Oulun yliopisto
Matemaattisten tieteiden laitos
Funktioiden estimointi
13. harjoitus, viikko 16, 2014

1. Todista luentojen s. 151 harjoitustehtäväksi jätetty seikka: matriisi

$$\Phi^T \Phi + n\lambda\Omega, \quad \lambda > 0$$

on positiivisesti definiitti. (Positiivisen semidefiniittisyyden todistaminen ei riitä tämän tehtävän ratkaisuksi!)

2. Tässä tehtävässä johdamme silottavalle kuutiolliselle splinille ristiinvalidoinnissa hyödyllisen identiteetin

$$\frac{1}{n}\sum_{i=1}^{n}[Y_i - \hat{m}_{n,-i}(x_i;\lambda)]^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{Y_i - \hat{m}_n(x_i;\lambda)}{1-[H(\lambda)]_{ii}}\right]^2, \qquad (*)$$

joka on voimassa, kun solmupisteet ovat eri suuria ja $n \geq 3$. Tässä $H(\lambda)$ on luentojen s. 152 määritelty matriisi, jolla on ominaisuus

$$\begin{bmatrix} \hat{m}_n(x_1;\lambda) \\ \vdots \\ \hat{m}_n(x_n;\lambda) \end{bmatrix} = H(\lambda) \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

ja $\hat{m}_n(x_i;\lambda)$ määritellään nyt hieman poikkeavasti lausekkeen $(**)$ minimoijana (ks. alla; siis normalisointitekijä on $1/n$ eikä $1/(n-1)$). Menettele seuraavasti

(a) Huomaa, että $H(\lambda)$ riippuu λ:sta ja solmupisteistä x_i, mutta ei arvoista Y_i.

(b) Huomaa, että $\hat{m}_{n,-i}(\cdot;\lambda)$ minimoi lausekkeen

$$\frac{1}{n}\sum_{\substack{j=1 \\ j\neq i}}^{n}(Y_j - g(x_j))^2 + \lambda \int [g''(x)]^2\,dx \qquad (**)$$

kaikkien C^2-funktioden g muodostamassa avaruudessa ja kulkee pisteen $(x_i, \hat{m}_{n,-i}(x_i;\lambda))$ kautta, ja päättele tästä, että

$$\hat{m}_{n,-i}(x_i;\lambda) = \sum_{j\neq i}[H(\lambda)]_{ij}Y_j + [H(\lambda)]_{ii}\,\hat{m}_{n,-i}(x_i;\lambda).$$

(c) Johda edellisestä tuloksesta identiteetti $(*)$.

3. Esitä funktio

$$m(x) = \begin{cases} 1, & 0 \le x \le \frac{1}{4} \\ 0, & \text{muuten} \end{cases}$$

Haarin kantafunktioiden $\{\varphi_0\} \cup \{\varphi_{jk} \mid j \ge 0, 0 \le k \le 2^j - 1\}$ avulla. (Kantafunktiot on määritelty luentojen s. 154.)

4. Osoita, että Haarin kantafunktioit ovat ortonormaaleja.

5. Olkoon $(\varphi_k)_{k \in \mathbb{N}}$ avaruuden $L^2([0,1])$ ortonormaali kanta, ja tarkastellaan regressiofunktion estimointia ortogonaalisarjakehtielmän avulla homoskedastisessa regressiotilanteessa

$$Y_i = m(X_i) + \sigma\varepsilon_i, \quad i = 1, \ldots, n$$

jossa parit (X_i, ε_i) ovat eri indekseillä riippumattomia ja samoin jakautuneita, jossa X_i ja ε_i ovat riippumattomia, X_i:n jakauma on välin $[0,1]$ tasainen jakauma, $E\varepsilon_i = 0$, $E\varepsilon_i^2 = 1$ ja $\sigma > 0$.

Oletetaan, että $m \in L^2([0,1])$. Tällöin se voidaan kehittää ortogonaalisarjaksi

$$m = \sum_{k=1}^{\infty} a_k \varphi_k,$$

ja eräs luonteva kertoimen a_k estimaattori on

$$\hat{a}_k = \frac{1}{n} \sum_{i=1}^{n} \varphi_k(X_i) Y_i.$$

Osoita, että $\hat{a}_k$ on harhaton.