Learning from data as an inverse problem

Věra Kůrková

Institute of Computer Science
Academy of Sciences of the Czech Republic

Generalization capability in learning from data can be investigated in terms of regularization, which has been used in many branches of applied mathematics to obtain stable solutions of inverse problems, i.e., problems of finding unknown causes (such as shapes of functions) of known consequences (such as measured data). It will be shown that supervised learning modeled in terms of minimizations of error functionals, the expected and the empirical one, can be reformulated as inverse problems with solutions in spaces of functions defined by kernels. Mathematical results from theory of inverse problems can be applied to construct optimal solutions of learning tasks, which can be used to design learning algorithms based on solutions of systems of linear equations.

Content:

Learning from data: minimization of the empirical error functional defined by a sample of data and minimization of the expected error functional defined by a probability distribution, optimizations of error functionals as best approximations, tools from approximation theory.

Generalization: philosophical concept of generalization, generalization in learning as a stability of solutions with respect to small changes of data, penalization of solutions with high-frequency oscillation.

Inverse problems: well and ill-posed problems, well and ill-conditioned problems, Moore-Penrose pseudosolution, measures of stability, regularization as improvement of stability, properties of optimal and regularized solutions.

Representation of learning as an inverse problem: typical operators defining inverse problems, tomography and Radon transform, operators defining inverse problems modeling learning, characterization of optimal and regularized solutions (the Representer Theorem).

Three reasons for using kernels in machine learning: kernels define a class of hypothesis spaces, where theory of inverse problems can be applied, kernels define stabilizers penalizing various types of high-frequency oscillations, kernels define transformations of input space geometry allowing more types of data to be separated linearly.

Learning algorithms based on the Representer Theorem: neural network learning as a solution of a system of linear equations, approximate optimization as complexity reduction, comparison with algorithms operating on networks with smaller number of units than the size of the sample of data.