STAT 824 Project

Each student will conduct independent research on a special topic of nonparametric statistics. The project has four components: 1) Writing a literature review, 2) giving a 10-minute proposal presentation to the class mid-semester, 3) writing a complete report for the project by the end of the semester, and 4) giving a 20-minute presentation at the end of the semester. The project comprises 80% of your course grade.

Choose a topic by class time Thursday, January 30th. The topic must be approved by me. There is a (far-from-exhaustive) list of suggestions at the end of this document. *You cannot choose a topic that another student has already chosen.*

Part 1 (20% of course grade) : *Submit in class on Thursday, March 6th.* Identify 6 important papers related to your topic. Among the 6 papers, there will be ideally 2 or more papers *which offer competing methods*, so that you can later compare the performance of the estimators in a simulation study. Submit a literature review based on these 6 papers; specifically, write a summary of at least 12 sentences for *each* of the 6 papers and submit these summaries.

Part 2 (20%): *Present in class Thursday, March 27th.* Based on your literature review, identify two methods which you can implement and compare with each other in a simulation study. Give a 10-minute presentation to the class introducing your topic to the class and summarizing what you have learned in your literature review. Indicate which methods you will focus on implementing and describe how you intend to compare the methods in simulation. Moreover, identify a data set on which you plan to implement the methods.

Part 3 (20%): *During the last two days of class, April the 22nd and 24th.* Give a 20-minute presentation to the class, covering the content you will include in the final report.

Part 4 (20%): *Submit by midnight Thursday, April 29th via Blackboard.* Turn in a written report of at least 12 pages (in LaTeX) with these sections:

1. **Abstract**: Describe the setting of your project and your findings.
2. **Introduction**: Describe the topic, giving its background and motivation, and provide a review of relevant papers in the area. Your literature review here should be in paragraph form and have a logical flow.
3. **Methodology**: Describe the two methods you implemented in detail (so that someone could implement them if they read your report).
4. **Data analysis**: Demonstrate the use of the two (or more) estimation methods on a real data set.
5. **Simulation study**: Compare the performances of the two (or more) methods on synthetic data sets.
6. **Conclusions**: Summarize the findings of your simulation study and offer any remarks you may have about which method you think should be used and in what settings.

Some topics you might find interesting – but feel free to propose others!

1. Mixed model approach to nonparametric regression.
2. Estimation of a monotone regression function, i.e. $Y = m(X) + \varepsilon$ with m monotone.
3. Handling heteroscedastic variance in nonparametric regression, i.e. $Y = m(X) + \sigma(X)\varepsilon$.
4. Nonparametric regression with wavelets.
5. Nonparametric regression with orthogonal series (sums of sines and cosines).
6. Local bandwidth selection, i.e. selecting different bandwidth $h(x)$ for different x.
7. Projection pursuit regression.
8. K-nearest neighbors regression.
9. Gaussian process regression.
10. Nonparametric quantile regression.
11. Regression trees.
12. The single index model.
13. Bandwidth selection for kernel density estimation when $d \geq 2$.
14. Wavelets in nonparametric density estimation.
15. Kernel density estimation for circular data.
16. Empirical likelihood estimation.
17. Nonparametric maximum likelihood estimation.
18. The m-out-of-n bootstrap. Why is it better to resample $m < n$ out of n in some settings?
19. The jackknife.
20. Deconvolution of functions: $X \sim f$, $\varepsilon \sim g$, g known; estimate f having observed $X + \varepsilon$?
21. Nonparametric estimation of survival function with censored data.
22. Estimating the spectral density of a stationary time series with periodogram smoothing.
23. Optimal transport theory and methods.
24. Estimating the mode or the conditional mode with modal regression.
25. Anything in Bayesian nonparametrics.