Neural Machine Learning and Data Mining II

COMP / ELEC / STAT 602, Fall 2015

Elective course, 3 credits (3 contact hours per week)

Course home page: http://www.ece.rice.edu/~erzsebet/ANNcourseII.html

(This website is also linked from Owl Space.)

Students are responsible for being familiar with this syllabus, and with the contents of both the above and the Owl Space website, and follow the postings as the course proceeds.

Instructor: Erzsébet Merényi
Class meets: TTH 1:00 - 2:20pm, KCK 107
Office/Phone: DH 2040, 713-348-3595
Office hours: by appointment
Assistant: N/A
email: erzsebet@rice.edu

Short course description

Advanced topics in Artificial Neural Network theories, with a focus on learning high-dimensional manifolds with neural maps (Self-Organizing Maps and variants, Learning Vector Quantization, both unsupervised and supervised paradigms) (unsupervised learning in general). Application to clustering, classification, dimension reduction, sparse representation. Comparison with "gold standards" through examples from image and signal processing. The course will be a mix of lectures and seminar style discussions with active student participation, based on recent research publications. Students will have access to research software environment to do simulations.

Sample Course Outline

1. Introduction, orientation

2. Review part of COMP / ELEC / STAT 502, Neural Machine Learning I.

2.1. Review of Unsupervised Learning, Hebbian Learning, Self-Organizing Maps, LVQ
2.2. The Basic Kohonen SOM

3. Kohonen Maps (SOMs) and Their Interpretation

3.1. Visualization of SOM knowledge – basics: U-matrix and variations, density map
3.2. Visualization of SOM knowledge – advanced: Connectivity Matrix and graph representation
3.3. Finding clusters: interpretation of the visualized knowledge, and cluster extraction

3.4. Data compression, and coding aspects

4. Variants of Self-Organizing Maps

4.1. Kohonen SOM vs Conscience algorithm, neighborhood functions and metrics
4.2. Criteria of faithful topographic mapping; Measures of topology violation in SOM, monitoring of violations, fixes
4.3. Neural Gas, Growing Self-Organizing Maps
4.4. Magnification in SOMs
4.5. Distortion based and information based Self-Organizing Maps, density matching

5. Self-Organizing Maps for High-Dimensional and Complex Data

5.1. Issues related to high dimensionality and complexity of data spaces
5.2. Why and how some favorite traditional methods fail for complicated, high-dimensional data
5.3. How do SOMs deal with high-dimensional data; Applications, case studies

6. Unsupervised Learning as Support for Supervised Classification

6.1. Hybrid ANN architectures containing unsupervised and supervised learning components
6.2. Classification versus prediction of continuous parameters (underlying causes)
6.3. The use of unlabeled samples to boost performance of supervised learning (classification)

7. Evaluation of Clustering Quality and Classification Accuracy

7.1. Cluster validity indices (classics; and CONNindex)
7.2. Evaluation of classification accuracy: sampling requirements, k-fold cross-validation, ROC curves, Kappa statistics, Wilcoxon signed ranks
7.3. Case studies

8. Dimension Assessment and Non-linear Dimension Reduction

8.1. Generalized Relevance Learning Vector Quantization and (GRLVQ and GRLVQI)
8.2. Matrix GRLVQ

9. Metrics for Learning, and Learning of Metrics

9.1. Feature spaces: homogeneous and inhomogeneous feature vectors
9.2. Feature representation: homogeneous and inhomogeneous representations
9.3. Domain specificity in metric construction

Software: Simulations and exercises can be based on C or Matlab programming, and / or using my group's research software environment

The information contained in the course syllabus, other than the absence policies, may be subject to change with reasonable advance notice as deemed appropriate by the instructor.

Detailed Course Schedule

A detailed schedule of class topics will come on-line in a timely manner at the course web site http://www.ece.rice.edu/~erzsebet/ANNcourseII.html under Course Schedule along with reading, review (home work) and presentation assignments. The materials (s.a. lecture notes and home work assignments) indicated in the Course Schedule will be downloadable from Owl Space or from other designated web site.

Pre-requisites

ELEC / COMP / STAT 502 or equivalent, or instructor's permission.
Further details on the above courses can be found at the links provided at the course home page, http://www.ece.rice.edu/~erzsebet/ANNcourseII.html under Prerequisites.

Course Materials

The course will be based on Lecture Notes, assigned papers from literature, scheduled as described at the course web site http://www.ece.rice.edu/~erzsebet/ANNcourseII.html under Course Schedule . Assigned readings as well as further recommended readings will include selected parts of the following books. All are available at www.amazon.com, or at the Fondren Library at Rice. Copies of mandatory readings will be provided in class.

Simon Haykin: Neural Networks. A Comprehensive Foundation. McMillan, New Jersey, 1999. (2nd Edition)
Teuvo Kohonen: Self-organizing Maps (Springer Series in Information Sciences S.). Springer-Verlag, 2001 (3rd Edition, ISBN: 3540679219)

Objectives of the Course

1. Student understanding of concepts, and mastery, of neural manifold learning and data mining methods, and their applications to high-dimensional complex data.
2. Student mastery of neural computation, both theoretically and through simulations using neural network software and / or the students' own software algorithms.
3. Student competence in critiquing articles form literature, and communicating their own findings from neural computing exercises.

Assignments, Grading Policies and Other Logistic Requirements for STAT 615

Grades will be made up of the following components, with approximate weights as shown:
75% - Performance in class
15% - Mini Project
10% - Homework

1. Performance in class The instructor will give an introductory lecture for each major topic, after which students will take turns presenting and critiquing articles assigned by the instructor. Demonstration of thorough understanding, rigorous presentation of algorithms, evaluation of scope, significance, applicability will be expected. Demonstrations or evaluations of capabilities of published algorithms or possible improvements will, in many cases, require the presenter's own simulations. Every student will be required to read all articles/handouts and expected to contribute to the discussions. This occasionally will involve running simple experiments and discussing/comparing results. The presenter will be required to turn in their presentation in electronic form and will receive feedback and grade points. Others will write reviews of selected articles (see Home work). Details of roles and requirements will be discussed in class.

2. Mini Project There will be a short project on a focused topic, as a conclusion of the semester. Projects will be presented in the last class period. The exact schedule, sleection of

topics, and requirements will be determined and announced in class well ahead of time.

3. Homework Students will be required to turn in a short (approximately one page) review of selected papers. The requirements of the review will be explained in class. There will be no other homework to turn in.

3.1 Late Homework Policy Homework is due at the beginning of class on the due date. You can either bring it to class or drop it in the designated wall pocket next to my office door before class. After the due date, but before next class, homework can be turned in for 50% credit. You are encouraged to discuss your reading with classmates but you will hand in your own paper reviews which you are expected to understand, and you are responsible for the quality of the writing.

Missed assignments If you must miss (or be late with) an assignment (homework, presentation) due to an extraordinary circumstance please notify me as much ahead of time as possible, and make arrangements with me for completing the missed assignment. If, in extreme emergency, you are unable to provide advance notice, please let me know as soon as possible afterwards, and I will work with you on a solution accordingly.

Expectations Regarding Honor Code, Collaboration, and Citation

In preparing assignments (presentations, reviews, project), students are encouraged to consult freely any material and anyone. However, each individual will write and turn in his or her own write-up or presentation, which they are expected to understand. In all work, students are expected to be scrupulous about proper citation of sources (where applicable), as required both as a matter of integrity and formally as a part of the Rice Honor Code.

Class Attendance and Absence Policy

Students are expected to attend all classes since 75% of the grade is composed of participation and performance in class. Students who must miss a class or assignment because of unavoidable circumstances should consult with the instructor well in advance so that alternative arrangements may be made.

University Disability Accommodation Policy

The University seeks to foster an environment of broad access and feasibly equal opportunity to education. The Office of Disability Support Services (DSS; Allen Center, Room 111; 713-348-584; adarice@rice.edu) supports and implements federal guidelines under the Rehabilitation Act of 1973 and the Americans with Disabilities Act. Students with documented disabilities requiring accommodation under Rice's established policies should consult DSS and the instructor; all such consultations and accommodations will be held confidential to the extent feasible.

Use of Machines in the Classroom

Cell phones must be turned off – or rendered silent – within the classroom. If you need to take an urgent call, please set your phone to vibration and take the call outside the room. Laptops or other small devices may be used in class only for specific class purposes. If you have an urgent need to be online for other purposes during class time, feel free to do so . . . but elsewhere.