# Stat 511 Lecture 42

Note Title 4/30/2008

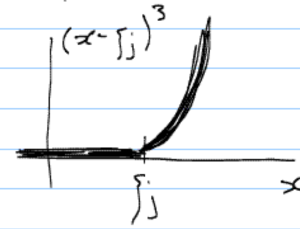
Smoothing / nonparametric regression

⋮

Polynomial Regression Splines ... (e.g. cubic regression splines) ... for "knot positions" $\xi_1 < \xi_2 < \cdots < \xi_k$ fit

$$y \approx \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{j=1}^{k} \theta_j (x - \xi_j)_+^3$$

for $(x-\xi_j)_+^3 = \begin{cases} (x-\xi_j)^3 & \text{if } x \geq \xi_j \\ 0 & \text{otherwise} \end{cases}$




If this via OLS (a regression program) there are $k+4$ parameters to fit — a fitted equation will be

$\theta$'s $\beta$'s

$$\hat{y} = b_0 + b_1 x + b_2 x^2 + b_3 x^3 \quad \text{for } x < \xi_1$$

$$\hat{y} = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \hat{\theta}_1 (x - \xi_1)_+^3 \quad \text{for } \xi_1 \leq x < \xi_2$$

i.e. this gives me fitted cubics on each interval that fit together nicely at knot positions –

## Cubic Smoothing Splines

Idea is to solve an optimization problem of choosing a function $f$ with 2 cont$^s$ derivatives that balances off 2 competing features of "niceness"

1) small $\sum_{i=1}^{n} (y_i - f(x_i))^2$ > penalty for poor fit

2) $\int_a^b (f''(t))^2 \, dt$ > penalty for wiggling

It's possible to optimize

$$\sum (y_i - f(x_i))^2 + \lambda \int_a^b (f''(t))^2 \, dt$$

"stiffness parameter"

over choices of $f$ — even "simple" software like JMP allows the fitting of such smoothers —

there is some theory available to guide choices of methods and parameters for these methods — see the early parts of Generalized Additive Models by Hastie & Tibshirani for refences —

Introduction to Bayes Analysis — for a quick look see slides on Vardeman's homepage for "Introduction to Bayes Analysis for Industry" —