Chapter 5 (Part II)

Basis Expansions and Regularization

Francisco, Yanjun, Benjamin, Allison

2002-9-26

What we have discussed

- Piecewise Polynomials and Splines
- Natural cubic splines
- Two examples about splines
- Filtering and Feature Extraction (Part)
  – Phoneme Recognition

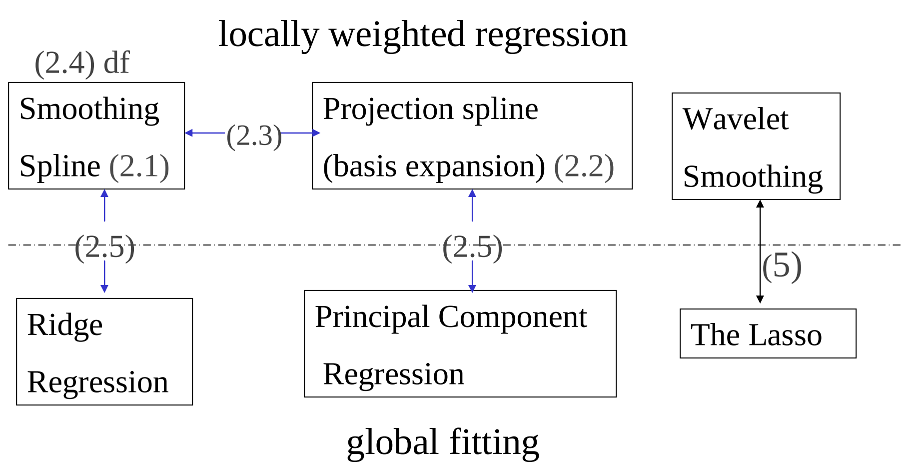
Today's Outline

- Smoothing Splines
  - Comparison with projection splines
  - Comparison with chapter 3 (shrinkage)
  - Automatic selection of smoothing parameters
- Multidimensional Splines

- Filtering and Feature Extraction (cont.)
- Nonparametric Logistic regression
- Wavelet Smoothing

2002-9-22 3

2. Smoothing Splines (roadmap)



2002-9-22 4

2.1 Smoothing Splines

- Why use smoothing splines? Intuition
  - Aim: to avoid the knot selection problem completely
  - So use a maximal set of knots –> N sample points

- Problem:
  - Among all functions f(x) with two continuous derivatives, to find the one minimize

$$RSS(f,\lambda) = \sum_{i=1}^{N} \{y_i - f(x_i)\}^2 + \lambda \int \{f''(t)\}^2 dt$$

2002-9-22 5

2.1 Smoothing Splines

- This RSS has a finite-dimensional unique minimizer
  - natural cubic spline having N knots just at the N sample points $f(x) = \sum_{j=1}^{N} N_j(x)\theta_j$

$$RSS(\theta,\lambda) = (y - N\theta)^T (y - N\theta) + \lambda\theta^T \Omega_N \theta$$

Solution: (Similar with Page. 60 – (3.43), (3.44))

$$\hat{\theta} = (N^T N + \lambda\Omega_N)^{-1} N^T y$$

2002-9-22 6

2.1 Smoothing Splines->Smoother Matrix

- Smoother Matrix S_λ
  - $\hat{\theta}$ are linear combination of y_i (linear smoother)
  - For N sample points: (Two N here: number /basis matrix)

$$\hat{f} = N(N^T N + \lambda\Omega_N)^{-1} N^T y = S_\lambda y$$

  - Smoother matrix depends only on x_i and λ

$$S_\lambda = N(N^T N + \lambda\Omega_N)^{-1} N^T$$

  - Important Property of Smoother matrix :
    - Positive semidefinite: all its eigenvalues >=0 (<=1 from 2.4)
    - Rank: N

2.2 Projection Matrix H_ξ

- Projection Matrix
  - Use M (M<<N) cubic-spline basis functions
  - Knot sequence ξ
  - B_ξ is N*M (N sample points, M basis functions)

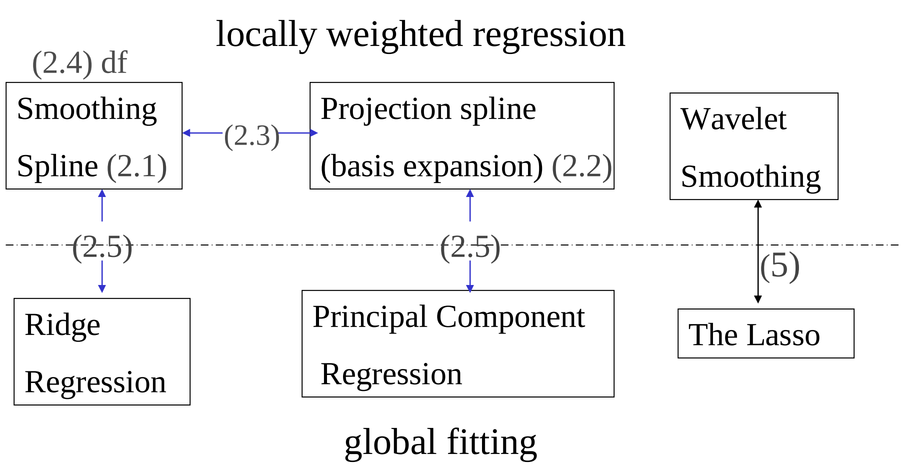
$$\hat{f} = B_\xi (B_\xi{}^T B_\xi)^{-1} B_\xi{}^T y = H_\xi y$$

  - Important Property
    - Projection Matrix is one kind of special matrix
    - Positive semidefinite: all its eigenvalues are 1 or 0
    - Rank M

2.2 Projection Matrix

- Projection Matrix: $rank = M = trace(H_\xi)$
  - The **trace** of a square matrix is: $tr(H) = \sum h_{i,i}$
  - Sum of eigenvalues are trace: $tr(H) = \sum_{i=1}^{N} \lambda_i$
  - The rank of an symmetric idempotent matrix is equal to its trace
  - Special properties of projection matrix - file

2. Smoothing Splines (roadmap)



2.3 Projection Matrix & Smoother Matrix

- Projection
  - symmetric
  - N*N
  - positive semidefinite
    - Eigenvalues 1 or 0

(The eigenvalues of a positive semidefinite matrix are non-negative)

  - Rank : M
  - Idempotent $H_\xi H_\xi = H_\xi$

- Smoother
  - symmetric
  - N*N
  - positive semidefinite
    - Eigenvalues >=0, <=1 (in 2.4)
  - Rank: N
  - Shrinking $S_\lambda S_\lambda \le S_\lambda$

For any symmetric A, the eigenvalues of A^2 are the square of those of A, and the eigenvectors are the same.

2.4 Smoother Matrix – Effective degrees of freedoms

- Projection
  - Dimension of projection space
    - $M = trace(H_\xi)$

- Smoother
  --* Effective degree of freedom
    - $df_\lambda = trace(S_\lambda)$

define Effective degrees of freedoms of a smoothing spline by analogy to projection matrix

???

Sum of eigenvalues are trace !