5) Nelinearna regresija

Ako je $f^*(x) = \mathbb{E}[Y \mid X = x]$ nelinearna f-ja, klasičan lin. model (penaliziran ili ne) može imati veliku pristranost

U nastavku, pretp. $\boxed{p = 1}$ (tj. $\boxed{X \in \mathbb{R}}$) te promatramo dva pristupa:

- splajnovi
- lokalna regresija

5.1 Splajnovi

za "bazne" f-je $h_m : \mathbb{R} \to \mathbb{R}$, $m = 1, \ldots, M$, jedan pristup je pretpostaviti da je

fiksne!

$$f^*(x) = \sum_{m=1}^{M} \beta_m h_m(x), \quad x \in \mathbb{R}, \qquad (5.1)$$

"linear basis expansion" za f^*

za neke $\beta = (\beta_1, \ldots, \beta_M) \in \mathbb{R}^M$.

→ zapravo lin. model s kovarijatama $h_1(X), \ldots, h_M(X)$!

[npr.] za $h_m(x) = x^{m-1}$, $m = 1, \ldots, d+1$ → "polinomijalna regresija reda d"

[Ipak, češće koristimo po dijelovima polinomijalnu aproksimaciju!]

za fiksne "čvorove" $\underset{=:t_0}{a} < t_1 < t_2 < \ldots < t_K < \underset{=:t_{K+1}}{b}$, $\boxed{\text{splajn}}$ reda $d \in \mathbb{N}$ (s čvorovima $t_1, \ldots, t_K$) je svaka f-ja $g : [a,b] \to \mathbb{R}$ t.d.

(i) g je polinom reda d na $C_k := [t_{k-1}, t_k]$, $\forall k = 1, \ldots, K+1$

(ii) $g^{(k)}$ je neprekidna na $\langle a, b \rangle$, $\forall k = 0, 1, \ldots, d-1$

(neprekidnost; k-ta derivacija)

→ zbog (i) dovoljno je provjeriti

$$g^{(k)}(t_i -) = g^{(k)}(t_i +), \quad \forall i = 1, \ldots, K, \ \forall k = 0, \ldots, d-1 \qquad (5.2)$$

Uočimo, splajnovi reda d s K čvorova čine <u>vektorski prostor</u> dimenzije

$$(K+1 \text{ intervala}) \cdot (d+1 \text{ parametar}) - (K \text{ čvorova}) \cdot (d \text{ uvjeta}) \quad (5.2)$$

$$= \boxed{d + K + 1} \quad (5.3)$$

$\boxed{\text{Pr. 5.1}}$ (Primjeri baza)

(i) $\boxed{d = 0}$ → po dijelovima <u>konstantne</u> f-je ("step-f-je")

↳ npr. $h_m(x) = \mathbb{1}_{C_m}(x)$, $m = 1, \dots, \underbrace{K+1}_{= d+K+1}$

$[d-1=0]$

(ii) $\boxed{d = 1}$ → <u>neprekidne</u> po dijelovima <u>lin.</u> f-je → $d+K+1 = \boxed{K+2}$

↳ npr. $h_1 \equiv 1$, $h_2(x) = x$,

$h_{2+m}(x) = (x - t_m)_+$, za $m = 1, \dots, K$

mijenjamo nagib u t_m

<u>(iii)</u> $\boxed{d = 3}$ → "kubični splajn" → $d+K+1 = \boxed{K+4}$

najčešće!

↳ npr. $h_1(x) = 1$, $h_2(x) = x$, $h_3(x) = x^2$, $h_4(x) = x^4$

$h_{4+m}(x) = (x - t_m)_+^3$, za $m = 1, \dots, K$

<u>Regresijski</u> splajn: za dani d i K, modeliramo $(M = d+K+1)$

$$f^*(x) = \sum_{m=1}^{M} \beta_m h_m(x),$$

za <u>neku</u> bazu $h_1, \dots, h_M$, tj. $\mathbb{F} := \begin{bmatrix} h_1(x^{(1)}) & \cdots & h_M(x^{(1)}) \\ \vdots & & \\ h_1(x^{(n)}) & \cdots & h_M(x^{(n)}) \end{bmatrix} \in \mathbb{R}^{n \times M}$ te

$$\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_M) := \underset{\beta \in \mathbb{R}^M}{\operatorname{argmin}} \|y - \mathbb{F} \cdot \beta\|_2^2 = \hat{\beta}^{ls}, \quad (5.4)$$

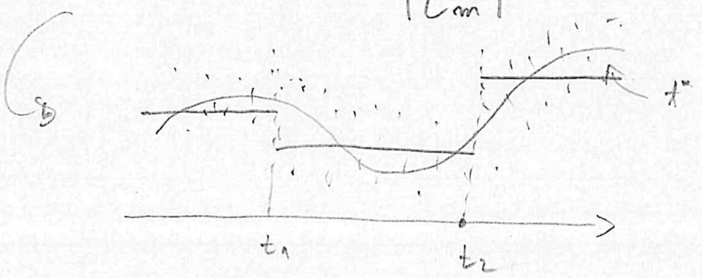
$$\hat{f}(x) := \sum_{m=1}^{M} \hat{\beta}_m \cdot h_m(x), \quad x \in [a, b]$$

Nap. 5.2 (i) $\hat{\beta}$ ovisi o bazi, ali $\hat{y}$ ne!

(ii) df (regr. splajn) $= M \; (= d+K+1)$ ∎

(DZ) Sve kao u Pr. 5.1 (i) → pokažite da je

$$\hat{\beta}_m = \frac{\sum_{i=1}^{n} y_i \mathbb{1}_{\{x^{(i)} \in C_m\}}}{|C_m|}, \quad m = 1, \dots, K+1.$$



Za x blizu a i b, $\mathrm{Var}(\hat{f}(x))$ može biti velika u slučaju kubičnog splajna, zbog malo podataka.

Prirodni kubični splajn je kubični splajn g koji je linearan na $\underbrace{[a, t_1]}_{C_0}$ i $\underbrace{[t_K, b]}_{C_K}$, tj. $g^{(k)} \equiv 0$ na C_0 i C_K, za $(k = 2, 3)$.

→ 4 dodatna uvjeta $\Rightarrow \underline{\underline{df}} = 3 + K + 1 - 4 = \underline{\underline{K}}$ → [za isti df, više unutarnjih čvorova!]

→ smanjuje $\mathrm{Var}(\hat{f}(x))$ za rubne x-eve (ali povećava pristranost)

→ g na C_0 (odn. C_K) potp. određena s g na C_1 (odn. C_{K-1}) (DZ) $\Rightarrow$ tipično $t_1 := \min_{i=1,\dots,n} x^{(i)}$, $t_K := \max_i x^{(i)}$.

Nap. 5.3 Ključan je odabir K i $t_1, \dots, t_K$ (ako smo fix. $d = 3$).

Npr. $t_1, \dots, t_K$ možemo postaviti na kvantile podataka $x^{(1)}, \dots, x^{(n)}$, a K (tj. df) odabrati CV metodom ∎

pristup (20)

(5.1.1) Smoothing spline

• $S[a,b] := \{ g: [a,b] \to \mathbb{R} : g$ dva puta neprekidno diferenc. $\}$

↳ Tražimo $f \in S[a,b]$ koja minimizira

$$RSS(f,\lambda) = \sum_{i=1}^{n} (y_i - f(x^{(i)}))^2 + \boxed{\lambda \int_a^b (f''(t))^2 \, dt} \quad (5.5)$$

Za $\lambda \geq 0$ fiksan.

više penaliziramo zakrivljenje f-je (f linearna $\Rightarrow f'' \equiv 0$)

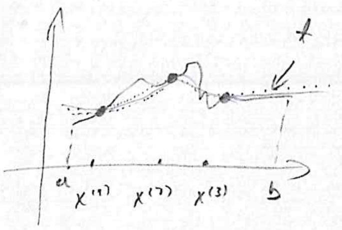
Tm. 5.4 | Pretp. da je $n \geq 2$ te $x^{(1)} < \dots < x^{(n)}$ (dakle, $x^{(i)} \neq x^{(j)}$, $\forall i \neq j$).

$\Rightarrow$ F-ja $f \in S[a,b]$ koja minimizira $RSS(f,\lambda)$ je nužno prirodni kubični splajn s čvorovima $\boxed{t_i = x^{(i)}, i = 1, \dots, n}$, $\forall \lambda > 0$.

Dokaz | Neka je $f \in S[a,b]$ proizvoljna. Može se pokazati da $\exists!$ prirodni kub. splajn $\tilde{f}$ s čvorovima $x^{(1)}, \dots, x^{(n)}$ t.d. $\tilde{f}(x^{(i)}) = f(x^{(i)}), \forall i = 1, \dots, n$ (jer $x^{(i)} \neq x^{(j)}, \forall i \neq j$ te je $\tilde{f}$ određena s n parametara.)

↳ $h := f - \tilde{f} \in S[a,b]$

$\Rightarrow \boxed{h(x^{(i)}) = 0, \forall i = 1, \dots, n}$



Tvrdnja slijedi ako pokažemo da

$$\boxed{\int_a^b (f''(t))^2 \, dt \geq \int_a^b (\tilde{f}''(t))^2 \, dt}, \quad (5.6)$$

i ćemu vrijedi jednakost akko $f \equiv \tilde{f}$.

$$\int_a^b (f''(t))^2 dt \underset{\boxed{f = h + \tilde{f}}}{=} \int_a^b \{ h''(t)^2 + \tilde{f}''(t)^2 + \underline{2 h''(t) \tilde{f}''(t)} \} \, dt . \quad (5.7)$$

$$\int_a^b h''(t) \tilde{f}''(t) \, dt = \begin{bmatrix} du = h'' & v = \tilde{f}'' \\ u = h' & dv = \tilde{f}^{(3)} \end{bmatrix} = \underbrace{h' \cdot \tilde{f}'' \Big|_a^b}_{= 0 \text{ je } \tilde{f}''(b) = \tilde{f}''(a) = 0} - \int_a^b h'(t) \tilde{f}^{(3)} \, dt$$

($\tilde{f}$ lin. na $[a, x^{(1)}]$, $[x^{(n)}, b]$)

$= [$ na svakom $[x^{(i)}, x^{(i+1)}]$, $i=1,\dots,n-1$, $\hat{f}^{(3)}$ je konstantna i jednaka $\hat{f}^{(3)}(x^{(i)}+)$, te $\hat{f}^{(3)} \equiv 0$ na $[a, x^{(1)}]$ i $[x^{(n)}, b]$ $]$

$$= -\sum_{i=1}^{n-1} \hat{f}^{(3)}(x^{(i)}+) \underbrace{\int_{x^{(i)}}^{x^{(i+1)}} h'(t)\,dt}_{= h(x^{(i+1)}) - h(x^{(i)}) = 0-0=0} = 0.$$

$$\Rightarrow \int_a^b f''(t)^2\,dt = \int_a^b \hat{f}''(t)^2\,dt + \underbrace{\int_a^b h''(t)^2\,dt}_{\geq 0} \qquad (5.7)$$

$\Rightarrow$ vrijedi (5.6) pri čemu imamo jednakost akko

$$0 = \int_a^b h''(t)^2\,dt \underset{h'' \text{ neprekidna}}{\iff} h'' \equiv 0 \text{ na } [a,b]$$

$$\iff h(t) = \alpha + \beta t \text{ za neke } \alpha, \beta \in \mathbb{R}$$

$$\underset{\substack{h(x^{(1)}) = h(x^{(2)}) = 0 \\ (n \geq 2)}}{\iff} h \equiv 0, \text{ tj. } f \equiv \hat{f} \text{ na } [a,b]. \qquad \blacksquare$$

Tm. 5.4 $\Rightarrow$ rješenje f od (5.5) tražimo u obliku $\boxed{f_\beta = \sum_{j=1}^{m} \beta_j \cdot h_j}$ gdje je $h_1, \dots, h_m : \mathbb{R} \to \mathbb{R}$, proizvoljna baza za prostor prirodnih splajnova s čvorovima $x^{(1)}, \dots, x^{(n)}$ [f je dakle prirodni splajn].

Za $\mathbb{F} \in \mathbb{R}^{n \times m}$ kao u (5.4), tj. $\mathbb{F}_{ij} = h_j(x^{(i)})$,

$$\sum_{i=1}^{n} (y_i - \underbrace{f_\beta(x^{(i)})}_{= \sum_{j=1}^{m} \beta_j h_j(x^{(i)})})^2 = \|y - \mathbb{F} \cdot \beta\|_2^2,$$

te

$$\boxed{\int_a^b f_\beta''(t)^2\,dt} = \int_a^b \Big(\sum_{j=1}^{m} \beta_j h_j''(t) \Big)^2 dt \qquad (5.8)$$

$$= \sum_{j,k=1}^{m} \beta_j \beta_k \underbrace{\int_a^b h_j''(t) h_k''(t)\,dt}_{=: \Omega_{jk}} =: \boxed{\beta^T \Omega \beta}$$

za $\Omega \in \mathbb{R}^{n \times n}$.

$\Rightarrow \forall \lambda \geq 0, \ \forall \beta \in \mathbb{R}^n,$

$$RSS(f_\beta, \lambda) = \| y - \mathbb{F}\beta \|_2^2 + \boxed{\lambda \beta^T \Omega \beta} \qquad (5.9)$$

generalizirana ridge regresija

$$\hat{\beta}_\lambda := \underset{\beta \in \mathbb{R}^n}{\operatorname{argmin}} \, RSS(f_\beta, \lambda) \overset{DZ}{=} (\mathbb{F}^T \mathbb{F} + \lambda \Omega)^{-1} \mathbb{F}^T y, \qquad (5.10)$$

a $f_{\hat{\beta}_\lambda}$ zovemo "<u>smoothing spline</u>" (SS) za dane $\mathcal{T}$ i λ.

Uočimo, opet je

$$\hat{y} := \underbrace{\mathbb{F} \cdot \hat{\beta}_\lambda}_{= f_{\hat{\beta}}(x)} = \underbrace{\mathbb{F} (\mathbb{F}^T \mathbb{F} + \lambda \Omega)^{-1} \mathbb{F}^T}_{=: S_\lambda \in \mathbb{R}^{n \times n}} y = \boxed{S_\lambda} \cdot y \qquad (5.11)$$

ne ovisi o y!

<u>Reinschova forma</u>

<u>Nap. 5.5</u> Za $\mathbb{F} = \begin{bmatrix} h_1(x^{(1)}) & \cdots & h_n(x^{(1)}) \\ \vdots & & \vdots \\ h_1(x^{(n)}) & \cdots & h_n(x^{(n)}) \end{bmatrix} \in \mathbb{R}^{n \times n}$ imamo

$\operatorname{rang}(\mathbb{F}) = n \Leftrightarrow \operatorname{Im}(\mathbb{F}) = \mathbb{R}^n \Leftrightarrow \forall y = (y_1, \ldots, y_n) \in \mathbb{R}^n$

$\exists \beta_1, \ldots, \beta_n \in \mathbb{R}$ t.d.

$$y = \sum_{m=1}^{n} \beta_m \cdot \begin{bmatrix} h_m(x^{(1)}) \\ \vdots \\ h_m(x^{(n)}) \end{bmatrix}, \text{ tj.}$$

$$y_i = \sum_{m=1}^{n} \beta_m h_m(x^{(i)}), \quad \forall i = 1, \ldots, n \qquad (5.12)$$

$$= f_\beta(x^{(i)})$$

prirodni kubični splajn s čvorovima $x^{(1)}, \ldots, x^{(n)}$

$\Rightarrow$ Ako $\boxed{x^{(i)} \neq x^{(j)}, \ \forall i \neq j}$, $\forall y$ $\exists\, \hat{\beta}$ t.d. vrijedi (5.12)

$\Rightarrow$ u tom slučaju $\operatorname{rang}(\mathbb{F}) = n$, te je specijalno $\mathbb{F}^T\mathbb{F}$ pozitivno definitna. ("$\mathbb{F}^T\mathbb{F} > 0$")

Nadalje, Ω je pozit. semidefinitna (vidi (5.8)) pa je u slučaju $x^{(i)} \neq x^{(j)}, \forall i \neq j$, $\boxed{\mathbb{F}^T\mathbb{F} + \lambda\Omega \text{ poz. defin. } \forall \lambda \geq 0.}$

$\Rightarrow \exists\, (\mathbb{F}^T\mathbb{F} + \lambda\Omega)^{-1}$. ■

Uočimo,

$$
\begin{aligned}
S_\lambda &= \mathbb{F}(\mathbb{F}^T\mathbb{F} + \lambda\Omega)^{-1}\mathbb{F}^T \\
&= \mathbb{F}\left(\mathbb{F}^T\left(I_n + \lambda(\mathbb{F}^T)^{-1}\Omega\,\mathbb{F}^{-1}\right)\mathbb{F}\right)^{-1}\mathbb{F}^T \\
&= \cancel{\mathbb{F}\cdot\mathbb{F}^{-1}}\left(I_n + \lambda(\mathbb{F}^T)^{-1}\Omega\,\mathbb{F}^{-1}\right)^{-1}\cancel{(\mathbb{F}^T)^{-1}\mathbb{F}^T} \\
&= (I_n + \lambda K)^{-1}, \qquad (5.13)
\end{aligned}
$$

Reinschova forma

gdje $K := (\mathbb{F}^T)^{-1}\Omega\,\mathbb{F}^{-1} \in \mathbb{R}^{n\times n}$.

$\Rightarrow$ $K^T = K$ te $K \geq 0$ ($\Omega \geq$) pa $\exists\, U \in \mathbb{R}^{n\times n}$ t.d. $U^TU = I_n$ te $D = \operatorname{diag}(d_1, \ldots, d_n) \in \mathbb{R}^{n\times n}$ t.d. $0 \leq d_1 \leq d_2 \leq \cdots \leq d_n$, t.d. $K = UDU^T$.

$\Rightarrow$

$$S_\lambda \overset{(5.13)}{=} U(I_n + \lambda D)^{-1}U^T, \ \forall \lambda \geq 0 \qquad (5.14)$$

$\Rightarrow$ stupci $u_1, \ldots, u_n \in \mathbb{R}^n$ od U su svojstveni vektori od S_λ ($\forall \lambda$!), a odgovarajuće svojstv. vrijednosti su

$$\rho_j(\lambda) := \frac{1}{1 + \lambda d_j}, \ j = 1, \ldots, n \qquad (5.15)$$

- ONB $u_1, \dots, u_n$ za $\mathbb{R}^n$ je tzv. <u>Demmler-Reinschova baza</u>, a $\forall j = 1, \dots, n$, o vektoru u_j razmišljamo kao o <u>f-ji</u> sa $\mathbb{R}$ u $\mathbb{R}$ evaluiranoj u točkama $x^{(1)}, \dots, x^{(n)}$.

$$\forall \lambda \geq 0 \quad \underset{\text{smoothing spline}}{\hat{y}} = S_\lambda y \overset{(5.14)}{=} \sum_{j=1}^{n} \overbrace{\underline{\underline{\frac{1}{1+\lambda d_j}}} u_j^T y}^{\in \mathbb{R}} \cdot u_j \, . \quad (5.16)$$

veće smanjenje za <u>veće</u> d_j

odgovara f-jama u_j koje su <u>zakrivljenije</u>!

- $df(SS_\lambda) = tr(S_\lambda) \overset{(5.14)}{=} \sum_{j=1}^{n} \frac{1}{1+\lambda d_j}, \quad \forall \lambda \geq 0 \quad (5.17)$

$$\Rightarrow \quad df(SS_\lambda) \xrightarrow{\lambda \to 0} n \quad (\text{interpolacija})$$

$$df(SS_\lambda) \xrightarrow[\lambda \to \infty]{} \sum_{j=1}^{n} \mathbb{1}\{d_j \neq 0\} = 2$$

uvjeti: $\boxed{0 = d_1 = d_2} < d_3 \leq \dots$

$(\lambda = +\infty \Rightarrow f_{\hat{B}}'' \equiv 0$, tj. $f_{\hat{B}}$ je <u>linearna</u>!)

Odabir λ [za razliku od regresijskog splajna, potrebno odabrati samo λ!]

→ CV metoda

Može se pokazati da je LOOCV procjena greške

$$L_{CV}^{(n)}(f_{\hat{\beta}_\lambda}) := \frac{1}{n}\sum_{i=1}^{n}(y_i - f_{\hat{\beta}_\lambda}^{-i}(x^{(i)}))^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - f_{\hat{\beta}_\lambda}(x^{(i)})}{1-(S_\lambda)_{i,i}}\right) \quad (5.18)$$

→ $\boxed{\underline{\text{ne}} \text{ moramo računati } f_{\hat{\beta}_\lambda}^{-i}, \ \forall i = 1, \dots, n}$

($f_{\hat{\beta}_\lambda}^{-i}$ je SS dobiven iz $T \setminus \{(x^{(i)}, y_i)\}$)

Nap. 5.6 | (5.18) je istina za mnoge algoritme oblika $\hat{y} = S \cdot y$ (tzv. "linear smoother") → npr. vrijedi za LS lin. regresiju i ridge regresiju.

S ne ovisi o y

Nekad se (5.18) aproksimira s

$$GCV(f_{\hat{\beta}_\lambda}) := \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - f_{\hat{\beta}_\lambda}(x^{(i)})}{1-\frac{tr(S_\lambda)}{n}}\right)^2 \quad (5.19)$$

"generalizirana" unakrsna validacija

nekad jednostavnije za izračunati nego $(S_\lambda)_{ii}$, $\forall i = 1, \dots, n$.

→ splajnovi_ISLR.R

→ DR_bozo.R