Questions

- What is the shrinking nature of S?
- P 132 What does “complexity” refer to?
  - Fan Li
- Why are first two eigenvalues always one?
  - Lots of people

Questions

- In general, I would appreciate discussion clarifying the explanation of the trace as effective degrees of freedom on p.130. Specifically, why are we converting to Reinsch form (an explanation of what it is and K would help)?
  - Paul Bennet
- What is the intuition of effective degree of freedom df? What is the numeric relationship between M and df, (bigger, equal or smaller)?
  - Yan Lui

2.4 Smoother Matrix – Effective degrees of freedoms

Effect of Penalty on S

- S in Reinsch form

$$S_\lambda = (I + \lambda K)^{-1}$$

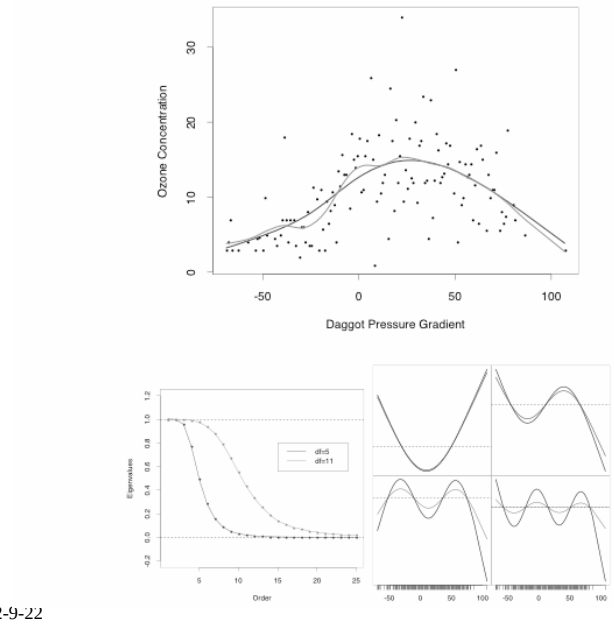
$$\min_f (y - f)^T (y - f) + \lambda f^T K f$$

- Eigen-decomposition

$$S_\lambda = \sum_{k=1}^{N} \rho_k(\lambda) u_k u_k^T$$

$$\rho_k(\lambda) = \frac{1}{1 + \lambda d_i}$$

- S is symmetric and pos semidefinite, so it has a real eigen-decomposition
- d's are the eigen values of the penalty matrix K
- Eigenvectors are not affected by changes in penalty



Shrinking and Projection Smoothers

- The smoothing spline decomposes y w.r.t. the basis u and shrinks each contribution

$$S_\lambda y = \sum_{k=1}^{N} u_k \rho_k(\lambda) \langle u_k, y \rangle$$

- In contrast, the basis regression method leaves each component unchanged or shrinks it to zero (H is a rank M projection matrix)
- Smoothing splines are analogous to ridge regression, basis regression is analogous to PCA

2.4 Smoother Matrix – Effective degrees of freedoms

Smoothing Splines

- u's ordered by decreasing eigenvalue increase in complexity (fig 5.7)
- First two eigenvalues are always one, eigenspace of functions linear in x
- ρ's are an inverse function of the penalty matrix K, λ controls rate of decrease to zero

$$\min_\theta \left\| y - U\theta \right\|^2 + \lambda \theta^T D \theta$$

2.4 Smoother Matrix – Effective degrees of freedoms

Limits of λ

- As $\lambda \to 0$, df $\to$ N, and S $\to$ I, the N-dimensional identity matrix (always returns y, zero training error)
- As $\lambda \to \infty$, df $\to$ 2, and S $\to$ H, the hat matrix for linear regression on x
- Local fitting method (fig 5.8)