Statistics 860 Lecture 6

Second Variational Problem:

$\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$

$\{\phi_1, \cdots, \phi_M\}$ span $\mathcal{H}_0$, $\mathcal{H}_0 \perp \mathcal{H}_1$

$$f = f_0 + f_1, \qquad \begin{array}{l} f \ \in \mathcal{H} \\ f_0 \in \mathcal{H}_0 \\ f_1 \in \mathcal{H}_1 \\ < f_0, \ f_1 >= 0 \end{array}$$

data $y = (y_1, \cdots, y_n)$,
$L_1, \cdots, L_n$ bounded linear functionals in $\mathcal{H}$.

Find $f_\lambda \in \mathcal{H}$ to min

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - L_i f)^2 + \lambda \|P_1 f\|^2 \qquad (*)$$

where P_1 is the orthogonal projection onto $\mathcal{H}_1$.

©G. Wahba 2016

Example: $W_m = \{f \in \mathcal{L}_2[0,1], f^{(m)} \in \mathcal{L}_2[0,1]\}$

Constructing the spline kernel:
Let $k_l(x) = B_l(x)/l!$, B_l are the Bernoulli polynomials.

Consider $K(s,t) = (-1)^{m-1} k_{2m}([s-t]) \quad (*)$

This is the RK for $W_m(per)$ with the constant functions removed. The square norm is

$$\|f\|^2 = \int_0^1 (f^{(m)})^2 du.$$

The elements in this space satisfy

$$\int_0^1 f(u)du = 0$$

in adition to the periodic boundary conditions up to $f^{(m-1)}$.

Now, add the one-dimensional space spanned by multiples of $k_m(x)$ to this space, with rank 1 kernel $k_m(s)k_m(t)$.

Verify that

$$\|f\|^2 = \int_0^1 \left(f^m(u)\right)^2 du$$

is the squared norm in the space with RK

$$k_m(s)k_m(t) + (-1)^{m-1}k_{2m}(\,[s-t]\,) \qquad (**)$$

We have taken the constant function out of $W_m(per)$ and added k_m to $W_m(per)$.

$\int_0^1 k_m^{(m)}(u)f^{(m)}(u)\,du = 0$ for $f \in$ the space with RK (*) (slide 2) since $k_m^{(m)}(x) = 1$ and $f^{(m-1)}(1) - f^{(m-1)}(0) = 0$.

Call this space H_1.

Let

$$M_0 f = \int_0^1 f(u)\,du$$

$$M_\nu f = f^{(\nu)}(1) - f^{(\nu)}(0), \qquad \nu = 1, 2, \ldots, m-1$$

$f \in H_1 \Rightarrow M_\nu f = 0,\ \nu = 0, 1, \ldots$

Let H_0 be the m-dimensional space spanned by $k_0, k_1, \ldots, k_{m-1}$ with

$$\langle f, g\rangle = \sum_{\nu=0}^{m-1} (M_\nu f)(M_\nu g).$$

$M_\nu k_\mu = 1$ for $\nu = \mu,\ = 0$ otherwise so that $k_0, k_1, \ldots, k_{m-1}$ are an orthonormal basis.

Let $W_m = H_0 \oplus H_1$.

$$\|f\|^2_{W_m} = \sum_{\nu=0}^{m-1} (M_\nu f)^2 + \int_0^1 \left(f^{(m)}(u)\right)^2 du$$

H_0 and H_1 are orthogonal subspaces since $\left(f^{(m)}(u)\right)^2 = 0$ for $f \in H_0$ and $M_\nu f = 0$ for $f \in H_1$.

$$J(f) = \int_0^1 \left(f^{(m)}(u)\right)^2 du = \|P_1 f\|^2$$

where P_1 is the orthogonal projection in W_m onto H_1.

$J(f)$ is a *semi-norm* on $W_m[0,1]$

Solve the second variational problem: Then it can be applied to the spline smoothing problem
by letting:

$f_1 = P_1 f, \quad \|P_1 f\|^2 = \int_0^1 (f^{(m)}(u))^2 du$

$L_i f = f(t_i)$

FIND $f \in W_m$ to min

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - f(t_i))^2 + \lambda \int_0^1 (f^{(m)}(u))^2 du$$

Notation: Here we let H_0 be of dimension M and spanned by $\phi_1, \cdots, \phi_M$. For the particular spline case, $M = m$, the $k_\nu, \nu = 0, ..., m-1$ will span H_0, and $L_i f = f(x_i)$. Let $L_i f = < \eta_i, f >, \quad f \in \mathcal{H}$
and let

$$< \eta_i, \phi_\nu >= t_{i\nu}; \quad T_{n \times M} = \{t_{i\nu}\}$$

be of rank M. This means that if we assume $f \in H_0$, the least squares problem has a unique solution: if $f = \sum_{\nu=1}^{M} d_\nu \phi_\nu$, then

$$min_{d_\nu} \sum_{i=1}^{n} (< \eta_i, \sum_{\nu=1}^{M} d_\nu \phi_\nu > - y_i)^2$$

has a unique minimizer

$$min \|Td - y\|^2 \text{ unique}$$

$$d = (T'T)^{-1} T' y$$

THEOREM

Let T have full column rank. Then the second variational problem (*) has a unique minimizer in $\mathcal{H}$ for every $\lambda > 0$ and f_λ has a representation

$$f_\lambda = \sum_{\nu=1}^{M} d_\nu \phi_\nu + \sum_{i=1}^{n} c_i \xi_i$$

where $\xi_i = P_1 \eta_i, i = 1, \cdots, n$.

In the case of the example $T = \{t_{i\nu}\}$ with

$$t_{i\nu} = k_{(\nu-1)}(x_i), \quad \nu = 1, \cdots, m$$

T will be of full rank m if there are at least m distinct values of the x_i's. (polynomial interpolation is unique)

ARGUMENT

Claim: $f_\lambda = \sum_{\nu=1}^{M} d_\nu \phi_\nu + \sum_{i=1}^{n} c_i \xi_i$
$< \phi_\nu, \xi_i >= 0$, since $\xi_i = P_1 \eta_i \in \mathcal{H}_1 \perp \mathcal{H}_0$.

Let $\Sigma_{n\times n} = \{< \xi_i, \xi_j >\}$ and suppose $\Sigma \succ 0$, then

$$\phi_1, \cdots, \phi_M, \xi_1, \cdots, \xi_n$$

span an $n + M$ dimensional subspace of $\mathcal{H}$, and any $f \in \mathcal{H}$ can be written

$$f = \sum_{\nu=1}^{M} d_\nu \phi_\nu + \sum_{i=1}^{n} c_i \xi_i + \rho$$

for some $d = (d_1, \cdots, d_M)'$, $(c_1, \cdots, c_n)' = c$, with $< \rho, \phi_\nu >= 0 =< \rho, \xi_i >$, all i, ν.

$P_1 f = \sum_{i=1}^{n} c_i \xi_i + \rho$
(since $\rho \perp \phi_\nu, \nu = 1, \cdots, M$)

$\|P_1 f\|^2 = c' \Sigma c + \|\rho\|^2$ since $< \rho, \xi_i >= 0$.

Let P_0 be the orthogonal projection onto $\mathcal{H}_0$. Then

$$
\begin{aligned}
< \eta_i, \xi_j > &= < \eta_i - P_0 \eta_i, \xi_j > \\
&= < \xi_i, \xi_j > \qquad : \Sigma = \{< \xi_i, \xi_j >\} \\
< \eta_i, \phi_\nu > &= t_{i\nu}, \qquad T = \{t_{i\nu}\}.
\end{aligned}
$$

The second variational problem can then be written

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - L_i f)^2 + \lambda \|P_1 f\|^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} (y_i - < \eta_i, \sum_{\nu=1}^{M} d_\nu \phi_\nu + \sum_{j=1}^{n} c_j \xi_j + \rho >)^2$$

$$+ \lambda [c' \Sigma c + \|\rho\|^2] \qquad \text{note } \rho \perp \xi_i$$

$$= \frac{1}{n} \|y - Td - \Sigma c\|^2 + \lambda [c' \Sigma c + \|\rho\|^2]$$

$\|\rho\|^2 = 0$, find d, c to minimize

$$\frac{1}{n}\|y - Td - \Sigma c\|^2 + \lambda c'\Sigma c$$

Differentiate with respect to d and c,

let $M = (\Sigma + n\lambda I)$

$$c = M^{-1}(I - T(T'M^{-1}T)^{-1}T'M^{-1})y \quad (*)$$

$$d = (T'M^{-1}T)^{-1}T'M^{-1}y \quad (**)$$

from Kimeldolf and Wahba (1971).

DONT USE THIS TO COMPUTE!!

multiply left and right of (*) by M to get

$$\boxed{\text{Mc}} = y - T(T'M^{-1}T)^{-1}T'M^{-1}y = \boxed{\text{y - Td}}$$

$$\boxed{\text{T'c}} = T'M^{-1}y - T'M^{-1}T(T'M^{-1}T)^{-1}T'M^{-1}y \equiv \boxed{0}$$

$$\boxed{\begin{aligned}(\Sigma + n\lambda I)c + Td &= y\\ T'c &= 0\end{aligned}}$$

$n + M$ equations in $n + M$ unknowns.

[DONT NEED $\Sigma \succ 0$].

To solve for c:

$$T = \begin{pmatrix} \overset{M}{\,} & & \overset{n-M}{\,} \\ Q_1 & : & Q_2 \\ & & \end{pmatrix} \begin{pmatrix} \overset{M}{R} \\ - - - \\ 0 \end{pmatrix} \begin{matrix} M \\ \\ n-M \end{matrix}$$

$Q = (Q_1 : Q_2)$, $Q'Q = I$ (orthogonal).

THE $Q - R$ decomposition

R is upper triangular

span$\{\tau_1, \tau_2, \cdots, \tau_M\}$ $\quad$ τ_ν columns of T

= span$\{$columns of $Q_1\}$

$c = Q_2\gamma$, for some $\gamma \in E_{n-M}$

since columns of Q_2 are $\perp$ to columns of Q_1 and (hence) $\perp$ to columns of T' $\quad \Rightarrow \quad T'c = 0$.

$$(\Sigma + n\lambda I)c + Td \;=\; y$$

Let $c = Q_2\gamma$

$$\begin{aligned}(\Sigma + n\lambda I)Q_2\gamma + Td &= y \\ Q_2'(\Sigma + n\lambda I)Q_2\gamma &= Q_2'y\end{aligned}$$

$$\gamma = [Q_2'(\Sigma + n\lambda I)Q_2]^{-1}Q_2'y$$

$$c = Q_2[Q_2'(\Sigma + n\lambda I)Q_2]^{-1}Q_2'y$$

$$\begin{aligned}Td &= y - Mc \\ Q_1'Td &= Q_1'(y - Mc)\end{aligned}$$

Use $T = Q_1R$ to get

$$Rd = Q_1'(y - Mc)$$

Find $f \in W_m$ to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda \int_0^1 (f^{(m)}(u))^2 du$$

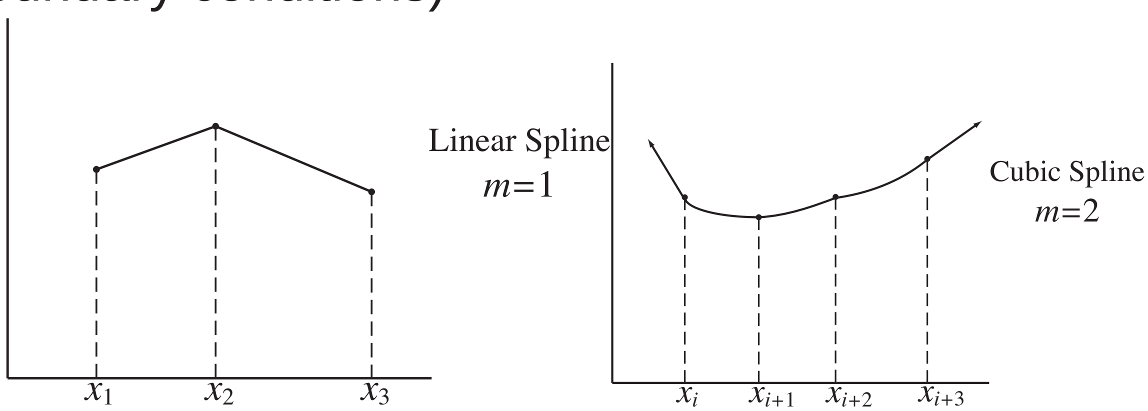
$x_i \in [0, 1]$

I. Schoenberg: (1940's): Solution f_λ is a "natural polynomial spline" of degree $2m - 1$.

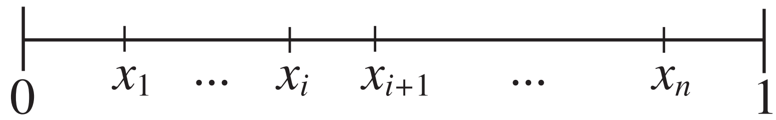
(1) $f_\lambda(s) \in \pi_{2m-1}$ in each interval $[x_i, x_{i+1}], i = 1, \cdots, n$

(2) $f_\lambda \in C^{2m-2}$ ($2m - 2$ continuous derivatives)

(3) $f_\lambda \in \pi_{m-1}$ for $x \leq x_1$ and $x \geq x_n$ (the "natural" boundary conditions)



To match up coefficients and data



Coefficients of the piecewise polynomials

$$m + \underbrace{2m + \cdots + 2m}_{n-1} + m = 2nm$$

$f \in C^{2m-2}$, $f^{(\nu)}(x_{i+}) - f^{(\nu)}(x_{i\_}) = 0$
$\nu = 0, \cdots, 2m-2$
This gives $n(2m-1)$ conditions.

The “natural” polynomial spline of degree $2m-1$

has	$2mn$	coefficients
satisfying	$(2m-1)n$	conditions
	n	

It will be determined given its values at n points (Theorem) (assuming least squares in π_{m-1} is unique).

Polynomial splines, the hard way:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - L_i f)^2 + \lambda\|P_1 f\|^2$$

$L_i f = f(x_i), \|P_1 f\|^2 = \int_0^1 (f^{(m)}(u))^2 du$

$\mathcal{H} = W_m = H_0 + H_1$ where H_0 is spanned by $k_0, k_1, \cdots, k_{m-1}$ and the RK for H_1 is

$$\|f\|^2 = \int_0^1 \left(f^m(u)\right)^2 du$$

is the squared norm in the space with RK

$$k_m(s)k_m(t) + (-1)^{m-1}k_{2m}(\,[s-t]\,) \qquad (**)$$

Although it looks like the spline will be of degree $2m$, it can be shown that the condition $T'c = 0$ will guarantee that it is a piecewise polynomial of degree at most $2m - 1$

MATRIX DECOMPOSITIONS
(Golub and Van Loan)

$\Sigma_{n\times n} \succeq 0$	$\Gamma D \Gamma'$	Eigenvalue-Eigenvector Decomposition
$X_{n\times p}$	$\underset{n\times n\ n\times p\ p\times p}{UDV^T}$ $UU^T = I_n, VV^T = I_p$	Singular Value Decomposition

$$p < n, \quad D = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_p \\ 0 & \cdots & 0 \end{pmatrix}$$

$$T_{n\times M} \quad = \quad Q_{n\times n} \quad R_{n\times M}, \quad \text{Q-R}$$

$$\begin{pmatrix} Q \end{pmatrix} \begin{pmatrix} R \\ -- \\ 0 \end{pmatrix} \qquad QQ' = I_n$$

Q is orthogonal, R is upper triangular.

$$\Sigma \succ 0 \quad = \quad LL'$$

where L is lower triangular (Cholesky Factorization).

$A(\lambda)$ – the influence matrix

Very important

$$\hat{y} = \begin{pmatrix} L_1 f_\lambda \\ \vdots \\ L_n f_\lambda \end{pmatrix} = A(\lambda) y$$

Definition of the influence matrix.

$f_\lambda = \sum_{\nu=1}^{M} d_\nu \phi_\nu + \sum_{i=1}^{n} c_i \xi_i$

$$\begin{pmatrix} L_1 f_\lambda \\ \vdots \\ L_n f_\lambda \end{pmatrix} = \begin{pmatrix} < \eta_1, f_\lambda > \\ \vdots \\ < \eta_n, f_\lambda > \end{pmatrix} = Td + \Sigma c$$

From the equations for c, d

$$\begin{array}{llll} (\Sigma + n\lambda I)c & + \; Td & = & y \\ \Sigma c & + \; Td & = & A(\lambda) y \end{array}$$

So

$$n\lambda c = (I - A(\lambda)) y$$

$$n\lambda c = (I - A(\lambda))y$$

From earlier

$$T = \left(\begin{array}{ccc} Q_1 & | & Q_2 \end{array} \right) \left(\begin{array}{c} R \\ -- \\ 0 \end{array} \right)$$

$$c = Q_2[Q_2'(\Sigma + n\lambda I)Q_2]^{-1}Q_2'y$$

so

$$(I - A(\lambda)) = n\lambda Q_2[Q_2'(\Sigma + n\lambda I)Q_2]^{-1}Q_2'$$

Note $A(\lambda)T = T_{n\times M}$ since

$(I - A(\lambda))Q_1 = 0_{n\times M}$ since $Q_2'Q_1 = 0_{(n-M)\times M}$

columns of T are eigenvectors of A with eigenvalue 1.

What are the remaining eigenvalues of $I - A(\lambda)$?

$$
\begin{aligned}
I - A &= n\lambda Q_2[\underbrace{Q_2'\Sigma Q_2} + n\lambda I_{n-M}]^{-1}Q_2' \\
& \qquad Q_2'\Sigma Q_2 = UDU'_{(n-M)\times(n-M)} \\
& \qquad D_{n-M} \text{ has eigs. } \{d_\nu\},\ \nu = 1, \cdots, n-M \\
&= n\lambda Q_2[U(D + n\lambda I_{n-M})^{-1}U^T]Q_2' \\
&= \underbrace{Q_2 U}\ [diag(\tfrac{n\lambda}{n\lambda + d_\nu})]_{n-M} U' Q_2' \\
& \qquad n \times (n-M) \text{ orthogonal}
\end{aligned}
$$

Eigenvalues of $I - A$ are $\overbrace{0, \cdots, 0}^{M}$, $\{\frac{n\lambda}{n\lambda + d_\nu}\}_{\nu=1}^{n-M}$.

The eigenvalues of A:

$$\overbrace{1, \cdots, 1}^{M}, \ \frac{d_\nu}{n\lambda + d_\nu}, \quad \nu = 1, \cdots, n - M$$

A is a "smoother" matrix:

$$0 \preceq A(\lambda) \preceq I \quad \text{ALWAYS}$$

as $\lambda \to 0$, $A(\lambda) \to I$
as $\lambda \to \infty$, $A(\lambda) \to$ projection

operator onto columns of T

$AT = T \Rightarrow$ If $y = T\theta$ for some θ, then $Ay = y$

If you give it "EXACT" data from some $f \in \mathcal{H}_0$, it will give you f back, any λ.