DA5030 | Intro to Machine Learning & Data Mining

Northeastern University
Khoury College of
Computer Sciences

Practice Problem 8

Problem 1 (60 Points)

Build an R Notebook of the social networking service example in the textbook on pages 296 to 310. Show each step and add appropriate documentation.

Problem 2 (40 Points)

Provide 100-300 word answers to each of the following interview questions:

1. (10 Points) What are some of the key differences between SVM and Random Forest for classification? When is each algorithm appropriate and preferable? Provide examples.
2. (10 Points) Why might it be preferable to include fewer predictors over many?
3. (10 Points) You are asked to provide R-Squared for a kNN regression model. How would you respond to that request?
4. (10 Points) How can you determine which features to include when building a multiple regression model?

Submission Details

- Practice Problems are for learning and practice and therefore are not graded and no submission is required. You are encourage to discuss and review them with your peers. Additionally, they are reviewed during weekly recitations. If you desire, you may ask for individual feedback from the instructional staff during office hours. Completing practice problems will prepare you for the graded practicums and their completion is critical to doing well on the practicums and the final project.
- Answer each interview question in your R Notebook and discuss with your peers or the instructional staff either in-person or through the discussion forum or recitation.

Useful Resources

- R Markdown Notebooks
- SNS Data Set

Learning

Blackboard
Lynda.com
Data Camp

Support

Contact Instructor
Virtual Office
Book Appointment

Search