BIOST/STAT 527: Non-parametric Regression and Classification

Class Meeting Times:
MW 10:00-11:20am (in THO 125)

In addition the class will be streamed on Zoom:
Zoom ID- 923 1520 8305
Zoom link- https://washington.zoom.us/j/92315208305

Office Hours:
(Noah) Tuesday 10:00am - 11:00am online (same zoom info as above)
(Anand) TBD

Instructor: Noah Simon
Office: 3rd floor of Hans Rosling Center for Population Health (likely much easier to contact via email during the pandemic)
Email: nrsimon@uw.edu

TA: Anand Hemmady Email: anandh@uw.edu

Course description

This course is an introduction to the tools and theory of non-parametric regression and classification. As part of the formal course objectives students are expected to learn: 1) Methods of non-parametric regression (eg. kernel smoothing, projection estimators, basis expansions, additive models, GAMs, tree-based models, and boosting/bagging). 2) Ideas behind the bias/variance tradeoff 3) Uses of cross-validation for model selection 4) Some of the frameworks for proving consistency/rates for various types of non-parametric estimators (eg. kernel smoothers, projection estimators, penalized estimators, sieves)

Prerequisites: Graduate students in Biostatistics/Statistics can register for the course, however familiarity with parametric modeling (and rates) at the graduate level would be useful (as well as familiarity/comfort with matrix algebra and multivariate taylor expansions). Some basic background in programming will also be useful (though programming in the course will not be extensive). Contact the instructor for permission to take the course if you are not a graduate student in Biostat/Stat and if you are not certain whether this course is right for you.

Computing Software: R and Python Python and R are freely available computing packages. They are available on departmental computers. Both can also be downloaded on your personal computer:
R from http://cran.r-project.org/ Python from https://www.python.org/

There are a number of textbooks on non-parametric estimation. We will not formally use any of these texts, but the following will be useful references:

Introduction to statistical learning (James, Witten, Hastie, Tibshirani) The elements of statistical learning (Friedman, Hastie, Tibshirani) Introduction to non-parametric estimation (Tsybakov) High Dimensional Statistics (Wainwright) Statistics for high dimensional data (Buhlmann and van de geer) Empirical processes in M estimation (van de geer)

A number of these texts are available for free online (potentially through springer-link).
While I definitely welcome students to use these books as a reference, class discussions are designed to be mostly self-contained and there is no official textbook for the course.

Slides for each lecture will be posted on the webpage before the class so that you can print them and bring your own copy. Hard copies will not be provided by the instructor.

**Assessment:**

Homework are assigned every 2 weeks (announced in class and on the webpage). Homeworks will consist of a combination of theoretical calculations and programming assignments. Students will need to submit your solutions (as instructed in each HW) electronically. There will be a final project (but no final exam).

**Important Notes:**

1. Class material, including lecture notes, homework assignments, and other course-related information will be posted on the webpage. Printed course materials will not be provided by the instructor. Please check the webpage regularly and print/copy the notes.
2. The instructor reserves the right to modify the course plan and the syllabus as conditions require.
3. Questions and discussions are welcome, and encouraged throughout the class; keep in mind that if there is something that is not clear to you, it most likely is unclear to others as well.
4. In the case of another spike in covid-19 cases, the course may move entirely online (*fingers crossed* that this won’t happen!)

**Tentative Course Outline:**

1. Parametric vs non-parametric models
2. The bias/variance tradeoff
3. Kernel smoothers, and local polynomial regression (theory and practice)
4. Projection estimators and sieves
5. Quantifying complexity in infinite dimensional (non-parametric) families
6. Penalties vs constraints
7. GLMs/M-estimators
8. Multivariate problems
9. Curse of dimensionality and additive models
10. Tree-based models
11. Boosting and Bagging

**Academic Integrity Statement**

Students at the University of Washington (UW) are expected to maintain the highest standards of academic conduct, professional honesty, and personal integrity. The UW School of Public Health (SPH) is committed to upholding standards of academic integrity consistent with the academic and professional communities of which it is a part. Plagiarism, cheating, and other misconduct are serious violations of the University of Washington Student Conduct Code (WAC 478-120). We expect you to know and follow the university’s policies on cheating and plagiarism, and the SPH Academic Integrity Policy. Any suspected cases of academic misconduct will be handled according to University of Washington regulations. For more information, see the University of Washington Community Standards and Student Conduct website.

**UW Disability Statement – Access and Accomodations**

Your experience in this class is important to me. If you have already established accommodations with Disability Resources for Students (DRS), please communicate your approved accommodations to me at your earliest convenience so we can discuss your needs in this course. If you have not yet established services through DRS, but have a temporary health condition or permanent disability that requires accommodations (conditions include but not limited to: mental health, attention-related, learning, vision, hearing, physical or health impacts), you are welcome to contact DRS at 206-543-8924 or uwdrs@uw.edu or disability.uw.edu. DRS
offers resources and coordinates reasonable accommodations for students with disabilities and/or temporary health conditions. Reasonable accommodations are established through an interactive process between you, your instructor(s) and DRS. It is the policy and practice of the University of Washington to create inclusive and accessible learning environments consistent with federal and state law.

**UW Inclusion Statement**

The UW School of Public Health seeks to ensure all students are fully included in each course. We strive to overcome systemic racism by creating an environment that reflects community and mutual caring, while we ally with others in combating all forms of social oppression. This is a work in progress, as transformation is rarely a fully-completed project. In this course, we will look for opportunities to improve our performance as we seek to break down institutional racism. This can include course readings, class interactions, faculty performance, and/or the institutional environment. We encourage students to talk to your faculty member and/or the program director if you have concerns about classroom climate. DCinfo@uw.edu is a resource for students with classroom climate concerns.

**Religious Accommodations**

Washington state law requires that UW develop a policy for accommodation of student absences or significant hardship due to reasons of faith or conscience, or for organized religious activities. The UW’s policy, including more information about how to request an accommodation, is available at Religious Accommodations Policy (https://registrar.washington.edu/staffandfaculty/religious-accommodations-policy/). Accommodations must be requested within the first two weeks of this course using the Religious Accommodations Request form (https://registrar.washington.edu/students/religious-accommodations-request/).