue working on the remaining code snippets from Group 1 and Group 2 labs.
- After you finish them, make sure to push your code to the GitHub Repository.

Scripts – work through these

Reminder to finish these code examples See in folder group2/ Lab1

Go over the following scrips,

Lab1_bronx1.R.

Lab1_bronx2.R

Lab1_ctree2.R

Lab1_kknn1.R

Lab1_kknn2.R

Lab1_kknn3.R

Lab1_kmeans1.R

Lab1_nyt.R

Search before you ask! You might need to search your code errors online when you are debugging your code!

script fragments in R available on the web site:
https://rpi.box.com/s/2xx9ul1fmc6bf5ff8h4jreae69emikmf

NOTE: you are allowed to work in small groups and discuss during this lab.

Scripts – work through these

Next…

See in folder group2/ Lab3

Go over the following scrips,

Lab3_ctree1.R

Lab3_ctree2.R

Lab3_ctree3.R

…..

And the remaining code snippets in group2/Lab 2 and Lab3

Search before you ask! You might need to search your code errors online when you are debugging your code!

script fragments in R available on the web site:

https://rpi.box.com/s/lu00cugurbk5mdvr0u58ztwsjufy57n2

NOTE: you are allowed to work in small groups and discuss during this lab.

Scripts – work through these

Next…

See in folder group2 and group3/ Labs

Go over the following scrips,

Lab3_ctree1.R

Lab3_ctree2.R

Lab3_ctree3.R

…..

And the remaining code snippets in group2/Lab 2 and Lab3

Search before you ask! You might need to search your code errors online when you are debugging your code!

script fragments in R available on the web site:
https://rpi.box.com/s/2xx9ul1fmc6bf5ff8h4jreae69emikmf

NOTE: you are allowed to work in small groups and discuss during this lab.

Trees for the Titanic

data(Titanic)

rpart, ctree, hclust for:

Survived ~ .

Read the titanic dataset documentation in Rdocumentation:
https://www.rdocumentation.org/packages/titanic/versions/0.1.0

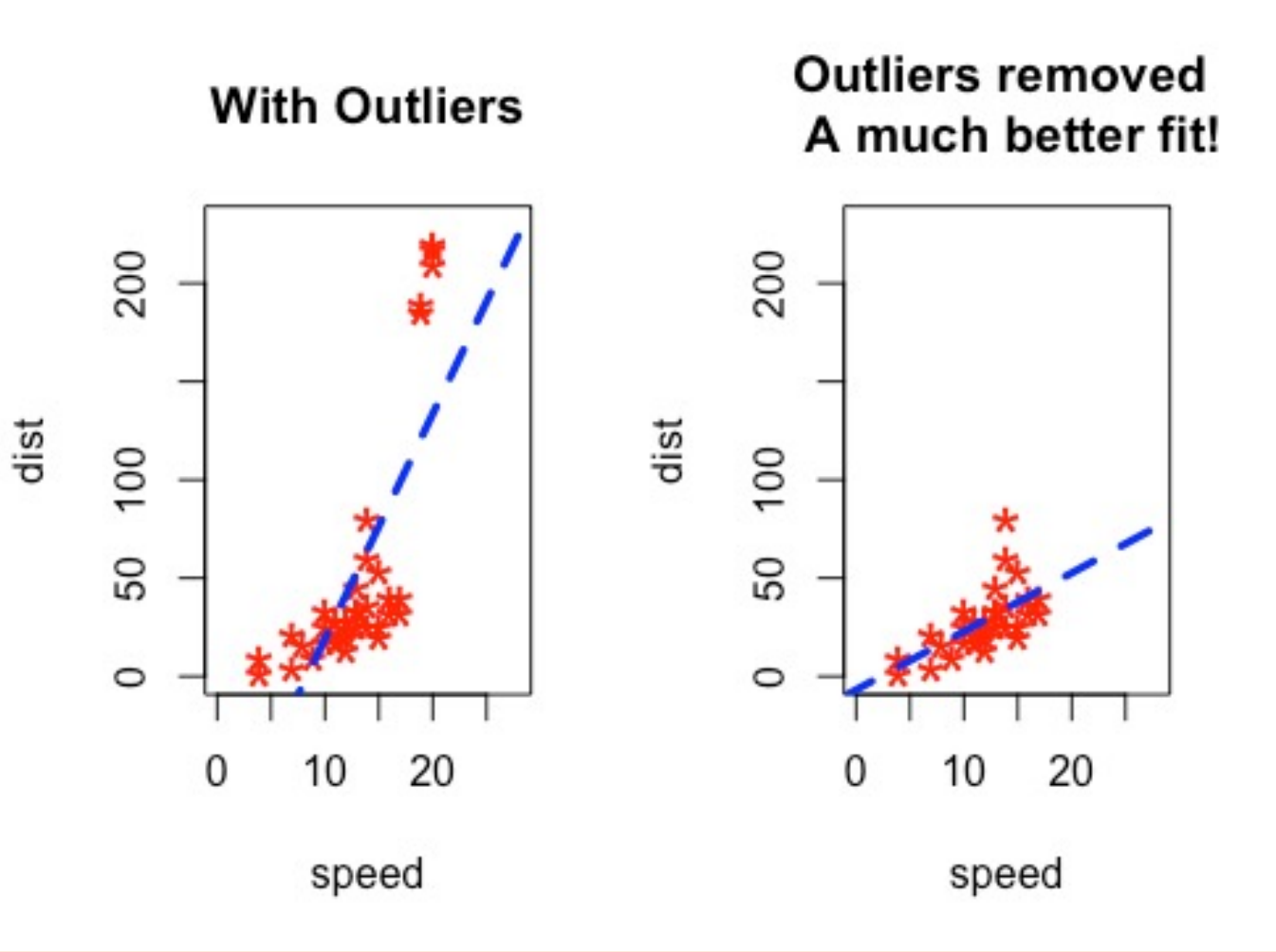
Outliers in Data: Example

```
# Outlier Examples
# Cars dataset is built in Rstudio.
# you need to load the cars dataset first.
cars1 <- cars[1:30,] # first 30 rows of the original cars dataset.
head(cars1)
# Now we will introduce some additional data points that are outliers.
cars_outliers <- data.frame(speed=c(19,19,20,20,20), dist=c(190,186,210,220,218)) # introduced the outliears
head(cars_outliers)
cars2 <- rbind(cars1, cars_outliers)
help(par) # Set or Query Graphical Parameters, read the RStudio documentation for "par" function.


par(mfrow=c(1, 2))
plot(cars2$speed, cars2$dist, xlim=c(0, 28), ylim=c(0, 230), main="With Outliers", xlab="speed", ylab="dist", pch="*", col="red", cex=2)
abline(lm(dist ~ speed, data=cars2), col="blue", lwd=3, lty=2)


# Plot of original data without outliers. Note the change in slope (angle) of best fit line.
plot(cars1$speed, cars1$dist, xlim=c(0, 28), ylim=c(0, 230), main="Outliers removed \n A much better fit!", xlab="speed", ylab="dist", pch="*", col="red", cex=2)
abline(lm(dist ~ speed, data=cars1), col="blue", lwd=3, lty=2)
```

Outliers Example …



KNN & KMeans Examples

- Work on the additional Code Snippets provided in LMS (under this week): Examples on KNN and KMeans.
- These two exercises are from the Textbook Introduction to Statistical Learning With R~ 7th Edition.

Project One-on-One

- Today, during the class we will do the project One-on-One to document your datasets.