IIT ROORKEE

NPTEL ONLINE CERTIFICATION COURSE

MULTIPLE LINEAR REGRESSION Part-4

LECTURES 25

DR. GAURAV DIXIT

DEPARTMENT OF MANAGEMENT STUDIES

MULTIPLE LINEAR REGRESSION

- Variable Selection
  - Availability of large no. of variables for selecting a set of predictors
  - Main idea is to select most useful set of predictors for a given outcome variable of interest
  - Selecting all the variables in the model is not recommended
    - Data collection issues in future
    - Measurement accuracy issues for some variables
    - Missing values
    - Parsimony

MULTIPLE LINEAR REGRESSION

- Variable Selection
  - Selecting all the variables in the model is not recommended
    - Multicollinearity: two or more predictors sharing the same linear relationship with the outcome variable
    - Sample size issues: Rule of thumb

$$n > 5*(p+2)$$

Where n=no. of observations

And p=no. of predictors

    - Variance of predictions might increase due to inclusion of predictors which are uncorrelated with the outcome variable
    - Average error of predictions might increase due to exclusion of predictors which are correlated with the outcome variable

MULTIPLE LINEAR REGRESSION

- Bias-variance trade-off
  - too few vs. too many predictors
    - Few predictors -> higher bias -> lower variance
  - Drop variables with ‘coefficient < std. dev. of noise’ and with moderate or high correlation with other variables
    - Lower variance
- Steps to reduce the no. of predictors
  - Domain knowledge
  - Practical reasons

MULTIPLE LINEAR REGRESSION

- Steps to reduce the no. of predictors
  - Summary statistics and graphs
  - Statistical methods using computational power
    - Exhaustive search: all possible combinations
    - Partial-iterative search: algorithm based
- Exhaustive Search
  - Large no. of subsets
  - Criteria to compare models
    - Adjusted R^2

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…