IIT ROORKEE

NPTEL ONLINE CERTIFICATION COURSE

PERFORMANCE METRICS Part-3

LECTURE 18

DR. GAURAV DIXIT

DEPARTMENT OF MANAGEMENT STUDIES

PERFORMANCE METRICS

- Performance Metrics in presence of a class of interest

$$\text{sensitivity} = \frac{n_{11}'}{n_{10}'+n_{11}'} = \text{true positive fraction}$$

$$\text{specificity} = \frac{n_{00}'}{n_{00}'+n_{01}'} = \text{true negative fraction}$$

- ROC (receiver operating characteristic) curve
  - Used to plot {sensitivity, 1-specificity} points as the cutoff value increases
  - Top left corner points reflect wanted performance

PERFORMANCE METRICS

- Open Excel and RStudio
- Rank Ordering of records for class of interest
  - Based on estimated probabilities of class membership

- Lift curve is used to display the effectiveness of the model in rank ordering of cases
  - Constructed using validation partition scores

PERFORMANCE METRICS

- Cumulative lift curve or gains chart
  - Used to plot cumulative no. of cases on x-axis and cumulative no. of true positive cases on y-axis
  - Plot displays the lift value of the model for a given no. of cases w.r.t the random selection (probability value of class membership determines the reference line)
- Open Excel and RStudio
- Decile Chart
  - Alternative plot to convey the same information as gains chart

PERFORMANCE METRICS

- Open RStudio
- Asymmetric Misclassification Costs
  - When misclassification error for a class of interest is more costly than for the other class
  - Example, misclassifying a customer as false positive who is actually likely to respond to the promotional offering
    - Opportunity cost of foregone sale vs. costs of making an offer (profit of ₹20 for a ₹ 100 item vs. ₹1 scenario)
  - Misclassification rate is not appropriate metric in this case

PERFORMANCE METRICS

- Asymmetric Misclassification Costs
  - Other considerations
    - Costs of analyzing data
    - Actual net value impact per record
    - New Goal : minimization of costs or maximization of profits
- Open Excel
- How to improve actual classifications by incorporating asymmetric misclassification costs?
    - Change the rules of classification e.g. cutoff value

PERFORMANCE METRICS

- Performance Metrics based on asymmetric misclassification costs

$$\text{average misclassification cost} = \frac{c_0 n_{0,1} + c_1 n_{1,0}}{n}$$

  - Measures average cost of misclassification per observation
  - Where c_i is cost of misclassifying a class i observation

Key References

- Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data by EMC Education Services (2015)
- Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner by Shmueli, G., Patel, N. R., & Bruce, P. C. (2010)

Thanks…