Sample Course Outline, COMP / ELEC / STAT 602
Neural Machine Learning and Data Mining II.
Approximately 13 x 3 = 39 hours, 3 credits

1. Review part of COMP / ELEC / STAT 502, Neural Machine Learning I.
- 1.1. Review of prototype-based unsupervised learning of manifold structure with neural maps: Hebbian Learning, Self-Organizing Maps (SOM)
- 1.2. The basic Kohonen SOM; and basic Learning Vector Quantizers (LVQs)

2. Interpretation of Kohonen Maps for Extraction of Manifold Structure
- 2.1. Visualization of SOM knowledge – basics: U-matrix and variations, density map
- 2.2. Visualization of SOM knowledge – advanced: Connectivity Matrix and graph representation, underlying theories
- 2.3. Finding clusters: interpretation of the visualized knowledge, and structure extraction
- 2.4. Data compression and coding aspects, sparsity

3. Advanced Variants of Neural Maps and Measures of Mapping Quality
- 3.1. Kohonen vs Conscience SOM, neighborhood functions and metrics
- 3.2. Criteria of topology-preserving mapping; Measures of topology violation
- 3.3. Visualization and monitoring of violations, fixes
- 3.4. Neural Gas, Growing Self-Organizing Maps
- 3.5. Magnification in neural maps, and explicit control for different optimality criteria
- 3.6. Preferential discovery of small clusters with controlled magnification in SOMs
- 3.7. Distortion based and information based Self-Organizing Maps, density matching

4. Self-Organizing Maps for High-Dimensional and Complex Data
- 4.1. Issues related to high dimensionality and complexity of data spaces
- 4.2. Why and how some favorite traditional methods fail for complicated, high-dimensional data
- 4.3. How do SOMs deal with high-dimensional data; Applications, case studies
- 4.4. Comparison with classics (PCA, MDS variants, LLE-s, ISOMAP)

5. Unsupervised Learning as Support for Supervised Classification
- 5.1. Hybrid ANN architectures containing unsupervised and supervised learning components
- 5.2. Classification versus regression (parameter inference, learning underlying causes)
- 5.3. The use of unlabeled samples to boost performance of supervised learning (classification)

6. Evaluation of Clustering Quality and Classification Accuracy
- 6.1. Cluster validity indices: classics; and advanced ones, CONNindex
- 6.2. Evaluation of classification accuracy: sampling requirements, k-fold cross-validation, ROC curves, Kappa statistics, Wilcoxon signed ranks
- 6.3. Case studies

7. Intrinsic Dimensionality, Non-linear Dimensionality Reduction
- 7.1. Classics
- 7.2. Generalized Relevance Learning Vector Quantization
- 7.3. Neural ICA

8. Similarity Metrics for Learning, and Learning of Metrics
- 8.1. Feature spaces: homogeneous and inhomogeneous (mixed) feature vectors
- 8.2. Feature representation: homogeneous and inhomogeneous representations
- 8.3. Domain specificity in metric construction (e.g., divergences for functional data)

The exact course contents will be shaped from the above menu emphasizing the collective interest of the participants.

Prerequisites: ELEC / COMP / STAT 502 or equivalent, or instructor's permission

This course is a specialized seminar/lecture course, emphasizing active student participation and research. Since most classes will consist of presentations and critical discussions of papers and book chapters by students, homework will be in the form of reading assignments, and simulation experiments. The emphasis is on deep understanding of the chosen topics, and therefore on the presentation quality and critical discussion of papers. Good coding skills in MATLAB, R or C are assumed.
Grades will be composed of performance in class (presentations, discussions, and experiments), home works, and a 2-week class project. Detailed descriptions and requirements are included in the full Syllabus and also posted at the course web site http://www.ece.rice.edu/~erzsebet/NMLcourseII.html .

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