



IIT ROORKEE



NPTEL ONLINE
CERTIFICATION COURSE

MACHINE LEARNING TECHNIQUE

k-NEAREST NEIGHBORS (k-NN)- Part 2

LECTURE 29

DR. GAURAV DIXIT

DEPARTMENT OF MANAGEMENT STUDIES



k-NEAREST NEIGHBORS (k-NN)

- k-NN
 - Choosing appropriate value of k
 - $k=1$: powerful for large no. of records in training partition
 - $k>1$: smoothing effects (control overfitting issues)
 - Low value of k -> more likely to fit the noise
 - High value of k -> more likely to ignore the local patterns in the data
 - Trade-off between benefits from local pattern vs global effects
 - $k=n$: naïve rule

k-NEAREST NEIGHBORS (k-NN)

- k-NN
 - Value of k: depends on nature of the data as well
 - Low value of k for data with complex and irregular structures
 - Typical value of k: between '1-20'
 - Odd value of k is preferred to avoid ties in majority class decisions
- Best value of k
 - Classification performance on validation partition
- Open RStudio



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...

