Topics in Statistical Theory (M16)

Richard J. Samworth

This course will provide an introduction to the theory behind a selection of statistical problems that play a key role in modern statistics. Most undergraduate statistics courses are restricted to the study of parametric models; here we will no longer assume that our distributions belong to finite dimensional classes and will instead study fundamental nonparametric problems such the estimation of a distribution function, a density function or a regression function. We will consider the canonical machine learning problem of classification, and may also cover some extreme value theory including anologues of the Central Limit Theorem for the maxima and minima of a sample. Minimax lower bounds are studied as a way of quantifying the intrinsic difficulty of a statistical problem, and provide limits on how well any estimator can perform in a given situation.

A tentative outline of the course is as follows:

- An introduction to nonparametric statistics: the basics of empirical process theory, Glivenko–Cantelli theorem, Dvoretzky–Kiefer–Wolfowitz theorem, order statistics, quantile estimation and associated asymptotic distribution theory.

- Kernel density estimation: histograms, bias and variance expansions, aymptotically optimal bandwidth, canonical kernels, higher order kernels, bandwidth selection, multivariate density estimation.

- Nonparametric regression: kernel nonparametric regression, bias and variance expansions. Cubic splines, natural cubic smoothing splines, choice of smoothing parameter, other splines, equivalent kernel. Classification problems, the Bayes classifier, nearest neighbour classifiers.

- Minimax theory: notion of information-theoretic lower bounds, distance and divergence between distributions, optimal rates, Le Cam's two points lemma.

- Extreme value theory: the extremal types theorem, domains of attraction, max-stability.

Prerequisites

A good background in undergraduate probability theory, elements of linear algebra and real analysis. Measure theory is not necessary but may be helpful; similarly for a preliminary course in mathematical statistics. Though the material in the Modern Statistical Methods course will not be needed here, the two courses complement each other well.

Literature

No book will be explicitly followed, but some of the material is covered in

L. Devroye, L. Györfi, G. Lugosi, *A Probabilistic Theory of Pattern Recognition*, Springer 1996.

A. Tsybakov, *Introduction to Nonparametric Estimation*, Springer 2009.

M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge University Press, 2019.

Additional support

Three example sheets will be provided and associated examples classes will be given. There will be a revision class in the Easter Term.