San José State University
Department of Mathematics & Statistics
Math 251 Statistical and Machine Learning Classification (Fall 2022)

Course and Contact Information

Instructor:	Dr. Guangliang Chen
Email:	guangliang.chen@sjsu.edu
Class Days/Time:	TR 10:30-11:45am
Meeting mode:	In person (MH 234)
Office Hours:	TR 12:20-1:20pm, W 4-5pm (on Zoom: 422 306 1605) and by appointment
Piazza:	https://piazza.com/class/l6qzsfu3n7p7gi
Webpage:	http://www.sjsu.edu/faculty/guangliang.chen/math251.html
Prerequisites:	Math 164 and Math 250

Course Description

Dimensionality reduction, instance-based classification, discriminant analysis, logistic regression, support vector machine, kernel methods, ensemble learning, neural networks and deep learning, classification of nonnumeric data. 3 units.

Course Goals

- Introduce the machine learning field of classification and its applications
- Present the ideas and theory of major classification methods in the literature
- Teach how to use specialized software to perform classification tasks while adequately addressing the practical challenges (e.g., parameter tuning, memory and speed)
- Provide students with valuable first-hand experience in handling big, complex data

Course Learning Outcomes (CLO)

Upon successful completion of this course, students will be able to:

- Use the terminology associated with classification properly
- Determine the nature of various classifiers (linear or nonlinear, distribution or optimization-based, etc.)
- State clearly the mathematical formulation of each classifier
- Apply dimensionality reduction techniques to preprocess the data for classification
- Perform classification on data sets assisted with software
- Properly set the parameters associated to each classifier
- Assess the classification output through measures such as confusion matrix, training/testing error

Required Texts / Readings

There is no required textbook for this course. Some recommended readings are the following:

- Hastie, Tibshirani, and Friedman (2009), "*The Elements of Statistical Learning: Data Mining, Inference, and Prediction*", 2nd edition, Springer-Verlag. Freely available online at http://statweb.stanford.edu/~tibs/ElemStatLearn/index.html
- James, Witten, Hastie and Tibshirani (2015), "*An Introduction to Statistical Learning with Applications in R*", 6th edition, Springer. ISBN 978-1-4614-7137-0, ISBN 978-1-4614-7138-7 (eBook), DOI 10.1007/978-1-4614-7138-7. Freely available online at http://www-bcf.usc.edu/~gareth/ISL/
- Nielson (2015), "*Neural Networks and Deep Learning*", Determination Press. Freely available online at http://neuralnetworksanddeeplearning.com/
- Goodfellow, Bengio, and Courville (2016), "*Deep Learning*", MIT Press. Freely available online at http://www.deeplearningbook.org.

Technology and Equipment Requirements

The course will make intensive use of specialized software such as MATLAB and Python to perform various computing tasks. Therefore, familiarity with at least one of the programming languages is required.

Health Statement

Students registered for a College of Science (CoS) class with an in-person component should view the CoS COVID-19 and Monkeypox Training slides for updated CoS, SJSU, county, state and federal information and guidelines, and more information can be found on the SJSU Health Advisories website. By working together to follow these safety practices, we can keep our college safer. Failure to follow safety practice(s) outlined in the training, the SJSU Health Advisories website, or instructions from instructors, TAs or CoS Safety Staff may result in dismissal from CoS buildings, facilities or field sites. Updates will be implemented as changes occur (and posted to the same links).

Course Requirements and Assignments

Course requirements include regular homework assignments, a midterm exam, and a course project.

Homework will be assigned regularly in Canvas. The homework assignments will typically contain both theory and programming questions.

The course has an in-class midterm covering the conceptual and mathematical aspects of the course. More instructions will be given later in class.

You are expected to attend all meetings and actively participate in classroom discussions. Additionally, you are expected to spend at least 6 hours per week outside of class time on this course.

Final Examination or Evaluation

The course requires a project that is to be selected and completed by students in small groups (of size 2 or 3). During the semester, the students need to give an oral presentation to report their project problems. At the end of the semester, they need to report their final results. Additionally, each group needs to submit a report to present all the details.

Grading Information

For the theory questions in homework, you can write your work on paper or a tablet (in the former case you need to be able to scan your work). Once completed, submit a legible, electronic copy of your work to Canvas (as a single file attachment).

Note that it is your work (in terms of correctness, completeness, and clarity), not just your answer, that is graded. Thus, correct answers with no or poorly written supporting steps may receive very little credit.

For the programming questions in homework, you need to present your results in an organized, meaningful way, interpret them carefully, and attach the code you used to obtain the results.

Students may collaborate on homework but must write independent solutions according to their own understanding and styles. Copying and other forms of cheating will not be tolerated and may result in a failing grade for the course, combined with appropriate disciplinary actions from the university.

You must submit homework on time to receive full credit. Late submissions within 24 hours of the due time can still be accepted but will receive a penalty of 10% of the total grade. Submissions that are late for more than one day (24 hours) will not be accepted for any reason.

No make-up exam will be given if you miss the sole midterm exam unless you have a legitimate excuse (such as illness or other personal emergencies) and can provide documented evidence.

The course project requires two presentations, the first on the data set and the second on results, as well as a comprehensive report. Both the presentations and report will be graded based on clarity, depth, accuracy, and completeness. More details are given in class.

The weights in determining the semester total are:

- Homework: 30%
- Midterm exam: 35%
- Project: 35% (10% = 5%+5% oral, 25% report)

Course grades will be assigned by combining the following cutoffs and the actual class distribution (I reserve the right to slightly adjust these percentages in the end):

Grade	*Percentage*	*Grade*	*Percentage*	*Grade*	*Percentage*	*Grade*	*Percentage*
A plus	*97 to 100%*	*B plus*	*86 to 89%*	*C plus*	*73 to 75%*		
A	*93 to 96%*	*B*	*80 to 85%*	*C*	*68 to 72%*	*D*	*60 to 64%*
A minus	*90 to 92%*	*B minus*	*76 to 79%*	*C minus*	*65 to 67%*	*F*	*0 to 59%*

Academic Dishonesty

Students who are suspected of cheating during an exam will be referred to the Student Conduct and Ethical Development office and depending on the severity of the conduct, will receive a zero on the assignment or a grade of F in the course.

University Policies

Per University Policy S16-9 *(http://www.sjsu.edu/senate/docs/S16-9.pdf)*, relevant university policy concerning all courses, such as student responsibilities, academic integrity, accommodations, dropping and adding, consent for recording of class, etc. and available student services (e.g. learning assistance, counseling, and other resources) are listed on Syllabus Information web page (http://www.sjsu.edu/gup/syllabusinfo), which is hosted by the Office of Undergraduate Education. Make sure to visit this page to review and be aware of these university policies and resources.

Fall 2022 Approximate Course Schedule

This schedule is subject to change, with fair notice made in class, if there is a delay in progress.

Class #	Date		Topics
1	AUG 23	T	Course introduction
2	25	R	kNN classification I
3	30	T	kNN classification II
4	SEP 1	R	Dimensionality reduction for classification
5	6	T	Evaluation criteria for classifiers
6	8	R	Bayes classifiers I
7	13	T	Bayes classifiers II
8	15	R	Bayes classifiers III
9	20	T	Logistic regression I
10	22	R	Logistic regression II
11	27	T	Logistic regression III
12	29	R	Logistic regression IV
13	OCT 4	T	Project presentations: Data sets
14	6	R	Support vector machine I
15	11	T	Support vector machine II
16	13	R	Support vector machine III
17	18	T	Kernel methods
18	20	R	Review
19	25	T	Midterm exam
20	27	R	Classification trees
21	NOV 1	T	Ensemble learning I
22	3	R	Ensemble learning II
23	8	T	Ensemble learning III
24	10	R	Neural networks I
25	15	T	Neural networks II
26	17	R	Neural networks III
27	22	T	Neural networks IV
28	29	T	Introduction to deep learning I
29	DEC 1	R	Introduction to deep learning II
30	6	T	Last class
Final exam day	13	T	Project presentations: Results (9:45am-12pm)