Course personnel:

• Professor: Dr. Florentina Bunea
• Office: OSB 209 D
• Phone/e-mail: 644-8775, flori@stat.fsu.edu
• Office Hours: Tue, Th 1:00 - 2:00 and by appointment.

Administrative Information

• Place: Room 110 OSB
• Time: Tue, Th 2:00 – 3:15

Course Objective: To become familiar with the most important and up to date methods of model selection and aggregation. To understand their theoretical and computational merits and limitations in a large array of statistical models, e.g.: parametric and non-parametric regression and generalized linear models, density estimation, classification, functional data analysis, models for case-control studies. To gain hands on experience by implementing and comparing these methods in concrete settings.

Course Outline: This class is based on very recent research materials, and there are not, as yet, any textbooks that cover it. A very comprehensive list of research papers and related books will be provided. Handouts of the material covered in this course and relevant research papers will be given at the beginning of every class. A detailed (tentative) outline of this material follows.
1 Variable selection in high dimensional parametric regression models

1.1 Linear regression models

1.1.1 Variable selection methods based on optimizing a penalized criterion

(I) Convex penalties: the Lasso (the \( \ell_1 \)) penalty. (4–5 lectures).
- Overview, background and discussion of computational merits. Link with soft thresholded estimators for orthogonal design regression models.
- General design regression models, general errors (beyond Gaussian models): adaptation to unknown sparsity, goodness of fit, consistency of subset selection. Detailed discussion of finite sample results: for a given sample size, under what conditions can we recover the target variable set, at a given level of confidence?
- How to choose the tuning parameter: Hoeffding and Bernstein’s inequalities.
- A discussion of the merits of the \( \ell_1 + \ell_2 \) (elastic net) penalty.

(II) The Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC): the \( \ell_0 \) penalty. Other non-convex penalties. Rationale, theoretical properties and computational limitations. (1-2 lectures).

1.1.2 Variable selection methods based on testing

- Discussion of methods designed to control the False Discovery Rate (FDR). (1-2 lectures).

1.2 Generalized linear models

1.2.1 Logistic regression

- Adaptation to unknown sparsity and consistency of subset selection via Lasso-type penalized criteria. A discussion of the FDR-type methods. (2 lectures)
2 Model selection techniques in nonparametric models

2.1 Nonparametric regression models

- Introduction, projection estimators in non-parametric regression, risk bounds, oracle inequalities, the concept of (minimax) adaptive estimation. Lasso/BIC/AIC/Mallow’s \( C_p \) in nonparametric regression; connections with regularized empirical risk minimization. (2–3 lectures).

- Kernel smoothing methods. Choice of tuning parameters via cross-validation: practical issues and theoretical properties. (1–2 lectures)

2.2 Functional data models

- Estimation of the mean of a stochastic process from an ensemble of curves: background. Mean estimation via adaptive soft and hard thresholding. (1-2 lectures)

- Adaptive confidence bands for the mean. (1 lecture)

3 Aggregation of estimators

3.1 Mathematical targets of aggregation

- Aggregation (”averaging”) of arbitrary estimators of a conditional mean (regression models, classification models). Mathematical targets: what do we expect to achieve by combining models? Is it better to ”select” than to ”average”? (1 lecture)

3.2 Aggregation via BIC and Lasso

- Discussion of methods and theoretical properties of the aggregate estimators. (1 lecture)

3.3 On-line aggregation via exponential weighting

- Discussion of methods and theoretical properties of the aggregate estimators, with emphasis on improved prediction accuracy. Applications to any method dependent on tuning parameters. (2 lectures)
4 Additional topics

4.1 Dimension reduction and variable selection in case-control studies

- The discussion and analysis of Lasso-type estimators for case-control (retrospective sampling) design. Consistent estimates of the odds ratio and consistent subset selection. (2 lectures)

4.2 Density estimation

- The Lasso estimate in mixture models. Applications to the estimation of the components of a mixture of normal densities. (1-2 lectures)

4.3 Classification

Grading Policies. This class will be Pass/Fail. Students will be given full liberty in selecting a project topic either from the list above or on a related topic. Here is a list of possible types of projects: (i) A computational comparison of methods for a particular model with applications to a real data set; (ii) A theoretical comparison of methods for a particular model; (iii) Improvements/extensions of an existing method; (iv) A summary of a very recent research paper (either listed as ”to appear” or printed in the last 6 months) including either an implementation of the newly proposed method and/or a clear report stating the theoretical results, filling in the gaps in the proofs; (v) Any reasonable combination of (i) - (iv). A project topic must be chosen (in consultation with me, if needed) and will be discussed in class at the beginning of February. A brief progress report will be presented March, the 17th and the final project will be presented April 21st and 23rd.

References


[34] Juditsky, A., Rigollet, Ph. and Tsybakov, A. (2005) Learning by mirror averaging. Prépublication du Laboratoire de Probabilités et Modèles Aléatoires, Universités Paris 6 and Paris 7. [https://hal.ccsd.cnrs.fr/ccsd-00014097](https://hal.ccsd.cnrs.fr/ccsd-00014097)


