

HIGH-DIMENSIONAL CLASSIFICATION BY SPARSE LOGISTIC REGRESSION

FELIX ABRAMOVICH

Abstract: In this talk we consider high-dimensional classification. We discuss first high-dimensional binary classification by sparse logistic regression, propose a model/feature selection procedure based on penalized maximum likelihood with a complexity penalty on the model size and derive the non-asymptotic lower and upper bounds for the resulting misclassification excess risk. Implementation of any complexity penalty-based criterion, however, requires a combinatorial search over all possible models. To find a model selection procedure computationally feasible for high-dimensional data, we consider logistic Lasso and Slope classifiers and show that they also achieve the optimal order. We extend further the proposed approach to multiclass classification by sparse multinomial logistic regression. In particular, we show that there exists a phase transition corresponding to small and large number of classes. We discuss also various possible types of sparsity in the multiclass setup.

This is a joint work with Vadim Grinshtein and Tomer Levy.