DSCI6003: Machine Learning and Data Analysis

Instructors:

Imran Younus : imran@galvanize.com
Mike Tamir : mtamir@galvanize.com
Jonathan Dinu : jonathan@galvanize.com (Guest Lecture)
Jared Thompson : jared.thompson@galvanize.com

Class Location: 44 Tehama St, 3rd Floor, gU Classroom
Class Time: 9:30 - 10:50 PM, M-T-W-Th
Lab Time: 10:50 - 11:50 PM, M-T-W-Th
Office Hours: Friday by Appointment

Description of the Course

Essential elements of Machine Learning, with a focused introduction to core supervised and unsupervised learning algorithms, statistical modeling, and key best practice techniques for building well trained models. Designed with coding lab practice to develop implementation skills.

Supplementary Materials:

There is no textbook for this course. Some books may come in handy, however:

- *An Introduction to Statistical Learning*, by Gareth James *et. al.*, (Available to download from http://www-bcf.usc.edu/~gareth/ISL/)
- *Machine Learning: A Probabilistic Perspective,* by Kevin Murphy
- *Machine Learning in Action*, by Peter Harrington

Course Requirements

Attendance

Students are expected to be present and on time for all class meetings.

Class Participation

You will learn more easily and enjoyably if you actively participate. Student contribution to class discussions is highly valued and is critical to the learning process. Students will be asked to participate in class activities designed to encourage open conversation about and involvement in course material.

Lab Exercises

Participation in and completion of lab exercises is a requirement for this course. Each unit includes exercises to provide practice applying techniques discussed in class and to reveal deficiencies in understanding in preparation for skills tests.

Evaluative Exams

There will be three mid-term skills test evaluative to demonstrate comprehension of unit material as well as a two-hour comprehensive final exam at the end of the term.

There will be no make-up exams without proper documentation approved by the administration.

Late assignments

All assignments are due *at the beginning* of class on the due date. If you are late to class for any reason, you will miss out on what is covered in the beginning of class, and disrupt classmates when you come in. Take whatever measures are necessary to ensure that you will be ready to turn your assignment in on time (*e.g.*, backup all your work on a memory stick, *etc.*) *1/2 of the points for the assignment will be deducted each calendar day the assignment is not handed in—regardless of the reason.*

Academic Integrity

As per the University's Academic Integrity Policy and Procedures: > The University expects that all students, graduate and undergraduate, will learn in an environment where they work independently in the pursuit of knowledge, conduct themselves in an honest and ethical manner and respect the intellectual work of others. Each member of the University community has a responsibility to be familiar with the definitions contained in, and adhere to, the Academic Integrity Policy. Students are expected to be honest in their academic work.

Violations of the Academic Integrity Policy include (but are not limited to):

1. **Cheating** -- *i.e.* Don't read off of your neighbors exams
2. **Collusion** -- Group work is encouraged *except on evaluative exams*. When working together (on exercises, *etc.*) acknowledgement of collaboration is required.
3. **Plagiarism** -- Reusing code presented in labs and lectures is expected, but copying someone else's solution to a problem is a form of plagiarism (even if you change the formatting or variable names).

Students who are dishonest in any class assignment or exam will receive an "F" in this course.

Grading

Type	Number	Points for each	Total Points
Python Exercises	7	3	21
Skills Tests	3	18	54

Final Project	1	25	25
TOTAL			100

Passing grades are greater than or equal to 70 points.

Tentative Schedule

(*subject to change*)

Week	Day	Topic	Assessment
1	1	Introduction to ML	
1	2	K Nearest Neighbours Classifiers	
1	3	Dimensionality Reduction	
1	4	Gradient Descent	
2	1	Naive Bayes Classifier	Exercise Set 1 Due
2	2	Cross Validation	
2	3	Logistic Regression	
2	4	Review for skills test	
3	1	Regularization	Exercise Set 2 Due Skills Test 1 on Lessons 1.1-2.2
3	2	Support Vector Machines	
3	3	Evaluation Metrics	
3	4	Tuning Best Practices	
4	1	Decision Trees	Exercise Set 3 Due
4	2	Random Forests	
4	3	Random Forest Cont.	
4	4	Review for skills test	
5	1	Bagging	Exercise Set 4 Due. Skills Test 2 on Lessons 3.1 through 4.4
5	2	Boosting	
5	3	Clustering Algorithms	
5	4	Methods for Evaluating Clustering Algorithms	
6	1	Generalized Methods of Moments and Expectation Maximization	Exercise Set 5 Due

6	2	Proximity Clustering	
6	3	Gap Statistic	
6	4	Review for skills test	
7	1	Forward and Backward Selection	Exercise Set 6 Due. Skills Test 3 on Lessons 5.1 through 6.3
7	2	Bias and Variance Tradeoff	
7	3	Recommendation Engines 1	
7	4	Recommendation Engines 2	
8	1	**review for final prep**	Exercise Set 7 Due
8	2	**Final TBD**	