workshop 4.8 solution

Veerasak Kritsanapraphan

1/22/2020

library(class)  
library(caret)  
library(e1071)  
library(stringr)  
library(neuralnet)  
library(fastDummies)  
library(dplyr)  
telco <- read.csv("WA\_Fn-UseC\_-Telco-Customer-Churn.csv", header=TRUE)

str(telco)

## 'data.frame': 7043 obs. of 21 variables:  
## $ customerID : Factor w/ 7043 levels "0002-ORFBO","0003-MKNFE",..: 5376 3963 2565 5536 6512 6552 1003 4771 5605 4535 ...  
## $ gender : Factor w/ 2 levels "Female","Male": 1 2 2 2 1 1 2 1 1 2 ...  
## $ SeniorCitizen : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Partner : Factor w/ 2 levels "No","Yes": 2 1 1 1 1 1 1 1 2 1 ...  
## $ Dependents : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 2 1 1 2 ...  
## $ tenure : int 1 34 2 45 2 8 22 10 28 62 ...  
## $ PhoneService : Factor w/ 2 levels "No","Yes": 1 2 2 1 2 2 2 1 2 2 ...  
## $ MultipleLines : Factor w/ 3 levels "No","No phone service",..: 2 1 1 2 1 3 3 2 3 1 ...  
## $ InternetService : Factor w/ 3 levels "DSL","Fiber optic",..: 1 1 1 1 2 2 2 1 2 1 ...  
## $ OnlineSecurity : Factor w/ 3 levels "No","No internet service",..: 1 3 3 3 1 1 1 3 1 3 ...  
## $ OnlineBackup : Factor w/ 3 levels "No","No internet service",..: 3 1 3 1 1 1 3 1 1 3 ...  
## $ DeviceProtection: Factor w/ 3 levels "No","No internet service",..: 1 3 1 3 1 3 1 1 3 1 ...  
## $ TechSupport : Factor w/ 3 levels "No","No internet service",..: 1 1 1 3 1 1 1 1 3 1 ...  
## $ StreamingTV : Factor w/ 3 levels "No","No internet service",..: 1 1 1 1 1 3 3 1 3 1 ...  
## $ StreamingMovies : Factor w/ 3 levels "No","No internet service",..: 1 1 1 1 1 3 1 1 3 1 ...  
## $ Contract : Factor w/ 3 levels "Month-to-month",..: 1 2 1 2 1 1 1 1 1 2 ...  
## $ PaperlessBilling: Factor w/ 2 levels "No","Yes": 2 1 2 1 2 2 2 1 2 1 ...  
## $ PaymentMethod : Factor w/ 4 levels "Bank transfer (automatic)",..: 3 4 4 1 3 3 2 4 3 1 ...  
## $ MonthlyCharges : num 29.9 57 53.9 42.3 70.7 ...  
## $ TotalCharges : num 29.9 1889.5 108.2 1840.8 151.7 ...  
## $ Churn : Factor w/ 2 levels "No","Yes": 1 1 2 1 2 2 1 1 2 1 ...

## Convert Data

### Dummy Column

telcodum <- select(telco, Churn, gender,SeniorCitizen,Partner,tenure,PhoneService, MultipleLines,OnlineBackup,Contract,PaperlessBilling,PaymentMethod, MonthlyCharges ,TotalCharges)  
telcodum <- dummy\_cols(telcodum)  
names(telcodum) <- str\_replace\_all(names(telcodum), c(" " = "." , "," = ""))  
str(telcodum)

## 'data.frame': 7043 obs. of 36 variables:  
## $ Churn : Factor w/ 2 levels "No","Yes": 1 1 2 1 2 2 1 1 2 1 ...  
## $ gender : Factor w/ 2 levels "Female","Male": 1 2 2 2 1 1 2 1 1 2 ...  
## $ SeniorCitizen : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Partner : Factor w/ 2 levels "No","Yes": 2 1 1 1 1 1 1 1 2 1 ...  
## $ tenure : int 1 34 2 45 2 8 22 10 28 62 ...  
## $ PhoneService : Factor w/ 2 levels "No","Yes": 1 2 2 1 2 2 2 1 2 2 ...  
## $ MultipleLines : Factor w/ 3 levels "No","No phone service",..: 2 1 1 2 1 3 3 2 3 1 ...  
## $ OnlineBackup : Factor w/ 3 levels "No","No internet service",..: 3 1 3 1 1 1 3 1 1 3 ...  
## $ Contract : Factor w/ 3 levels "Month-to-month",..: 1 2 1 2 1 1 1 1 1 2 ...  
## $ PaperlessBilling : Factor w/ 2 levels "No","Yes": 2 1 2 1 2 2 2 1 2 1 ...  
## $ PaymentMethod : Factor w/ 4 levels "Bank transfer (automatic)",..: 3 4 4 1 3 3 2 4 3 1 ...  
## $ MonthlyCharges : num 29.9 57 53.9 42.3 70.7 ...  
## $ TotalCharges : num 29.9 1889.5 108.2 1840.8 151.7 ...  
## $ Churn\_No : int 1 1 0 1 0 0 1 1 0 1 ...  
## $ Churn\_Yes : int 0 0 1 0 1 1 0 0 1 0 ...  
## $ gender\_Female : int 1 0 0 0 1 1 0 1 1 0 ...  
## $ gender\_Male : int 0 1 1 1 0 0 1 0 0 1 ...  
## $ Partner\_Yes : int 1 0 0 0 0 0 0 0 1 0 ...  
## $ Partner\_No : int 0 1 1 1 1 1 1 1 0 1 ...  
## $ PhoneService\_No : int 1 0 0 1 0 0 0 1 0 0 ...  
## $ PhoneService\_Yes : int 0 1 1 0 1 1 1 0 1 1 ...  
## $ MultipleLines\_No.phone.service : int 1 0 0 1 0 0 0 1 0 0 ...  
## $ MultipleLines\_No : int 0 1 1 0 1 0 0 0 0 1 ...  
## $ MultipleLines\_Yes : int 0 0 0 0 0 1 1 0 1 0 ...  
## $ OnlineBackup\_Yes : int 1 0 1 0 0 0 1 0 0 1 ...  
## $ OnlineBackup\_No : int 0 1 0 1 1 1 0 1 1 0 ...  
## $ OnlineBackup\_No.internet.service : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Contract\_Month-to-month : int 1 0 1 0 1 1 1 1 1 0 ...  
## $ Contract\_One.year : int 0 1 0 1 0 0 0 0 0 1 ...  
## $ Contract\_Two.year : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ PaperlessBilling\_Yes : int 1 0 1 0 1 1 1 0 1 0 ...  
## $ PaperlessBilling\_No : int 0 1 0 1 0 0 0 1 0 1 ...  
## $ PaymentMethod\_Electronic.check : int 1 0 0 0 1 1 0 0 1 0 ...  
## $ PaymentMethod\_Mailed.check : int 0 1 1 0 0 0 0 1 0 0 ...  
## $ PaymentMethod\_Bank.transfer.(automatic): int 0 0 0 1 0 0 0 0 0 1 ...  
## $ PaymentMethod\_Credit.card.(automatic) : int 0 0 0 0 0 0 1 0 0 0 ...

### Data Sampling

index <- sample(2, nrow(telcodum), replace=TRUE, prob=c(0.7,0.3) )  
traindata <- telcodum[index==1,]  
testdata <- telcodum[index==2,]  
sprintf("Number of Record in Training Dataset is %d" , nrow(traindata))

## [1] "Number of Record in Training Dataset is 4896"

sprintf("Number of Record in Testing Dataset is %d" , nrow(testdata))

## [1] "Number of Record in Testing Dataset is 2147"

### Train Model

telco\_train <- traindata[, c(14,15,3,5,16:36)]  
telco\_test <- testdata[, c(14,15,3,5,16:36)]   
str(telco\_train)

## 'data.frame': 4896 obs. of 25 variables:  
## $ Churn\_No : int 1 1 0 0 1 1 0 1 1 1 ...  
## $ Churn\_Yes : int 0 0 1 1 0 0 1 0 0 0 ...  
## $ SeniorCitizen : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ tenure : int 1 34 2 8 22 10 28 13 16 58 ...  
## $ gender\_Female : int 1 0 1 1 0 1 1 0 0 0 ...  
## $ gender\_Male : int 0 1 0 0 1 0 0 1 1 1 ...  
## $ Partner\_Yes : int 1 0 0 0 0 0 1 1 0 1 ...  
## $ Partner\_No : int 0 1 1 1 1 1 0 0 1 0 ...  
## $ PhoneService\_No : int 1 0 0 0 0 1 0 0 0 0 ...  
## $ PhoneService\_Yes : int 0 1 1 1 1 0 1 1 1 1 ...  
## $ MultipleLines\_No.phone.service : int 1 0 0 0 0 1 0 0 0 0 ...  
## $ MultipleLines\_No : int 0 1 1 0 0 0 0 1 1 0 ...  
## $ MultipleLines\_Yes : int 0 0 0 1 1 0 1 0 0 1 ...  
## $ OnlineBackup\_Yes : int 1 0 0 0 1 0 0 0 0 0 ...  
## $ OnlineBackup\_No : int 0 1 1 1 0 1 1 1 0 1 ...  
## $ OnlineBackup\_No.internet.service : int 0 0 0 0 0 0 0 0 1 0 ...  
## $ Contract\_Month-to-month : int 1 0 1 1 1 1 1 1 0 0 ...  
## $ Contract\_One.year : int 0 1 0 0 0 0 0 0 0 1 ...  
## $ Contract\_Two.year : int 0 0 0 0 0 0 0 0 1 0 ...  
## $ PaperlessBilling\_Yes : int 1 0 1 1 1 0 1 1 0 0 ...  
## $ PaperlessBilling\_No : int 0 1 0 0 0 1 0 0 1 1 ...  
## $ PaymentMethod\_Electronic.check : int 1 0 1 1 0 0 1 0 0 0 ...  
## $ PaymentMethod\_Mailed.check : int 0 1 0 0 0 1 0 1 0 0 ...  
## $ PaymentMethod\_Bank.transfer.(automatic): int 0 0 0 0 0 0 0 0 0 0 ...  
## $ PaymentMethod\_Credit.card.(automatic) : int 0 0 0 0 1 0 0 0 1 1 ...

nn\_model <- neuralnet(Churn\_Yes + Churn\_No ~   
 SeniorCitizen+tenure+  
 gender\_Female+gender\_Male  
 #+Partner\_Yes+Partner\_No+PhoneService\_No+PhoneService\_Yes+MultipleLines\_No.phone.service+MultipleLines\_No+MultipleLines\_Yes+OnlineBackup\_Yes+OnlineBackup\_No+OnlineBackup\_No.internet.service+Contract\_Month-to-month+Contract\_One.year+Contract\_Two.year+PaperlessBilling\_Yes+PaperlessBilling\_No+PaymentMethod\_Electronic.check+PaymentMethod\_Mailed.check  
 #+PaymentMethod\_Bank.transfer.(automatic)+PaymentMethod\_Credit.card.(automatic)  
 , data=telco\_train, hidden=2)  
plot(nn\_model)

### Evaluate Testing Model

#prediction <- predict(svm\_model, input\_telco\_test, propability=T)  
#confusionMatrix(prediction, output\_telco\_test)