

Statistical learning

Instructor: Vilmos Prokaj

Term: Fall

Weeks: 8-14

Contact hours: 3

Credits: 6

Aim and scope:

To understand the statistical methods of regression and classification, the variance-bias trade-off. The students also get an insight into the modern notion of regularization. The lectures are accompanied by practices, where the learned methods are illustrated on real datasets.

Syllabus:

Overview of supervised problems, regression and classification, linear model, logistic regression. Linear regression, subset selection, ridge regression and lasso. Linear and quadratic discriminant analysis, logistic regression and their comparison.

Regularization. Splines. Regression and classification.

Local regression. Bias–variance trade-off, AIC, cross-validation, bootstrap.

Maximum likelihood estimation, Bayesian method, EM algorithm,

Generalized additive model, regression and classification trees.

Boosting, AdaBoost. Gradient boosting, ensemble learning

Support vector machine.

Grading: exam (incorporating the solution of homeworks)

Literature:

Hastie, T.; Tibshirani, R. & Friedman, J. (2009), The elements of statistical learning, Springer, New York.

James, G.; Witten, D.; Hastie, T. & Tibshirani, R. (2013), An introduction to statistical learning, Vol. 103, Springer, New York.