# MATH 447: MATHEMATICS OF *MACHINE LEARNING*
## *FALL 2023*

INSTRUCTOR: STANISLAV MINSKER, KAP 406E
EMAIL: MINSKER@USC.EDU
TEACHING ASSISTANT: ROBERT HUANG, EMAIL: HUANGROB@USC.EDU
CLASSROOM: KAP 163, TIME: MWF 2 – 2:50PM
OFFICE HOURS: MONDAY 9:30-11AM, OR BY APPOINTMENT.

---

**General Information:**

This course provides an introduction to the mathematical foundations and concepts of *Machine Learning*. Machine learning algorithms are often analyzed using the language of probability theory and mathematical statistics, giving rise to the "statistical learning" framework. How does machine learning differ from the classical methods of mathematical statistics? Statistical learning framework does not assume that the data we observe strictly follows the underlying model (e.g., Gaussian distribution). Instead, (quoting L. Breiman), [1], *"The approach is that nature produces data in a black box whose insides are complex, mysterious, and, at least, partly unknowable. What is observed is a set of $x$'s that go in and a subsequent set of $y$'s that come out. The problem is to find an algorithm $f(x)$ such that for future $x$ in a test set, $f(x)$ will be a good predictor of $y$."*

Throughout the course, we will try to understand when the approach of statistical learning is advantageous, and what are its limitations. We will look at the basic underlying principles that naturally lead to some of the popular machine learning algorithms, and will introduce mathematical tools needed to understand their performance. While the main focus of the course is on the theoretical foundations, practical applications will be considered as well.

Please note: the course material will be communicated via the (chalk/white)board, and the lectures will not be recorded. Important course information, assignments, selected solutions, and grades will be posted on Blackboard. If you are new to Blackboard, please check https://studentblackboardhelp.usc.edu/ for help and useful information.

**Prerequisites:**

Good knowledge of linear algebra, calculus and probability theory: Math 226 or 229, Math 208 (preferably Math 407), Math 225. Additional required background material will be covered during lectures.

**Grading:**

- (40%) Homework assignments, given every other week. You are not allowed to search for assistance/solutions online but are encouraged to ask me for help if you get stuck (points will not be deducted for using hints).
- (25%) Midterm exam; the exam will take place during the class on **Monday, October 16.**

[1] Leo Breiman (2001). "Statistical Modeling: The Two Cultures", *Statistical Science*, Vol. 16, No. 3.

- (35%) Final exam; the exam takes place on **Friday, December 8, between 2 and 4 p.m.**
- Late submissions of homework assignments will not be accepted.
- Please see the registration calendar for additional information, including the last day to drop the course: https://classes.usc.edu/term-20233/calendar/.

Grades will be computed on the following scale: $[92\%, 100\%] = A$, $[89\%, 92\%) = A-$, $[86\%, 89\%) = B+$, $[82\%, 86\%) = B$, $[80\%, 82\%) = B-$, $[78\%, 80\%) = C+$, $[73\%, 78\%) = C$, $[70\%, 73\%) = C-$, $[68\%, 70\%) = D+$, $[65\%, 68\%) = D$, $[62\%, 65\%) = D-$, $[0\%, 62\%) = F$. The final cutoffs for the grades may be changed (but only in **your favor**) based on the overall performance of the class.

**Reminder**: D is a minimum passing in undergraduate courses.

**"Extra credit" opportunity:**

This course is partners with the Joint Educational Project (JEP, https://dornsife.usc.edu/joint-educational-project) at USC, a service-learning program. JEP participation involves placement as a volunteer in a community setting where you will be a mentor or tutor and learn about the neighborhood around USC as well as the city in which you live. You will be trained and supervised by JEP and its staff, and will complete 8 weeks of service starting from week 6 of the semester. Evidence of the successful completion of your placement will be provided to the instructor by JEP. As an incentive to participate, **I will drop the lowest homework score** once you complete your JEP assignment. Please reach out to me and to JEP directly if you are interested in getting involved.

**Books and useful references:**

The course material does not follow a single book. The main reference for the course is the book "Understanding Machine Learning: From Theory to Algorithms" by Shai Shalev-Shwartz and Shai Ben-David, available online at https://www.cs.huji.ac.il/ shais/UnderstandingMachineLearning/. Note that we may (and often will) use notations different from the ones used in the book.

Another useful resource is the book "The Elements of Statistical Learning" by T. Hastie. R. Tibshirani and Jerome Friedman. Authors generously made it available online: https://hastie.su.domains/ElemStatLearn/.

**(Approximate) list of covered topics:**

We are going to cover material that (approximately) corresponds to Chapters 2-6, 9-10, 12, 15-16, not necessarily in this order, of the book "Understanding Machine Learning". Some specific topics inculde

- Differences between the approaches of machine learning and statistics; the curse of dimensionality.
- Introduction to binary classification and empirical risk minimization; overfitting.
- Linear separators, kernel trick, and Reproducing Kernel Hilbert spaces; Support Vector Machines; Linear and Logistic regression.
- PAC learning model and bounds for the error of learning algorithms.
- Vapnik-Chervonenkis combinatorics.
- Convex learning problems and voting algorithms (AdaBoost).
- Neural networks.

---

**Important dates:**

- September 4: Labor Day, university holiday;

- September 8: last day to drop a class without a mark of "W," except for Monday-only classes, and receive a refund for Session 001; last day to change enrollment option to Pass/No Pass or audit for Session 001;
- October 6: last day to drop a course without a mark of "W" on the transcript. Mark of "W" will still appear on student record and STARS report and tuition charges still apply; last day to change between letter grade or Pass/No pass in a letter graded course for Session 001;
- Fall recess: October 12-13, Thanksgiving break: November 22-26.

**Students Requiring Special Accommodation:**

I would be happy to discuss any special accommodations at the beginning of the course. Any student requesting academic accommodations based on special needs is required to register with OSAS. A letter of verification for approved accommodations can be obtained from OSAS as well. Please be sure the letter is delivered to me as early in the semester as possible. OSAS is located in GFS 120 and is open 8:30 a.m. till 5:00 p.m., Monday through Friday. The phone number for OSAS is (213) 740-0776 and the email is OSASFrontDesk@usc.edu.

**Academic Integrity:**

General principles of academic honesty include the concept of respect for the intellectual property of others, the expectation that **individual work will be submitted unless otherwise allowed by an instructor**, and the obligations both to protect one's own academic work from misuse by others as well as to avoid using another's work as one's own. All students are expected to understand and abide by the principles. In particular, plagiarism - presenting someone else's ideas and work as your own, either verbatim or recast in your own words - is a serious academic offense with equally serious consequences.

The Student Guidebook contains the Student Conduct Code (SCampus, Part B: https://policy.usc.edu/scampus/). Students will be referred to the Office of Student Judicial Affairs and Community Standards for further review, should there be any suspicion of academic dishonesty. The Review process can be found at: https://sjacs.usc.edu/students/academic-integrity/.