Neural Networks for Pattern Recognition – Statistical foundation, perspective and alternatives, 36125651

Graduate course, Semester Aleph 2020, Wednesdays, 17:00-20:00
Instructor: Prof. Mayer Aladjem (Retired Associate Professor), http://www.ee.bgu.ac.il/~aladjem
Office hours: 1 hour per week-after the lecture (Wednesdays, 20:00-21:00)

This course is intended to be largely self-contained.
It is suitable for all fields of engineering specializations.

General prerequisites: Basic undergraduate mathematical courses. Programming Matlab.

Course Description: The course consists of 13 topics that introduce neural computing procedures. In the recent years, neural computing has emerged as a practical technology, with successful applications in many fields. The majority of these applications are concerned with problems in pattern recognition known as **"Machine Learning"** also.

From the perspective of multivariate statistical methods for data analysis, neural networks can be regarded as extension of many conventional techniques, which have been developed over several decades. The learned material provides comprehensive information, including algorithms and procedures, for designing effective, practical neural networks systems. The course addresses some conventional procedure (stochastic approximation, Gaussian mixture models, EM algorithm, exploratory projection pursuit, principal components analysis (PCA)) in the perspective of their extensions in the framework of the neural networks.

The students will work on individual project covering the learned methods. Each of them will study the methods on specific synthetic data sets generated by him/her. The instructor will define the type and parameters of the data for each student.

Upon successful completion of the course, the students should be able to exploit the learned methods in the specific data analysis studies in the student's research and/or work and understand methods published in the neural network and statistical journals.

The lecture handouts will be available before the class in
http://moodle2.bgu.ac.il/?lang=en

Attendance regulation: 5% of the final grade is compulsory attendance in at least 80% of the classes. The attendance in the lectures is critical for student's success in the project and the course.

Assessment:

Compulsory attendance in at least 80% of the classes: 5%

Project: 95% (including interview on the project and course topic, which will be around 30 min. with each student)

The project contains two parts formulated during the teaching of topics 6,8 and 10 respectively (see course topics on the next page).

Learning outcomes: The student will have a thorough understanding of:

1. Parametric methods for probability density estimation.
2. Robbins-Monro (stochastic approximation) algorithm.
   Successive maximum likelihood (ML) estimate of density function.
3. Semi-parametric density estimation- mixture models. Estimation by SQP procedure for non-linear optimization under constrains and EM algorithm.
4. Model based strategy for mixture density estimation. BIC, NEC, K-fold-CV, Monte Carlo CV criteria for model selection.
5. PCA and independent component analysis (ICA) dimensionality reduction.
6. Projection pursuit mixture density estimation.
7. Assessment the accuracy of the density estimation in the simulations – F-measure and PVE.
8. RBF and ML networks for classification and regression. Deep architectures.
9. Generalization by weight decay, early stopping of the training, training with noise.
10. Restricted estimators – L2 ridge and L1 Lasso penalties.

11. Course topics:

1) Key concepts in neural networks for pattern recognition. Classification and regression analysis. Pre-processing and feature extraction. Generalization. Over-fitting. Control the effective complexity of the model by regularization.

2) Hypothesis testing. Explanation of the procedures for generation synthetic data sets used in the project of the course.

3) Probability density estimation. Parametric methods. Maximum Likelihood (ML) estimate of the parameters. Properties and characteristics of ML estimators.

4) Robbins-Monro (stochastic approximation) algorithm. Successive ML estimate of density function.

5) Geometrical perspective of Lagrange multipliers. Semi-parametric density estimation. Mixture models. Analytic expression for the gradient of the error function. Procedures for non-linear optimization under constrains. Sequential quadratic programming (SQP) procedure.

6) Expectation Maximization (EM) algorithm for mixture density estimation. Various parameterizations for the covariance matrices of Gaussian mixture models (GMMs). Initialization of the EM algorithm – random subset data points, clustering algorithms. Explanation the first part of the project and discussions.

7) Bayesian information criterion (BIC). Conventions for calibration BIC differences. Model based strategy for mixture density estimation. Model selection by normalized entropy criterion (NEC), K-fold cross-validation (CV), Monte Carlo CV.

8) Assessment the accuracy of the density estimation in the simulations – F-measure for clustering quality of the GMMs and percentage of variance explained (PVE). Continuation the explanation of the first part of the project.

9) Single Layer networks for classification. Logistic discrimination. Solution by Singular Value Decomposition (SVD). Exact interpolation. Radial Basis Function (RBF) networks.

10) Two-stage training of RBF networks. Basis function optimization – clustering algorithms, GMMs. Supervised training. PBF properties – best approximation, best basis functions for controlling the effective complexity of the model by regularization and other. Explanation the second part of the project and discussions.

11) Multi-Layer (ML) networks. Deep architectures. Batch and sequential training. Random initialization. Weight-space symmetries. Recursive training. Optimization in practice.

12) Error functions. Learning and generalization. Weight decay, early stopping of the training, training with noise. Restricted estimators – L2 ridge and L1 Lasso penalties. Deep learning.

13) Projection pursuit mixture density estimation. Relations to Independent Component Analysis (ICA). Dimensionality reduction by PCA and ICA.

Required reading:

1. C.M.Bishop, "Neural networks for pattern recognition", Oxford University Press, 1995.
2. I.T.Nabney "NETLAB Algorithms for Pattern Recognition", Springer, 2001
3. Selected journal papers.

Additional literature:

1. C.M.Bishop, "Pattern Recognition and Machine Learning", Springer, 2006.
2. T. Hastie, R. Tibshirani, J. Friedman, "The elements of statistical learning – Data Mining, Inference, and Prediction", Springer, 2001.