USC Viterbi School of Engineering

CSCI 467: Introduction to Machine Learning (Spring 2020)

Units: 4

Instructor: Mohammad Reza Rajati, PhD
PHE 412
rajati@usc.edu – Include CSCI 467 in subject

Office Hours: Wednesday 1:30 pm –3:00 pm

Webpage: Personal Homepage at Intelligent Decision Analysis

TA(s): Soumyaroop Nandi
soumyarn@usc.edu – Include CSCI 467 in subject

Office Hours: TBD

Course Producer: Roger Lin
ruizhel@usc.edu – Include CSCI 467 in subject

Office Hours: TBD

Lecture: Monday, Wednesday, 3:30 pm –4:50 pm in VKC 156

Discussion 1: Tuesday 4:00 pm –4:50 pm in VKC 256

Discussion 2: Thursday 4:00 pm –4:50 pm in VKC 256

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

– All HWs, handouts, solutions will be posted in PDF format

Prerequisite: (CSCI 270 and MATH 225) and 1 from (EE 364 or MATH 407 or BUAD 310).

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

Tentative Grading: Programming Assignments (Labs) 35%
Problem Sets 15%
Midterm Exam 20%
Final Exam 30%
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). You may post it anonymously if you wish. Often times, if one student has a question/comment, other also have a similar question/comment. Use e-mail with the professor, TA, graders only for issues that are specific to you individually (e.g., a scheduling issue or grade issue).

Catalogue Description: Methods for building intelligent and adaptive systems from statistical analyses; theoretical understanding of such methods and the computational implications. .

Course Description: This is an introductory undergraduate course on Machine Learning with a focus on applications. The primary approach of instruction in this course is *Learing by Doing.* The focus of the course is to provide the students with basic understanding of Machine Learning algorithms and to make them use the algorithms to analyze data and convert them into information for decision-making.

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 feedforward neural networks and deep learning.

Exam Dates:

- **Midterm Exam:** Wednesday March 11, 3:30-4:50 PM.
- **Final Exam:** Friday, May 8, 2:00 PM- 4:00 PM as set by the university.

Textbooks:

- **Required Textbook:**

  1. Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, *An Introduction to Statistical Learning with Applications in R*, Springer, 2013. (ISLR)
     Available at http://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.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
     Author: 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 grade in problem sets and your lowest grade in programming assignments (Labs) will be dropped from the final grade. Lab 0 will not be graded.

- *Participation on Piazza has up to 5% extra credit, which is granted on a competetive basis *at the discretion of the instructor.*

- **Homework Policy**
  - Homework is assigned on an approximately weekly basis. A one-day grace period can be used for each homework with 10% penalty. *Absolutely no late homework will be accepted after the grace period. A late assignment results in a zero grade.*
  - Homework solutions 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 or family emergency, a signed letter from your manager or counselor or physician has to be submitted. This letter must include the contact of your physician or counselor or manager.
  - Midterm and final 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. One letter size cheat sheet (back and front) is allowed for the midterm. Two letter size cheat sheets (back and front) are allowed for the final.
  - All exams are cumulative, with considerable 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

Monday	Wednesday
Jan 13th **1** **Introduction to Statistical Learning** (ISLR Chs.1,2, ESL Chs.1,2) Motivation: Big Data Supervised vs. Unsupervised Learning	15th **2** **Introduction to Statistical Learning** (ISLR Chs.1,2, ESL Chs.1,2) Regression, Classification
20th Martin Luther King Day	22nd **3** **Introduction to Statistical Learning** (ISLR Chs.1,2, ESL Chs.1,2) The Regression Function Nearest Neighbors **Lab 0 Due (Not Graded)**
27th **4** **Introduction to Statistical Learning** (ISLR Chs.1,2, ESL Chs.1,2) Model Assessment The Bias-Variance Trade-off No Free Lunch Theorem	29th **5** **Linear Regression** (ISLR Ch.3, ESL Ch. 3) Estimating Coefficients Estimating the Accuracy of Coefficients **Lab 1 Due**
Feb 3rd **6** **Linear Regression** (ISLR Ch.3, ESL Ch. 3) Variable Selection and Hypothesis Testing Multiple Regression Analysis of Variance and the F Test	5th **7** **Linear Regression** (ISLR Ch.3, ESL Ch. 3) Stepwise Variable Selection Qualitative Variables **PS 1 Due**
10th **8** **Classification** (ISLR Ch. 4, ESL Ch. 4) Multi-class and Multi-label Classification Logistic Regression Class Imbalance Hypothesis Testing and Variable Selection	12th **9** **Classification** (ISLR Ch. 4, ESL Ch. 4) Subsampling and Upsampling SMOTE Multinomial Regression **Lab 2 Due**
17th President's Day	19th **10** **Classification** (ISLR Ch. 4, ESL Ch. 4) Bayesian Linear Discriminant Analysis **PS 2 Due**
24th **11** **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	26th **12** **Resampling Methods** (ISLR Ch. 5, ESL Ch. 7) Model Assessment Validation Set Approach Cross-Validation The Bias-Variance Trade-off for Cross-Validation **PS 3 Due**

MONDAY	WEDNESDAY
Mar 2nd **13** **Resampling Methods** (ISLR Ch. 5, ESL Ch. 7) Cross-Validation The Bootstrap Bootstrap Confidence Intervals	4th **14** **Linear Model Selection and Regularization** (ISLR Ch.6, ESL Ch. 3) Subset Selection AIC, BIC, and Adjusted R^2 **Lab 3 Due**
9th **15** **Linear Model Selection and Regularization** (ISLR Ch.6, ESL Ch. 3) Shrinkage Methods Ridge Regression	11th **16** **Midterm** **PS 4 Due**
16th Spring Recess	18th Spring Recess
23rd **17** **Linear Model Selection and Regularization** (ISLR Ch.6, ESL Ch. 3) The LASSO Elastic Net Dimension Reduction Methods*	25th **18** **Tree-based Methods** (ISLR Ch. 8, ESL Chs. 9, 10) Regression and Classification Trees **Lab 4 Due**
30th **19** **Tree-based Methods** (ISLR Ch. 8, ESL Chs. 9, 10, 16) Bagging, Random Forests, and Boosting*	Apr 1st **20** **Support Vector Machines** (ISLR Ch. 9, ESL Ch. 12) Maximal Margin Classifier Support Vector Classifiers Support Vector Machines The Kernel Trick L1 Regularized SVMs Multi-class and Multilabel Classification The Vapnik-Chervonenkis Dimension* Support Vector Regression* **PS 5 Due**
6th **21** **Neural Networks and Deep Learning** (ESL Ch. 11, DL Ch. 6) The Perceptron Feedforward Neural Networks	8th **22** **Neural Networks and Deep Learning** (ESL Ch. 11, DL Ch. 6) Feedforward Neural Networks Backpropagation and Gradient Descent Overfitting **Lab 5 Due**
13th **23** **Neural Networks and Deep Learning** (DL Chs. 6, 7) Regularization Early Stopping and Dropout	15th **24** **Neural Networks and Deep Learning** (DL Chs. 9, 10) Convolutional Neural Networks **PS 6 Due**

MONDAY	WEDNESDAY
20th **25** **Unsupervised Learning** (ISLR Ch. 10, ESL Ch. 14) K-Means Clustering Hierarchical Clustering	22nd **26** **Unsupervised Learning** (ISLR Ch. 10, ESL Ch. 14) Practical Issues in Clustering **PS 7 Due**
27th **27** **Unsupervised Learning** (ISLR Ch. 10, ESL Ch. 14) Principal Component Analysis* Anomaly Detection*	29th **28** **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 **Lab 6 Due**

Notes:

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

Statement on Academic Integrity: USC seeks to maintain an optimal learning environment. 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 these principles. SCampus, the Student Guidebook, contains the University Student Conduct Code (see University Governance, Section 11.00), while the recommended sanctions are located in Appendix A. See: `http://scampus.usc.edu`.

Emergency Preparedness/Course Continuity in a Crisis In case of a declared emergency if travel to campus is not feasible, USC executive leadership will announce an electronic way for instructors to teach students in their residence halls or homes using a combination of Blackboard, teleconferencing, and other technologies. See the university's site on Campus Safety and Emergency Preparedness: `http://preparedness.usc.edu`

Statement for Students with Disabilities: Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me (or to TA) as early in the semester as possible. DSP is located in STU 301 and is open 8:30 a.m.–5:00 p.m., Monday through Friday. Website: `http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html`

(213) 740-0776 (Phone), (213) 740-6948 (TDD only), (213) 740-8216 (FAX) ability@usc.edu.