Statistics 860 Lecture 11

Smoothing Spline ANOVA - to handle many (heterogenous) variables, different than rbf's.

Must read: Spline Models, Chapter 10
`wahba.wang.gu.95.pdf`
`chiang.wahba.tribbia.johnson.99.pdf`

Yuedong Wang: Smoothing Splines: Methods and Applications (Chapman & Hall/CRC Monographs on Statistics & Applied Probability) 2011 Examples keyed to the `assist` package in R.

Chong Gu, Smoothing Spline ANOVA Models Springer, 2002. Examples are keyed to `gss` code in R.

ANOVA = ANalysis of VAriance

ANOVA

Elementary version is taught in engineering statistics. Here it is generalized to RKHS and scattered, noisy data.

$t = (t_1, \cdots, t_d)$, $t_\alpha \in \mathcal{T}^{(\alpha)}, \alpha = 1, 2, \cdots, d$.

$\mathcal{E}_\alpha$ is an average operator on $\mathcal{T}^{(\alpha)}$, if for $f_\alpha(t_\alpha)$ defined on $\mathcal{T}^{(\alpha)}$,

$$\mathcal{E}_\alpha f_\alpha = \int_{\mathcal{T}^{(\alpha)}} f_\alpha(t_\alpha) d\mu_\alpha$$

where $d\mu_\alpha$ is some probability measure with support on $\mathcal{T}^{(\alpha)}$,

So $\{\mathcal{E}_\alpha, \mathcal{T}^{(\alpha)}, d\mu_\alpha\}$ define an averaging operator, which is well defined for all functions on $\mathcal{T}^{(\alpha)}$ integratable with respect to $d\mu_\alpha$.

Let $1^{(\alpha)}$ be the function which is 1 on $\mathcal{T}^{(\alpha)}$, then

$$\mathcal{E}_\alpha 1^{(\alpha)} = \int_{\mathcal{T}^{(\alpha)}} 1^{(\alpha)} d\mu_\alpha = 1^{(\alpha)}$$

$\mathcal{E}_\alpha$ is IDEMPOTENT, that is

$$\mathcal{E}_\alpha(\mathcal{E}_\alpha f_\alpha) = \mathcal{E}_\alpha f_\alpha$$

Examples

Ex1:
$\mathcal{T}^{(\alpha)} = [0, 1]$, $d\mu_\alpha$ = Lebesgue measure (i.e. the uniform density.)

$\mathcal{E}_\alpha f_\alpha = \int_0^1 f_\alpha(t_\alpha) d\mu_\alpha(t_\alpha) \cdot 1^{(\alpha)}$ (a constant times $1^{(\alpha)}$)

Ex2:
$\mathcal{T}^{(\alpha)} = E^2$, the two dimensional plane.
Let $s_\alpha(j) = (x_{1\alpha}(j), x_{2\alpha}(j)), j = 1, 2, \cdots, N$ be N points in E^2. Let $d\mu_\alpha$ assign weights $w_1, \cdots, w_N$ to these N points where $w_k > 0$ and $\sum_{k=1}^N w_k = 1$. This defines a probability distribution and hence an averaging operator on $\mathcal{T}^{(\alpha)}$ by

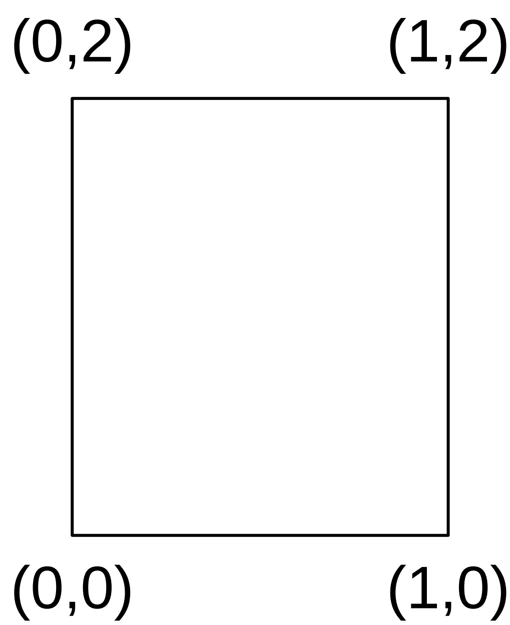
$$\mathcal{E}_\alpha f_\alpha = \sum_{j=1}^N w_j f_\alpha(s_\alpha(j)) \cdot 1^{(\alpha)}$$

This will be useful in combining thin plate spline penalty functionals in one variable with other penalty functionals on other variables. (Since it is not convenient (or desireable) to average over all of E^2.)

Note that this averaging operator is supported on a finite subset of E^2, but it will still be useful.

Next, we extend the domain of $\mathcal{E}_\alpha$ from $\mathcal{T}^{(\alpha)}$ to $\mathcal{T}^{(1)} \otimes \mathcal{T}^{(2)} \cdots \otimes \mathcal{T}^{(d)}$, let $t = (t_1, \cdots, t_d)$, $t_\alpha \in \mathcal{T}^{(\alpha)}, \alpha = 1, \cdots, d$.

Thus, $t \in \mathcal{T}^{(1)} \otimes \cdots \otimes \mathcal{T}^{(d)} \equiv \mathcal{T}$. For example, $\mathcal{T}^{(1)} = [0, 1], \mathcal{T}^{(2)} = [0, 2]$, then $\mathcal{T} = \mathcal{T}^{(1)} \otimes \mathcal{T}^{(2)}$ is the rectangle in E^2 with corners at $(0, 0), (1, 0), (0, 2)$ and $(1, 2)$. ("Cartesian Product" of $[0, 1]$ and $[0, 2]$).



Let $f(t)$, $t \in \mathcal{T} = \mathcal{T}^{(1)} \otimes \cdots \otimes \mathcal{T}^{(d)}$, then $\mathcal{E}_\alpha f = \int f(t_1, \cdots, t_{\alpha-1}, s_\alpha, t_{\alpha+1}, \cdots, t_d) d\mu_\alpha(s_\alpha)$ resulting in what looks like a a function which only depends on $t_\beta, \beta \neq \alpha$. We want to think of it as a function defined on $\mathcal{T} \in \mathcal{H}$, which, however, does not depend on the αth coordinate t_α.

$$(\mathcal{E}_1, \cdots, \mathcal{E}_d)f = \int_{\mathcal{T}^{(1)}} \cdots \int_{\mathcal{T}^{(d)}} f(s_1, \cdots, s_d) d_{\mu_1} \cdots d_{\mu_d} \cdot 1(\mathcal{T})$$

Where $1(\mathcal{T})$ is the function which is 1 on $\mathcal{T}$.

Here is the ANOVA decomposition, for $d = 2$, a function of 2 variables: NOTE that each variable t_α has its own domain $\mathcal{T}^{(\alpha)}$ and we have not assumed ANYTHING about this domain other than the existence of <u>some</u> probability measure on it.

$$I = (\mathcal{E}_1 + (I - \mathcal{E}_1))(\mathcal{E}_2 + (I - \mathcal{E}_2))$$

where I is the identity operator $If = f$.

$$I = \mathcal{E}_1\mathcal{E}_2 + \mathcal{E}_2(I - \mathcal{E}_1) + \mathcal{E}_1(I - \mathcal{E}_2) + (I - \mathcal{E}_1)(I - \mathcal{E}_2)$$

Note that AVERAGING OPERTORS as defined commute: $\mathcal{E}_1\mathcal{E}_2 = \mathcal{E}_2\mathcal{E}_1$.

Thus

$$f \equiv \mathcal{E}_1\mathcal{E}_2 f + \mathcal{E}_2(I - \mathcal{E}_1)f + \mathcal{E}_1(I - \mathcal{E}_2)f + (I - \mathcal{E}_1)(I - \mathcal{E}_2)f$$

$$f(t) \equiv \mu + f_1(t_1) + f_2(t_2) + f_{12}(t_1, t_2) \quad (*)$$

where

$$\begin{array}{lcl} \mu & = & \mathcal{E}_1\mathcal{E}_2 f \\ f_1 & = & \mathcal{E}_2(I - \mathcal{E}_1)f \\ f_2 & = & \mathcal{E}_1(I - \mathcal{E}_2)f \\ f_{12} & = & (I - \mathcal{E}_1)(I - \mathcal{E}_2)f \end{array}$$

Note that $\mathcal{E}_1 f_1 = \mathcal{E}_2 f_2 = \mathcal{E}_1 f_{12} = \mathcal{E}_2 f_{12} = 0 \quad (**)$ since $\mathcal{E}_1(\mathcal{E}_2(I - \mathcal{E}_1)) = \mathcal{E}_1\mathcal{E}_2 - \mathcal{E}_1\mathcal{E}_2\mathcal{E}_1 = 0$ etc.

So for any f on $\mathcal{T} = \mathcal{T}^{(1)} \otimes \mathcal{T}^{(2)}$ such that the averaging operators are well defined, f has a unique decomposition (*) satisfying the side condition (**).

The general d-dimensional ANOVA decomposition:

$$
\begin{aligned}
I &= \prod_{\alpha=1}^{d}(\mathcal{E}_\alpha + (I-\mathcal{E}_\alpha)) \\
&= \prod_{\alpha=1}^{d}\mathcal{E}_\alpha + \sum_{\alpha=1}^{d}(I-\mathcal{E}_\alpha)\prod_{\beta\neq\alpha}\mathcal{E}_\beta \\
&\quad + \sum_{\alpha<\beta}(I-\mathcal{E}_\alpha)(I-\mathcal{E}_\beta)\prod_{\gamma\neq\alpha\beta}\mathcal{E}_\gamma \\
&\quad + \cdots \\
&\quad + \prod_{\alpha=1}^{d}(I-\mathcal{E}_\alpha)
\end{aligned}
$$

$$
\begin{aligned}
f(t_1,\cdots,t_d) &= \mu + \sum_\alpha f_\alpha(t_\alpha) + \\
&\quad + \sum_{\alpha<\beta} f_{\alpha\beta}(t_\alpha, t_\beta) \\
&\quad + \cdots \\
&\quad + f_{12\cdots d}(t_1,\cdots,t_d)
\end{aligned}
$$

To deal with higher dimensions and avoid the "curse of dimensionality": truncate the decomposition somewhere.

Additive models (MAIN EFFECTS MODELS)

$$f(t_1, \ldots, t_\alpha) \simeq \mu + \Sigma_{\alpha=1}^{d} f_\alpha(t_\alpha)$$

Hastie and Tibshirani:

$$f(t_1, \ldots, t_d) = \mu + \sum_{\alpha=1}^{d} f_\alpha(x_\alpha) + f_{\alpha_1\alpha_2}(t_{\alpha_1}, t_{\alpha_2})$$

Main effects model plus a single 2-factor interaction

The model selection problem starts with deciding which terms to drop.

ANOVA-DECOMPOSITION of RKHS

$\mathcal{H}^{\alpha}$: RKHS of functions on $\mathcal{T}^{(\alpha)}$ with $1 (= 1^{(\alpha)}) \in \mathcal{H}^{\alpha}$

$1^{(\alpha)}$ is the function on $\mathcal{T}^{(\alpha)}$ which is 1 on $\mathcal{T}^{(\alpha)}$.

$$\mathcal{H}^{\alpha} = \underbrace{[1^{(\alpha)}]}_{\text{a 1-dimensional subspace}} \oplus \mathcal{H}^{(\alpha)}$$

$f \in \mathcal{H}^{(\alpha)} \Rightarrow \int_{\mathcal{T}^{(\alpha)}} f_{\alpha}(s_{\alpha}) d\mu_{\alpha}(s_{\alpha}) = 0$

Let $P_{\{1^{(\alpha)}\}}$ be the projection operator in $\mathcal{H}^\alpha$ onto $1^{(\alpha)}$ defined by

$$P_{\{1^{(\alpha)}\}} f = \int_{\mathcal{T}^{(\alpha)}} f_\alpha(s_\alpha) d\mu_\alpha(s_\alpha) \cdot 1^{(\alpha)} \qquad \in \mathcal{H}^\alpha$$

(maps $f \in \mathcal{H}^\alpha$ onto its "mean" times the constant function)

LEMMA (Prove for yourself)

If $R^\alpha(s_\alpha, t_\alpha)$ is the RK for $\mathcal{H}^\alpha$, then $R^{(\alpha)}(s_\alpha, t_\alpha) =$

$$\left(I - P_{\{1^{(\alpha)}\}(s_\alpha)}\right)\left(I - P_{\{1^{(\alpha)}\}(t_\alpha)}\right) R^\alpha(s_\alpha, t_\alpha)$$

is the RK for $\mathcal{H}^{(\alpha)}$. Here $P_{\{1^{(\alpha)}\}(s_\alpha)}$ means that $P_{\{1^{(\alpha)}\}}$ is applied to what follows as a function of s_α.

$$R^{(\alpha)}(s_\alpha, t_\alpha) =$$

$$\left(I - P_{\{1^{(\alpha)}\}(s_\alpha)}\right)\left(I - P_{\{1^{(\alpha)}\}(t_\alpha)}\right) R^\alpha(s_\alpha, t_\alpha)$$

$$\equiv \left(I - \mathcal{E}_{\alpha(s_\alpha)}\right)\left(I - \mathcal{E}_{\alpha(t_\alpha)}\right) R^\alpha(s_\alpha, t_\alpha)$$

Note that

$$\mathcal{E}_{\alpha(s_\alpha)} R^{(\alpha)}(s_\alpha, t_\alpha) = 0 \quad \mathcal{E}_{\alpha(t_\alpha)} R^{(\alpha)}(s_\alpha, t_\alpha) = 0$$

$\mathcal{H}^\alpha = [1^{(\alpha)}] \oplus \mathcal{H}^{(\alpha)}$ and $1^{(\alpha)}$ will be $\perp$ to $\mathcal{H}^{(\alpha)}$ with the square norm

$$\|f\|^2_{new} = |P_{\{1^{(\alpha)}\}} f|^2 + \|(I - P_{\{1^{(\alpha)}\}}) f\|^2_{\mathcal{H}^\alpha}$$

(This is not necessarily the norm you started with.)

EXAMPLE:

$$\mathcal{T}^{(\alpha)} = [0,1] \qquad \mathcal{E}_\alpha f = \int_0^1 f(\ldots x_\alpha \ldots) dx_\alpha$$

$\mathcal{H}^\alpha = W_m(per)$ periodic functions on $[0,1]$ with $f^{(m)} \in \mathcal{L}_2$

$$\|f\|^2 = \left[\int_0^1 f(x)dx\right]^2 + \int_0^1 (f^{(m)}(x))^2 dx$$

$$\mathcal{H}^\alpha = 1^{(\alpha)} \oplus \mathcal{H}^{(\alpha)} = \mathcal{H}_0 \oplus \mathcal{H}_1$$

$$\mathcal{H}^{(\alpha)} = \{f : \; f^{(\nu-1)}(1) - f^{(\nu-1)}(0) = 0,$$

$$\nu = 1, 2, \ldots, m, \text{ and } f^{(m)} \in \mathcal{L}_2\}$$

$$R(x,x') = 1 + \frac{(-1)^{m-1}}{(2m)!} B_{2m}([x - x']) = 1 + R_1(x,x')$$

(p. 22 of book)

$$B_{2m}(x) = (-1)^{m-1} 2 \cdot (2m)! \sum_{\nu=1}^{\infty} \frac{1}{(2\pi\nu)^{2m}} \cos 2\pi\nu x$$

Easy case since $(\mathcal{E}_\alpha f)1^{(\alpha)}$ is the projection of f onto $\mathcal{H}_0$

$$\mathcal{E}_{\alpha(x)} R_1(x, x') = \mathcal{E}_{\alpha(x')} R_1(x, x') = 0$$

Decomposition of

$$\begin{aligned}\mathcal{H} &= \prod_\alpha \mathcal{H}^\alpha \quad = \quad \prod_\alpha([1^{(\alpha)}] \oplus \mathcal{H}^{(\alpha)}) \\ &= [1] + \Sigma_\alpha \mathcal{H}^{(\alpha)} + \Sigma_{\alpha<\beta} \mathcal{H}^{(\alpha)} \otimes \mathcal{H}^{(\beta)} + \\ &\quad \ldots + \prod_\alpha \otimes \mathcal{H}^{(\alpha)}\end{aligned}$$

with some abuse of notation—strictly speaking, should use

$$\prod_{\beta \neq \alpha} [1^{(\beta)}] \otimes \mathcal{H}^{(\alpha)}$$

which is a subspace of $\mathcal{H}$: $f(t_1, \ldots, t_d) \in \mathcal{H}^{(\alpha)}$ means that it only depends on t_α

FIRST ANOVA VARIATIONAL PROBLEM IN $\mathcal{H}$:
Let $d = 2, \quad t = (t_1, t_2)$

$$\mathcal{H} = [1] \oplus \mathcal{H}^{(1)} \oplus \mathcal{H}^{(2)} \oplus (\mathcal{H}^{(1)} \otimes \mathcal{H}^{(2)})$$

Given the RK's for $\mathcal{H}^{(1)}$ and $\mathcal{H}^{(2)}$

$R^{(1)}(s_1, t_1) =$ the RK for $\mathcal{H}^{(1)}$
$R^{(2)}(s_2, t_2) =$ the RK for $\mathcal{H}^{(2)}$
$R^{(1)} \otimes R^{(2)} =$ the RK for $\mathcal{H}^{(1)} \otimes \mathcal{H}^{(2)}$
$(R^{(1)} \otimes R^{(2)})(s, t) = R^{(1)}(s_1, t_1) R^{(2)}(s_2, t_2)$

GOAL to find

$$f = \mu + f_1 + f_2 + f_{12}$$

with $f_\alpha \in \mathcal{H}^{(\alpha)},\ \alpha = 1, 2$ and $f_{12} \in \mathcal{H}^{(1)} \otimes \mathcal{H}^{(2)}$
to minimize

$$\Sigma(y_i - L_i f)^2 + \lambda_1 \|f_1\|^2_{\mathcal{H}^{(1)}} + \lambda_2 \|f_2\|^2_{\mathcal{H}^{(2)}} + \lambda_{12} \|f_{12}\|^2_{\mathcal{H}^{(1)} \otimes \mathcal{H}^{(2)}}$$

$\lambda_{12} = \infty \longrightarrow$MAIN EFFECTS MODEL

To come: Spline ANOVA and the second variational problem: Allows a bigger null space than just the constant function, so multivariate parametric models will be a special case, as smoothing parameters get large. See Lin, Wahba, Zhang, Gao, Klein and Klein (2000), `lin:wahba:zhang:gao:klein:klein2000.pdf`. Note that there are yes/no variables there. In that case $\mathcal{T}^{(\alpha)}$ for one of those variables consists of two points, ± 1.