Math 407: Statistical Machine Learning, Spring 2019

Instructor: Dr. Rosanna Overholser, 178 Boivin Hall, 885-1475, *rosanna.overholser@oit.edu*

Web Page: *http://math.oit.edu/~overholserr*. Please check this site frequently for links to the course materials, daily calendar and homework assignments. Grades will be posted in Blackboard.

Office Hours: 4-4:50 MWF, 11-11:50 Th., and by appointment.

Classroom: Lecture on MWTh in Owens 217, computer lab on F in Boivin Hall 126

Prerequisites: (MATH 361 and MATH 252) or (MATH 465)

Credits: 4

Please check with your advisor to see if this course can be counted as an upper division math elective for your major. Please consider using this course towards a math or statistics minor.

Course Description:

An introduction to machine learning with an emphasis on statistical theory. Supervised (discriminative and generative models) and unsupervised learning for categorical and numerical outcomes. Model selection and assessment.

Modeling techniques covered include multiple linear regression, logistic regression, linear and quadratic discriminant analysis, k-nearest neighbors, shrinkage methods, additive models, tree-based methods, support vector machines, neural networks and ensemble learning. Dimension reduction techniques for dealing with high dimensional data. Applications will be based on clean "textbook" datasets as well as some messier "real" datasets.

Course Outcomes:

After completing this course, students will be able to

- Describe the assumptions behind a variety of machine learning algorithms,
- Derive or simulate the bias and variance of a machine learning algorithm
- Select appropriate machine learning methods based on the type of data available and the needs of the anticipated user,
- Clean and transform data as needed for training,
- Train and assess the performance of a particular model,
- Select the "best" model from several trained models while guarding against overfitting,
- Use a model to make a new prediction.

Textbooks:

We will primarily use the first book listed below, occasionally diving into the second for a more in-depth look at an algorithm. Both books are available for free online.

- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*. Vol. 112. New York: springer, 2013.
- Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*. Vol. 1. New York: Springer series in statistics, 2001.

Software:

You may choose either R or Python 3 to 1) clean, visualize, and summarize datasets, 2) fit models, and 3) create tables and figures for presentation of results. Both these softwares are available for free online. Our main textbook has labs at the end of each chapter with R code, but these labs are also available online in Python.

Grading:

- Weekly computer labs and homework assignments - 35%
- Midterm (week 6) – 20 %
- Team Project (week 10) – 25%
- Final Exam – 20 %

Other Things of Importance:

Incomplete Grades: An incomplete grade can only be assigned to you under the following circumstances:

1. You have/had a grade of 70% or better (including zeros for any work not done) by the date to withdraw with a W.

2. You have a SERIOUS problem that begins after the withdraw date and prevents you from being able to complete the term.

Illness: In the event that you are ill, please stay home and rest! You can always read the lecture notes online afterwards or ask a classmate to fill you in.

Disabilities: Students with a documented disability who require assistance or academic accommodations should contact the office of Disability Services immediately to discuss eligibility. Disability Services staff are located on both the Klamath Falls and Wilsonville campuses, however arrangements can be made to meet with a student on any campus. Meetings are by appointment only, so please contact the Disability Services office at the campus closest to you: Klamath Falls (541) 885-1790 and Portland-Metro (503) 821-1305. Specific information and Disability Services forms can be found at *www.oit.edu*, then go to "Academics" and click on "Student Success Center" and then "Disability Services." This link leads to the department's website: *http://www.oit.edu/academics/ssc/disability-services*.

Calendar: Below are some important dates for the term.

    April 1th - Classes begin
    April 12th - Last day to withdraw without a ``W"
    May 6th - Midterm
    May 17th - Last day to withdraw with a ``W" from this course
    May 27th – Holiday
    June 7th – Project Due
    Monday, June 10th, 2-4pm - Final Exam (in our regular room, OW217)

Tentative Schedule:

Wk	Topic	Ch.	Lab
1	Intro to statistical learning	ISL 2	Intro to python/R for statistical applications
2	Linear regression, K nearest neighbors	ISL 3	Linear regression
3	Classification methods	ISL 4	Logistic regression, LDA, QDA, K nearest neighbors
4	Model assessment and selection	ISL 5	Test and training sets, cross validation and bootstrap
5	Techniques for high dimensional data	ISL 6	Lasso and ridge regression, principal components and partial least squares
6	Non-linear models	ISL 7	Smoothing splines, generalized additive models
7	Tree based methods	ISL 8	Decision trees, random forests
8	Support vector machines	ISL 9	Support vector machines
9	Neural networks	ESLII 11	Intro to Keras and Tensorflow
10	Ensemble learning	ESLII 16	Ensemble learning