USC Viterbi School of Engineering

# DSCI 552: Machine Learning for Data Science (Summer 2022)

---

**Units:** 4

**Instructor:** Mohammad Reza Rajati, PhD
PHE 414
rajati@usc.edu – Include DSCI 552 in subject

**Office Hours:** Online, by appointment

**TA(s):** TBD
@usc.edu – Include DSCI 552 in subject

**Office Hours:** TBD

**Course Producer(s):** TBD
@usc.edu – Include DSCI 552 in subject

**Lecture:** Tuesday, Wednesday, Thursday 3:00 pm –4:50 pm RTH 217

**Webpages:** Piazza Class Page for everything except grades
and USC DEN Class Page for grades
and GitHub for code submission

– All HWs, handouts, solutions will be posted in PDF format
– *Student has the responsibility to stay current with webpage material*

**Prerequisite:** Prior courses in multivariate calculus, linear algebra, probability, and statistics.
– This course is a prerequisite to DSCI 558.

**Other Requirements:** Computer programming skills.
Using Python is mandatory.
Students must know Python or must be willing to learn it.

**Tentative Grading:** Assignments 45%
Midterm 1 20%
Midterm 2 25%
Final Project 10%
Participation on Piazza* 5%

**Letter Grade Distribution:**

| | | | |
|---|---|---|---|
| $\geq$ 93.00 | A | 73.00 - 76.99 | C |
| 90.00 - 92.99 | A- | 70.00 - 72.99 | C- |
| 87.00 - 89.99 | B+ | 67.00 - 69.99 | D+ |
| 83.00 - 86.99 | B | 63.00 - 66.99 | D |
| 80.00 - 82.99 | B- | 60.00 - 62.99 | D- |
| 77.00 - 79.99 | C+ | $\leq$ 59.99 | F |

**Disclaimer:** Although the instructor does not expect this syllabus to drastically change, he reserves every right to change this syllabus any time in the semester.

**Note on e-mail vs. Piazza:** If you have a question about the material or logistics of the class and wish to ask it electronically, please post it on the piazza page (not e-mail). Often times, if one student has a question/comment, other also have a similar question/comment. Use private Piazza posts with the professor, TA, graders only for issues that are specific to your individually (e.g., a scheduling issue or grade issue). Try minimizing the use of email to the course staff.

**Catalogue Description:** Practical applications of machine learning techniques to real-world problems. Uses in data mining and recommendation systems and for building adaptive user interfaces.

**Course Description:** This is a foundational course with the primary application to data analytics, but is intended to be accessible both to students from technical backgrounds such as computer science, computer engineering, electrical engineering, or mathematics; and to students from less technical backgrounds such as business administration, communication, accounting, various medical specializations including preventative medicine and personalized medicine, genomics, and management information systems. A basic understanding of engineering and/or technology principles is needed, as well as basic programming skills, sufficient mathematical background in probability, statistics, and linear algebra.

**Course Objectives:** Upon successful completion of this course a student will

- Broadly understand major algorithms used in machine learning.
- Understand supervised and unsupervised learning techniques.
- Understand regression methods.
- Understand resampling methods, including cross-validation and bootstrap.
- Understand decision trees, dimensionality reduction, regularization, clustering, and kernel methods.
- Understand hidden Markov models and graphical models.
- Understand feedforward and recurrent neural networks and deep learning.

**Exam Dates:**

- **Midterm Exam 1:** Friday June 24, 3-4:50 PM. (May be changed to a a different hour on the same day)
- **Midterm Exam 2:** Friday, July 22, 3:00 - 4:50 PM (May be changed to a different hour on the same day)
- **Final Project Due (no grace period):** Tuesday July 26, 3:00 PM

**Textbooks:**

- **Required Textbook:**

  1. Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, *An Introduction to Statistical Learning with Applications in R*, Springer, 2021. (ISLR)
     Available at https://web.stanford.edu/~hastie/ISLRv2_website.pdf

- **Recommended Textbooks:**

  1. *Applied Predictive Modeling*, 1st Edition
     **Authors:** Max Kuhn and Kjell Johnson; Springer; 2016. **ISBN-13:** 978-1-4614-6848-6
  2. *Machine Learning: A Concise Introduction*, 1st Edition
     **Author:** Steven W. Knox; Wiley; 2018. **ISBN-13:** 978-1-119-43919-6
  3. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edition
     **Authors:** Trevor Hastie, Robert Tibshirani, and Jerome Friedman; Springer; 2008. (ESL) **ISBN-13:** 978-0387848570
  4. *Machine Learning: An Algorithmic Perspective*, 2nd Edition
     **Author:** Stephen Marsland; CRC Press; 2014. **ISBN-13:** 978-1-4614-7137-0
  5. *Deep Learning*, 1st Edition
     **Authors:** Ian Goodfellow, Yoshua Bengio, and Aaron Courville; MIT Press; 2016. (DL) **ISBN-13:** 978-0262035613
  6. *Neural Networks and Learning Machines*, 3rd Edition
     **Author:** Simon Haykin; Pearson; 2008. **ISBN-13:** 978-0131471399
  7. *Neural Networks and Deep Learning: A Textbook*, 1st Edition
     **Authors:** Charu Aggrawal; Springer; 2018. **ISBN-13:** 978-3319944623
  8. *Introduction to Machine Learning*, 2nd Edition
     **Author:** Ethem Alpaydine; MIT Press; 2010. (AL) **ISBN-13:** 978-8120350786
  9. *Machine Learning*, 1st Edition
     **Authors:** Tom M. Mitchell; McGraw-Hill Education; 1997. **ISBN-13:** 978-0070428072

**Grading Policies:**

- The letter grade distribution table guarantees the *minimum* grade each student will receive based on their final score. When appropriate, relative performance measures will be used to assign the final grade, at the discretion of the instructor.
  - Final grades are non-negotiable and are assigned at the discretion of the instructor. If you cannot accept this condition, you should not enroll in this course.
- Your lowest homework grade and half of your second lowest homework grade will be dropped from the final grade. For example, if you received 90, 85, 10, 95, 65, 80, 100, 100 your homework score will be $\frac{0.5\times65+80+85+90+95+100+100}{6.5} = 89.62$ instead of $\frac{10+65+80+85+90+95+100+100}{8} = 78.13$.

- Homework 0 will not be graded.
- *Participation on Piazza has up to 5% extra credit, which is granted on a competitive basis *at the discretion of the instructor.*

- **Homework Policy**
  - Homework is assigned on an approximately weekly basis. Homework due dates are mentioned in the course outline, so mark your calendars. A two-day grace period can be used for each homework with 10% penalty per day. *Absolutely no late homework will be accepted after the grace period. A late assignment results in a zero grade.*
  - Assignments are *project-style*; therefore, *we do not provide solutions to the assignments.* We do not provide sample code/ code skeleton for the assignments. This is a firm rule.
  - Poor internet connection, failing to upload properly, or similar issues are NOT acceptable reasons for late submissions. If you want to make sure that you do not have such problems, submit homework eight hours earlier than the deadline. Please do not ask the instructor to make individual exceptions.
  - Homework solutions and simulation results should be typed or *scanned* using scanners or mobile scanner applications like CamScan and uploaded (photos taken by cell-phone cameras and in formats other than pdf will NOT be accepted). Programs and simulation results have to be uploaded on GitHub as well.
  - Students are encouraged to discuss homework problems with one another, but each student must do their own work and submit individual solutions written/ coded in their own hand. Copying the solutions or submitting identical homework sets is written evidence of cheating. The penalty ranges from F on the homework or exam, to an F in the course, to recommended expulsion.
  - Posting the homework assignments and their solutions to online forums or sharing them with other students is strictly prohibited and infringes the copyright of the instructor. Instances will be reported to USC officials as academic dishonesty for disciplinary action.
- **Exam Policy**
  - **Make-up Exams:** No make-up exams will be given. If you cannot make the above dates due to a class schedule conflict or personal matter, you must drop the class. In the case of a required business trip or a medical emergency, a signed letter from your manager or physician has to be submitted. This letter must include the contact of your physician or manager.
  - Exams will be closed book and notes. Calculators are allowed but computers and cell-phones or any devices that have internet capability are not allowed, except tablets for writing the solutions if exams are online. One letter size cheat sheet (back and front) is allowed for the first midterm. Two letter size cheat sheets (back and front) are allowed for the second midterm. Cheat sheets cannot be electronic.
  - All exams are cumulative, with an emphasis on material presented since the last exam.
- **Attendance:**

  - Students are required to attend all the lectures and discussion sessions and actively participate in class discussions. Use of cellphones and laptops is prohibited in the classroom. If you need your electronic devices to take notes, you should discuss with the instructor at the beginning of the semester.

**Important Notes**:

- Textbooks are secondary to the lecture notes and homework assignments.
- Handouts and course material will be distributed.
- Please use your USC email to register on Piazza and to contact the instructor and TAs.

## Tentative Course Outline

| Wednesday | Thursday | Tuesday |
|---|---|---|
| May 18th **1**<br>**Introduction to Statistical Learning** (ISLR Chs.1,2, ESL Chs.1,2)<br>Motivation: Big Data<br>Supervised vs. Unsupervised Learning | 19th **2**<br>**Introduction to Statistical Learning** (ISLR Chs.1,2, ESL Chs.1,2)<br>Regression, Classification<br>The Regression Function<br>Nearest Neighbors | 24th **3**<br>**Introduction to Statistical Learning** (ISLR Chs.1,2, ESL Chs.1,2)<br>Model Assessment<br>The Bias-Variance Trade-off<br>No Free Lunch Theorem |
| 25th **4**<br>**Linear Regression** (ISLR Ch.3, ESL Ch. 3)<br>Estimating Coefficients<br>Estimating the Accuracy of Coefficients | 26th **5**<br>**Linear Regression** (ISLR Ch.3, ESL Ch. 3)<br>Variable Selection and Hypothesis Testing<br>Multiple Regression<br>Analysis of Variance and the F Test | 31st **6**<br>**Linear Regression** (ISLR Ch.3, ESL Ch. 3)<br>Stepwise Variable Selection<br>Qualitative Variables |
| June 1st **7**<br>**Classification** (ISLR Ch. 4, ESL Ch. 4)<br>Multi-class and Multi-label Classification<br>Logistic Regression<br>Class Imbalance<br>Hypothesis Testing and Variable Selection | 2nd **8**<br>**Classification** (ISLR Ch. 4, ESL Ch. 4)<br>Subsampling and Upsampling<br>SMOTE<br>Multinomial Regression<br>Bayesian Linear Discriminant Analysis | 7th **9**<br>**Classification** (ISLR Ch. 4, ESL Ch. 4)<br>Measures for Evaluating Classifiers<br>Quadratic Discriminant Analysis*<br>Comparison with K-Nearest Neighbors<br>The Naïve Bayes' Classifier<br>Text Classification<br>Feature Creation for Text Data<br>Handling Missing Data |
| 8th **10**<br>**Resampling Methods** (ISLR Ch. 5, ESL Ch. 7)<br>Model Assessment<br>Validation Set Approach<br>Cross-Validation<br>The Bias-Variance Trade-off for Cross-Validation<br>The Bootstrap<br>Bootstrap Confidence Intervals | 9th **11**<br>**Linear Model Selection and Regularization** (ISLR Ch.6, ESL Ch. 3)<br>Subset Selection<br>AIC, BIC, and Adjusted $R^2$)<br>Shrinkage Methods<br>Ridge Regression | 14th **12**<br>**Linear Model Selection and Regularization** (ISLR Ch.6, ESL Ch. 3)<br>The LASSO<br>Elastic Net<br>Dimension Reduction Methods* |

| Wednesday | Thursday | Tuesday |
|---|---|---|
| 15th **13**<br>**Tree-based Methods** (ISLR Ch. 8, ESL Chs. 9, 10)<br>Regression and Classification Trees<br>Cost Complexity Pruning | 16th **14**<br>**Tree-based Methods** (ISLR Ch. 8, ESL Chs. 9, 10, 16)<br>Bagging, Random Forests, and Boosting | 21st **15**<br>**Support Vector Machines** (ISLR Ch. 9, ESL Ch. 12)<br>Maximal Margin Classifier<br>Support Vector Classifiers |
| 22nd **16**<br>**Support Vector Machines** (ISLR Ch. 9, ESL Ch. 12)<br>The Kernel Trick<br>Support Vector Machines<br>L1 Regularized SVMs<br>Multi-class and Multilabel Classification<br>The Vapnik-Chervonenkis Dimension*<br>Support Vector Regression | 23rd **17**<br>**Unsupervised Learning** (ISLR Ch. 10, ESL Ch. 14)<br>K-Means Clustering<br>Hierarchical Clustering | 28th **18**<br>**Unsupervised Learning** (ISLR Ch. 10, ESL Ch. 14)<br>Practical Issues in Clustering |
| 29th **19**<br>**Unsupervised Learning** (ISLR Ch. 10, ESL Ch. 14)<br>Principal Component Analysis<br>Anomaly Detection*<br>Association Rules*<br>Mixture Models and Soft K-Means* | 30th **20**<br>**Active and Semi-Supervised Learning**<br>Semi-Supervised Learning<br>Self-Training<br>Co-Training<br>Yarowsky Algorithm Refinements<br>Active vs. Passive Learning<br>Stream-Based vs. Pool-Based Active Learning<br>Query Selection Strategies | July 5th **21**<br>**Neural Networks and Deep Learning** (ESL Ch. 11, DL Ch. 6)<br>The Perceptron<br>Feedforward Neural Networks<br>Backpropagation and Gradient Descent<br>Overfitting |
| 6th **22**<br>**Neural Networks and Deep Learning** (DL Chs. 6, 7)<br>Autoencoders and Deep Feedforward Neural Networks<br>Regularization<br>Early Stopping and Dropout<br>Adversarial Training* | 7th **23**<br>**Neural Networks and Deep Learning** (DL Chs. 9, 10)<br>Convolutional Neural Networks<br>Sequence Modeling<br>Recurrent Neural Networks | 12th **24**<br>**Neural Networks and Deep Learning** (DL Ch. 10)<br>Sequence-to-Sequence Modeling*<br>Long Short Term Memory (LSTM) Neural Networks |

| Wednesday | Thursday | Tuesday |
|---|---|---|
| 13th **25**<br>**Hidden Markov Models**<br>(AL Ch. 15)<br>Principles<br>The Viterbi Algorithm | 14th **26**<br>**Reinforcement Learning***<br>Definitions<br>Task-Reward-Policy Formulation<br>Total Discounted Future Reward<br>Optimal Policy<br>Value Function<br>Q-Function<br>The Bellman Equation<br>Q-Learning<br>Exploration- Exploitation<br>Temporal Difference Learning<br>Extensions to Stochastic Environments and Rewards<br>Deep Reinforcement Learning | 19th **27**<br>**Fuzzy Systems***<br>Fuzzy Sets<br>Set Operations<br>T-norms, T-conorms, and Fuzzy complements<br>Cylindrical Extensions and Fuzzy Relations<br>Fuzzy If-Then Rules as Association Rules<br>Inference from Fuzzy Rules<br>Fuzzification and Defuzzification<br>Learning Fuzzy Rules from Examples<br>The Wang-Mendel Algorithm<br>Fuzzy C-Means Clustering |
| 20th **28**<br>**Review*** | 21st **29**<br>**Guest Lecture*** | 26th **30**<br>**Final Project** |

**Notes:**

- Items marked by * will be covered only if time permits.

## Homework and Project Due Dates

| Monday | |
|---|---|
| May 23rd<br>**Homework 0 Due (not graded)** | **1** |
| 30th<br>**Homework 1 Due (Moved to Tuesday May 31)** | **2** |
| June 6th<br>**Homework 2 Due** | **3** |
| 13th<br>**Homework 3 Due** | **4** |
| 20th<br>**Homework 4 Due** | **5** |
| 27th<br>**Homework 5 Due** | **6** |
| July 4th<br>**Homework 6 Due (moved to Tuesday July 5th)** | **7** |
| 11th<br>**Homework 7 Due** | **8** |
| 18th<br>**Homework 8 Due** | **9** |
| 25th<br>**Final Project Due (moved to Tuesday July 26th)** | **10** |

## Statement on Academic Conduct and Support Systems

**Academic Conduct:**

Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, "Behavior Violating University Standards" policy.usc.edu/scampus-part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on Research and Scholarship Misconduct.

**Students and Disability Accommodations:**

USC welcomes students with disabilities into all of the University's educational programs. The Office of Student Accessibility Services (OSAS) is responsible for the determination of appropriate accommodations for students who encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each course. The LOA must be given to each course instructor by the student and followed up with a discussion. This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at osas.usc.edu. You may contact OSAS at (213) 740-0776 or via email at osasfrontdesk@usc.edu.

**Support Systems:**

*Counseling and Mental Health - (213) 740-9355 – 24/7 on call*
studenthealth.usc.edu/counseling
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

*National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call*
suicidepreventionlifeline.org
Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

*Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-9355(WELL), press "0" after hours – 24/7 on call*
studenthealth.usc.edu/sexual-assault
Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

*Office for Equity, Equal Opportunity, and Title IX (EEO-TIX) - (213) 740-5086*
eeotix.usc.edu
Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

*Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298*
usc-advocate.symplicity.com/care_report
Avenue to report incidents of bias, hate crimes, and microaggressions to the Office for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

*The Office of Student Accessibility Services (OSAS) - (213) 740-0776*
osas.usc.edu
OSAS ensures equal access for students with disabilities through providing academic accommodations and auxiliary aids in accordance with federal laws and university policy.

*USC Campus Support and Intervention - (213) 821-4710*

campussupport.usc.edu
Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

*Diversity, Equity and Inclusion - (213) 740-2101*
diversity.usc.edu
Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

*USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call*
dps.usc.edu,emergency.usc.edu
Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

*USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-120 – 24/7 on call*
dps.usc.edu Non-emergency assistance or information.

*Office of the Ombuds - (213) 821-9556 (UPC) / (323-442-0382 (HSC)*
ombuds.usc.edu
A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.

*Occupational Therapy Faculty Practice - (323) 442-3340 or otfp@med.usc.edu*
chan.usc.edu/otfp
Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.