Classification Model Metrics

Ben Arancibia

May 2, 2015

1) Read in the csv

```
data <- read.csv("/users/bcarancibia/CUNY_IS_621/classificationmethods/classificationoutput.csv")
```

There are three key columns: * class: the actual class of the observation * Scored.Class: the predicted class of the observation * Scored.Probability: the predicted probability of success for the observation

2) Use the table() function to get the raw confusion matrix.

```
table(data$class,data$Scored.Labels)
```

```
##
##        0    1
##   0  119    5
##   1   30   27
```

```
(119+27) / 181
```

```
## [1] 0.8066298
```

The columns represent the actual class and the rows represent the predicted class.

What the confusion matrix is, is an easy to quickly show two types of errors. The two types of errors are those errors which are false negatives or false positives. The overall error percentage is quite low, around 80% was scored correctly, but the errors for each predicted class are different. For example, the number of actual scored 0s is 149, but the predicted 0s is only 119, error rate of 20%. For 1s, it is an error rate of 16.6% (5/30).

3) Write a function that takes the dataset as a dataframe with actual and predicted classifications identified. Return the accuracy of the predictions.

$Accuracy = (TP + TN)/(TP + FP + TN + FN)$