USC Viterbi
School of Engineering

DSCI 552: Machine Learning for Data Science (Spring 2023)

Units: 4

Instructor: Mohammad Reza Rajati, PhD
PHE 412
rajati@usc.edu – Include DSCI 552 in subject.

Office Hours: Right after the lecture, by appointment

Webpage: Personal Homepage at Intelligent Decision Analysis

TA(s): Will be introduced on Piazza.

Lecture 1: Tuesday, Thursday, 10:00 am –11:50 am GFS 116 & Online

Lecture 2: Tuesday, Thursday, 3:30 pm –5:20 pm OHE 122 & Online

Webpages: Piazza Class Page for discussions, announcements, and course materials
and USC DEN Class Page for exams and 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). Minimize the use of email to the course staff and only use it when *absolutely necessary.*

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 1 (Both Sections):** Friday March 10, 10:00 AM-11:50 AM
- **Midterm 2 (Both Sections):** Friday April 28, 10:00 AM-12:00 Noon
- **Final Project Due (Both Sections):** Tuesday May 9, 4:00 PM as set by the university. **Grace period**: the project can be submitted until 11:59 PM of the same day with 30% penalty. Any change in the project after the deadline is considered late submission. The project is graded based on *when it was submitted, not when it was finished.*

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 biweekly basis. Homework due dates are mentioned in the course outline, so mark your calendars. A three-day grace period can be used for each homework with 10% penalty per day. Any change in homework after the deadline makes it a late submission. *Absolutely no late homework will be accepted after the grace period. A late assignment results in a zero grade.*
  - Late Days: No late homework will be accepted after the three day grace period. One second after the deadline is considered late. However, students are allowed to use *six* late days for homework for any reason (including sickness, family emergencies, overwhelming workload, exams, etc) *without incurring the 10% penalty.* Beyond that, no individual extension will be granted to anyone for any reason whatsoever.

    Example: A student can submit six assignments, one day late each, without any penalty. Or three assignments, two days late each, without penalty, or two assignments three days late each. A student cannot use four late days for one assignment, and two late days for another assignment. An assignment submitted four days late will receive a zero grade, although its grade will be dropped as the lowest homework grade, according to the above grading policies.
  - Use your six late days strategically and only if you absolutely need them. Always remember that later in the semester, you might become sick or have heavy workload in other courses and might need to use your late days.
  - Assignments are *project-style*; therefore, *we do not provide solutions to 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 is graded based on *when it was submitted, not when it was finished.*
  - 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. One important (but not exclusive) instance of cheating is having access to other students' solutions. Claims of " being inspired" by other students' codes, or using them as "sample code" are not acceptable. Asking questions from your peers and exchanging tips about coding are highly encouraged and should not be confused with outright cheating.

  - 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.
  - An excused absence supported by documents in the first midterm can be made up by using the second midterm's grade in lieu of the first midterm. An excused absence in the second midterm results in an IN (incomplete) grade.
  - Exams will be closed book and notes. Calculators are allowed but computers and cellphones or using any devices that have internet capability are not allowed, except for writing the solutions or being proctored are not allowed. One letter size cheat sheet (back and front) is allowed for Midterm 1. Two letter size cheat sheets (back and front) are allowed for Midterm 2.
  - All exams are cumulative, with an emphasis on material presented since the last exam.
  - For several reasons, including unauthorized circulation of previous exams, we DO NOT provide exam solutions. This is a firm rule.

- **Project**
  - The final project is more like a slightly extended Homework that will be assigned after Midterm 2 as the final summative experience.
  - The project topic and steps will be provided to students, similar to homework assignments.
  - Projects must be finished *individually*.
  - A short grace period of a few hours after the project deadline will be given to students for 30% penalty. Late submissions will be graded zero. One second late is late.
  - Project is graded based on *when it was submitted, not when it was finished.*
  - Homework late days *cannot* be used for project in any circumstances.

- **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.

Tuesday	Thursday
$\boxed{\text{Jan 10th}}$ **1** **Introduction to Statistical Learning** (ISLR Chs.1,2, ESL Chs.1,2) Motivation: Big Data Supervised vs. Unsupervised Learning	12th **2** **Introduction to Statistical Learning** (ISLR Chs.1,2, ESL Chs.1,2) Regression, Classification The Regression Function Nearest Neighbors
17th **3** **Introduction to Statistical Learning** (ISLR Chs.1,2, ESL Chs.1,2) Model Assessment The Bias-Variance Trade-off No Free Lunch Theorem	19th **4** **Linear Regression** (ISLR Ch.3, ESL Ch. 3) Estimating Coefficients Estimating the Accuracy of Coefficients
24th **5** **Linear Regression** (ISLR Ch.3, ESL Ch. 3) Variable Selection and Hypothesis Testing Multiple Regression Analysis of Variance and the F Test	26th **6** **Linear Regression** (ISLR Ch.3, ESL Ch. 3) Stepwise Variable Selection Qualitative Variables
31st **7** **Classification** (ISLR Ch. 4, ESL Ch. 4) Multi-class and Multi-label Classification Logistic Regression Class Imbalance Hypothesis Testing and Variable Selection	$\boxed{\text{Feb 2nd}}$ **8** **Classification** (ISLR Ch. 4, ESL Ch. 4) Subsampling and Upsampling SMOTE Multinomial Regression Bayesian Linear Discriminant Analysis
7th **9** **Classification** (ISLR Ch. 4, ESL Ch. 4) Measures for Evaluating Classifiers Quadratic Discriminant Analysis* Comparison with K-Nearest Neighbors The Naïve Bayes' Classifier Text Classification Feature Creation for Text Data Handling Missing Data	9th **10** **Resampling Methods** (ISLR Ch. 5, ESL Ch. 7) Model Assessment Validation Set Approach Cross-Validation The Bias-Variance Trade-off for Cross-Validation The Bootstrap Bootstrap Confidence Intervals
14th **11** **Linear Model Selection and Regularization** (ISLR Ch.6, ESL Ch. 3) Subset Selection AIC, BIC, and Adjusted R^2) Shrinkage Methods Ridge Regression	16th **12** **Linear Model Selection and Regularization** (ISLR Ch.6, ESL Ch. 3) The LASSO Elastic Net Dimension Reduction Methods*

Tuesday	Thursday
21st **13** **Tree-based Methods** (ISLR Ch. 8, ESL Chs. 9, 10) Regression and Classification Trees Cost Complexity Pruning	23rd **14** **Tree-based Methods** (ISLR Ch. 8, ESL Chs. 9, 10, 16) Bagging, Random Forests, and Boosting
28th **15** **Support Vector Machines** (ISLR Ch. 9, ESL Ch. 12) Maximal Margin Classifier Support Vector Classifiers	Mar 2nd **16** **Support Vector Machines** (ISLR Ch. 9, ESL Ch. 12) The Kernel Trick Support Vector Machines L1 Regularized SVMs Multi-class and Multilabel Classification The Vapnik-Chervonenkis Dimension* Support Vector Regression
7th **17** **Unsupervised Learning** (ISLR Ch. 12, ESL Ch. 14) K-Means Clustering Hierarchical Clustering	9th **18** **Unsupervised Learning** (ISLR Ch. 12, ESL Ch. 14) Practical Issues in Clustering
14th Spring Recess	16th Spring Recess
21st **19** **Unsupervised Learning** (ISLR Ch. 12, ESL Ch. 14) Principal Component Analysis Anomaly Detection* Association Rules* Mixture Models and Soft K-Means*	23rd **20** **Active and Semi-Supervised Learning** Semi-Supervised Learning Self-Training Co-Training Yarowsky Algorithm Refinements Active vs. Passive Learning Stream-Based vs. Pool-Based Active Learning Query Selection Strategies
28th **21** **Neural Networks and Deep Learning** (ISLR Ch. 10, ESL Ch. 11, DL Ch. 6) The Perceptron Feedforward Neural Networks Backpropagation and Gradient Descent Overfitting	30th **22** **Neural Networks and Deep Learning** (DL Chs. 6, 7) Autoencoders and Deep Feedforward Neural Networks Regularization Early Stopping and Dropout Adversarial Training*

Tuesday	Thursday
Apr 4th **23** **Neural Networks and Deep Learning** (ISLR Ch. 12, DL Chs. 9, 10) Convolutional Neural Networks Sequence Modeling Recurrent Neural Networks	6th **24** **Neural Networks and Deep Learning** (ISLR Ch. 12, DL Ch. 10) Sequence-to-Sequence Modeling* Long Short Term Memory (LSTM) Neural Networks
11th **25** **Hidden Markov Models** (AL Ch. 15) Principles The Viterbi Algorithm	13th **26** **Reinforcement Learning*** Definitions Task-Reward-Policy Formulation Total Discounted Future Reward Optimal Policy Value Function Q-Function The Bellman Equation Q-Learning Exploration- Exploitation Temporal Difference Learning Extensions to Stochastic Environments and Rewards Deep Reinforcement Learning
18th **27** **Reinforcement Learning*** Definitions Task-Reward-Policy Formulation Total Discounted Future Reward Optimal Policy Value Function Q-Function The Bellman Equation Q-Learning Exploration- Exploitation Temporal Difference Learning Extensions to Stochastic Environments and Rewards Deep Reinforcement Learning	20th **28** **Fuzzy Systems*** Fuzzy Sets Set Operations T-norms, T-conorms, and Fuzzy complements Cylindrical Extensions and Fuzzy Relations Fuzzy If-Then Rules as Association Rules Inference from Fuzzy Rules Fuzzification and Defuzzification Learning Fuzzy Rules from Examples The Wang-Mendel Algorithm Fuzzy C-Means Clustering

Tuesday	Thursday
25th **29** **Fuzzy Systems*** Fuzzy Sets Set Operations T-norms, T-conorms, and Fuzzy complements Cylindrical Extensions and Fuzzy Relations Fuzzy If-Then Rules as Association Rules Inference from Fuzzy Rules Fuzzification and Defuzzification Learning Fuzzy Rules from Examples The Wang-Mendel Algorithm Fuzzy C-Means Clustering	27th **30** **Invited Lecture***

Notes:

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

Homework Due Dates & Exams

Friday	
Jan 13th -	1
20th -	2
27th **Homework 0 Due (not graded)**	3
Feb 3rd **Homework 1 Due**	4
10th **Homework 2 Due**	5
17th -	6
24th **Homework 3 Due**	7
Mar 3rd **Homework 4 Due**	8
10th **[Midterm 1]**	9
17th **Spring Recess**	10
24th **Homework 5 Due**	11
31st **Homework 6 Due**	12
Apr 7th **Homework 7 Due**	13
14th -	14
21st **Homework 8 Due**	15
28th **[Midterm 2]**	16

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.