Binary Classification Accuracy Measures

Data Mining
Prof. Dawn Woodard
School of ORIE
Cornell University

Outline

1 Announcements

2 Changing the Prediction Threshold

3 Practice Questions

- Questions?
- Step-through debugging in R is provided by the “debug” function. For info, call “help(debug)”
  - Define your function (e.g. “myFun”)
  - Call “debug(myFun)”
  - Run your function: “myFun(args)”

Threshold

- In the case of naive Bayes where Y takes two values (for instance, 0 or 1), to predict Y we calculated $Pr(Y = 1|X_1, \ldots, X_K)$
- If this was greater than 0.5, we classified as $Y = 1$. Why?
  -
- If $Pr(Y = 1|X_1, \ldots, X_K) <= 0.5$ we classified as $Y = 0$
- Note: what we do for the boundary case $Pr(Y = 1|X_1, \ldots, X_K) = 0.5$ is an arbitrary choice

Threshold

- We talked last time about why it makes more sense in some contexts to use thresholds other than 0.5
- Note: for an outcome with > 2 categories the concept of the decision threshold is more complicated. Today we focus on **OUTCOMES WITH 2 CATEGORIES**
- These categories are typically named 0 and 1. $Y = 1$ is also called a "positive," while $Y = 0$ is called a "negative." For an outcome with values "Yes" and "No", we take $Y = 1$ to mean "Yes" and $Y = 0$ to mean "No" by convention.

Accidents Data

- Applying naive Bayes to the accidents data and using a decision threshold of 0.5 on the test data, the classification % accuracy is 59.5. The classification table for the test data is:

	Prediction = 0	Prediction = 1
Actual = 0	399	116
Actual = 1	289	196

- Here we are making two types of errors
- How many times did we predict $Y = 0$ when actually $Y = 1$?
- How many times did we predict $Y = 1$ when actually $Y = 0$?

Accidents Data

	Prediction = 0	Prediction = 1
Actual = 0	399	116
Actual = 1	289	196

- For what percentage of the records with $Y = 1$ did we predict $Y = 0$?
  - This is called the False Negative Rate
- For what percentage of the records with $Y = 0$ did we predict $Y = 1$?
  - This is called the False Positive Rate

Accidents Data

	Prediction = 0	Prediction = 1
Actual = 0	399	116
Actual = 1	289	196

- For what percentage of the records with $Y = 1$ did we predict $Y = 1$?
  - This is called the True Positive Rate

- For what percentage of the records with $Y = 0$ did we predict $Y = 0$?
  - This is called the True Negative Rate

Accidents Data

- If we lower the threshold from 0.5 to 0.4, should our false positive rate increase or decrease?

  A. Increase
  B. Decrease

- If we raise the threshold from 0.5 to 0.7, should our true positive rate increase or decrease?

  A. Increase
  B. Decrease

Accidents Data

...so the threshold controls the tradeoff between the false positive and false negative rates. You can decrease one by changing the threshold, but the other one will increase.

Accidents Data

For the accidents data:

- With threshold = 0.5 we have

	Prediction = 0	Prediction = 1
Actual = 0	399	116
Actual = 1	289	196

- With threshold = 0.4 we have

	Prediction = 0	Prediction = 1
Actual = 0	210	305
Actual = 1	136	349

Accidents Data

For the accidents data:

- With threshold = 0.5 we have

	Prediction = 0	Prediction = 1	Total
Actual = 0	399	116	515
Actual = 1	289	196	485

- With threshold = 0.4 we have

	Prediction = 0	Prediction = 1	Total
Actual = 0	210	305	515
Actual = 1	136	349	485

Why doesn’t the threshold affect the row sums?

Accidents Data

- With threshold = 0 what table do we get? (with neighbor)

	Prediction = 0	Prediction = 1	Total
Actual = 0			515
Actual = 1			485

- What are our false positive and true positive rates?

Accidents Data

- With threshold = 1 what table do we get?

	Prediction = 0	Prediction = 1	Total
Actual = 0			515
Actual = 1			485

- What are our false positive and true positive rates?

Accidents Data

- Varying the threshold we get the following false positive and true positive rates:

Threshold	False Positive Rate	True Positive Rate
0	1.00	1.00
0.4	0.59	0.72
0.5	0.23	0.40
0.6	0.07	0.17
1	0.00	0.00

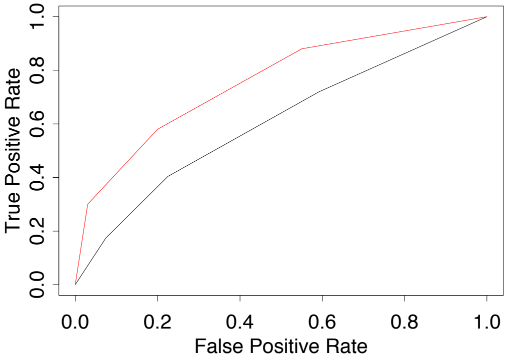
- Sketch a plot with the false positive rate on the x-axis and the true positive rate on the y-axis
- The plot should not have threshold on either axis

ROC Curves

This type of plot is called a Receiver Operating Characteristic (ROC) curve:

ROC Curves

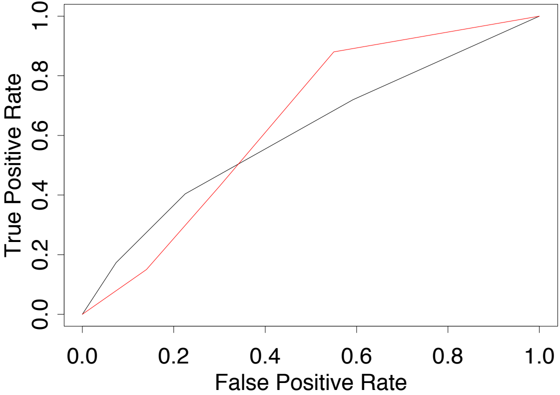
What if one classifier has the ROC curve in black and another has the ROC curve in red:



Is one of these classifiers uniformly better than the other? Why?

ROC Curves

What if the two classifiers have these curves; is one uniformly better?



ROC Curves

- Say we have a classifier, called the "coin-flipping" classifier, that flips a coin to decide whether to predict 0 or 1

- What are the false positive and true positive rates for this classifier?

- Say the coin can be biased, so that it has probability p of being heads

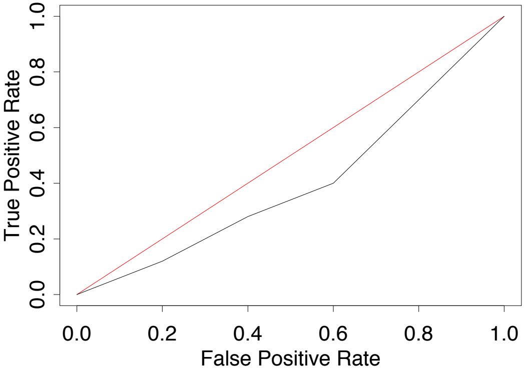
- What are the false positive and true positive rates for this classifier?

ROC Curves

What does the ROC curve look like for this classifier?

ROC Curves

What does it mean if we have a classifier with the ROC curve in black?



ROC Curves

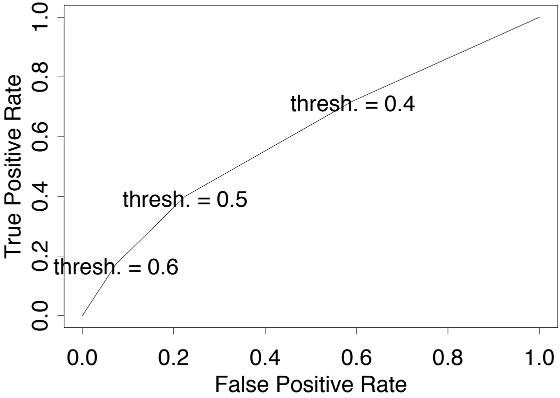
- For any classification method for binary outcome that gives us a prediction of $\Pr(Y = 1|X_1, \ldots, X_K)$ (i.e. any statistical classification method), we can alter the threshold in this way to obtain a range of true positive and false positive rates.

- We cannot do this with any classification method that only gives us a prediction of $Y = 1$ or $Y = 0$, i.e. many heuristic classification methods.

- This is a big benefit of statistical methods for classification of binary outcomes.

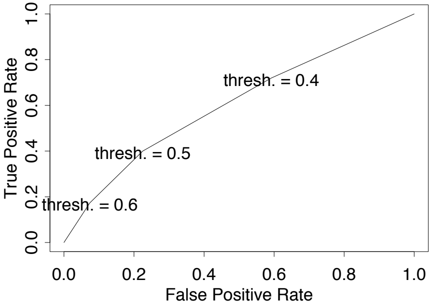
ROC Curves

Say we have a classifier with the following ROC curve on the test data:



ROC Curves

Say false positive errors are much more costly than false negative errors for this problem.



What threshold might you want to use for prediction?

A. 1.0

B. 0.6

C. 0.5

D. 0.4

ROC Curves

What are the FPR and FNR at that threshold?

ROC Curves

- Give an example of a problem where false positive errors are much less costly than false negative errors.

- Give an example of a problem where false positive errors are much more costly than false negative errors.

Error Rates

Distinction between false negative probability and false negative rate, etc.:

$$FNP = P(\text{Predict } Y = 1 \mid \text{Actually } Y = 0)$$

$$FPR = \#\text{obs w/ prediction 1 and Y actually} = 0 / \text{obs}$$

$$\text{w/ actually } Y = 0$$

Error Rates

- For a test data set having 10 records with $Y = 1$ and 15 records with $Y = 0$, what is the chance of correctly classifying all of them, as a function of the true positive probability t, the false positive probability f, the true negative probability g, and the false negative probability h? Assume independence across observations.

Error Rates

- Say the frequency of $Y = 1$ in the general population is 67%. We select a random person and predict the value of Y for that person, based on their covariates. What is the probability that person has $Y = 1$ and we predict $Y = 0$, as a function of our true positive probability t, false positive probability f, true negative probability g, and false negative probability h?