IIT ROORKEE

NPTEL ONLINE CERTIFICATION COURSE

DIMENSION REDUCTION TECHNIQUES Part-3

LECTURE 15

DR. GAURAV DIXIT

DEPARTMENT OF MANAGEMENT STUDIES

DIMENSION REDUCTION TECHNIQUES

- Principal Component Analysis (PCA)
  - Used for reducing the no. of predictors
  - Used for quantitative variables
  - Highly correlated variable subsets
  - Main idea is to find a set of new variables that contains most of the information of original variables
  - Eliminating covariation and multicollinearity
  - Redistribution of variability
- Open RStudio

DIMENSION REDUCTION TECHNIQUES

- Principal Component Analysis (PCA)
  - Data Mining Process
    - Apply PCA to the training partition
    - Predictors would now be principal score columns
    - Apply the principal weights obtained from training partition to the variables in the validation partition to obtain the scores
  - Relationship between predictors and output variable is ignored

DIMENSION REDUCTION TECHNIQUES

- Data Mining Techniques
  - Subset selection procedures using Regression models
    - Linear regression for prediction
    - Logistic regression for classification
    - Regression models can also be used for combining categories (using p-values)
  - Classification and Regression Tree (CART)
    - Classification tree for classification
    - Regression tree for prediction (Using tree diagram)

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…