Forecast Intervals for the Area Under the ROC Curve with a Time-varying Population

Lealand Morin

Department of Economics
College of Business Administration
University of Central Florida

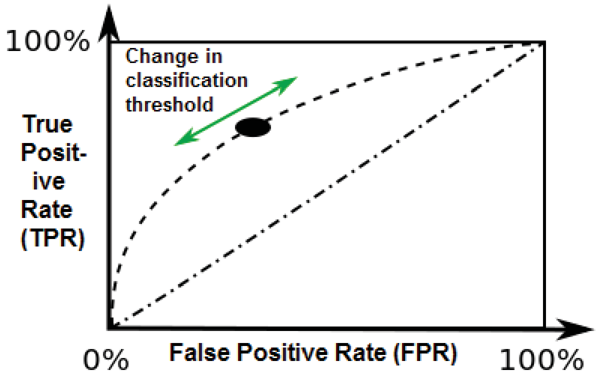
October 19, 2018

Predicting Performance of Classification Models

- **What:** Method for calculating a forecast interval for the Area Under the ROC curve (AUROC)
  - Area under the Receiver Operating Characteristic (ROC) curve: quality of a signal for predicting an outcome
  - Predictive modeling: performance of a classification model
- **Why:** Characterize the likely range of model performance while a model is used for prediction
  - In practice, businesses will use model until:
    - Performance (AUROC) degrades
    - Population changes
- **How:** Measure the variation in AUROC in terms of the variation in the underlying distribution of predictive variables
  - Not only from sampling variation from a fixed distribution

Measuring Predictive Value of Classification Models

Receiver Operating Characteristic Curve:
True Positive Rate vs. False Positive Rate



Measuring Predictive Value of Classification Models

Definition of AUROC

- Direct definition: Calculation of area by integration
  - $\int_{-\infty}^{\infty} TPR(t)[-FPR'(t)]dt$
- Direct definition: Pairwise comparison of correct ordering of predictions for all pairs of predictions
  - $\hat{A} = \hat{\Pr}\{y > x\} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{\{y_j > x_i\}}$
- In words: If you were to pick a pair of predictions, drawn randomly from predictions corresponding to pairs of the positive (y) and negative (x) outcomes, the AUROC is the probability that these predictions are correctly ordered.

An Important Correspondence for Classification Models

Predicting Outcomes vs. Measuring Difference

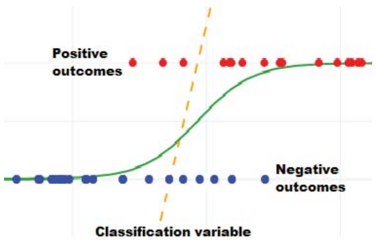


Figure: Predictive value of classification variables

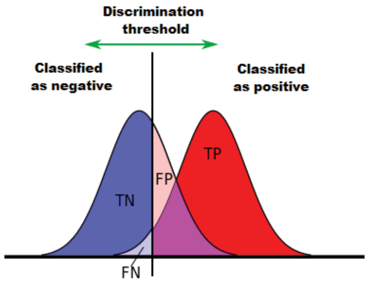


Figure: Difference in the distributions of variables

Predicting Performance of Classification Models

Calculation of forecast intervals:

1. Build model from entire sample and measure AUROC
2. Measure distance between distributions
   - by dividing sample into a series of subsamples
   - by specifying a model for the evolution of the distributions
3. Calculate extreme AUROC values that correspond to movements a specified distance away from the full sample

Optimization Problem

Find the highest value of the AUROC, $A^{(U)}$, a specified distance from observed distribution
$\max_{\mathbf{u},\mathbf{v}} \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} u_i v_j I_{\{y_j > x_i\}}$

- subject to $D(\mathbf{u} \otimes \mathbf{v}, \mathbf{f} \otimes \mathbf{g}) = \bar{D}$,
- unit mass constraints $\sum_{i=1}^{m} u_i = 1, \quad \sum_{j=1}^{n} v_j = 1,$
- nonnegativity constraints $\{u_i \geq 0\}_{i=1}^{m}, \quad \{v_j \geq 0\}_{j=1}^{n}$
- where $\mathbf{f}$ and $\mathbf{g}$ are the observed distributions of classification variables for positive and negative cases, respectively, while $\mathbf{u}$ and $\mathbf{v}$ are the weights with distance $\bar{D}$

Optimum

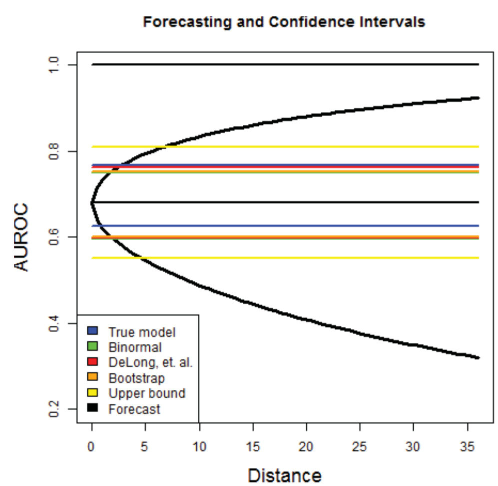
Fixed points from first order conditions

- $\frac{dD(\mathbf{u},\mathbf{f})}{du_i} = \lambda \sum_{j=1}^{n} v_j I_{\{y_j > x_i\}} + \gamma_x + \delta_{x,i}, i = 1, \ldots, m$
- $\frac{dD(\mathbf{v},\mathbf{g})}{dv_j} = \lambda \sum_{i=1}^{m} u_i I_{\{y_j > x_i\}} + \gamma_y + \delta_{y,i}, j = 1, \ldots, n$
- Solved with the recurrence relations (for a particular distance function)
  - $u_i^{(t+1)} = k_x f_i^{1+u_i^{(t)}} \exp\left\{\lambda \sum_{j=1}^{n} v_j^{(t)} I_{\{y_j > x_i\}}\right\}$
  - $v_j^{(t+1)} = k_y g_j^{1+v_j^{(t)}} \exp\left\{\lambda \sum_{i=1}^{m} u_i^{(t)} I_{\{y_j > x_i\}}\right\}$
- k_x and k_y are normalizing constants and Lagrange multiplier λ is the step size.

Competing Procedures

- Parametric models:
  - Binormal model: $[\Phi(\tilde{z}_{\alpha/2}), \Phi(\tilde{z}_{1-\alpha/2})]$
  - Biexponential model: $[\hat{A} \pm z_{1-\alpha/2}\hat{\sigma}_A]$, with
    $\sigma_A^2 = \frac{1}{mn}\{A(1-A) + (n-1)(P_{yyx} - A^2) + (m-1)(P_{yxx} - A^2)\}$,
    $P_{yyx} = A/(2-A)$, $P_{yxx} = 2A^2/(1+A)$
- Empirical distribution (DeLong et. al.):
  $P_{yyx} = \frac{1}{mnn}\sum_i \sum_j \sum_k I_{\{y_j > x_i \cap y_k > x_i\}}$
  $P_{yxx} = \frac{1}{mmn}\sum_i \sum_j \sum_k I_{\{y_j > x_i \cap y_j > x_k\}}$
- Bootstrap: $[\hat{A}^*_{\alpha/2}, \hat{A}^*_{1-\alpha/2}]$
- Upper bound of variance: $\sigma^2_{max} = \frac{A(1-A)}{\min\{m,n\}} \left(\leq \frac{1}{4\min\{m,n\}} \right)$
- Fixed error rate: Interval depends on a specified error rate.

Prediction Intervals Expanding with Distance



Predicting Performance of Classification Models

Structure of Simulation

- Regime-switching model
  - 2 states, high- and low-AUROC regimes, equally probable
  - past regimes known, future unknown
- Measure AUROC from both regimes
- Measure distance between distributions in regimes and full sample
- Calculate extreme AUROC values that correspond to movements away from full sample, using distances between distributions

Simulation Results

Coverage Rates
True values: $A_L = 0.75, A_H = 0.80$

Method	Coverage Rate	Correct Forecast Rate
Bi-normal	0.6805	0.5912
DeLong et. al.	0.6845	0.5979
Bootstrap	0.6730	0.5889
Upper Bound	0.9665	0.8654
Forecast	0.9940	0.9451

Prediction Intervals $[A_L, A_U]$

Solving for extreme values of A for a particular distance $\hat{D}$ (measured from sample)

- Record estimate of AUROC ($\rightarrow \hat{A}$)
- Solve distance minimization problem for a particular candidate A_0 ($\rightarrow \bar{D}$)
- Search on A_0 above $\hat{A}$ until $\bar{D} = \hat{D}$ ($\rightarrow A_U$)
- Repeat for A_0 below $\hat{A}$ ($\rightarrow A_L$)

A Practical Solution

In practice

- Appetite to compare AUROC stats for classification models
  - between samples: indicate drop potential
  - between models: comparison of predictive value
- Often surprising how far AUROC can move over time
- Answered the question:
  Can we predict likely range for *future* AUROC?

Future Research

Next steps:

- Using distance to specify a confidence interval
  - requires mapping to 95% confidence interval
- Bootstrap test statistic
  - Shift weight to closest distribution with $A = A_0$
  - Simulating from this distribution will satisfy the null hypothesis
  - Reject null if actual statistic is in tails of simulated distribution
- Extend to multiple samples
  - Classification variables from same population
  - Need to account for covariance