

Statistical Machine Learning

Course Description

Modern computing power has enabled the development of powerful tools for uncovering complex high dimensional relationships. These tools form a basic component of the interrelated areas known as statistics, Machine Learning, big data, artificial intelligence, and data science. In this course we will cover the methods which are most important in application. Fundamental methods are L1 and L2 regularized regression, ensemble tree methods, and deep learning. Our focus is intuitive understanding and practical application of the methods. While this is not a theoretical course, we will look at some details of the algorithms in order to understand them. Examples will illustrate the methods in python and R. Useful background knowledge is basic computing skills (e.g. R or python but other languages can be used), linear algebra, and calculus (preferable some multivariate). The course is meant to be accessible to a wide variety of students. If your calculus is rusty to can get by, but you have to know or learn some environment like R or python to do the applied work.

Key topics covered will be:

1. Probability Review and Naive Bayes
2. K Nearest Neighbors and the Bias-Variance Tradeoff
3. Linear Regression and Regularization
4. Regularized Generalized Linear Models (logistic regression, multinomial regression)
5. Classification Metrics
6. Deep Neural Nets
7. Trees, Random Forests, and Boosting
8. Support Vector Machines
9. Clustering, Hierarchical and k-means
10. Dimension Reduction, Principal Components and the Autoencoder
11. Mixture Modeling, Latent Variables and the EM algorithm
12. State Space Models and Hidden Markov Models

13. LDA, Latent Dirichlet Allocation for text analysis

We may not get to all the topics but the essential ones are the first 10 (up to and including Dimension reduction).

Books

The most important reference is the the first one, “An Introduction to Statistical Learning”. We will cover much of the material in this book with emphasis on chapters 5,6,8, and 10.

“An Introduction to Statistical Learning, second edition“ by James, Witten, Hastie, and Tibshirani.

“The Elements of Statistical Learning“ by Hastie, Tibshirani, and Friedman.

“Computer Age Statistical Inference” (Efron and Hastie)

“Machine Learning” (Kevin Murphy)

“Deep Learning” (Goodfellow, Bengio, Courville)

The first book (An Introduction ...) is at a lower level, but still is a great book.

Other useful advanced books are:

“Machine Learning” (Kevin Murphy)

“Pattern Recognition and Machine Learning” (Christopher M. Bishop)

“Bayesian Reasoning and Machine Learning” (David Barber)

“Statistical Learning with Sparsity” (Hastie, Tibshirani, Wainwright)

Three more applied and very useful books are:

“Applied Predictive Modeling” (Kuhn and Johnson)

“Machine Learning with R” (Larntz)

R books:

“The Art of R Programming” (Matloff)

“Advanced R” (Wickham)

“R for Data Science” (Wickham, Grolemund)

Python Books:

“Python Distilled” (Beazley)

“Introducing Python” (Lubanovic)

“Python Data Science Handbook” (VanderPlas)

“Python for Data Analysis” (Wes McKinney)

“Hands-On Machine Learning with Scikit-Learn and TensorFlow” (Geron)

“Python Machine Learning Cookbook” (Joshi)

Course Materials

Course materials (including as much of the slides as I can) will be available on the web:

http://www.rob-mcculloch.org/2021_ml/webpage/index.html

Software

We will mostly use R and Python. For most topics, I give working examples in both R and python. That being said, you can use any software environment you want as long as it supports basic topics such as boosting and neural nets (deep learning). For example, I know that recently one student did everything in Julia.

Grades

Grades will be based on homework and final Projects which may (should!!!) be done in groups. By far the most important thing is your project.

Mostly the homework will be applied things you will do in R and/or Python (or whatever software you want).

Academic Dishonesty

Academic honesty is expected of all students in all examinations, papers, laboratory work, academic transactions and records. The possible sanctions include, but are not limited to, appropriate grade penalties, course failure (indicated on the transcript as a grade of E), course failure due to academic dishonesty (indicated on the transcript as a grade of XE), loss of registration privileges, disqualification and dismissal. For more information, see <http://provost.asu.edu/academicintegrity>.

Students with Disabilities

Disability Accommodations: Qualified students with disabilities who will require disability accommodations in this class are encouraged to make their requests to me at the beginning of the semester either during office hours or by appointment. Note: Prior to receiving disability accommodations, verification of eligibility from the Disability Resource Center (DRC) is required. Disability information is confidential.

Establishing Eligibility for Disability Accommodations

Students who feel they will need disability accommodations in this class but have not registered with the Disability Resource Center (DRC) should contact DRC immediately. Their office is located on the first floor of the Matthews Center Building. DRC staff can also be reached at: 480-965-1234 (V), 480-965-9000 (TTY). For additional information, visit: www.asu.edu/studentaffairs/ed/drc. Their hours are 8:00 AM to 5:00 PM, Monday through Friday.

Policy on Threatening Behavior

All incidents and allegations of violent or threatening conduct by an ASU student (whether on-or off campus) must be reported to the ASU Police Department (ASU PD) and the Office of the Dean of Students. If either office determines that the behavior poses or has posed a serious threat to personal safety or to the welfare of the campus, the student will not be permitted to return to campus or reside in any ASU residence hall until an appropriate threat assessment has been completed and, if necessary, conditions for return are imposed. ASU PD, the Office of the Dean of Students, and other appropriate offices will coordinate the assessment in light of the relevant circumstances.

Classroom behavior: Make sure you arrive on time for class

Excessive tardiness will be subject to sanctions. Under no circumstances should you allow your cell phone to ring during class. Any disruptive behavior, which includes ringing cell phones, listening to your mp3/iPod player, text messaging, constant talking, eating food noisily, reading a newspaper will not be tolerated. The use of laptops (unless for lecture note taking), cell phones, MP3, IPOD, etc are strictly prohibited during class. Students who engage in disruptive classroom behavior may be subject to various sanctions. The procedures for initiating a disruptive behavior withdrawal can be found at <https://clas.asu.edu/resources/disruptive-behavior>.

Absences related to religious observances/practices

If you will be absent from class due to a religious observance or practice, it is your responsibility to inform the instructor during the first week of class. Your instructor will work with you on alternative and reasonable arrangements for any time missed.

Absences related to university sanctioned events and activities:

If you will be absent from class due to participation in a university sanctioned event/activity, it is your responsibility to inform the instructor during the first week of class. Your instructor will work with you on alternative and reasonable arrangements for any time missed.

Inclusion

The School of Mathematical and Statistical Sciences encourages faculty to address and refer to students by their preferred name and gender pronoun. If your preferred name is different than what appears on the class roster, or you would like to be addressed using a specific pronoun, please let your instructor know.

Policy on Sexual Discrimination

Arizona State University is committed to providing an environment free of discrimination, harassment, or retaliation for the entire university community, including all students, faculty members, staff employees, and guests. ASU expressly prohibits discrimination, harassment, and retaliation by employees, students, contractors, or agents of the university based on any protected status: race, color, religion, sex, national origin, age, disability, veteran status, sexual orientation, gender identity, and genetic information.

Title IX:

Title IX is a federal law that provides that no person be excluded on the basis of sex from participation in, be denied benefits of, or be subjected to discrimination under any education program or activity. Both Title IX and university policy make clear that sexual violence and harassment based on sex is prohibited. An individual who believes they have been subjected to sexual violence or harassed on the basis of sex can seek support, including counseling and academic support, from the university. If you or someone you know has been harassed on the basis of sex or sexually assaulted, you can find information and resources at <https://sexualviolenceprevention.asu.edu/faqs>.

Note:

As a mandated reporter, I am obligated to report any information I become aware of regarding alleged acts of sexual discrimination, including sexual violence and dating violence. ASU

Counseling Services, <https://eoss.asu.edu/counseling>, is available if you wish discuss any concerns confidentially and privately.