Northeastern University
Khoury College of
Computer Sciences

Practice Problems 3

1. Download the data set for the tutorial.
2. Follow this tutorial on applying kNN to prostate cancer detection and implement all of the steps in an R Notebook. Make sure to explain each step and what it does. (*Note*: The data set provided as part of this assignment has been slightly modified from the one used in the tutorial, so small deviations in the result can be expected.)
3. Once you've complete the tutorial, try another *kNN* implementation from another package, such as the ***caret*** package. Compare the accuracy of the two implementations.
4. Try the *confusionMatrix* function from the ***caret*** package to determine the accuracy of both algorithms.

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.

Useful Resources

- R Markdown Notebooks
- Prostate Cancer Data Set

Hints

- Occasionally packages are updated. For example, an update on Jan 31 2020 to **dplyr** caused an incompatibility with the **caret** package. In such scenarios you can often wait for an update to the package, not install an update, or downgrade to an earlier version of a package. Here's how to downgrade a package: webpage. To check the current version of a package use *sessionInfo*, e.g., *sessionInfo("dplyr")*.

Northeastern University
Khoury College of Computer Sciences

Learning

Blackboard
Lynda.com
Data Camp

Support

Contact Instructor
Virtual Office
Book Appointment

Search

Created by Martin Schedlbauer, PhD