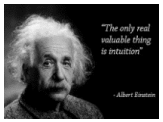
Classification Trees

Analytics Modeling Options

	Modeling Method		
	Structured		Visual, Text, Unstructured, etc.
Descriptive	Cluster analysis, correlation, market basket analysis, sample statistics, ANOVA		Bubble charts, network diagrams, natural language processing, clustering dendograms, etc.
Predictive	Association	Decision Tree	Charts
Quantitative Value	Regression	Regression Trees	Regression plots, scatter plots, Tableau diagrams, trend charts, etc.
Classification	Logistic Regression; Other Categorical Regression Models	Classification Trees	Tree maps, interactive diagrams,
Prescriptive	Operations research, decision modeling, optimization, linear programming		Simulations, etc.



Classification Trees: Intuition

- Classification trees are very **similar** in concept to **regression trees**
- Except that the **outcome** is **qualitative** consisting of **K** possible **classes** and the classification is based on the **highest probability** of classifying observation in the **correct class**
- So, it is an **alternative** method to Logistic Regression, LDA and QDA
- **Instead** of selecting nodes and pruning based on minimizing the **ESS**, classification trees **minimize classification error**.

Classification Tree Errors

- If the **data** is pretty **differentiated** across **classes**, the training **error** will be relatively **low**. But if the data is somewhat **commingled** across classes the tree model will **misclassify** some training observations
- We can inspect each data point in each region m and see if the tree model classified it correctly to the most predominant class k (with the highest proportion of observations) in that region:
- The **Error Rate** is the fraction of observations in region m that do not belong to the **most common** class k in that region:

$$Error\ Rate = E = 1 - Max\ (p_{mk})$$

R Tips

`tree(){tree}` → Function in the `{tree}` package to fit regression and classification trees

`classtree.fit=tree(y~x1+x2+etc.,data=dataName)` → Fits a classification tree when y is binomial

`plot(classtree.fit)` → Plots the tree.fit tree object

`text(classtree.fit,pretty=0)` → To add labels and text to the tree

```
cv.classtree=cv.tree(classtree.fit,
                FUN=prune.misclass)
```
→ computes the CV for various pruned trees; `FUN=prune.misclass` uses the misclassification error for cross validation and pruning; otherwise, the default is deviance

```
prune.classtree.fit=
    prune.misclass(classtree.fit,best=9)
```
→ Prunes the `classtree.fit` object to `9` terminal nodes

KOGOD SCHOOL
of
BUSINESS