Preface

Math 253: Statistical Computing & Machine Learning

Starting out in R

Math 253 and the Macalester statistics curriculum

Math 253, Statistical Computing and Machine Learning, is a course at
Macalester College. The course comes early in our curriculum for the
statistics major. Currently, that curriculum looks like:

1. Comp 110, Data Computing, a no-prerequisite course in data wran-
   gling and visualization. The main tools in the course are R, dplyr,
   and ggplot.

2. Math 125, Epidemiology. This course also has no pre-requisites. It's
   designed to teach quantitative and statistical literacy in the context
   of public health and decision making. Unlike the other courses in
   our curriculum, Math 125 is *not* computer intensive. This is an
   elective for the statistics major.

3. Math 155, Introduction to Statistical Modeling. This is our main
   entry-level statistics course. It also has no pre-requisites, but we en-
   courage students to take a course in our calculus sequence: Applied
   Multivariate Calculus I, II, and III. The course covers important
   concepts in modeling: model architectures and fitting using mainly
   linear and logistic regression, covariation and adjustment, interpre-
   tation of model coefficients, inference techniques such as analysis of
   variance and co-variance and the ways these techniques can inform
   the decisions needed to build useful models, causality including
   techniques for making reasonable conclusions about causation from
   observational data. The course makes very extensive and intensive
   use of R and the `mosaic` package for R.

4. This course, **Math 253, Statistical Computing and Machine
   Learning**. Math 253 introduces a broader set of model architec-
   tures (e.g. those associated with "machine learning" such as support
   vector machines) and the trade-offs that make machine learning a
   human skill rather than a push button mechanism. The main text
   is Introduction to Statistical Learning in R. Computing in R is used
   intensively in the course. Math 155 is a pre-requisite. The comput-
   ing used in that course is the only pre-requisite. A small part of
   Math 253 is given to basic programming in R, but the large major-
   ity is about the mathematical and statistical concepts involved in
   learning from data and the exercise of a variety of architectures for
   classification and regression models.

Math 253 is placed intentionally in the middle of our statistics curriculum to make it accessible to many students who are interested and may have use for the techniques, but whom are not primarily interested in statistics *per se.*

5. Upper level courses including:

   - Math 454: Bayesian statistics
   - Math 453: Biostatistics
   - Math 454: Mathematical statistics

6. Supporting courses including:

   - Math 354: Probability
   - Math 135: Applied Multivariate Calculus I
   - Math 236: Linear algebra
   - Comp 123: Core conceptsin computer science

7. Electives such as

   - Math 432: Mathematical modeling
   - Math/Comp 365: Numerical linear algebra
   - Comp 302: Introduction to database management systems

These notes are written in Bookdown

The document uses an elaboration on R/Markdown, called "Bookdown." I don't yet know if this will be a good way to maintain and distribute class notes.

Some advantages:

- All the notes are in one place.
- Students and other instructors can clone the notes for their own uses.
- People spotting mistakes can contribute corrections via the "pull request" mechanism supported by Bookdown working with GitHub.
- Multiple instructors can contribute to the notes via GitHub. This is the stated purpose behind Bookdown — allowing book-length publications to be authored by many authors.

Disadvantages:

- The notes are *always* a work in progress. Don't be mislead by the polish that RMarkdown and Bookdown give to the notes.