

Homework 3

UP TO AND INCLUDING MODULE 7

This homework will be graded on effort, like every other homework in the class. You must submit your homework as **one single PDF file**, containing a clear description of everything you did, and including as many screenshots from Jupyter Notebook as required to illustrate the work you did. You may not submit any other files to support your solutions – anything we should know should be included in your report. This is as much a test of technical communication as technical skill.

Problem 1

Tahoe Healthcare has been approached by a healthcare analytics company, Xaltra, about a new system for managing readmissions. The CEO of Tahoe believes the initial success with the CareTracker system could be enhanced through the use of better predictive analytics and is intrigued by Xaltra's approach. The Xaltra system merges data from a variety of hospital systems to provide state-of-the-art predictions of readmissions risk. It uses up to 50,000 variables and generates its predictions using advanced machine learning algorithms. Xaltra has been adopted by a number of hospitals throughout the country. The price for the Xaltra system is \$250,000 in up-front license and integration fees plus an ongoing \$135,000 per year fee for maintenance and support. The expected lifetime of the system is 10 years.

Xaltra requested a sample of data to test their system's performance against Tahoe's current logistic regression method (the one we developed together in class). Tahoe provided Xaltra with the same dataset and segment used in our class session, `Tahoe_Data.xlsx`, which includes admissions and outcome data for all AMI patients treated by Tahoe in the last year.

The ROC curve results for the Xaltra test are provided in the tab `Pb1_ROC` in the `homework4.xlsx` file. For convenience, these are shown alongside the ROC curve for Tahoe's current logistic prediction system.

- a) Given these ROC test results, what is your estimate of the total readmissions and CareTracker costs for AMI patients for the past year if Tahoe had used the Xaltra system? Explain your estimate.
- b) What would have been the reduction in cost relative to Tahoe's current system, over the last year? Do the savings justify the fees Xaltra is charging? Why or why not.

Problem 2

In an effort to get to know the Columbia neighborhood better, your cluster has decided to conduct a systematic review of restaurants in the area. Each week you will select a pair of students to try out a local restaurant and post their reviews and ratings of it. To make these reviews as informative as possible, however, you would like to select student reviewers whose taste in food most closely matches the preferences of the cluster. At the same time, it is desirable to select a pair of students who complement each other in the sense of having different preferences in cuisine. Being deeply immersed in Business Analytics this term, you have decided to use the cluster survey data and analytics to design your review teams.

- a) Food ratings data from your cluster is available in the `homework4.xlsx` file. Using this data, find the person in your learning team whose food preference most closely matches that of the overall cluster. Provide the name of your learning team reviewer and explain how you selected them. (Only consider students who are present in the data.)

Hint: Compute the "food rating distance" of the candidate reviewer to every other student in your cluster by accounting for the food types they jointly rated. The learning team reviewer you will select will be the one with the smallest average food rating distance to students in your cluster.

- b) Next, we want to find a partner for the reviewer you selected in part (a). For the person selected to be your learning team's reviewer, find the fellow student in your cluster whose taste in food is least similar to your reviewer. Provide their name and explain how you selected them.