STAT 380: Classification Technique: Evaluating Performance of a Classification Technique

UAEU

Outlook of the unit

- **Prediction and Classification Approaches**
  - Classification Techniques
    - Logistic regression
    - Discriminant analysis
  - **Evaluating Performance of a Classification Technique**
  - Tree-based methods: Decision trees
    - Classification trees
    - Regression trees

Evaluating Performance of a Classification Technique

- A natural criterion for judging the performance of a classifier is the probability of making a **misclassification** error.

- Misclassification means that the record belongs to one class but the model classifies it as a member of a different class..

- Is there a minimal probability of misclassification that we should require of a classifier?

- A classifier that makes no errors would be perfect - unrealistic.

Propensities and cut-off for classification

First step in most classification algorithms is to estimate the probability π (propensity) that a unit belongs to each of the classes.

If overall classification accuracy is of interest, the unit can be assigned to the class with the highest probability.

- It is possible, however, to use a cutoff that is either higher or lower than 0.5. Two examples:
  - unequal misclassification costs
  - unequal importance of classes.

The default **cutoff** value in two-class classifiers is 0.5.

In many records, a single class is of special interest, so we will focus on that particular class. It may make sense in such cases to consolidate classes so that you end up with two: the class of interest and all other classes.

Confusion Table, Accuracy and and Error Rate

Classification Matrix/Confusion Table

	True Value of Response (Y) from Data	
Predicted Response(Y) From the Model	**Y=1**	**Y=0**
$\hat{Y}$=1	**True Positive (n_{11})**	**False Positive (n_{12})**
$\hat{Y}$=0	**False Negative (n_{21})**	**True Negative (n_{22})**

- Classification matrix summarizes the correct and incorrect classifications that a classifier produced.
- Rows and columns of the confusion matrix correspond to the predicted and true (actual) classes.
- Example:

		Actual class	
		Y= 1	Y= 0
Predicted class	Y=1	2600	100
	Y= 0	100	200

- Diagonal cells give the number of correct classifications.
- Off-diagonal cells give counts of misclassification.
- Classification matrix gives estimates of the true classification and misclassification rates.

Predictive measures derived from a confusion matrix:

TN= TRUE Negative

TP= TRUE Positive

FN= FALSE Negative

FP= FALSE Positive

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

$$\textbf{Error Rate:=}1 - Accuracy$$

Specificity and Sensitivity

Specificity and Sensitivity

- Sensitivity and specificity are statistical measures of the performance of a binary classification test.
- Sensitivity(true positive rate) measures the proportion of actual positives which are correctly identified.
- Specificity (true negative rate) measures the proportion of negatives that are correctly identified.

Specificity and Sensitivity

TN= TRUE Negative

FN= FALSE Negative

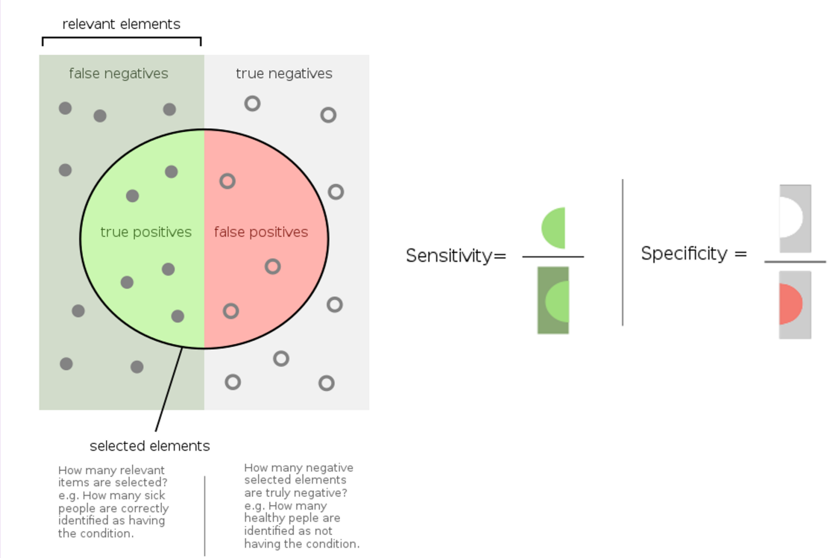
TP= TRUE Positive

FP= FALSE Positive

$$\textbf{Sensitivity:=}\frac{TP}{TP+FN}$$

$$\textbf{Specificity:=}\frac{TN}{TN+FP}$$

Sensitivity and Specificity



Specificity and Sensitivity: An Example from Health Science

- **Sensitivity:** Ability of a Testing Procedure to identify those who have disease. The sensitivity of a test is the probability of a positive test result given the presence of the disease,

$$P(\text{Test Result} = \textit{Positive} \mid \text{Diseased})$$

- **Specificity:** Ability of a test to exclude those who don't have the disease The specificity of a test is the probability of a negative test result given the absence of the disease,

$$P(\text{Test Result} = \textit{Negative} \mid \text{Not Diseased})$$

Example

*Scenario*1 : Imagine we have a sample of 100 points, 50 from ‘Class 0’ (For example: healthy) and the others ‘Class 1 (For Example: Not Healthy). Let us asume ans extreme case where a classification technique (Medical Testing) can identify all the ‘Class 0’ correctly (positive for all Diseased) and also correctly idensifies all ’Class1’ objects (negative for all the healthy).

Sensitivity= Specificity=100% Prevalence $= \frac{50}{100} = 50\%$

Example

Scenario2 : Imagine we have a sample of 150 points, 70 from 'Class 0' (For example: healthy) and the others from 'Class 1 (For Example: Not Healthy). Let us asume that a classification technique (Medical Testing) can correctly identify 65 out of the 70 belongs to the 'Class 0' (positive for all Diseased) while it can correctly detect 77 of objects that belongs to the 'Class1'(Not Healthy).

- Construct the classification table and calculate the Sensitivity, Specificity, Accuracy and misclassification rate of the testing procedure based on the classification table that you obtained.

Receiver Operating Characteristic (ROC)

Example from Logistic Regression

A logistic regression is fitted based on a training data set of a ranodm sample of 50 individuals aged more than 40. The objective of the study is to model whether a randomly selected individual from the population of study is Diabetes based on their Age and Gender. A logistic regression is fitted with the response, Y indicative of the presence of Diabetes. The covariates considered are Age and Gender. The predicted probabilities for a test sample of size 10 along with their corresponding covariates are provided below.

Example from Logistic Regression

Age	Gender	True Y	Predicted Probability
49	1	0	0.572
40	0	0	0.447
48	1	0	0.559
48	1	1	0.559
49	1	1	0.572
53	0	1	0.793
48	0	1	0.559
45	1	1	0.518
51	0	0	0.597
51	1	1	0.597

True Y	Predicted Probability	y_hat_5	y_hat_55	y_hat_6
		whether Prob>.5	whether Prob>.55	whether Prob>.6
0	0.572	1	1	0
0	0.447	0	0	0
0	0.559	1	1	0
1	0.559	1	1	0
1	0.572	1	1	0
1	0.793	1	1	1
1	0.559	1	1	0
1	0.518	1	0	0
0	0.597	1	1	0
1	0.597	1	1	0

Prob Cutoff 0.5			
		True Y	
		Y=1	Y=0
Predicted Y	Y_hat=1	1	0
	Y_hat=0	3	6

Prob Cutoff 0.55			
		True Y	
		Y=1	Y=0
Predicted Y	Y_hat=1	1	1
	Y_hat=0	3	5

Prob Cutoff 0.57			
		True Y	
		Y=1	Y=0
Predicted Y	Y_hat=1	2	3
	Y_hat=0	2	3

Receiver Operating Characteristic (ROC)

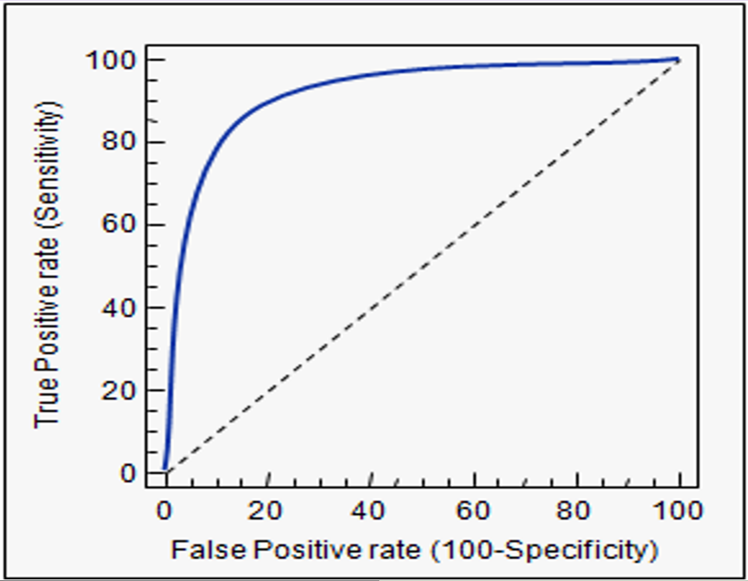
The **R**eceiver **O**perating **C**haracteristic (ROC) curve is a way to visualize interrelationship between sensitivity and specificity

- In a ROC curve the true positive rate (Sensitivity) is plotted as function of the false positive rate (100-Specificity) for different cut-off points. Each point on the ROC curve represents a (sensitivity, specificity) pair corresponding to a specific decision threshold.

- A test with perfect discrimination (no overlap in the two distributions) has a ROC curve that passes through the upper left corner (100% sensitivity, 100% specificity).

- Therefore, the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test.

ROC



True Positive rate (Sensitivity)
0
20
40
60
80
100
False Positive rate (100-Specificity)

- If $0.7 \leq AUC < 0.8$, this is considered acceptable discrimination.
- If $0.8 \leq AUC < 0.9$, this is considered excellent discrimination.
- If $0.9 \leq AUC$, this is considered outstanding discrimination.

- AUC (area under curve) indicates model goodness, 1 being a perfect model and below 0.5 (yellow line) a useless model (worse then a coin flip).

Youden Index

If sensitivity and specificity are equally important, the Youden index(YJ) will indicate the performance at a given cutoff(c).

$$YJ(c) = \text{Sensitivity}(c) + \text{Specificity}(c) - 1$$

$$c_{\text{optimal}} = \operatorname*{Arg\,max}_{c} YJ(c)$$