PREDICTION PERFORMANCE

LECTURE 21

DR. GAURAV DIXIT

DEPARTMENT OF MANAGEMENT STUDIES

PERFORMANCE METRICS

- Prediction Performance
  - Continuous outcome variable
  - Predictive accuracy vs. goodness of fit
  - E.g., goodness of fit measures in regression modeling
    - R^2 and std. err estimate
    - Residual analysis
  - Prediction Error on validation partition
  - Benchmark criterion: average

PERFORMANCE METRICS

- Prediction Error
  - For a record i, prediction error = actual value - predicted value
  - $e_i = y_i - \hat{y}_i$

- Predictive Accuracy Measures
  - Average Error

$$\frac{1}{n}\sum_{i=1}^{n} e_i$$

  - On average, indicates over or under prediction

PERFORMANCE METRICS

- Predictive Accuracy Measures
  - Mean Absolute Error (MAE) or Mean Absolute Deviation (MAD)

$$\frac{1}{n}\sum_{i=1}^{n}|e_i|$$

    - On average, magnitude of error
  - Mean Absolute Percentage Error (MAPE)

$$100\% \times \frac{1}{n}\sum_{i=1}^{n}|e_i/yi|$$

    - On average, percentage deviation from actual values

PERFORMANCE METRICS

- Predictive Accuracy Measures
  - Root Mean Squared Error (RMSE)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2}$$

    - Similar to std. err estimate computed on validation partition
    - Measured in same unit as the outcome variable

PERFORMANCE METRICS

- Predictive Accuracy Measures
  - Total Sum of Squared Errors (Total SSE or SSE)

$$\sum_{i=1}^{n} e_i^2$$

- These Predictive Accuracy Measures are used to
  - Compare the candidate models
  - Degree of prediction accuracy
  - Outlier issues

PERFORMANCE METRICS

- Outlier influence in accuracy measures
    - By comparing median based measures and mean based measures
    - Histogram or boxplot of residuals
- Model with high predictive accuracy may or may not be same as
    - model with best fit of data
- Evaluation using visualization techniques
    - Lift curve
        - Relevant when records with highest predicted values are sought

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...