Computer Exercise 4

Oct. 24, 2019

1. k-NN

Find a publicly available package of k-NN for classification. Study its usage and study the algorithm it used for searching nearest neighbors. Design your own experiment based on the data used in Exercises 1-3 to try the package, and observe the influence of the choice of k on the training process and on the test performance. Write a report on the experiment. The report should at least contain the following contents:

- Description of the package, including its source and platform, the algorithm(s) it used for searching nearest neighbors, and the hyper-parameters and options users need to choose.
- The experiment design. You can do multiple ways of experiment design to experience the performance of the method in different scenarios. Explain why you design your experiment in this way.
- Experiment results, with discussions on the observations in the experiment.

2. Random Forest (RF)

Find a publicly available package of random forest (RF) for classification. Study its usage and algorithm. Especially, study the method it used for assessing the relative contribution or importance of each feature in the decision. Design your own experiment based on the data used in Exercises 1-3 to try the package, and explore the behavior of the method in different experiment scenarios and different choices in the algorithm. Write a report on the experiment. The report should at least contain the following contents:

- Description of the package, including its source and platform, the algorithm(s), and the hyper-parameters and options users need to choose.
- Write a special section on the method to assess feature contributions in RF.
- The experiment design. You can do multiple ways of experiment design to experience the performance of the method in different scenarios. Explain why you design your experiment in this way.
- Experiment results, with discussions on the observations in the experiment.

Report due date: Nov. 3 (Sunday), 23:59