

Untitled

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(ggplot2)
library(maps)
library(lattice)
library(caret)
library(gbm)

## Loaded gbm 2.1.5

library(gridExtra)

data(scat)
str(scat)

## 'data.frame': 110 obs. of 19 variables:
## $ Species : Factor w/ 3 levels "bobcat","coyote",...: 2 2 1 2 2 2 1 1 1 1 ...
## $ Month : Factor w/ 9 levels "April","August",...: 4 4 4 4 4 4 4 4 4 4 ...
## $ Year : int 2012 2012 2012 2012 2012 2012 2012 2012 2012 2012 ...
## $ Site : Factor w/ 2 levels "ANNU","YOLA": 2 2 2 2 2 2 1 1 1 1 ...
## $ Location : Factor w/ 3 levels "edge","middle",...: 1 1 2 2 1 1 3 3 3 2 ...
## $ Age : int 5 3 3 5 5 5 1 3 5 5 ...
## $ Number : int 2 2 2 2 4 3 5 7 2 1 ...
## $ Length : num 9.5 14 9 8.5 8 9 6 5.5 11 20.5 ...
## $ Diameter : num 25.7 25.4 18.8 18.1 20.7 21.2 15.7 21.9 17.5 18 ...
## $ Taper : num 41.9 37.1 16.5 24.7 20.1 28.5 8.2 19.3 29.1 21.4 ...
## $ TI : num 1.63 1.46 0.88 1.36 0.97 1.34 0.52 0.88 1.66 1.19 ...
## $ Mass : num 15.9 17.6 8.4 7.4 25.4 ...
## $ d13C : num -26.9 -29.6 -28.7 -20.1 -23.2 ...
## $ d15N : num 6.94 9.87 8.52 5.79 7.01 8.28 4.2 3.89 7.34 6.06 ...
## $ CN : num 8.5 11.3 8.1 11.5 10.6 9 5.4 5.6 5.8 7.7 ...
## $ ropey : int 0 0 1 1 0 1 1 0 0 1 ...
## $ segmented: int 0 0 1 0 1 0 1 1 1 1 ...
## $ flat : int 0 0 0 0 0 0 0 0 0 0 ...
## $ scrape : int 0 0 1 0 0 0 1 0 0 0 ...

sum(is.na(scat))

## [1] 47

df = subset(scat)
#Set the Species column as the target/outcome and convert it to numeric
df$Species<-as.numeric(factor(df$Species))

#Remove the Month, Year, Site, Location features
df = subset(scat, select = -c(Month,Year,Site, Location) )

#Check if any values are null. If there are, impute missing values using KNN
sum(is.na(scat))

## [1] 47

preProcValues <- preProcess(df, method = c("knnImpute","center","scale"))

#Converting every categorical variable to numerical
library('RANN')
train_processed <- predict(preProcValues, df)
sum(is.na(train_processed))

## [1] 0

dmy <- dummyVars("~ .", data = train_processed,fullRank = T)
train_transformed <- data.frame(predict(dmy, newdata = train_processed))

str(train_transformed)

## 'data.frame': 110 obs. of 16 variables:
## $ Species.coyote : num 1 1 0 1 1 1 0 0 0 0 ...
## $ Species.gray_fox: num 0 0 0 0 0 0 0 0 0 0 ...
## $ Age : num 1.207 -0.252 -0.252 1.207 1.207 ...
## $ Number : num -0.433 -0.433 -0.433 -0.433 0.968 ...
## $ Length : num 0.0587 1.3679 -0.0867 -0.2322 -0.3777 ...
## $ Diameter : num 1.8396 1.7623 0.0622 -0.1181 0.5516 ...
## $ Taper : num 0.961 0.642 -0.726 -0.182 -0.487 ...
## $ TI : num 0.0283 -0.1406 -0.7171 -0.24 -0.6277 ...
## $ Mass : num 0.388 0.583 -0.458 -0.571 1.469 ...
## $ d13C : num 0.00468 -1.26856 -0.85947 3.12113 1.66403 ...
## $ d15N : num -0.165 0.807 0.359 -0.546 -0.141 ...
## $ CN : num 0.0276 0.7922 -0.0816 0.8468 0.6011 ...
## $ ropey : num -1.131 -1.131 0.876 0.876 -1.131 ...
## $ segmented : num -1.131 -1.131 0.876 -1.131 0.876 ...
## $ flat : num -0.239 -0.239 -0.239 -0.239 -0.239 ...
## $ scrape : num -0.217 -0.217 4.562 -0.217 -0.217 ...

sum(is.na(train_transformed))

## [1] 0
```

Including Plots

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.