Model Evaluation

Amin Fesharaki

11/14/2021

```
#Import Datasets
adult_test <- read.csv(file = "/Users/datascience/Desktop/ADS 502 Data Sets/Website Data Sets/adult_ch6_
adult_train <- read.csv(file = "/Users/datascience/Desktop/ADS 502 Data Sets/Website Data Sets/adult_ch6
library(C50)
library(rpart); library(rpart.plot)
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

23. Using the training dataset,create aC5.0model (Model1) to predict a customer's Income using Marital Status and Capital Gains and Losses. Obtain the predicted responses.

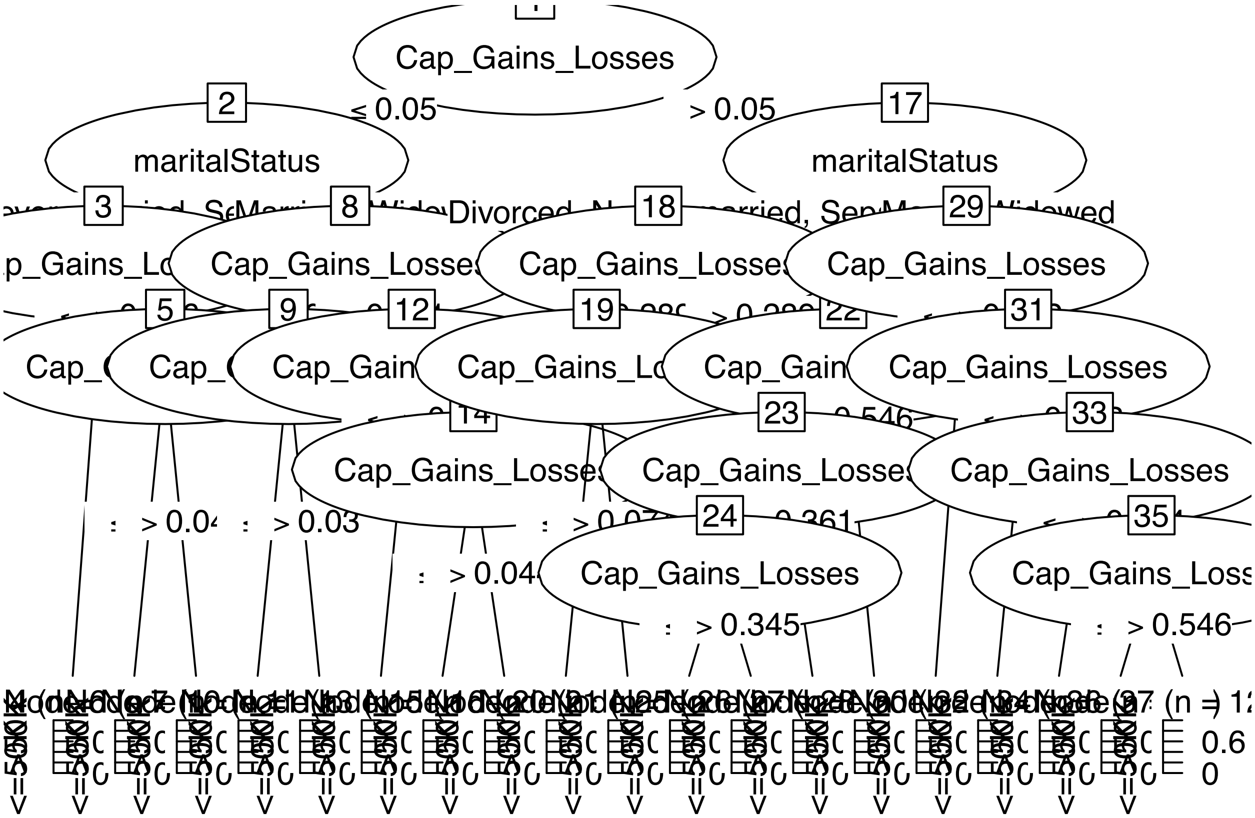
```
# Change the name to eliminate spaces
colnames(adult_train)[1] <- "maritalStatus"

# Change categorical variables into factors
adult_train$Income <- factor(adult_train$Income)
adult_train$maritalStatus <- factor(adult_train$maritalStatus)

# Predictions
X = data.frame(maritalStatus = adult_train$maritalStatus, Cap_Gains_Losses = adult_train$Cap_Gains_Losse


# C5 model
c5 <- C5.0(formula = Income ~ maritalStatus + Cap_Gains_Losses, data = adult_train)

# Graph
plot(c5)
```



```r
# Predict
Predictions = predict(object = c5, newdata = X)
```

24. Evaluate Model 1 using the test data set. Construct a contingency table to compare the actual and predicted values of Income.

```r
# Change categorical variables into factors
colnames(adult_test)[1] <- "maritalStatus"
adult_test$Income <- factor(adult_test$Income)
adult_test$maritalStatus <- factor(adult_test$maritalStatus)

# Predictions
x_test = data.frame(maritalStatus = adult_test$maritalStatus, Cap_Gains_Losses = adult_test$Cap_Gains_Lo
predx <- predict(object = c5, newdata = x_test)

t1 <- table(adult_test$Income, predx)
row.names(t1) <- c("Actual: 0", "Actual: 1")
colnames(t1) <- c("Predicted: 0", "Predicted: 1")
t1 <- addmargins (A = t1, FUN = list(Total = sum), quiet = TRUE)
t1
```

```
##            predx
##             Predicted: 0 Predicted: 1 Total
##   Actual: 0         4658           16  4674
##   Actual: 1         1057          424  1481
##   Total             5715          440  6155
```

25. For Model1,recapitulate Table7.4from the text,calculating all of the model evaluation measures shown in the table. Call this table the Model Evaluation Table. Leave space for Model 2.

26. Clearly and completely interpret each of the Model 1 evaluation measures from the Model Evaluation Table.
27. Create a cost matrix, called the 3x cost matrix, that specifies a false positive is four times as bad as a false negative

```
cost.C5 <- matrix(c(0,4,1,0), byrow =TRUE, ncol =2)
#dimnames(cost.C5) <- list (c("0", "1"), c("0", "1"))
```

28. Using the training data set, build a C5.0 model (Model 2) to predict a customer's Income using Marital Status and Capital Gains and Losses, using the 3x cost matrix.
29. Evaluate your predictions from Model 2 using the actual response values from the test data set. Add Overall Model Cost and Profit per Customer to the Model Evaluation Table. Calculate all the measures from the Model Evaluation Table.

```
t2 <- table(adult_test$Income, CostP)
row.names(t2) <- c("Actual: 0", "Actual: 1")
colnames(t2) <- c("Predicted: 0", "Predicted: 1")
t2 <- addmargins (A = t2, FUN = list(Total = sum), quiet = TRUE)
t2
```

```
##            CostP
##             Predicted: 0 Predicted: 1 Total
##   Actual: 0         3129         1545  4674
##   Actual: 1          128         1353  1481
##   Total             3257         2898  6155
```

```
##MODEL 2##
# Assigning General Form of Table to matrix values
TN2 <- t2[1,1]
FN2 <- t2[2,1]
FP2 <- t2[1,2]
TP2 <- t2[2,2]
TAN2 <- t2[1,3]
TAP2 <- t2[2,3]
TPN2 <- t2[3,1]
TPP2 <- t2[3,2]
GT2 <- t2[3,3]

# Accuracy
Acc2 <- (TN2 + TP2) / (GT2)
# Error Rate
Error2 <- 1 - Acc2
# Sensitivity / Recall
Sens2 <- TP2 / TAP2
#Specificity
Spec2 <- TN2/TAN2
# Precision
Prec2 <- TP2/TPP2
# F1
F12 <- 2*((Prec2*Sens2)/(Prec2+Sens2))
# F2
F22 <- 5*((Prec2*Sens2)/(4* Prec2 +Sens2))
cat ("MODEL 2", "\nAccuracy = ", Acc2, "\nError =", Error2, "\nSensitivity/Recall =", Sens2, "\nSpecifi
```

```
## MODEL 2
```

```
## Accuracy =  0.7281885
## Error = 0.2718115
## Sensitivity/Recall = 0.9135719
## Specificity = 0.669448
## Precision = 0.4668737
## F1 = 0.6179493
## F2 = 0.7668329
```

30. Compare the evaluation measures from Model 1 and Model 2 using the 3x cost matrix. Discuss the strengths and weaknesses of each model.

The accuracy for model 1 is higher (82% versus 72%) and the error rate for model 1 is lower (17% versus 27%). Having better accuracy and lower error rate is a strength for model 1. The strength for model 2 would be emphasizing the negative cost since it is 4 times higher than the false positive cost, therefore potentially saving money/resources/etc. However the error costs for model 2 are unequal which represents one of its weaknesses. Creating two models allows the user to select a model that minimizes/maximises the model cost per record.