Teaching Statement

Michael Schweinberger

Background

I have developed a well-received graduate-level course, which introduces students to three popular streams of research in statistics, machine learning, and artificial intelligence: **low- and high-dimensional graphs representing**

- data structure: e.g., network data;
- model structure: e.g., conditional independence structure;
- mathematical operations: e.g., neural networks and deep learning.

I have taught the course 7 times since 2014 and intend to teach its 8th edition in the spring of 2022. My latest student-based teaching evaluations from the spring 2021 are 1.00 on a scale from 1 = "Outstanding" to 5 = "Poor" (13 students), despite the fact that I taught the course online due to the COVID-19 pandemic. All student-based teaching evaluations since 2017 can be found at the end of my teaching statement (24 pages).

In addition, I have taught classes on statistical learning at all levels 2012–present, at Penn State University (PSU), Rice University (Rice), and the University of Missouri, Columbia (MU):

- Bachelor's level (8 times): Stat 401 at PSU (2 times); Stat 310 at Rice (5 times); and Stat 4710 and 7710 at MU (1 time);
- Advanced Bachelor's level (3 times): Stat 419 at Rice (3 times);
- Master's level (3 times): Stat 519 at Rice (3 times);
- Ph.D. level (1 time): Stat 532 at Rice (1 time).

Stat 419 and 519 at Rice shared lectures, although the homeworks and tests were not the same.

Last, but not least, as a result of teaching the core course on the foundations of statistical learning from data (Stat 519) at Rice, I have been in charge of preparing and grading the written Ph.D. qualifying exam in statistics at Rice and leading oral examinations in 2018, 2019, and 2020.

Teaching statistics in the age of data science

Statistics provides the guiding principles for learning from data with theoretical guarantees and making well-informed decisions and predictions in the face of uncertainty, which is a critical skill in the complex, interdependent and interconnected world of the twenty-first century.

As such, statistics has much to offer to students, but the age of data science presents both opportunities and challenges for teaching statistics. On the one hand, it presents opportunities in that statistics courses have witnessed a surge of demand and statisticians have the opportunity to help shape an important slice of the world's future workforce. On the other hand, it presents challenges as statistics courses have more students, and the background and career goals of those students is more heterogeneous than in the past.

I have attempted to make my courses useful for all students, regardless of background and career goals, and have placed a strong emphasis on

- connecting with students by choosing examples that appeal to them, such as examples from artificial intelligence and machine learning;
- leaving no student behind, making sure that all students understand the key ideas of how to learn from data.

Connecting with students

In all of my courses, I attempt to connect with students by choosing examples that appeal to them. These examples are taken from my research, which intersects with random graphs and graphical models along with Ising models and Gibbs measures in physics, Markov random fields in image processing and machine learning, and Markov logic networks and Boltzmann machines in artificial intelligence. While the mentioned models come in countless forms and shapes, many of them are exponential families for dependent random variables, with or without missing data. As a consequence, these exponential-family models can be used to demonstrate the main ideas of statistical learning from data, because exponential-family models provide an important source of examples in courses on statistical learning. The mentioned models in artificial intelligence and machine learning are popular examples of them and may appeal to students, helping them understand the key ideas of how to learn from data. I therefore use simple versions of those models as examples in all of my courses: e.g., to demonstrate maximum likelihood estimation without missing data, I use a small-scale Markov random field as a model of the human brain, and to demonstrate maximum likelihood estimation with missing data, I use a small-scale Boltzmann machine from artificial intelligence.

Leaving no student behind

I have designed my courses so that no student is left behind, regardless of background and career goals. To leave no student behind, I place a strong emphasis on the key ideas of how to learn from data. I believe that the key ideas are simple and that every student can understand them, regardless of background. I explain the key ideas as clearly and simply as possible, and support them by careful mathematical arguments that start simple and proceed step by step, without skipping steps that may seem trivial to me but may not be obvious to students. I encourage students to ask questions in the classroom, so that I can address gaps in understanding on the spot.

Undergraduate-level courses

In introductions to the mathematical foundations of statistical learning from data at the undergraduate level and the advanced undergraduate level, I have found it useful to exploit analogies, such as analogies with criminal investigations and criminal trials. For example, to introduce statistical estimators, I use analogies with criminal investigations: Statistical estimators resemble detectives in criminal investigations, using a trail of evidence $X_1, \ldots, X_n$ to track down the source θ of the evidence. To explain statistical tests, I use analogies with criminal trials.

Graduate-level courses

In graduate-level courses, my goal is to make key ideas as accessible as possible, so that all students understand them.

Example 1: What do data scientists estimate when minimizing functions of data and parameters? Many data scientists minimize functions $f(X_1, \ldots, X_n;\, \theta)$ of data $X_1, \ldots, X_n$ and parameters θ (e.g., sums of squares, negative loglikelihood functions, and penalized versions of them). While most of them know how to compute minimizers of random functions $f(X_1, \ldots, X_n; \theta)$ of interest, computing minimizers of $f(X_1, \ldots, X_n; \theta)$ raises an important question:

What do minimizers of random functions $f(X_1, \ldots, X_n; \theta)$ estimate?

To understand what minimizers of random functions $f(X_1, \ldots, X_n; \theta)$ estimate, I impress on students the need to understand that

- many random functions $f(X_1, \ldots, X_n; \theta)$ are approximations of non-random functions, e.g., expectations $\mathbb{E}\, f(X_1, \ldots, X_n; \theta)$;
- minimizers of random functions $f(X_1, \ldots, X_n; \theta)$ estimate minimizers of non-random functions, e.g., expectations $\mathbb{E}\, f(X_1, \ldots, X_n; \theta)$.

I explain these ideas by using simple examples and plotting estimating functions in `R`, along with careful mathematical arguments.

Example 2: What do functions of data and parameters approximate? To understand what functions $f(X_1, \ldots, X_n; \theta)$ of data and parameters approximate and hence what minimizers of $f(X_1, \ldots, X_n; \theta)$ estimate, it is important to gain insight into the concentration-of-measure phenomenon (Talagrand, Annals of Probability, 1996). The concentration-of-measure phenomenon refers to the mathematical observation that the probability mass of smooth functions $f(X_1, \ldots, X_n; \theta)$ of many independent or weakly dependent random variables $X_1, \ldots, X_n$ is strongly concentrated around the expectation $\mathbb{E}\, f(X_1, \ldots, X_n; \theta)$. As a consequence, if we know the expectation $\mathbb{E}\, f(X_1, \ldots, X_n; \theta)$ or can approximate it (e.g., by bounding it), then we have a good idea of how $f(X_1, \ldots, X_n; \theta)$ will behave when data $X_1, \ldots, X_n$ are generated. These statements can be quantified by using concentration inequalities, e.g., Subgaussian concentration inequalities. To help students understand what $f(X_1, \ldots, X_n; \theta)$ approximates and hence what minimizers of $f(X_1, \ldots, X_n; \theta)$ estimate, I introduce them to concentration inequalities, which are indispensible tools in machine learning and high-dimensional statistics (e.g., Wainwright, 2019).

Facilitating learning and assessing learning outcomes

My courses on the mathematical foundations of statistical learning from data at the undergraduate level, the Master's level, and the Ph.D. level are more mathematical than many other statistics courses. In those courses, learning occurs by solving problems. I facilitate learning by solving well-chosen and well-motivated problems in the classroom, with greatest possible clarity and attention to detail; by providing students with problems from past semesters in the form of past exams; and by assigning homework problems. The assessment of learning outcomes in those courses is based on homeworks and exams.

In my course on statistical learning with networks, which attracts students with many different backgrounds and career goals, learning occurs by solving problems in homeworks and by applying the ideas learned in the course in final projects (which may be carried out by teams of students instead of individual students). I give students wide leverage to choose final projects that appeal to them. For example, I allow students with bioinformatics and genetics backgrounds to choose final projects that are more methodological in nature and involve statistical analyses of data sets from genetics, whereas students with mathematical skills and ambitions can choose final projects that are theoretical. The assessment of learning outcomes in those courses is based on homeworks, exams, and final projects.

Supplement

All student-based teaching evaluations since 2017 can be found in the supplement (24 pages), ordered from most recent to least recent.