Evaluating Classification & Predictive Performance

Objectives

- Understand error rate
- How to read the confusion matrix
- Understand classifier performance metrics such as precision, recall, sensitivity, and specificity
- How to read an ROC curve
- How to read/interpret a Lift /Decile chart

Why Evaluate?

- Multiple methods are available to classify or predict
- For each method, multiple choices are available for settings
- To choose best model, need to assess each model's performance

Error Rate

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	0
1	201	85
0	25	2689

Overall error rate

= (25+85)/3000 = 3.67%

Accuracy

= 1 – err = (201+2689)/3000 = 96.33%

Alternate Performance Measures

- If 1 is the important class, 0 is the other class

- **Sensitivity** = % of class 1 correctly classified
- **Specificity** = % of class 0 correctly classified
- **False positive rate** = % of predicted 1 that were actually 0
- **False negative rate** = % of predicted 0 that were actually 1

Precision and Recall

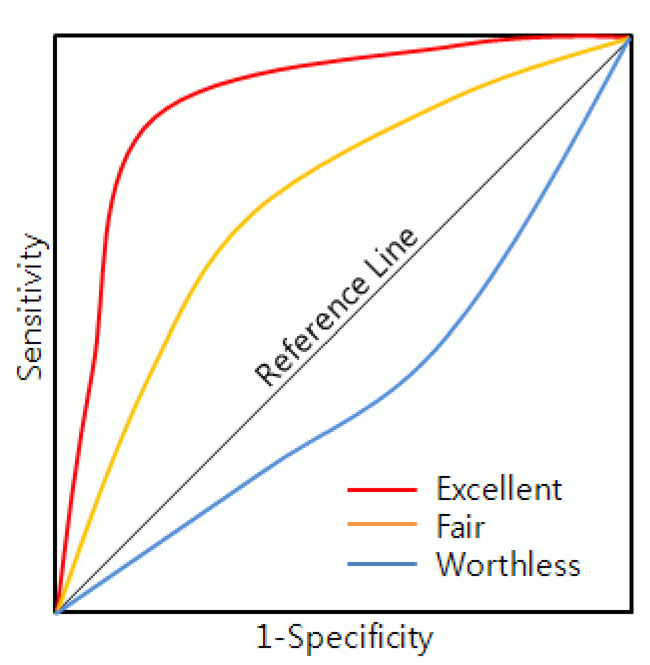
- Confusion matrix

Actual	Predicted: Positive, 1	Predicted: Negative, 0	
Positive, 1	True Positive	False Negative	**Recall** *Sensitivity*
Negative, 0	False Positive	True Negative	Recall_0 *Specificity*
	Precision	Precision_0	

- Some common metrics:
  - **Accuracy** = (True positive + true negative) / (total)
  - **Precision** = (True positive) / (predicted positive)
  - **Recall** = (True positive) / (actual positive) – also known as sensitivity
- Precision and Recall metrics can be calculated for any class
  - **Precision_0** = (True negative) / (Predicted negative)
  - **Recall_0** = (True negative) / (actual negative) -- also known as specificity

ROC Curve

- ROC (receiver operating characteristics) curve
  - Plot the pair {Sensitivity, 1-Specificity} as the cutoff value increases from 0 to 1
- The diagonal line reflects the performance of the naïve rule using varying cutoff values



The larger the area under the ROC curve, the better is the classifier.

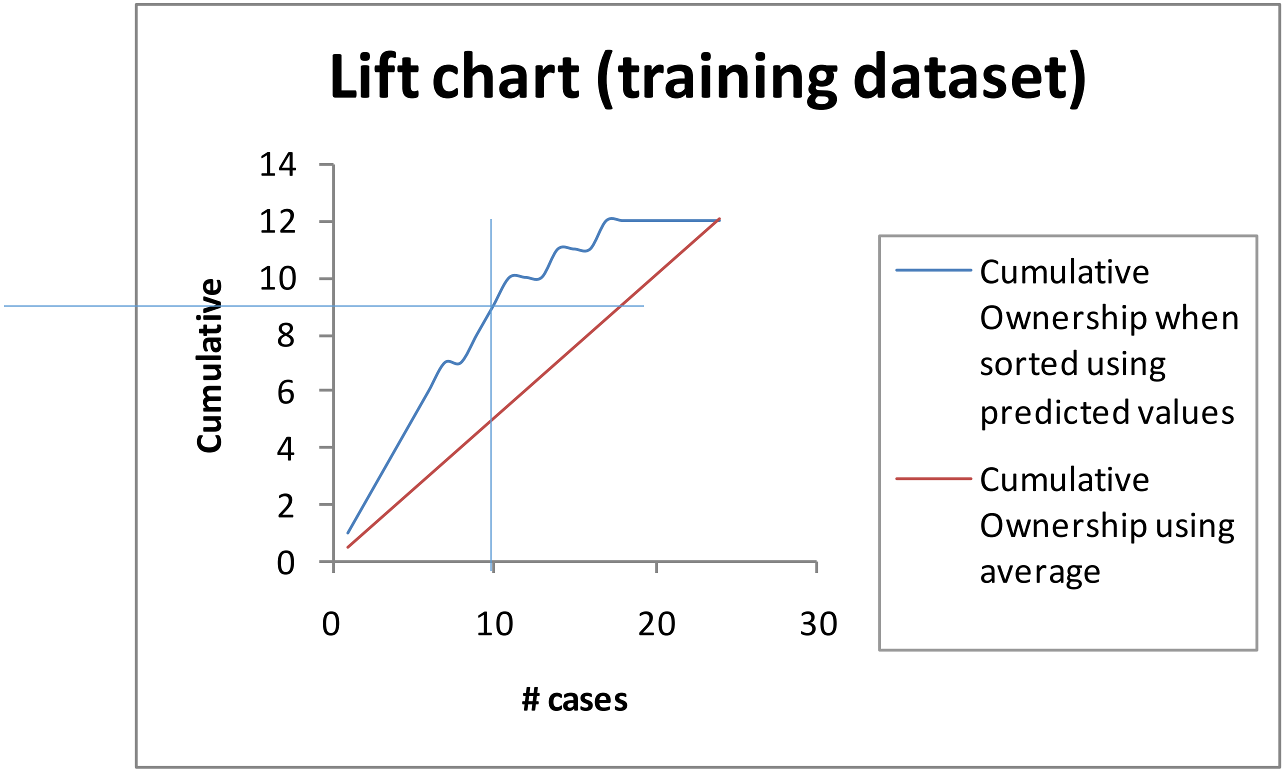
Lift and Decile Charts: Goal

- Useful when classifier gives a probability (rather just yes/no)
  - Allow us to order the cases by their probability of class membership.
- Goal is to select a relative small number of cases and get a relative large portion of the members in the important class.
  - **Skim the cream**
- Helps evaluate, e.g.,
  - How many tax records to examine
  - How many loans to grant
  - How many customers to mail offer to

Lift and Decile Charts – Cont.

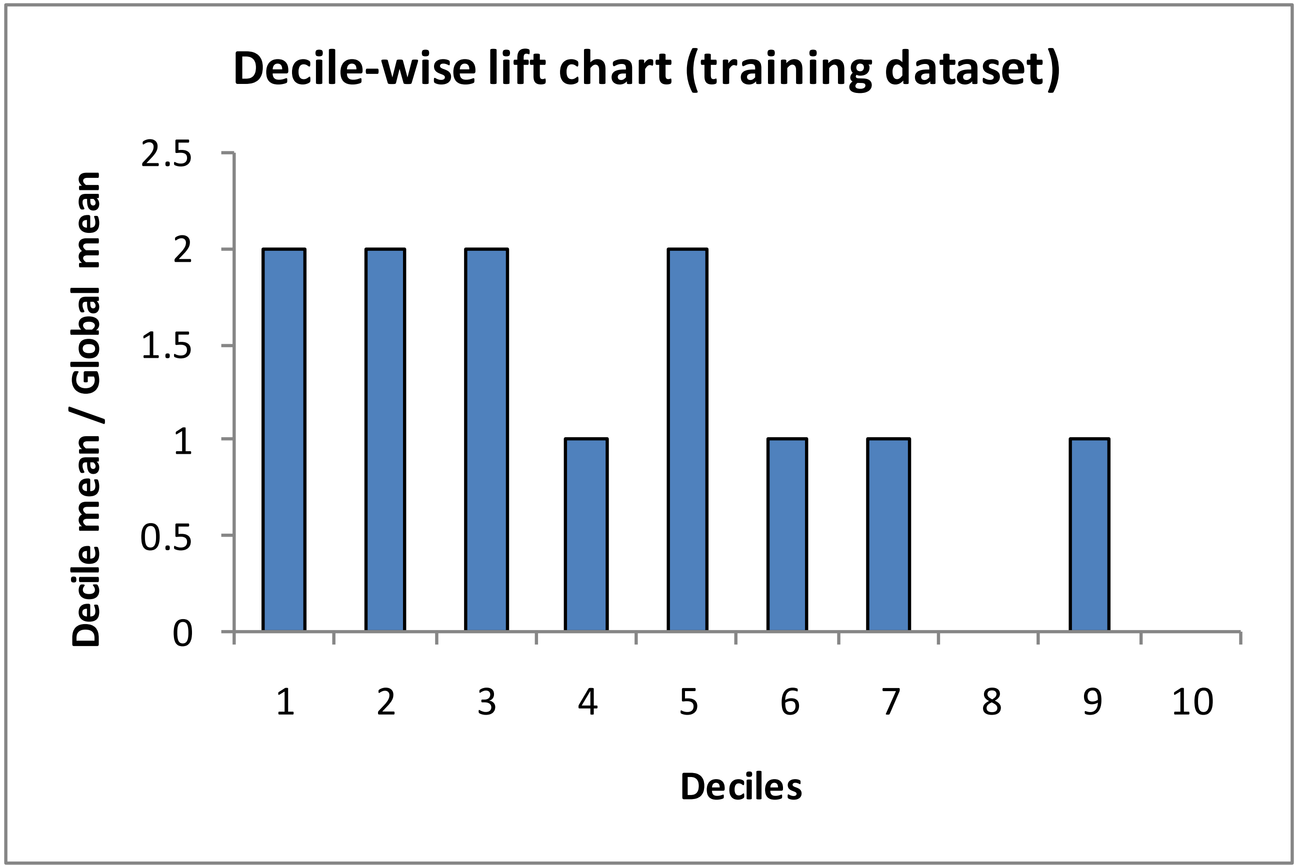
- Compare performance of DM model to "no model, pick randomly"
- Measures ability of DM model to identify the important class, relative to the naïve rule
  - In lift chart: compare step function to straight line
  - In decile chart: compare to ratio of 1

Lift Chart – Cumulative Performance



After examining (e.g.,) 10 cases (x-axis), 9 owners (y-axis) have been correctly identified

Decile Chart



In "most probable" (top) decile, model is twice as likely to identify the important class (compared to avg. prevalence)

Measuring Predictive error

- Not the same as "goodness-of-fit"

- We want to know how well the model predicts **new data**, not how well it fits the data it was trained with

- Key component of most measures is difference between actual *y* and predicted *y* ("error")

Some measures of error

MAE or MAD: Mean absolute error (deviation)
Gives an idea of the magnitude of errors

Average error
Gives an idea of systematic over- or under-prediction

MAPE: Mean absolute percentage error

RMSE (root-mean-squared-error): Square the errors, find their average, take the square root

Total SSE: Total sum of squared errors

Lift Chart for Predictive Error

Similar to lift chart for classification, except…

Y axis is cumulative value of numeric target variable (e.g., revenue), instead of cumulative count of "responses"

Lift chart example – spending