Non-parametric Regression

① Problem Setting

a) $\boxed{\text{Random Design}}$: $Y = r(X) + \varepsilon$

interest

$\begin{cases} \boxed{r(x) = \mathbb{E}[Y|X=x]} \\ \varepsilon = Y - \mathbb{E}[Y|X] \end{cases}$

can achieve by: given RV (X, Y)

Then do the $\boxed{\text{decomposition}}$:

$$\begin{cases} Y = \mathbb{E}[Y|X] + (Y - \mathbb{E}[Y|X]) \\ := \mathbb{E}[Y|X] + \xi \end{cases} \qquad \boxed{\xi := Y - \mathbb{E}[Y|X]}$$

Ob: $\mathbb{E}[\xi] = \mathbb{E}[Y - \mathbb{E}[Y|X]] = 0$

b) $\boxed{\text{fixed design}}$

$Y_i = r(x_i) + \varepsilon_i \qquad \boxed{\mathbb{E}[\varepsilon_i] = 0}$

equivalent

assume $\boxed{\mathbb{E}[Y_i | X_i = x_i] = r(x_i)}$

$\boxed{\text{our interest}}$

② our interest is: $r(x) = \mathbb{E}[Y|X=x]$

find an estimator $\hat{r}_n(x)$ based on $\mathcal{D} = \{(x_i, Y_i)\}_{i=1}^{n}$

③ Linear Smoother :

a) $\hat{r}_n(x) = \sum_{i=1}^{n} l_i(x) Y_i$

b) $l_i(x)$ only depends on $X_1, \ldots, X_n$ for $i=1,2,\ldots,n$

→ A toy-example for **Linear Smoother Analysis**

$\{ Y_i = r(x_i) + \varepsilon$

$$\hat{r}(x_i) = \sum_{j=1}^{n} l_j(x_i)\cdot Y_j$$

$\Rightarrow$ training error: $\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{r}(x_i)) = \frac{1}{n}\|Y-\hat{r}\|_2^2$

expected training error : $\frac{1}{n}\mathbb{E}_Y[\|Y-\hat{r}\|_2^2]$

average prediction risk : $\frac{1}{n}\mathbb{E}_{Y,Y^*}[\|Y^*-\hat{r}\|_2^2]$

Result : $\boxed{\text{average prediction risk} - \text{expected training error} = \frac{2}{n}\sigma^2 \operatorname{Tr}(L)}$

Assume: if we fix $(x_1,\ldots,x_n)$, and enforce the randomness on $(Y_1,\ldots,Y_n)$

④ One Construction of Linear Smoother

N-W estimator

$$\hat{r}_{NW}(x) = \sum_{i=1}^{n}\frac{K(\frac{x-X_i}{h})}{\sum_{j=1}^{n}K(\frac{x-X_j}{h})}\cdot Y_i$$

$$= \sum_{i=1}^{n}\frac{\frac{1}{n}\cdot\frac{1}{h}K(\frac{x-X_i}{h})\cdot Y_i}{\boxed{\frac{1}{n}\cdot\sum\frac{1}{h}(\frac{x-X_i}{h})} \rightarrow \hat{p}_n(x)}$$

$\int y\,\hat{p}_n(x,y)\,dy$

$= \int y \sum_{i=1}^{n} K(\frac{x-X_i}{h})K(\frac{y-Y_i}{h})\,dy\,\frac{1}{nh^2}$

$= \sum_{i=1}^{n} K(\frac{x-X_i}{h}) \underbrace{\int y K(\frac{y-Y_i}{h})\,dy}_{Y_i\cdot h}\,\frac{1}{nh^2}$

$= \sum_{i=1}^{n} K(\frac{x-X_i}{h})\cdot Y_i\cdot\frac{1}{nh}$

$$\approx \frac{\int y\,\hat{p}_n(x,y)\,dy}{\hat{p}_n(x)} \approx \mathbb{E}[Y\mid X=x] \rightarrow \text{intuitive explanation}$$

• MSE / MISE Analysis: $\longrightarrow$ Bias & Variance Analysis

$$\hat{r}_{NW}(x) = \frac{\frac{1}{n}\sum_i Y_i\frac{1}{h}K(\frac{x-X_i}{h})}{\hat{\ }\ldots}$$

$p_n(x)$

a) $\mathbb{E}[\hat{r}_{NW}(x)] \approx \frac{1}{p(x)} \cdot \mathbb{E}[Y \cdot \frac{1}{h} K(\frac{x-X}{h})]$

$$= \frac{1}{p(x)} \cdot \iint y \, \frac{1}{h} K\left(\frac{u-x}{h}\right) p(u,y)\, du\, dy$$

$$= \frac{1}{p(x)} \int r(u) \cdot p(u) \cdot \frac{1}{h} K\left(\frac{u-x}{h}\right) du$$

$$\boxed{u = x+th}$$

$$= \frac{1}{p(x)} \int K(t) \cdot r(x+th)\, p(x+th)\, dt$$

$$\boxed{\text{Taylor}}$$

$$= \frac{1}{p(x)} \cdot \left\{ r(x)p(x) + \frac{ch^2}{2}\left[r(x)p''(x) + 2r'(x)p'(x) + r''(x)p(x) \right] \right\}$$

$$= r(x) + \frac{ch^2}{2}\left[r(x) \cdot \frac{p''(x)}{p(x)} + 2r'(x) \cdot \frac{p'(x)}{p(x)} + r''(x) \right]$$

b) variance can also be computed with similar manner

⑤ CV approximation

Theoretically, we can achieve optimal $\hat{h}$ by minimizing MISE

$\boxed{\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{r}_{(-i)}(X_i))^2}$

infeasible to compute

good estimate

$\Rightarrow$ use CV estimation

$\boxed{\mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} (\hat{r}_n(x_i) - r_n(x_i))^2\right)}$

poor estimate

training error

$\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{r}_n(X_i))^2$

Analysis: $\mathbb{E}[(Y_i - \hat{r}_{(-i)}(X_i))^2] = \mathbb{E}[(\varepsilon_i + r(X_i) - \hat{r}_{-i}(X_i))^2]$

$$= \mathbb{E}[\varepsilon_i^2] + \mathbb{E}[(r(x_i) - \hat{r}_{(-i)}(x_i))^2]$$

$$+ 2\mathbb{E}[\boxed{\varepsilon_i}(r(x_i) - \boxed{\hat{r}_{-i}(x_i)})]^2$$

r.v. with respect to Y_i $\xleftrightarrow{\text{indpt}}$ r.v. with respect to $Y_{(-i)}$

$$= \underbrace{\mathbb{E}[\varepsilon_i^2]}_{\sigma^2} + \underbrace{\mathbb{E}[(r(x_i) - \hat{r}_{(-i)}(x_i))^2]}_{\boxed{\text{MSE term}}}$$

(approximately)