SSC 383D: Statistical Modeling II
Course syllabus
Spring 2014

Course overview

SSC 383D is a capstone graduate-level course on multivariate statistical modeling. It is intended for Ph.D students in statistics, but also suitable for graduate students in other disciplines with (1) substantial previous statistical training, and (2) an interest in designing bespoke statistical methods for their own data sets.

We will focus on statistical inference arising from formal probability models. The course is taught primarily, although not exclusively, from a Bayesian perspective. Major topics to be covered include: theory of the multivariate normal distribution; mixture models and other latent-variable models; density estimation; hierarchical models; generalized linear models; and "non-iid" models incorporating, for example, spatial or temporal dependence. Examples will be taken from across the social, biological, and physical sciences.

Structure of the course

The course revolves around independent inquiry. Each week I will give you a set of exercises. These will consist of results for you to prove, problems for you to solve, and (more in the second half of the course) data sets for you to analyze. When you come to class, you will present your solutions, and we will discuss them. Not everyone will present every day, but the presentation load will even out over the course of the semester. There are no lectures and no exams. The idea is for you to derive the course material on your own, with a bit of guidance from me. As you will gather from this description, the work load is substantial. Expect to spent between 5-10 hours per week, or more, completing the exercises. I encourage you to work on your own, so that you may experience the intellectual thrill of re-inventing statistics solely through your own efforts. But you are allowed to discuss things with classmates if you wish.

Your course grade will be determined by two things: (1) your level of day-to-day involvement in presenting the exercises in class; and (2) the complete packet of completed exercises that you will turn in at the end of the course. If you are consistently unprepared when called upon to

present, your grade will suffer correspondingly.

Prerequisites

This is a graduate course in statistics, and you will need to do your own heavy lifting, both mathematically and computationally. Because of this, I assume that you have a solid background in the relevant material. The formal prerequisite is SSC 384C, although this is easily waived for students with adequate preparation. The substantive prerequisite is that you are familiar with the following material.

(1) Linear algebra and multivariable calculus.

(2) Simple programming skills in a language such as R or Matlab. The official language of instruction for the course is R, although you are free to use any language you wish.

(3) Basic probability at the level of Chapters 1–4 of *Introduction to Probability* by Bertsekas and Tsitsiklis. You will not need measure theory.

(4) Basic mathematical statistics at the level of Casella and Berger's *Statistical Inference*. This covers a lot of ground, so to be more specific, you will need to understand: sufficiency, maximum likelihood, point estimation, hypothesis testing, confidence intervals, sampling distributions, and basic large sample theory. You will not need: ancillarity, completeness, the Rao-Blackwell theorem, most powerful tests.

(5) An elementary understanding of linear regression, including: the fitted values and residuals of a regression, the interpretation of regression coefficients, the least-squares estimate of coefficients, confidence intervals and hypothesis tests based on normality assumptions, and simple ideas from ANOVA (F tests, decomposition of variance, and so forth).

(6) Exposure to Bayesian inference in some simple context, e.g. finite spaces ("What is the probability that the patient has the disease, given a positive test?") or simple conjugate families.

If you are ill-prepared in more than one of these areas, then this course is probably not for you, at least yet. I have listed the areas in rough order of (my own subjective assessment of) the difficulty with which they can be picked up on the fly. Thus if you are missing an item from 1–3, you will have a tough time. But if you know 1–3, you stand a good chance of picking up one of 4–6 pretty quickly.