10-11.5.22 תרגול #7 מודלים ב'

Local Polynomial Estimators:

* עבור מודל i.i.d. $(X_1,Y_1),\dots,(X_n,Y_n)$ עם

$$Y_j = f(X_j) + \varepsilon_j \quad ; \quad \varepsilon_j \overset{iid}{\sim} \quad ; \quad \mathbb{E}\varepsilon_j = 0$$

$$\left(\sigma^2 = 1 \text{ נניח } \text{..בהמשך}\right) \longleftarrow \mathbb{E}\varepsilon_j^2 = \sigma^2$$

אומד LP ל-f מסדר l יהיה:

$$\hat{f}(x) := \hat{\theta}(x)^T U(0)$$

$$U(t) = \left(1, t, \frac{1}{2!}t^2, \dots, \frac{1}{l!}t^l\right)^T \quad ; \quad \hat{\theta}(x) = B_n(x)^{-1} a_n(x)$$

(חיובי K)

$$a_n(x) = \frac{1}{nh}\sum_{j=1}^{n} Y_j U\left(\frac{X_j - x}{h}\right) K\left(\frac{x - X_j}{h}\right)$$

$$B_n(x) = \frac{1}{nh}\sum_{j=1}^{n} U\left(\frac{X_j - x}{h}\right) U\left(\frac{X_j - x}{h}\right)^T K\left(\frac{x - X_j}{h}\right)$$

ובאופן שקול ניתן לכתוב:

$$\hat{f}(x) = \sum_{j=1}^{n} Y_j W_{nj}(x) \qquad \text{כאשר}$$

$$W_{nj}(x) = \frac{1}{nh} U^T(0) B_n(x)^{-1} U\left(\frac{X_j - x}{h}\right) K\left(\frac{x - X_j}{h}\right)$$

* בנוסף, אם חישבנו חסם ל-MSE ובדרך לשם היה צריך להוכיח כמה טענות טכניות ולבצע המון חישובים.

* למה 2.1: עבור כל ערכי $X_1,\dots,X_n$ ופולינום Q עם $\deg(Q) \le l$,

$$\sum_{j=1}^{n} Q(X_j) W_{nj}(x) = Q(x) \quad ; \quad x \in [0,1]$$

ובפרט:

$$\sum_{j=1}^{n} W_{nj}(x) = 1 \quad ; \quad \sum_{j=1}^{n} (X_j - x)^k W_{nj}(x) = 0 \quad ; \quad 1 \le k \le l$$

הנחות:

(LP1) קיימים $n_0\in\mathbb{N}$, $\lambda_0>0$ כך ש-

$x\in[0,1]$, $\forall n\ge n_0$ $\quad \lambda_{min}(B_n(x))\ge\lambda_0$ ע"ע קטן ביותר.

(LP2) קיים $a_0>0$ כך שלכל קטע $A\subseteq[0,1]$, $n\ge1$:

$$\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}_{\{X_j\in A\}}\le a_0\cdot\left(\frac{1}{n}\vee Leb(A)\right)$$

(LP3) $supp(K)\subseteq[-1,1]$; $\|K\|_\infty<\infty$.

למה 2.2: אם (LP1)-(LP3) מתקיימות, אז לכל $n\ge n_0$, $h\ge\frac{1}{2n}$ ו- $x\in[0,1]$:

1. $|W_{nj}(x)|\le\frac{C}{nh}$

2. $\sum_{j=1}^{n}|W_{nj}(x)|\le C$

3. $W_{nj}(x)=0$ for $|X_j-x|>h$.

3:

ואז נגדיר $LP(l)$ עבור הנגזרת $f^{(s)}$ עם $s \le l$ של הפונק' הלא ידועה f בנקודה החדשה:

$$\hat{f}_{n,s}(x) = U^{(s)}(0)^T \hat{\theta}_n(x) h^{-s}$$

כאשר $U^{(s)}$ הוא הווקטור שהקורדינטות שלו הן הנגזרת ה-s של הקורדינטות המתאימות בווקטור: $U(t) = \left(1, t, \frac{t^2}{2!}, \dots, \frac{t^l}{l!}\right)$.

סעיף ב':

בהנחה ש- $B_n(x) > 0$, הראו כי האומד $\hat{f}_{n,s}(x)$ הוא ליניארי מהצורה:

$$\hat{f}_{n,s}(x) = \sum_{j=1}^{n} Y_j W^s_{nj}$$

ומצאו ביטוי ל- W^s_{nj}.

פתרון:

~~נזכור את שהבנינו בסעיף הקודם, הקורדינטה ה-s של $\hat{\theta}_n(x)$ תהיה הנגזרת ה-s מוכפלת ב-h^s~~

$$\cancel{\theta(x) = \left(f(x), f'(x)h, \dots, f^{(l)}(x)h^l\right)^T}$$

~~ו- נגזור~~ $\cancel{U'(t) = \left(0, 1, \frac{2t}{2!}, \frac{3t^2}{3!}, \dots, \frac{lt^{l-1}}{l!}\right)}$

$$\cancel{U^{(s)}(t) = (\underbrace{0, \dots, 0}_{s-1}, 1, \dots)}$$

$$\cancel{U^{(s)}(0) = e_s \quad (s \text{ unit vector})}$$

אם $B_n(x)$ הפיכה, אז:

$$\hat{\theta}_n(x) = B_n^{-1}(x) a_n = B_n^{-1}(x) \cdot \frac{1}{nh} \sum_{j=1}^{n} Y_j U\left(\frac{X_j - x}{h}\right) K\left(\frac{x - X_j}{h}\right)$$

$$= \sum_{j=1}^{n} Y_j \frac{1}{nh} B_n^{-1}(x) U\left(\frac{X_j - x}{h}\right) K\left(\frac{x - X_j}{h}\right)$$

$$\Rightarrow \hat{f}_{n,s}(x) = U^{(s)}(0)^T \hat{\theta}_n(x) h^{-s} = \sum_{j=1}^{n} Y_j \cdot \underbrace{\frac{1}{nh^{s+1}} \cdot U^{(s)}(0)^T B_n(x)^{-1} U\left(\frac{X_j - x}{h}\right) K\left(\frac{x - X_j}{h}\right)}_{W^s_{nj}}$$

סעיף ב:

הראו שהאומד $\hat{f}_{n,s}$ משחזר את הנגזרת מסדר s של פולינומים, מדרגה l לכל היותר ולכן המשקולות מספקות את' מקיימות:

$$\sum_{i=1}^{n}(X_i-x)^k W_{ni}^{s}(x)=\begin{cases}0 & ,\ k\in\{0,\dots,l\}\setminus\{s\}\\ s! & ,\ k=s\end{cases}$$

הערה: כשאנחנו פולינומים הכוונה היא שכאשר מכניסים את $\hat{f}_{n,s}$ על $Y_j=Q(X_j)$ כך ש- $\deg(Q)\le l$ פולינום Q אז האומד מחזיר בדיוק את הנגזרת ה- s.

פתרון:

אם Q פולינום מדרגה l לכל היותר,

$$Q(X_i)=Q(x)+Q'(x)(X_i-x)+\dots+\frac{1}{l!}Q^{(l)}(x)(X_i-x)^l=q(x)^T U\left(\frac{X_i-x}{h}\right)$$

כאשר

$$q(x)=\left(Q(x),Q'(x)h,\dots,Q^{(l)}(x)h^l\right)$$

לכן, כלומר המינימייזר יתקבל עבור $\hat{\theta}_n(x)=q(x)$.

$$\left\{\hat{\theta}(x)=\underset{\theta\in\mathbb{R}^{l+1}}{\arg\min}\sum_{j=1}^{n}\left(Y_j-\theta^T U\left(\frac{X_j-x}{h}\right)\right)^2 K\left(\frac{x-X_j}{h}\right)\quad :\text{ואז}\quad Y_j=Q(X_j)\ \text{כי}\right\}$$

נרצה להראות ש

$$\sum_{j=1}^{n}Q(X_j)W_{nj}^{s}(x)=Q^{(s)}(x)$$

אבל לפי סעיף א,

$$\sum_{j=1}^{n}Q(X_j)W_{nj}^{s}(x)=U^{(s)}(0)^T\underbrace{\hat{\theta}_n(x)}_{=q(x)}h^{-s}$$

נחשב את $U^{(s)}(0)$:

$$U_j^{(s)}(x)=\frac{d^s}{dx^s}\frac{1}{j!}x^j=\begin{cases}0 & s>j\\ \frac{1}{(j-s)!}x^{j-s} & ,\ s\le j\end{cases}=\begin{cases}0 & s>j\\ 1 & s=j\\ \frac{x^{j-s}}{(j-s)!} & s<j\end{cases}$$

$$\frac{j(j-1)\cdots(j-s+1)}{j!}=\frac{1}{(j-s)!}$$

$\Leftarrow$ $U_j^{(s)}(0)=e_s$.

אז מ.ק. קיבלנו:

$$\sum_{j=1}^{n}Q(X_j)W_{nj}^{s}(x)=e_s^T q(x)h^{-s}=Q^{(s)}(x)$$

ניקח $Q(u)=(u-v)^k$, קיים יחידים ואז נציב $u=x$.

$$\sum_{i=1}^{n}(X_i-x)^k W_{ni}^{s}=Q^{(s)}(x)=\begin{cases}0 & k\in\{0,\dots,l\}\setminus\{s\}\\ s! & k=s\end{cases}$$

Scanned with CamScanner

דג:

נב על מודל הרגרסיה עם $\sigma^2=1$. יהי $T(f)$ פונקציונל מהצורה

$$T(f)=\int\psi(x)f(x)\,dx$$

f לא ידועה, ψ ידועה וחסומה. אומד ההצבה יהיה

$$T(\hat f_n)=\int\psi(x)\underbrace{\hat f_n(x)}_{\text{אומד LP ל-}f}\,dx$$

בהנחה של (LP1)-(LP3) מתקיים, הוכיחו כי ניתן לבחור גרעין K ורוחב פס h כך ש:

$$\sup_{f:\|f\|_\infty\le M}\mathbb{E}_f\left(n^{1/2}\left(T(\hat f_n)-T(f)\right)\right)^2\le C(K,M)<\infty$$

הערות: זה אומר שניתן לאמוד את $T(f)$ בקצב $\sqrt{n}$, ללא תלות בחלקות של f.

רמז: הראו כי $\int_0^1|W_{nj}(x)|\,dx\le C\cdot\frac1n$ והשתמשו במשפט זה. (C לא תלוי ב-j,n).

פתרון:

נתחיל מהוכחת הרמז ונשתמש בו בהמשך:

$$\int_0^1|W_{nj}(x)|\,dx=\frac{1}{nh}\int_0^1\left|U^T(0)B_n(x)^{-1}U\left(\frac{X_j-x}{h}\right)\cdot K\left(\frac{x-X_j}{h}\right)\right|dx$$

$$\le\frac{1}{nh}\int_0^1\underbrace{\|U^T(0)\|_2}_{=1}\cdot\underbrace{\left\|B_n(x)^{-1}U\left(\frac{X_j-x}{h}\right)\right\|_2}_{\le\|\cdot\|_{2,2}\|\cdot\|_2\ \le\lambda^{-1}}\underbrace{\left|K\left(\frac{x-X_j}{h}\right)\right|}_{\le\max|K|=\|K\|_\infty}dx$$

$$\le\frac{1}{nh}\cdot\lambda_0^{-1}\cdot\|K\|_\infty\cdot\int_0^1\left\|U\left(\frac{X_j-x}{h}\right)\cdot\underbrace{1_{\left\{\left|\frac{X_j-x}{h}\right|\le1\right\}}}_{\text{מהתומך של }K\ (LP3)}\right\|_2dx$$

$u=\frac{X_j-x}{h}$ וכן h מצטמצם.

$$\le\frac1n\cdot\lambda_0^{-1}\|K\|_\infty\int_{\mathbb{R}}\|U(u)\cdot1_{\{|u|\le1\}}\|_2\,du$$

$\int_{\mathbb{R}}\|U(u)1_{\{|u|\le1\}}\|_2du\le\underbrace{\max_{v\in[-1,1]}\|U(v)\|}_{\text{מתקיים ב-}v=1}\cdot2$

$$\le\frac2n\lambda_0^{-1}\|K\|_\infty\cdot\sqrt{1+1+\frac{1}{(2!)^2}+\dots+\frac{1}{(\ell!)^2}}=C\cdot n^{-1}$$

וסיימנו.

נחזור כעת ל. MSE. נחסום ~~[illegible]~~ שונות והטיה בנפרד.

חסם לשונות:

$$T(\hat{f}_n) = \int_0^1 \psi(x)\hat{f}_n(x)\,dx = \int_0^1 \psi(x)\sum_{j=1}^{n} Y_j W_{nj}(x)\,dx = \sum_{j=1}^{n} Y_j \int_0^1 \psi(x) W_{nj}(x)\,dx$$

$$\Rightarrow \mathrm{Var}_f(T(\hat{f}_n)) = \mathrm{Var}_f\left(\sum_{j=1}^{n} Y_j \int_0^1 \psi(x) W_{nj}(x)\,dx\right) = \sum_{j=1}^{n}\left(\int_0^1 \psi(x) W_{nj}(x)\,dx\right)^2 \underbrace{\mathrm{Var}_f(Y_j)}_{=\mathrm{Var}_f(\varepsilon_j)=1}$$

$$= \sum_{j=1}^{n}\left(\int_0^1 \psi(x) W_{nj}(x)\,dx\right)^2 \le \|\psi\|_\infty \sum_{j=1}^{n}\left(\int_0^1 W_{nj}(x)\,dx\right)^2 \le \|\psi\|_\infty \cdot n\,(cn^{-1})^2$$

$$= c_1 \cdot n^{-1}.$$

הערה:

$$\mathrm{Var}_f(Y_j) = \mathbb{E}(Y_j - \mathbb{E}Y_j)^2 = \mathbb{E}\left(f(x_j) - \varepsilon_j - f(x_j)\right)^2 = \mathbb{E}\varepsilon_j^2 = \mathrm{Var}_f(\varepsilon_j)$$

חסם להטיה:

$$\mathbb{E}_f T(\hat{f}_n) - T(f) = \int_0^1 \psi(x)\,\mathbb{E}_f \hat{f}_n(x)\,dx - \int_0^1 \psi(x) f(x)\,dx$$

$$= \int_0^1 \psi(x)\left[\mathbb{E}_f \hat{f}_n(x) - f(x)\right]dx$$

$$* \ \mathbb{E}_f \hat{f}_n(x) = \mathbb{E}_f \sum_{j=1}^{n} Y_j W_{nj}(x) = \sum_{j=1}^{n} f(x_j) W_{nj}(x)$$

$$= \int_0^1 \psi(x)\left[\sum_{j=1}^{n} f(x_j) W_{nj}(x) - f(x)\right]dx$$

$\sum_{j=1}^{n} W_{nj}(x) = 1$

$$= \int_0^1 \psi(x)\sum_{j=1}^{n} W_{nj}(x)\left(f(x_j) - f(x)\right)dx$$

$$\Rightarrow \left|\mathbb{E}_f T(\hat{f}_n) - T(f)\right| \le \int_0^1 |\psi(x)|\sum_{j=1}^{n}\left|W_{nj}(x)\left(f(x_j) - f(x)\right)\right|dx$$

Lemma 2.2 (3)

$$\le \|\psi\|_\infty \cdot 2\|f\|_\infty \cdot \int_0^1 \sum_{j=1}^{n}|W_{nj}(x)| \cdot \mathbb{1}_{\left\{\left|\frac{x_j - x}{h}\right| \le 1\right\}}\,dx$$

$x_j - h \le x \le x_j + h \Rightarrow x_j + h - (x_j - h) = 2h$

Lemma 2.2 (1) $\sum_{j=1}^{n}|W_{nj}(x)| < c$

$$\le 2\|\psi\|_\infty\|f\|_\infty \cdot 2ch = c_2 h \quad //$$

* אם כן, קיבלנו

$$= MSE \le c_1\frac{1}{n} + c_2^2 h^2$$

אם נבחר $h_n^* \le \frac{1}{\sqrt{n}}$ אז

$$MSE \le c_1\frac{1}{n} + c_2^2 h^2 \le c_1\frac{1}{n} + c_2^2\frac{1}{n}$$

$$\Rightarrow \sup_{f:\|f\|_\infty \le M} \mathbb{E}_f\left(n^{1/2}\left(T(\hat{f}_n) - T(f)\right)\right)^2 \le C(K, M) < \infty \quad \blacksquare$$

5: $(C_p \text{ cross-validation})$

וכעת לא נסמן ב-h קריטריון ה- C_p cross-validation לבחירת רוחב פס h עבור אומד לינארי מהצורה

$$\hat{f}_h(x) = \sum_{j=1}^{n} Y_j W_{nj}(x,h)$$

בעקבות ההגדרה ב- 1.1.

סעיף א': נסתכל על הגרסה הבדידה של MISE:

$$dMISE(\hat{f}_h, f) := \mathbb{E}_f \frac{1}{n}\sum_{j=1}^{n}\left(\hat{f}_n(x_j) - f(x_j)\right)^2 =: \mathbb{E}_f\|\hat{f}_n - f\|^2_{2,n} \qquad (2.7)$$

הראו כי ביטוי זה ממוזער ע"י אותו h כמו הביטוי

$$J(h) := \mathbb{E}_f\left(\|\hat{f}_n\|^2_{2,n} - \frac{2}{n}\sum_{j=1}^{n}\hat{f}_n(x_j)f(x_j)\right) \qquad (2.8)$$

פתרון: הסימון dMISE מקיים:

$$\mathbb{E}_f \frac{1}{n}\sum_{j=1}^{n}\left(\hat{f}_n(x_j) - f(x_j)\right)^2 = \mathbb{E}_f\left(\frac{1}{n}\sum_{j=1}^{n}\hat{f}_n(x_j)^2 - 2\frac{1}{n}\sum_{j=1}^{n}\hat{f}_n(x_j)f(x_j)\right) + \mathbb{E}\frac{1}{n}\sum_{j=1}^{n}f(x_j)^2$$

$$= \mathbb{E}_f\left[\|\hat{f}_n\|^2_{2,n} - \frac{2}{n}\sum_{j=1}^{n}\hat{f}_n(x_j)f(x_j)\right] + \mathbb{E}_f\|f\|^2_{2,n}$$

$$= J(h) + \mathbb{E}_f\|f\|^2_{2,n}$$

המחובר השני לא תלוי ב-h, ולכן המינימום של J(h) הוא המינימום של dMISE.

סעיף ב': נגדיר את הסטטיסטי

$$\hat{T}(h) = \frac{2}{n}\sum_{j=1}^{n} Y_j \hat{f}_h(x_j) - \frac{2\sigma^2}{n}\sum_{j=1}^{n} W_{nj}(x_j, h)$$

כאשר $\sigma^2 = Var(\epsilon_j)$.

הראו כי הסטטיסטי הנ"ל אומד עבור המחובר השני של J(h).

פתרון:

$$\mathbb{E}_f\hat{T}(h) = \mathbb{E}_f\frac{2}{n}\sum_{j=1}^{n}\left(f(x_j)+\epsilon_j\right)\hat{f}_h(x_j) - \frac{2\sigma^2}{n}\sum_{j=1}^{n}W_{nj}(x_j,h)$$

$$= \underbrace{\mathbb{E}_f\frac{2}{n}\sum_{j=1}^{n}f(x_j)\hat{f}_h(x_j)}_{\text{התוחלת של המחובר השני}} + \frac{2}{n}\mathbb{E}_f\Big(\underbrace{\sum_{j=1}^{n}\epsilon_j\hat{f}_n(x_j) - \sigma^2\sum_{j=1}^{n}W_{nj}(x_j,h)}_{\text{נראה שזה מתאפס, סיימנו.}}\Big)$$

אז כן,

$$\mathbb{E}_f \sum_{j=1}^n \varepsilon_j \hat{f}_h(x_j) = \mathbb{E}_f \sum_{j=1}^n \varepsilon_j \sum_{k=1}^n Y_k W_{nk}(x_j,h)$$

$$\cancel{\mathbb{E}_f \sum_{j=1} f(x_j)+}$$

$$= \mathbb{E}_f \sum_{j=1}^n \varepsilon_j \sum_{k=1}^n \left(f(x_k)+\varepsilon_k\right) W_{nk}(x_j,h)$$

כי עבור $j \neq k$ מתאפס
כי בלתי תלוי ומתפרק ל-
$\mathbb{E}\varepsilon_j^2 \cdot \mathbb{E}\varphi(x_j)$

$$= \sigma^2 \mathbb{E}_f \sum_{j=1}^n W_{nj}(x_j,h) \quad //$$

סעיף ד: מצאו אומד ל- $J(h)$.

פתרון:

כחישוב בסעיף ג הראנו ססטיסטי אומד לאיבר השני של $J(h)$ והראשון שלו
צריך לשנות שלו באיבר הראשון (כמו כי $\hat{f}$). לכן:

$$\hat{J}(h) = \|\hat{f}_h\|^2_{2,n} - \hat{T}(h) = \|\hat{f}_h\|^2_{2,n} - \frac{2}{n}\sum_{j=1}^n Y_j \hat{f}_h(x_j) + \frac{2\sigma^2}{n}\sum_{j=1}^n W_{nj}(x_j,h)$$

אומד ל- $J(h)$.

סעיף ה: נדרש ל- קריטריון ה- C_p:

$$C_p(h) := \frac{1}{n}\sum_{j=1}^n \left(Y_i - \hat{f}(x_j)\right)^2 + \frac{2\sigma^2}{n}\sum_{j=1}^n W_{nj}(x_j,h)$$

הוכיחו כי $C_p(h)-\sigma^2$ הוא אומד ל- dMISE.

הסיקו מכאן שהמנצח $\hat{h}$ של $C_p(h)$ מהווה הבחירה הדיונית עבור h.

פתרון: נשים לב כי

$$\cdot\ C_p(h) = \underbrace{\frac{1}{n}\sum_{j=1}^n Y_i^2 - \frac{2}{n}\sum_{j=1}^n Y_j \hat{f}_h(x_j) + \underbrace{\frac{1}{n}\sum_{j=1}^n \hat{f}_h(x_j)^2}_{\|\hat{f}_h\|^2_{2,n}} + \frac{2\sigma^2}{n}\sum_{j=1}^n W_{nj}(x_j,h)}_{\hat{J}(h)} = \hat{J}(h) + \frac{1}{n}\sum_{j=1}^n Y_j^2$$

לכן,

$$\cdot\ \mathbb{E}C_p(h) = \mathbb{E}\left(\hat{J}(h) + \frac{1}{n}\sum_{j=1}^n Y_j^2\right) = J(h) + \frac{1}{n}\sum_{j=1}^n \mathbb{E}_f\left(f(x_j)^2 + \underbrace{2f(x_j)\varepsilon_j}_{\substack{\mathbb{E}(\cdot)=0 \\ (\text{ind.})}} + \underbrace{\varepsilon_j^2}_{\mathbb{E}(\cdot)=\sigma^2}\right)$$

$$= J(h) + \mathbb{E}_f\|f\|^2_{2,n} + \sigma^2$$

אם נחסיר σ^2 משני האגפים, נקבל אומד ל- $J(h)+\mathbb{E}_f\|f\|^2_{2,n}$, שזה ה- dMISE, כנדרש.
יש עניין לבחור כך את $\hat{h}$ כיוון שזה שבעקבות למצוא את dMISE נמצא אומד שבוו.