

Statistical methods

Prof. Lorenzo Finesso¹

¹Istituto di Ingegneria Biomedica, ISIB-CNR, Padova
Email: lorenzo.finesso@isib.cnr.it

Timetable: 24 hrs. (two lectures of two hours each, per week). Class meets every Monday and Wednesday from 10:30 to 12:30. First lecture on Monday, June 16, 2014. Room DEI/G (3-rd floor, Dept. of Information Engineering, Via Gradenigo 6/a, Padova).

Course requirements: Basics of Probability Theory and Linear Algebra.

Examination and grading: homework assignments and take-home exam.

Aim: The course will present a survey of statistical techniques which are important in applications. The unifying power of the information theoretic point of view will be stressed.

Course contents:

Background material. The noiseless source coding theorem will be quickly reviewed in order to introduce the basic notions of entropy and informational divergence (Kullback-Leibler distance) of probability measures. The analytical and geometrical properties of the divergence will be presented.

Divergence minimization problems. Three basic minimization problems will be posed and, on simple examples, it will be shown that they produce the main methods of statistical inference: hypothesis testing, maximum likelihood, maximum entropy.

Multivariate analysis methods. Study of the probabilistic and statistical aspects of the three main methods: Principal Component Analysis (PCA), Canonical Correlations (CC) and Factor Analysis (FA). In the spirit of the course these methods will be derived also via divergence minimization. Time permitting there will be a short introduction to the Nonnegative Matrix Factorization method as an alternative to PCA to deal with problems with positivity constraints.

EM methods. The Expectation-Maximization method was introduced as an algorithm for the computation of the Maximum Likelihood (ML) estimator with partial observations (incomplete data). We will derive the EM method for the classic mixture decomposition problem and also interpret it as an alternating divergence minimization algorithm (à la Csiszár Tusnányi).

Hidden Markov models. We will introduce the simple yet powerful class of Hidden Markov models (HMM) and discuss parameter estimation for HMMs via the EM method.

The MDL method. The Minimum Description Length method of Rissanen will be presented as a general tool for model complexity estimation.

References: A set of lecture notes and a list of references will be posted on the web site of the course.