Classification Methods

model categorial outcome

classification rule = mathematical function that predicts the outcome of a new sample unit when the values of the covariates are known

split the data into a training and a test set

use the training dataset to build the classification rule

use the test dataset to evaluate how the classification rule labels new cases with known labels using ROC analysis

Evaluating Classification Rule

Misclassification Error Table		Predicted Class		Total True/False
		0	**1**	
True Class	**0**	n_{00} true negative	n_{01} false positive	n_{0+}
	1	n_{10} false negative	n_{11} true positive	n_{1+}
Predicted True/False		n_{+0}	n_{+1}	n

accuracy = rate of correctly classified labels in the test set $\frac{n_{00}+n_{11}}{n}$

false positive = predicted value is 1 and true value is 0 $\frac{n_{01}}{n_{0+}}$

false negative = predicted value is 0 and true value is 1 $\frac{n_{11}}{n_{1+}}$

sensitivity = true positive proportion $\frac{n_{11}}{n_{1+}} = 1 - \frac{n_{01}}{n_{0+}}$

specificity = true negative proportion $\frac{n_{00}}{n_{0+}} = 1 - \frac{n_{00}}{n_{0+}}$

positive predictive value (PPV) = proportion of true positives in predicted positives $\frac{n_{11}}{n_{+1}}$

negative predictive value (NPV) = proportion of true negatives in predicted negatives $\frac{n_{00}}{n_{+0}}$

Misclassification Error Table		Predicted Class 0	Predicted Class 1	Total True/False
True Class	0	0 true negative	n_{01} false positive	n_{0+}
True Class	1	0 false negative	n_{11} true positive	n_{1+}
Predicted True/False		n_{+0}	n_{+1}	n

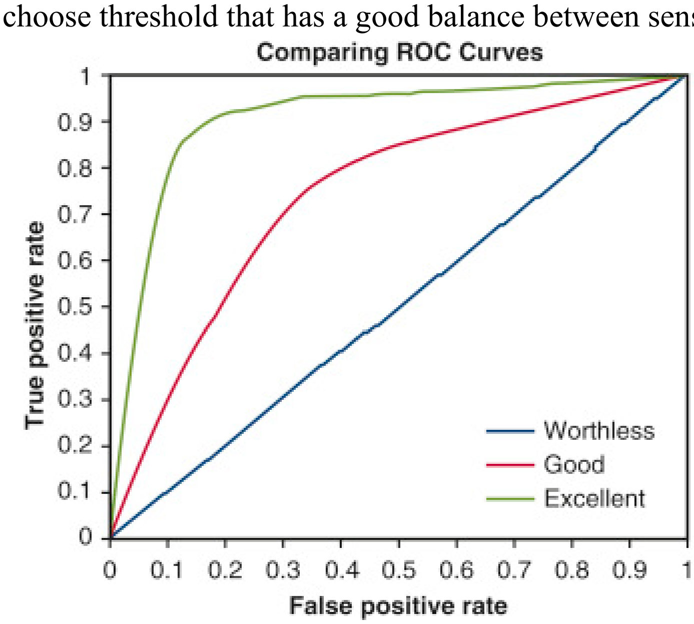
test with maximum sensitivity will always predict 1 so specificity is 0

Misclassification Error Table		Predicted Class 0	Predicted Class 1	Total True/False
True Class	0	n_{00} true negative	0 false positive	n_{0+}
True Class	1	n_{10} false negative	0 true positive	n_{1+}
Predicted True/False		n_{+0}	n_{+1}	n

test with maximum specificity will always predict 0 so sensitivity is 0
cannot maximize sensitivity and specificity simultaneously in one test

ROC Analysis

ROC curve simultaneously displays the two types of errors for all possible thresholds
area under ROC curve (AUC) shows overall performance of a classifier summarized over all possible thresholds
ideal ROC curve will hug the top left corner and have large AUC
a classifier that performs no better than random chance will have AUC of 0.5
compares different classifiers because can consider all possible thresholds
use ROC analysis to choose threshold that has a good balance between sensitivity and specificity



Methods to Generate Classification Rules

Logistic Regression

$$logit(p) = log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_k X_k$$

$$p = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_k X_k}}$$

estimate the probability of a binary outcome as a function of covariates
choose a threshold on the predicted value to decide how to classify
each threshold can be associated with a misclassification error table

Classification Trees

provides classification rule using a partition of the space of covariates
in each terminal node, the "default" classification rule is to assign the most common label, in the dataset
ROC analysis of the best classification tree identified by cross-validation

K-Nearest Neighbor

identify k closest observations in training set
assign x to the most common class of the k nearest neighbors
non-parametric approach
no assumptions are made about the shape
doesn't identify which predictors are important
standardize observations because the method is distance-based
choose k by ROC analysis

Discriminant Analysis

model the distribution of the predictors separately in each of the response classes
use Bayes' theorem to flip them around into estimates for $P(Y = k|X = x)$
linear discriminant analysis assumes equal variance between groups
quadratic discriminant analysis assumes unequal variances between groups, so each class has its own covariance matrix