CSE 455/555 Introduction to Pattern Recognition

SUNY at Buffalo

Syllabus for Spring 2013

Last updated: 3 Jan 2013

Instructor: Jason Corso (UBIT: jcorso)

Course Webpage: http://www.cse.buffalo.edu/~jcorso/t/CSE555 or http://www.cse.buffalo.edu/~jcorso/t/CSE455 but this is just a link to the first one.

Syllabus: http://www.cse.buffalo.edu/~jcorso/t/CSE555/files/syllabus.pdf.

Meeting Times: TR 11:00-12:20

Location: Davis 101

Teaching Assistants: David Johnson (UBIT: davidjoh)

Office Hours:

- Instructor: TR 12:20-2:00 (Davis 332)
- TA: TBD Check Course Webpage (Note also there are recitations this term.)

Final Exam: Friday 3 May 2013, 11:45–2:45 in Davis 101.

Mailing List: cse555-jcorso-list@listserv.buffalo.edu for both 455 and 555 students.

Student Updates: All updates will be posted to the course website and sent to the mailing list. There is no other official course announcement mechanism, such as a newgroup, blog, piazza or otherwise.

A Note On Contacting The Instructor: You are encouraged to contact the instructor or TA via the course-wide mailing list via email. You **must** 1) send the email from your UBIT account (since this is the approved one on the mailing list) and 2) include `[CSE555]` at the beginning of the command-line (even if you are in CSE455). Email that does not follow these conventions will not be read.

Main Course Material

Course Catalog Description: Foundations of pattern recognition algorithms and machines, including statistical and structural methods. Data structures for pattern representation, feature discovery and selection, classification vs. description, parametric and non-parametric classification, supervised and unsupervised learning, use of contextual evidence, clustering, recognition with strings, and small sample-size problems.

Prerequisites: It is assumed the students have a working knowledge of calculus, linear algebra, and probability theory. It is also assumed the students have some experience programming in a scientific computing environment.

Course Goals: After taking the course, the student should have a clear understanding of 1) the design and construction and a pattern recognition system and 2) the major approaches in statistical and syntactic pattern recognition. The student should also have some exposure to the theoretical issues involved in pattern recognition system design such as the curse of dimensionality. Finally, the student will have a clear working knowledge of implementing pattern recognition techniques and the scientific Python computing environment. These goals are evaluated through the course project, homeworks, and exams.

Textbooks: The main (required) textbook for the course is

- Duda, R.O., Hart, P.E., and Stork, D.G. *Pattern Classification*. Wiley-Interscience. 2nd Edition. 2001.

The textbook has a website: http://www.rii.ricoh.com/~stork/DHS.html.

Recommended supplemental textbooks are

- Bishop, C. M. *Pattern Recognition and Machine Learning*. Springer. 2007.
- Marsland, S. *Machine Learning: An Algorithmic Perspective*. CRC Press. 2009. (Also uses Python.)
- Theodoridis, S. and Koutroumbas, K. *Pattern Recognition*. Edition 4. Academic Press, 2008.
- Russell, S. and Norvig, N. *Artificial Intelligence: A Modern Approach*. Prentice Hall Series in Artificial Intelligence. 2003.
- Bishop, C. M. *Neural Networks for Pattern Recognition*. Oxford University Press. 1995.
- Hastie, T., Tibshirani, R. and Friedman, J. *The Elements of Statistical Learning*. Springer. 2001.
- Koller, D. and Friedman, N. *Probabilistic Graphical Models*. MIT Press. 2009.

In addition, students are strongly suggested to supplement the textbook and local lectures with the online materials in related courses, such as Ng's coursera machine learning course, https://www.coursera.org/course/ml (or in iTunes U), and other materials on videolectures.net. The website will be augmented with links to related lectures when available.

Course Work

The course work in this offering is quite different than the past few offerings from Prof. Corso. Based on the past four times this course has been offered and feedback thereof, there will be **NO** mandatory graded homeworks or projects. Instead, a homework assignment will be posted for each lecture/chapter with relevant problems (both analytical and programming oriented); solutions will also be posted for all problems. Programming problems will have a mix of Matlab and Python questions (based on what the instructor have available).

Despite not directly impacting the students' bottom line, they are strongly encouraged to work through the assignments, both analytical and programming problems, to deepen their understanding of the material.

Course Project: There is no course project. See above.

Programming Environment: Matlab or Scientific Python: Lecture and homework material will be given in matlab and python.

Students are encouraged to learn and use Python (i.e., SciPy, NumPy) in the course. Many programming materials given in lecture and programming aspects of the homeworks will be given in Python. No introduction to scientific Python will be given in the course, and it is the students' responsibility to get up to speed; hand-holding stopped when you signed up for this course. Additional python resources will be maintained at http://www.cse.buffalo.edu/~jcorso/t/CSE555/python_resources.html.

To allow for a common Python environment, the course will officially rely on the Enthough Python Distribution (EPD) http://www.enthought.com/products/epd.php, which is easy to get, free, and includes the packages needed for our material. The course will use EPD version 7.3. Students are encouraged to install it on their own computers, and it is also installed on the CSE network (see https://wiki.cse.buffalo.edu/services/content/enthought-python-distribution for more information).

Turning in Assignments: Paperless: No assignments will be turned in.

Course Evaluation

The following is a description of how students will be evaluated in the course. The instructor reserves the right to make minor adjustments as necessary.

Quizzes:

In place of the graded homeworks and a mid-term exam, the course will have *weekly* quizzes. Each quiz will have two problems on it: (1) a rote recall question, such as *What is the basic assumption on the distribution of the data in k-Means clustering?*, worth 2 points and (2) a more involved problem asking you to derive a formula, execute an algorithm, or propose a solution to a problem, worth 8 points.

Quizzes will be given weekly on either Tuesday or Thursday; you will not know in advance when the quiz will be given. In other words, you need to be in class. There is no chance to get higher than an F if you miss all or most of the quizzes, see below.

There will be 14 quizzes given throughout the semester. The lowest 2 will be dropped.

Grading Rubric:

A final percentage score will be calculated as a weighted average of the course work according to the following table:

- Quizzes (60%)
- Mid-Term Exam (0%) *there is no mid-term*
- Final Exam (40%)
- Homeworks (0%) *homeworks are not graded*
- Project (0%) *there is no project*

Letter grades will be given in the range of F to A (with minuses and pluses). Mapping of raw percentage scores to letter grades will be based on the following rubric: Letter grade A is given for raw percentage scores of 85 and higher for 555 and 80 and higher for 455. Remaining letter scores are graded based on a clustering of the students output with each cluster mean mapped to a letter grade in decreasing order (essentially, this means graded on a curve); this is based on overall class performance.

Distinctions 455 and 555 grading: 455 and 555 will be graded on separate curves and the mapping to grade A is different. In addition, 455 students will be required to solve fewer problems on the final exam.

Late Work and Missed Exam Policy: There is no work to turn in.

12 of 14 quizzes will count toward the grade. If more than 2 quizzes are missed, then zero grades will be averaged in. No make-up quizzes will be given.

Similarly, the date of the final exam is known far in advance (see above). Do not miss the exam. **No** make-up exams will be given other than for those University approved reasons. This is a firm policy. Do not expect special treatment.

Regrading: If you have a question about the grading of any piece of work, first consult with the teaching assistant who graded your work. If you cannot resolve your questions with the teaching assistant, you should consult with the instructor of the course.

Any questions about the grading of a piece of work must be raised within one week of the date that the work was returned by the teaching assistant or the instructor. In other words, if you do not pick up your work in a timely fashion, you may forfeit your right to question the grading of your work.

Incomplete ("I") Grades: Generally, incomplete ("I") grades are not given. However, very rarely, circumstances truly beyond a student's control prevents him or her from completing work in the course. In such cases the instructor can give a grade of "I." The student will be given instructions and a deadline for completing the work, usually no more than 30 days past the end of the semester. University and department policy dictate that "I" grades can be given only if the following conditions are met:

- An Incomplete will only be given for missing a small part of the course.
- An Incomplete will only be given when the student misses work due to circumstances beyond his/her control.
- An Incomplete will only be given when the student is passing the course except for the missed material.
- An Incomplete is to be made up with the original course instructor within the time specied by the appropriate University regulation (see appropriate document above), and usually within the following semester.
- An Incomplete will not be given to allow the student to informally retake the entire course, and have that grade count as the grade of the original course.

Incompletes can not be given as a shelter from poor grades. It is your responsibility to make a timely resignation from the course if you are doing poorly for any reason. The last day to resign the course is Friday, March 29 2013.

Course Outline

The following is the list of topics we will cover this semester. The selection of topics has been made to provide the student with both a fair sampling and an indepth, useful know-how of the big field of pattern recognition. This has required that we drop some topics completely (e.g., Neural Networks) to allow for more indepth discussion of other topics (e.g., Dimension Reduction). As many topics as possible will be grounded with real-world problems and data, and they will be presented both in terms of the mathematical theory as well as the algorithmic and programming aspects.

A calendar will be maintained on the course website and updated as the semester proceeds. This outline may change to adapt to interest and progress (or lack thereof). The flow of topics is also a different this term than previous offerings by Prof. Corso; the changes are based on feedback received from students and are in the interest of optimizing the effectiveness and interest of the course.

1. Introduction to Pattern Recognition
2. Tree Classifiers — *Getting our feet wet with real classifiers*
   (a) Decision Trees: CART, C4.5, ID3.
   (b) Random Forests
3. Bayesian Decision Theory — *Grounding our inquiry*
4. Linear Discriminants — *Discriminative Classifiers: the Decision Boundary*
   (a) Separability
   (b) Perceptrons
   (c) Support Vector Machines
5. Parametric Techniques — *Generative Methods grounded in Bayesian Decision Theory*
   (a) Maximum Likelihood Estimation
   (b) Bayesian Parameter Estimation
   (c) Sufficient Statistics
6. Non-Parametric Techniques
   (a) Kernel Density Estimators
   (b) Parzen Window
   (c) Nearest Neighbor Methods
7. Unsupervised Methods — *Exploring the Data for Latent Structure*
   (a) Component Analysis and Dimension Reduction
      i. The Curse of Dimensionality
      ii. Principal Component Analysis
      iii. Fisher Linear Discriminant
      iv. Locally Linear Embedding
   (b) Clustering
      i. K-Means
      ii. Expectation Maximization
      iii. Mean Shift
8. Classifier Ensembles
   (a) Bagging
   (b) Boosting / AdaBoost
9. Graphical Models — *The Modern Language of Pattern Recognition and Machine Learning*
   (a) Introductory ideas and relation back to earlier topics
   (b) Bayesian Networks
   (c) Sequential Models
      i. State-Space Models
      ii. Hidden Markov Models
      iii. Dynamic Bayesian Networks
10. Algorithm Independent Topics — *Theoretical Treatments in the Context of Learned Tools*
    (a) No Free Lunch Theorem
    (b) Ugly Duckling Theorem
    (c) Bias-Variance Dilemma
    (d) Jacknife and Bootstrap Methods
11. Other Items Time Permitting
    (a) Syntactic Methods
    (b) Neural Networks

Additional Information

Newsgroup: There is a newsgroup, sunyab.cse.555, for this course. You must learn how to read news and subscribe to this newsgroup. You are expected to read the newsgroup on a daily basis. There will often be important material posted there, such as supplementary course notes, homework and sample exam questions, and occasionally late breaking news. You may post general course related articles to the newsgroup. Use discretion in posting articles related to homework assignments and the project: when in doubt, e-mail the TA or instructor first.

All 455 students should use the 555 newsgroup as well.

The news (nntp) server you need to connect to is news.buffalo.edu. Note that you must authenticate using your UBIT name and password to use this news server, and you must be connecting from a UB IP address (i.e. if you are not using a university machine, you need to use VPN). For further information on accessing the newsgroup, refer to http://ubit.buffalo.edu/newsgroups/index.php.

Similar Courses at This and Other Institutions: (incomplete and in no important order)

- Professor Sargur Srihari at UB. http://www.cedar.buffalo.edu/~srihari/CSE555/index.html
- Professor Song-Chun Zhu at UCLA. http://www.stat.ucla.edu/~sczhu/Courses/UCLA/Stat_231/Stat_231.html
- Professor Anil Jain at MSU. http://www.cse.msu.edu/~cse802
- Professor Yann Lecun at NYU. http://www.cs.nyu.edu/~yann/2007f-G22-2565-001/index.html

General Notes

If you don't understand something covered in class, ask about it right away. The only silly question is the one which is not asked. If you get a poor mark on an assignment or exam, find out why right away. Don't wait a month before asking. The instructor and teaching assistant are available to answer your questions. Don't be afraid to ask questions, or to approach the instructor or TA in class, during office hours, through the newsgroup or through e-mail. This course is intended to be hard work, but it is also intended to be interesting and fun. We think pattern recognition is interesting and exciting, and we want to convince you of this.

Disabilities

If you have a diagnosed disability (physical, learning, or psychological) that will make it difficult for you to carry out the course work as outlined, or that requires accommodations such as recruiting note-takers, readers, or extended time on exams or assignments, you must consult with the Office of Disability Services (25 Capen Hall, Tel: 645-2608, TTY: 645-2616, Fax: 645-3116, http://www.student-affairs.buffalo.edu/ods/). You must advise your instructor during the rst two weeks of the course so that we may review possible arrangements for reasonable accommodations.

Counseling Center

Your attention is called to the Counseling Center (645-2720), 120 Richmond Quad. The Counseling Center staff are trained to help you deal with a wide range of issues, including how to study effectively and how to deal with exam-related stress. Services are free and condential. Their web site is http://www.student-affairs.buffalo.edu/shs/ccenter/.

Distractions In The Classroom - Behavioral Expectations

The following is the text of a policy adopted by the Faculty Senate on 5/2/2000. You are expected to know and adhere to this policy.

OBSTRUCTION OR DISRUPTION IN THE CLASSROOM - POLICIES
UNIVERSITY AT BUFFALO

To prevent and respond to distracting behavior faculty should clarify standards for the conduct of class, either in the syllabus, or by referencing the expectations cited in the Student Conduct Regulations. Classroom "etiquette" expectations should include:

- Attending classes and paying attention. Do not ask an instructor in class to go over material you missed by skipping a class or not concentrating.
- Not coming to class late or leaving early. If you must enter a class late, do so quietly and do not disrupt the class by walking between the class and the instructor. Do not leave class unless it is an absolute necessity.
- Not talking with other classmates while the instructor or another student is speaking. If you have a question or a comment, please raise your hand, rather than starting a conversation about it with your neighbor.
- Showing respect and concern for others by not monopolizing class discussion. Allow others time to give their input and ask questions. Do not stray from the topic of class discussion.
- Not eating and drinking during class time.
- Turning off the electronics: cell phones, pagers, and beeper watches.
- Avoiding audible and visible signs of restlessness. These are both rude and disruptive to the rest of the class.
- Focusing on class material during class time. Sleeping, talking to others, doing work for another class, reading the newspaper, checking email, and exploring the internet are unacceptable and can be disruptive.
- Not packing bookbags or backpacks to leave until the instructor has dismissed class.

Academic Integrity

A zero-tolerance policy on cheating will be adopted in this course. The following is the formal statement of academic integrity. Source: http://www.cse.buffalo.edu/graduate/policies_acad_integrity.php

The academic degrees and the research findings produced by our Department are worth no more than the integrity of the process by which they are gained. If we do not maintain reliably high standards of ethics and integrity in our work and our relationships, we have nothing of value to offer one another or to offer the larger community outside this Department, whether potential employers or fellow scholars.

For this reason, the principles of Academic Integrity have priority over every other consideration in every aspect of our departmental life, and we will defend these principles vigorously. It is essential that every student be fully aware of these principles, what the procedures are by which possible violations are investigated and adjudicated, and what the punishments for these violations are. Wherever they are suspected, potential violations will be investigated and determinations of fact sought. In short, breaches of Academic Integrity will not be tolerated.

University Statements on Academic Integrity

The University at Buffalo Department of Computer Science and Engineering endorses and adheres to the University policy on Academic Integrity. Students should be familiar with that policy, as expressed in the following documents.:

- UB Office of Judicial Affairs statement on Academic Dishonesty. http://www.ub-judiciary.buffalo.edu/art3a.shtml#integrity
- UB Undergraduate Catalog statement on Academic Integrity. http://undergrad-catalog.buffalo.edu/policies/course/integrity.shtml

Departmental Statement on Academic Integrity in Coding Assignments and Projects

The following statement further describes the specific application of these general principles to a common context in the CSE Department environment, the production of source code for project and homework assignments. It should be thoroughly understood before undertaking any cooperative activities or using any other sources in such contexts.

All academic work must be your own. Plagiarism, defined as copying or receiving materials from a source or sources and submitting this material as one's own without acknowledging the particular debts to the source (quotations, paraphrases, basic ideas), or otherwise representing the work of another as one's own, is never allowed. Collaboration, usually evidenced by unjustifiable similarity, is never permitted in individual assignments. Any submitted academic work may be subject to screening by software programs designed to detect evidence of plagiarism or collaboration.

It is your responsibility to maintain the security of your computer accounts and your written work. Do not share passwords with anyone, nor write your password down where it may be seen by others. Do not change permissions to allow others to read your course directories and files. Do not walk away from a workstation without logging out. These are your responsibilities. In groups that collaborate inappropriately, it may be impossible to determine who has offered work to others in the group, who has received work, and who may have inadvertently made their work available to the others by failure to maintain adequate personal security. In such cases, all will be held equally liable.

These policies and interpretations may be augmented by individual instructors for their courses. Always check the handouts and web pages of your course and section for additional guidelines.

Departmental Policy on Violations of Academic Integrity

Any student accused of a violation of academic integrity will be so notified by the course director. An informal review will be conducted, including a meeting between these parties. After this review and upon determination that a violation has occurred, the following sanctions will be imposed. **It is the policy of this department that, in general, any violation of academic integrity will result in an F for the course, that all departmental financial support including teaching assistantship, research assistantship or scholarships be terminated, that notification of this action be placed in the student's confidential departmental record, and that the student be permanently ineligible for future departmental financial support.** A second violation of academic integrity will cause the department to seek permanent dismissal from the major and bar from enrollment in any departmental courses. Especially flagrant violations will be considered under formal review proceedings, which may in addition to the above sanctions result in expulsion from the University.