Semiparametric Regression, Penalized Splines, and Mixed Models

David Ruppert
Cornell University

A semiparametric regression models combines parametric and nonparametric components. Penalized splines can model the nonparametric components using a pre-determined basis that is rich enough to avoid under-fitting. Over-fitting is prevented by a roughness penalty, and penalized splines include classical smoothing splines as a special case. A penalized splines can be viewed as a BLUP in a mixed model or as an empirical Bayes estimator. The mixed mixed viewpoint is especially convenient for applications because of its conceptual simplicity, because it allows the use of readily available software, and especially because it can incorporate random subject-specific effects as well. Additive models, single-index models, and nonlinear regression models can be fit relatively simply. Penalized spline methods are arguably the most effective methods known for nonparametric regression with covariate measurement errors.

The first part of this talk will be an overview of mixed-model splines for semiparametric regression circa 2003 when "Semiparametric Regression" was published by Ruppert, Wand, and Carroll. The second half will survey work that has been published since then.

One particularly interesting development in the past few years has been an asymptotic theory for penalized splines. There have been two parallel developments. In one the number of knots is a smoothing parameter and the asymptotics are similar to those of un-penalized least-squares splines. In the second, the number of knots increase sufficiently fast that it does not play the role of a smoothing parameter. In this case, the asymptotics are similar to those of smoothing splines and, somewhat surprisingly, the asymptotic distribution does not depend on the degree of the spline, only the order of the penalty.