Optimization methods for machine learning

Syllabus a.a. 2015-16

1. Introduction.
    - Definition of learning systems. Goals and applications of machine learning (classification and regression). Basics on statistical learning theory (Vapnik Chervonenkis bound). Underfitting and Overfitting. Use of data: training set, test set, validation set.
2. Artificial Neural Networks.
    - Neurons and biological motivation. Linear threshold units. The Perceptron and its learning algorithm (proof of convergence). Classification of linearly separable patterns.
    - Multi-Layer Feedforward Neural Networks. Gradient method: basics. Back-propagation (BP) algorithm. BP batch version: proof of convergence and choice of the learning rate. BP on-line version: incremental method, theorem of convergence. Momentum updating rule.
    - Radial-Basis function (RBF) networks: regularized and generalized RBF networks. Their use in interpolation and approximation. learning strategies and error functions. Unsupervised selection of center. Supervised selection of weights and centers: decomposition methods into two blocks and decomposition methods into more blocks. Convergence theory of decomposition methods.
    - Early stopping
3. Support Vector Machines (Kernel methods)
    - Soft and hard Maximum Margin Classifiers. Quadratic programming formulation of the soft/hard maximum margin separators. Kernels methods.
    - Dual formulation of the primal QP problem. Wolfe duality theory for QP. KKT conditions. Frank Wolfe method: basics. Decomposition methods: SMO-type algorithms, MVP algorithm, SVM^{light}, cyclic methods. Convergence theory.
    - Implementation tricks: Caching, shrinking.
    - Choosing parameters: k-fold cross-validation.
    - Multiclass SVM problems: one-against-one and one-against-all.
4. Practical use of learning algorithms.
5. Comparing learning algorithms from the optimization point of view.
6. Use of standard software (Weka, LIBSVM)