

**Course Outline**  
**STAT 841 / 441, CM 763**  
**Statistical Learning-Classification**  
**Fall 2015**

**Instructor:**

Ali Ghodsi  
Dept. of Statistics & Actuarial Science  
University of Waterloo  
Office: M3 4208  
E-mail: aghodsib@uwaterloo.ca  
Office hours: 4:00-5:00 T

**Lectures:** (11:30-12:50TTh, RCH 207)

**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 field.

**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
- Neural networks
- Radial basis function networks
- Naive Bayes
- Trees
- Assessing error rates and model selection
- Support vector machines
- Kernel methods
- k-nearest neighbors
- Deep learning
- Bagging
- Boosting
- Semi-supervised learning for classification
- Metric learning for classification

## **Evaluation:(tentative)**

Assignments	50%	(4 or 5 assignments)
Final project	50%	(10% Presentation) (40% Ranking and report)

## **Project:**

Final group project (presentation and reports up to 7 pages of PDF) are worth 50% of your final grade . You are encouraged to participate in the *Right Whale Recognition* kaggle competition as your final project. If you don't have access to adequate computational resources, you may chose other possible types of projects as follows:

- Another active kaggle completion.
- 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).

## **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. This could be construed as cheating, and you will be responsible as the poster. If you have questions about an assignment 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 341 material, information and email messages found on *Learn* and *Piazza* throughout the semester.

## **Important Dates:**

Oct 6 Final project proposal due (Use the link posted on Leran)  
Nov 17 Presentation begin (tentative)

**Academic Honesty:** In assignments, projects and wikicoursenote, if you use ideas, 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>