DSO530 Statistical Learning Methods

Lecture 4b: Classification V

Dr. Xin Tong
Department of Data Sciences and Operations
Marshall School of Business
University of Southern California

Review question 1

What are ROC space and ROC curve? How does one usually use ROC curves to compare classification methods?

Review question 2

If ROC curve for method A intersects with ROC curve for method B, what can you say about these two methods?

Review question 3

What does the “ROC” in ROC curve stand for?

Book reading

Read Chap 4 of ISLR up to and including 4.3.4

A word of caution

- Recall how ROC curve was constructed in Tutorial 4
- Suppose you have a type I error control target and choose one point on the ROC curve accordingly (which one makes sense?)
- You get a classifier whose type I error on a specific dataset is bounded from above by the target value
- However, it is not guaranteed that when you apply this classifier to the entire population, its performance is as you expected. This is a point that is usually ignored by practitioners.
- (Optional) How to get a classifier whose population type I error is controlled by a preferred target α with high probability $1 - \delta$? A Python implementation: https://pypi.org/project/nproc/ . You will have to specify the type I error control target α, and how much you want the high probability $1 - \delta$.