IIT ROORKEE

NPTEL ONLINE CERTIFICATION COURSE

# PERFORMANCE METRICS

## LECTURE 16

**DR. GAURAV DIXIT**

DEPARTMENT OF MANAGEMENT STUDIES

# PERFORMANCE METRICS

- Need for performance metrics
  - Usefulness of a model
  - Comparison of candidate models
- Classification Performance
  - Probability of misclassification
  - Naïve Rule: most prevalent class
    - Serves as benchmark
  - Class Separation
- Open RStudio

# PERFORMANCE METRICS

- Performance Metrics based on Naïve Rule
  - Multiple $R^2$
    - Distance between fit of model to data and fit of naïve rule to data
- Naïve rule equivalent for prediction
  - Sample mean
- Classification Matrix

  $n_{i,j}$: no. of class i cases classified as class j cases

| Classification Matrix | | |
|---|---|---|
| | Predicted Class | |
| Actual Class | 1 | 0 |
| 1 | $n_{1,1}$ | $n_{1,0}$ |
| 0 | $n_{0,1}$ | $n_{0,0}$ |

# PERFORMANCE METRICS

- Classification Performance
  - Validation partition classification matrix
  - Comparison of training partition classification matrix with validation partition classification matrix
    - Detect overfitting
- Performance Metrics based on classification matrix
  - Misclassification rate or error
  - Accuracy

# PERFORMANCE METRICS

- Performance Metrics based on classification matrix

$$\text{err} = \frac{n_{0,1} + n_{1,0}}{n}$$

$$\text{accuracy} = 1 - \text{err} = \frac{n_{0,0} + n_{1,1}}{n}$$

- Open RStudio
- Cutoff probability value
  - Accuracy for all the classes is important
    - A case is assigned to the class with the highest probability as estimated by the model

# PERFORMANCE METRICS

- Cutoff probability value
  - Accuracy for a particular class of interest is important
    - A case is assigned to the class of interest if probability for the class is above cutoff value
  - Default cutoff value for a two class model is 0.5 (principally similar to naïve rule)
- Open Excel
  - One-variable table

# PERFORMANCE METRICS

- Why change cutoff probability value from 0.5?
  - Class of interest
  - Asymmetric misclassification cost

- When to incorporate change in cutoff value?
  - After final model selection
  - Before model derivation

# 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…