Evaluation

Prof. Eric A. Suess

April 8, 2020

Introduction

We have primarily talked about **Classification methods**, such as, kNN, Naive Bayes, C5.0, RIPPER, CART, Logistic Regression, etc.

In the Classification setting we have used **Accuracy/Success Rate** to Evaluate the "usefulness" of an algorithm.

Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$

So we have looked at the Confussion Matrix.

acc <- mean(pred == testy)</pre>

Introduction

We have started to look at **Prediction methods**, such as, Linear Regression, Multiple Linear Regression, etc.

So we looked at "Accuracy" as the **correlation** between the test values of the **response** and the **predicted/fitted** values from the model.

When using Prediction methods a quantitative response is predicted.



But Logistic Regression is used for Classification, right?

Yes, but it uses the predicted probabilties.

In R we can classify using the **ifelse()** function to convert the probabilities into 0 and 1.

ifelse(prob < 0.5, 0, 1)

There are a number of values that can be calculated to evaluate accuracy using Classification algorithms.

Kappa - adjusts accuracy by accounting for the possibility of a correct prediction by chance alone. So should be a bit smaller than what we have discussed as Accuracy.

Sensitivity

Sensitivity = $\frac{TP}{TP+FN} \approx P(+|D)$

Specificity

Specificity =
$$\frac{TN}{TN+FP} \approx P(-|D^c)$$

Precision

$$Precision = \frac{TP}{TP + FP}$$

Recall

 $\textit{Recall} = \frac{\textit{TP}}{\textit{TP} + \textit{FN}}$

F-measure or F1 or F-score

 $\textit{Fmeasure} = \tfrac{2 \times \textit{Precision} \times \textit{Recall}}{\textit{Precision} + \textit{Recall}} = \tfrac{2 \times \textit{TP}}{2 \times \textit{TP} + \textit{FP} + \textit{FN}}$

The F-measure assumes equal weight for the Precision and Recall. This may not always be the case.

Visualizing Performance Tradeoffs - ROC

Visualizations can be very helpful for understanding how the performance of learning algorithms differ.

Useful for comparing two or more learners side-by-side.

The Receiver Operating characteristic (ROC) is commonly used. To use the ROC we need:

- 1. the class values/labels
- $2. \ the predicted probabilities of the <math display="inline">\ensuremath{\textbf{positive class}}$

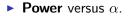
ROC - Sensitivity/Specificity plot

See page 312/332 for an example.

The ROC plots the **Sensitivity** versus **1** - **Specificity**. For the MS Statistics students this is:

True Positive Rate versus False Positive Rate

or



ROC - Sensitivity/Specificity plot

No predictive value, 45 degree line

Perfect predictive value, up and across. 100% true positives with no errors.

ROC - AUC

The **Area Under the Curve** (AUC) is commonly used to compare Classifiers.

Holdout Method

- Training
- Validation
- Testing

Repeated Holdout

k-fold cross validation

10-fold cross validation

Train on 9 of the folds and test on the last. Average the accuracy measure.

Random sample with replacement. Train on the sample and test on the remaining examples.

 $error = 0.632 \times error_{test} + 0.368 \times error_{train}$