Course Outline

STAT 841 / 441, CM 763
Statistical Learning-Classification

Fall 2018

Instructor:
Ali Ghodsi
Dept. of Statistics & Actuarial Science
University of Waterloo
Office: M3 4208
E-mail: `ali.ghodsi@uwaterloo.ca`

Prerequisites:
Grads: none for STATS/CS/ECE/SYDE grad students , instructor permission otherwise
Undergrads: STAT 341 or (STAT 330 and 340)

Course Description: Classification, also known as **pattern recognition** is the problem of predicting a discrete random variable Y from another random variable X. The random variable X may take many different forms from Digital image libraries and text corpora to gene expression microarrays and financial time series.

This course provides a comprehensive introduction to the problem of classification and pattern recognition and reflects recent developments in the filed.

Required Textbook: There is no required textbook for the class. Three recommended books that cover the similar material are:

- Hastie, Tibshirani, Friedman *Elements of Statistical Learning*

- Bishop, *Pattern Recognition and Machine Learning.*

- Murphy, *Machine Learning: a Probabilistic Perspective*

Tentative topics::

- Feature selection

- Feature extraction (dimensionality reduction)

- Error rates and the Bayes classifier

- Gaussian and linear classifier

- Linear regression and logistic regression

- Feed-forward Neural networks

- Radial basis function networks

- Naive Bayes

- Trees

- Assessing error rates and model selection

- Support vector machines

- Kernel methods

- k-nearest neighbors

- Introduction to Deep learning

- Convolutional neural networks

- Bagging

- Boosting

- Semi-supervised learning for classification

- Metric learning for classification

Evaluation:(tentative)

Assignments	40%	(Four Assignments)
Data challenge	20%	(Two Data challenges)
Paper presentation	10%	
Final project	30%	

Paper presentation:

Paper presentation starts on November 13 (tentative) You need to write a summary/critique of the paper that you would like to present on wikicoursenote.com (https://goo.gl/ZmthQN) by exactly one week before your presentation date.

attending all presentations is mandatory.

Project:

Final group project (presentation and reports up to 4 pages of PDF plus 1 page reference) are worth 30% of your final grade .

Your final project can be:

- A Kaggle completion. You may choose an *active* competition from *featured* or *research* categories. Kaggle Competitions in other categories (*in class, getting start* or *playground*) are not eligible for the final project.

- Develop a new algorithm. In this case, you will need to demonstrate (theoretically and/or empirically) why your technique is better (or worse) than other algorithms. (Note: A negative result does not lose marks, as long as you followed proper theoretical and/or experimental techniques).

- Application of classification to some domain. This could either be your own research problem, or you could try reproducing results of someone else's paper.

Note that you cannot borrow part of an existing thesis work, nor can you re-use a project from another course as your final project. Final project reports will be checked by Turnitin (Plagiarism detection software).

Final project presentation:

You will present your final project as a poster on Thursday, December 13th. Poster presentations will take place on this date from 9:00 AM to 4:00 PM at M3 Atrium.

Acceptable forms for the poster is full poster or slides printed on A4 /letter size papers.

In addition to the poster, you will have to turn in a four-page single-spaced, single-column report plus one page reference. This is due both in hard copy and electronically (PDF) by Thursday, December 13th.

Communication

All communication should take place using the *Piazza* discussion board.

Piazza is a good way to discuss and ask questions about the course materials, including assignments, in a public forum. It enables you to learn from the questions of others, and to avoid asking questions that have already been asked and answered. It also provides a forum for course personnel to make announcements and clarifications about assignments and other course-related topics. Students are expected to read Piazza on a regular basis.

Enrolling in Piazza

You will be sent an invitation to your UW email address. It will include a link to a web page where you may complete the enrollment process.

Piazza Guidelines

Here are some guidelines that you should keep in mind when posting items to Piazza:

1. Please remember that everything you post is public - everyone enrolled in this course will be reading it. As a result, in any posts you make, do not give away any details on how to do any of the assignments / data challenges. This could be construed as cheating, and you will be responsible as the poster. If you have questions about an assignment / data challenge that require you give specific details of your solution, you may still post to Piazza, but check *This is a private post - only visible to class instructors* (and TAs). If the instructors and/or TAs feels that posting it to everyone is appropriate, they will do so.

2. Keep posts related to the course, concise, and topical. As students are all expected to read Piazza on a regular basis, try not to waste the time of readers.

3. Please be diligent about attempting to find the answer before you post a question. Piazza includes excellent search facilities – use them! Scan all of the questions that have already been asked. Better yet, read them along with the answers. You'll learn lots! Please do all you can to avoid duplicates.

4. Make it easy for other students to find your question – just in case they have the same question and want to see the answer.

   - Use a meaningful subject heading. "Help" and even "Help for A3Q2" is not very meaningful. "Clarify parameter order for A3Q2" is much better.
   - Tag your post with all the applicable tags. Start a tag by typing the hash character (#). A drop-down list of tags that are currently in use will appear. Use one of them, if applicable. If not, create a new one. However, any tag you create should be applicable to many posts not just yours.

5. Please don't post things to the group that provide no useful information to readers. Posts like "I have the same question as this one just posted", or "I agree with this comment" serve no useful purpose, and waste people's time.

6. Keep complaints about the course out of Piazza or mark them with the *This is a private post - only visible to class instructors* checkbox. If you have a concern about anything to do with the course, the best way to deal with it, and to get results, is to take it to the course instructor. Piazza is not a complaint forum.

Assignments and grades will be handled through *Learn*. Please log on frequently to Piazza and Learn. You are responsible for being aware of all STAT 441 / 841 material, information and email messages found on *Learn* and *Piazza* throughout the semester.

Important Dates:

Assignments and Data challenges due dates:

Assignment 1	Sept 28, 1 pm
Assignment 2	Oct 12, 1pm
Data Challenge 1	Oct 26, 1pm
Assignment 3	Nov 9, 1pm
Assignment 4	Nov 23, 1pm
Data Challenge 2	Dec 3, 1pm

Proposal, Presentations, and Final project due dates:

Oct 7: Proposal
Form groups with up to four people and register your group here: https://goo.gl/kQPdZn

Nov 13: Paper Presentations begin (tentative) You need to write a summary/critique of the paper that you would like to present on wikicoursenote.com https://goo.gl/ZmthQN by exactly one week before your presentation date.

Dec 13: Final project

You will present your final project as a poster on Wednesday, December 13th. Acceptable forms for the poster is full poster or slides printed on A4 /letter size papers. In addition to poster presentation, the final project report (four-page paper, single-column, single-spaced plus one page reference) is due Dec 13th. You need to upload the electronic version to Turnitin Dropbox on Learn. and Hand in the hard copy of your project to me on the same day.

Academic Honesty: In all assignments, data challenges, and projects , if you use ideas, codes, plots, text and other intellectual property developed by someone else you have to cite the original source.

If you copy a sentence or a paragraph from work done by someone else, in addition of citing the original source you have to use quotation marks to identify the scope of the copied material.

Example: Plagiarism is an act of 'using ideas, plots, text and other intellectual property developed by someone else while claiming it is your original work.' [1]

Evidence of copying or plagiarism will cause a failing mark in the course.

Persons with Disabilities: The office for Persons with Disabilities (OPD), located in Needles Hall, Room 1132 collaborates with all academic departments to arrange appropriate accommodations for students with disabilities without compromising the academic integrity of the curriculum. If you require academic accommodations to lessen the impact of your disability, please register with OPD at the start of each academic term.

References

[1] Tec Encyclopedia. http://www.answers.com/topic/plagiarism