STAT489HW7

Nicholas Thompson

11/17/2021

#Question 1

Text

Description automatically generated

#Question 2

The output from this knit differs from what my output had, even when I set a seed. With that said, some numbers will be different in comment than they are in output.

set.seed(111519)  
train=sample(1:nrow(Default), 0.7\*nrow(Default))  
  
def.train=Default[train,]  
def.test=Default[-train,]  
  
  
#a  
#single hidden layer of ten units  
nnet.fit2a=nnet(default~ ., data=def.train, size=10)

## # weights: 51  
## initial value 2801.823839   
## iter 10 value 1031.947933  
## final value 1031.942679   
## converged

probs2a=predict(nnet.fit2a,def.test)  
pred2a= rep("No", 3000)  
pred2a[probs2a > 0.5] = "Yes"  
table(pred2a, def.test$default)

##   
## pred2a No Yes  
## No 2903 97

mean(pred2a!=def.test$default)

## [1] 0.03233333

#b  
log2b <- glm(default~balance + student + income, family = "binomial", data = def.train)  
  
probs=predict(log2b, def.test, type = "response")  
  
pred2b=rep("No", 3000)  
  
pred2b[probs >.5]="Yes"  
  
mean(pred2b!=def.test$default)

## [1] 0.02933333

#comparing testing errors, the logistic regression produces the smallest error (0.024) compared to the single hidden layer neural network  
  
#c  
h2o.init(nthreads = -1)

## Connection successful!  
##   
## R is connected to the H2O cluster:   
## H2O cluster uptime: 1 minutes 4 seconds   
## H2O cluster timezone: America/Chicago   
## H2O data parsing timezone: UTC   
## H2O cluster version: 3.34.0.3   
## H2O cluster version age: 1 month and 11 days   
## H2O cluster name: H2O\_started\_from\_R\_Nicholas\_rrk503   
## H2O cluster total nodes: 1   
## H2O cluster total memory: 1.94 GB   
## H2O cluster total cores: 4   
## H2O cluster allowed cores: 4   
## H2O cluster healthy: TRUE   
## H2O Connection ip: localhost   
## H2O Connection port: 54321   
## H2O Connection proxy: NA   
## H2O Internal Security: FALSE   
## H2O API Extensions: Amazon S3, Algos, AutoML, Core V3, TargetEncoder, Core V4   
## R Version: R version 4.1.1 (2021-08-10)

train=as.h2o(def.train)

## | | | 0% | |======================================================================| 100%

test=as.h2o(def.test)

## | | | 0% | |======================================================================| 100%

y="default"  
model2c=h2o.deeplearning(  
 y = y,   
 training\_frame =train,  
 distribution = "multinomial",  
 activation = "Rectifier",  
 hidden = c(10,5),  
 epochs = 10  
 )

## | | | 0% | |======================================================================| 100%

perf=h2o.performance(model2c,test)  
perf

## H2OBinomialMetrics: deeplearning  
##   
## MSE: 0.02403792  
## RMSE: 0.1550417  
## LogLoss: 0.08717146  
## Mean Per-Class Error: 0.247792  
## AUC: 0.9408859  
## AUCPR: 0.4359649  
## Gini: 0.8817718  
##   
## Confusion Matrix (vertical: actual; across: predicted) for F1-optimal threshold:  
## No Yes Error Rate  
## No 2841 62 0.021357 =62/2903  
## Yes 46 51 0.474227 =46/97  
## Totals 2887 113 0.036000 =108/3000  
##   
## Maximum Metrics: Maximum metrics at their respective thresholds  
## metric threshold value idx  
## 1 max f1 0.326411 0.485714 103  
## 2 max f2 0.085304 0.568862 216  
## 3 max f0point5 0.529306 0.513196 56  
## 4 max accuracy 0.860278 0.971000 13  
## 5 max precision 0.902774 0.900000 9  
## 6 max recall 0.000145 1.000000 396  
## 7 max specificity 0.979947 0.999656 0  
## 8 max absolute\_mcc 0.326411 0.468636 103  
## 9 max min\_per\_class\_accuracy 0.031411 0.877024 287  
## 10 max mean\_per\_class\_accuracy 0.030717 0.886277 288  
## 11 max tns 0.979947 2902.000000 0  
## 12 max fns 0.979947 97.000000 0  
## 13 max fps 0.000029 2903.000000 399  
## 14 max tps 0.000145 97.000000 396  
## 15 max tnr 0.979947 0.999656 0  
## 16 max fnr 0.979947 1.000000 0  
## 17 max fpr 0.000029 1.000000 399  
## 18 max tpr 0.000145 1.000000 396  
##   
## Gains/Lift Table: Extract with `h2o.gainsLift(<model>, <data>)` or `h2o.gainsLift(<model>, valid=<T/F>, xval=<T/F>)`

#test error for model with hidden layers c(45,30)=0.02333  
#test error for model with hidden layers c(250, 210)=0.028  
#test error for model with hidden layers c(500, 450)=0.026  
#test error hidden layers c(500, 450, 450)=0.024  
#hidden layer (50,45,45)=0.023  
#hidden layer (50,45,45,30,15,15,7)=0.96  
#hidden layer (30,20)=0.032  
#Increasing the amount of hidden layers and units in them only yields a minimum test error of .023 from others I tested. After about 5 hidden layers, the neural network began to increasingly fail no matter the units in the layers  
#from what it appears, large amounts of units in the hidden layers actually makes the testing error worse.  
#minimum testing error I achieved using this was 0.023 from a network with 2 hidden layers, one with 45 and one with 30 units  
#regardless of what I tested, adding multiple hidden layers showed better performance compared to single hidden layer.  
h2o.shutdown(prompt = FALSE)