COURSE:

"The Craft of Smoothing"

Brian Marx

11-12, July 2012

Abstract

Course Description:
In the course, we describe in detail the basics and use of P-splines, as a combination of regression on a B-spline basis and difference penalties (on the B-spline coefficients). Our approach is practical. We see smoothing as an everyday tool for data analysis and statistics. We emphasize the use of modern software and we provide functions for R/ S-Plus.

There will be six sessions:

Session 1 presents the idea of bases for regression. It will show why global bases, like power functions or orthogonal polynomials are ineffective and why local bases (Gaussian bell-shaped curves or B-splines) are attractive.

In Session 2, penalties are introduced, as a tool to give complete and easy control over smoothness. The combination of B-splines and difference penalties will be studied for smoothing, interpolation and extrapolation. In these first two sessions the data are assumed to be normally distributed around a smooth curve.

In Session 3, we extend P-splines to non-normal data, like counts or a binomial response. The penalized regression framework makes it straightforward to transplant most ideas from generalized linear models to P-spline smoothing. Important applications are density estimation and variance smoothing.

Any smoothing method has to balance fidelity to the data and smoothness of the fitted curve. An optimal balance can be found by cross-validation or AIC. This subject is studied in Session 4, as well as the computation of error bands of an estimated curve. We also show how optimal smoothing performs on simulated data, to give you confidence in that it makes the right choices.

In the first four sessions we only consider one-dimensional smoothing. When there are multiple explanatory variables, we can use generalized additive models, varying-coefficient models, or combinations of them. Tensor products of B-splines and multi-dimensional difference penalties

make an excellent tool for smoothing in two (or more) dimensions. This is the subject of session 5.

The final Session 6 looks at the use of P-splines in regression problems with very many variables, which are ordered, like in optical spectra. In the chemometric literature this is known as multivariate calibration.

In addition there will be two computer lab sessions, in which R software will be used to solve a number of smoothing problems. One session will concentrate on simple functions with limited goals. This will improve your understanding of what is going on "under the hood". The other session will apply smoothing to the generalized linear model and density estimation. An optional third lab will be provided that uses the {\tt mgcv} package, written by Simon Wood, a large but powerful tool that can handle a variety of situations.

Course schedule
11-12, July 2012

Day	Start	End	Type	Subject
Wed	9:00	9:50	Lecture	Regression and basis functions
	9:50	10:40	Lecture	The power of penalties
	10:40	11:00	Break	Coffee
	11:00	12:00	Lecture	Generalized linear smoothing
	12:00	13:00	Lunch	
	13:00	14:00	Lecture	Optimal smoothing in action
	14:00	15:30	Lab	Computer lab 1
Thur	9:30	10:30	Lecture	Multi-dimensional smoothing
	10:30	10:45	Break	Coffee
	10:45	11:30	Lecture	Penalized signal regression
	11:30	12:30	Lab	Computer lab 2 (+optional lab 3)