Assignment 2

Saipriya Gourineni

2022-10-02

R Markdown

library(‘caret’) library(‘ISLR’) library(‘dplyr’) library(‘class’)

Creating the new Dataset

UniversalBankData <- read.csv(“C:/Users/Saipr/Desktop/Machine learning 2/UniversalBank.csv”, sep = ‘,’)

UniversalBankData$ID < -NULLUniversalBankData$ZIP.Code <- NULL summary(UniversalBankData)

UniversalBankData$Personal.Loan = as.factor(UniversalBankData$Personal.Loan)

Normalized_model <- preProcess(UniversalBankData[, -8],method = c(“center”, “scale”)) Bank_normalized <- predict(Normalized_model,UniversalBankData) summary(Bank_normalized)

#dividing the data into 60% for training dataset and 40% for testing dataset

training_index <- createDataPartition(UniversalBankData$Personal.Loan, p = 0.6, list = FALSE) training.df = Bank_normalized[training_index,] validation.df = Bank_normalized[-training_index,]

#Prediction

To_Predict = data.frame(Age = 40, Experience = 10, Income = 84, Family = 2, CCAvg = 2, Education = 1, Mortgage = 0, Securities.Account = 0, CD.Account = 0, Online = 1, CreditCard = 1) print(To_Predict) To_Predict_Normalized <- predict(Normalized_model,To_Predict)

Prediction <- knn(train= training.df[,1:7,9:12], test = To_Predict_Normalized[,1:7,9:12], cl= training.df$Personal.Loan, k=1) print(Prediction)

#Question 2

set.seed(123) Bankcontrol <- trainControl(method= “repeatedcv”, number = 3, repeats = 2) searchGrid = expand.grid(k=1:10)

knn.model = train(Personal.Loan~., data = training.df, method = ‘knn’, tuneGrid = searchGrid,trControl = Bankcontrol)

knn.model

#Question 3

predictions <- predict(knn.model,validation.df)

confusionMatrix(predictions,validation.df$Personal.Loan)

#Question 4

To_Predict_Normalization = data.frame(Age = 40, Experience = 10, Income = 84, Family = 2, CCAvg = 2, Education = 1, Mortgage = 0, Securities.Account =0, CD.Account = 0, Online = 1, CreditCard = 1) To_Predict_Normalization = predict(Normalized_model, To_Predict) predict(knn.model, To_Predict_Normalization)

#Question 5

training_size = 0.5 training_index = createDataPartition(UniversalBankData$Personal.Loan, p = 0.5, list = FALSE) training.df = Bank_normalized[training_index,]

testing_size = 0.2 testing_index = createDataPartition(UniversalBankData$Personal.Loan, p = 0.2, list = FALSE) testing.df = Bank_normalized[testing_index,]

valid_size = 0.3 Validation_index = createDataPartition(UniversalBankData$Personal.Loan, p = 0.3, list = FALSE) validation.df = Bank_normalized[Validation_index,]

testingknn <- knn(train = training.df[,-8], test = testing.df[,-8], cl = training.df[,8], k =3) Validationknn <- knn(train = training.df[,-8], test = validation.df[,-8], cl = training.df[,8], k =3) trainingknn <- knn(train = training.df[,-8], test = train.df[,-8], cl = training.df[,8], k =3)

confusionMatrix(testingknn, testing.df[,8]) confusionMatrix(trainingknn, training.df[,8]) confusionMatrix(Validationknn, validation.df[,8])

#From the result, I can say that Training accuracy is slightly higher than the testing and validation sets. So the alogorithm is working accordingly. 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.

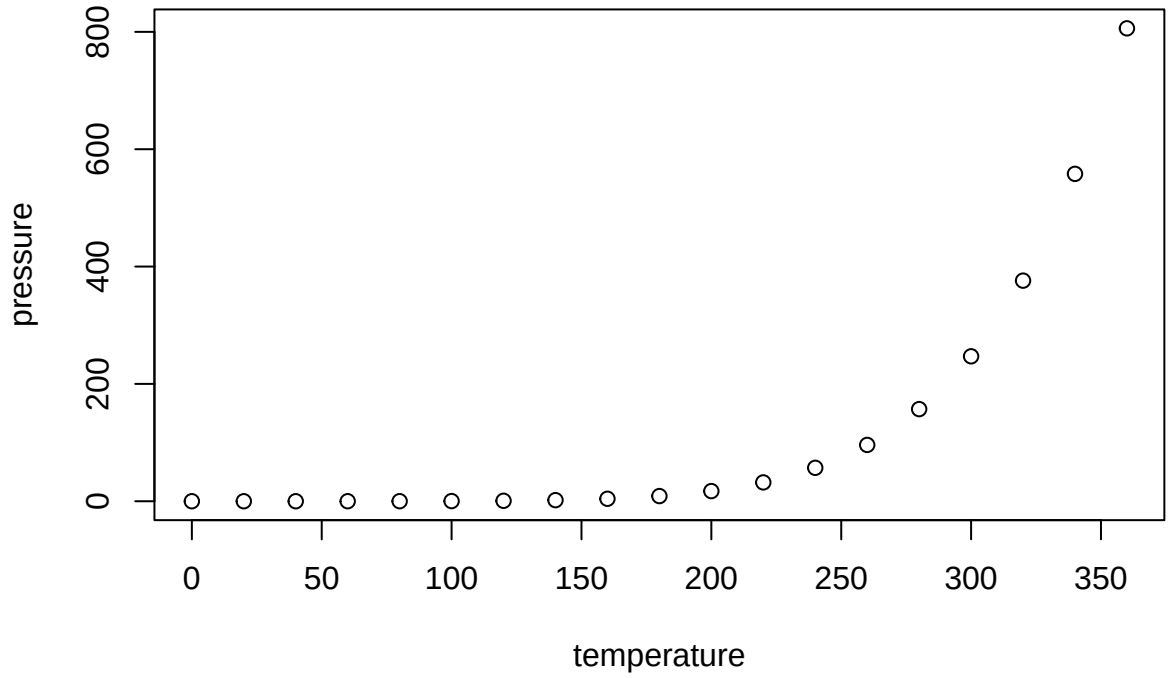
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

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.