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.

- 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	
		0	1
Predicted class	0	2600	100
	1	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.

- 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	
		0	1
Predicted class	0	2600	100
	1	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.

Accuracy measures - the classification matrix

- We summarize the classification for the validation data as follows.
- **Classification matrix**:

		Actual class	
		C_1	C_2
Predicted class	C_1	$n_{1,1}$	$n_{2,1}$
	C_2	$n_{1,2}$	$n_{2,2}$

- Estimated **misclassification rate**:

$$err = \frac{n_{1,2} + n_{2,1}}{n_v},$$

  where n_v is the total number of units in the validation data.
- **Estimated accuracy**:

$$accuarcy = 1 - err = \frac{n_{1,1} + n_{1,1}}{n_v}.$$

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

Evaluation Metrics

Predicted Response(Y) From the Model	True Value of Response (Y) from Data: 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})

Evaluation Metrics (1)

General form of a 2×2 confusion matrix

		Actual value		
		C_1	C_2	Row total
Predicted value	C_1'	**TruePositive**	**FalsePositive**	P′
	C_2'	**FalseNegative**	**TrueNegative**	N′
	Column total	P	N	

Note: C_1 is assumed to correspond to a positive class

Predictive measures derived from a confusion matrix:

Accuracy: $\dfrac{TP + TN}{TP + TN + FP + FN}$

Error rate : $1 - Accuracy$

Sensitivity : $\dfrac{TP}{TP + FN}$

Specificity : $\dfrac{TN}{TN + FP}$

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.

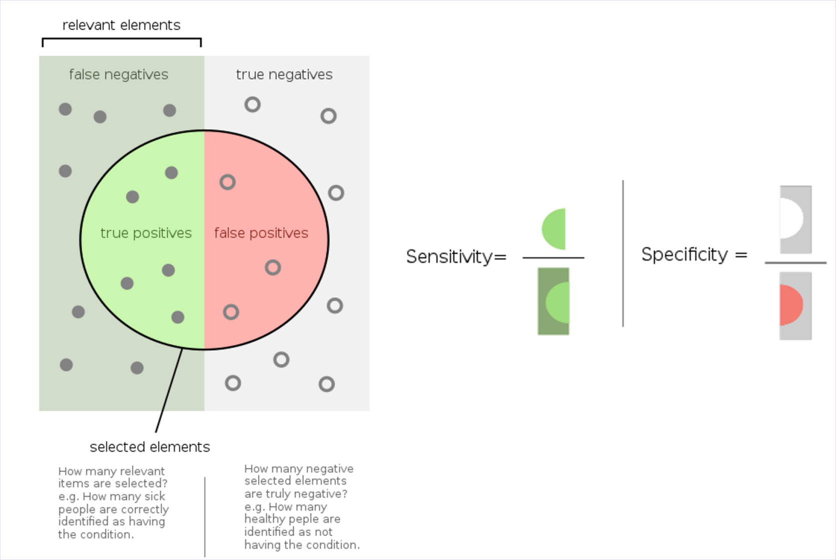
- Specificity (true negative rate) measures the proportion of negatives that are correctly identified.

Specificity and Sensitivity

Sensitivity: Ability of a test 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, $P(Test = + \mid Diseased) = \frac{a}{a+c}$

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, $P(Test = - \mid NotDiseased) = \frac{d}{b+d}$

Sensitivity and Specificity



Receiver Operating Characteristic (ROC)

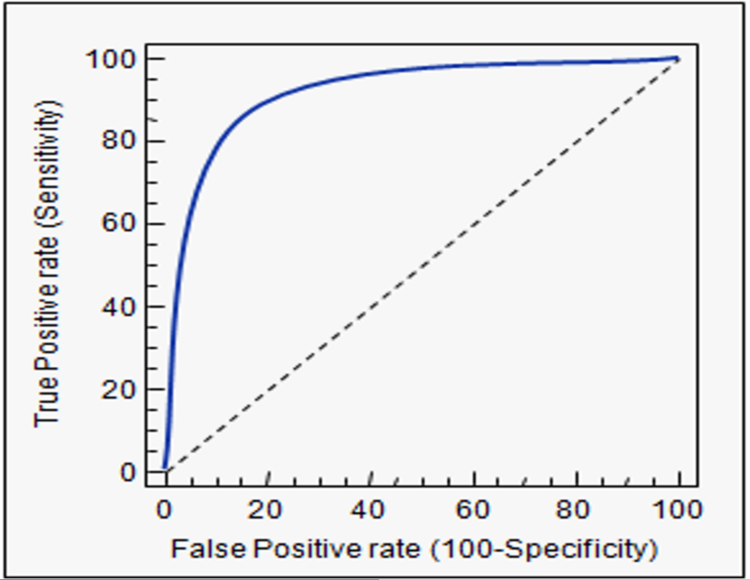
The **R**eceiver **O**perating **C**haracteristic (ROC) curve is a way to visualize interrelationship between sensitivity and specificity

- 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)
100
80
60
40
20
0
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