PAGE NO.:
DATE: / /

SML
Assignment - 3

Q1 → $R(f) = E[(Y-f(x))^2] = \int (y-f(x))^2 p(x,y)\, dx\, dy$

true Risk / Expected Risk

$p(x,y)$ = Joint probability distribution of X (input) and Y (output)

1) For optimal Function $[f^*(x)]$ → $f^*(x) = \underset{f(x)}{\arg\min}\ E[(Y-f(x))^2]$

as we want to minimize the expected Risk for our optimal function

also, $p(x,y) = p(y|x)\, p(x)$

$$R(f) = \int\int (y-f(x))^2\, p(x,y)\, dy\, dx$$

$$\Rightarrow \int\int (y-f(x))^2\, p(y|x)\, p(x)\, dy\, dx$$

$$\Rightarrow \int \left[\int (y-f(x))^2\, p(y|x)\, dy\right] p(x)\, dx$$

$$E[(Y-f(X))^2 / X=x] = \int (y-f(x))^2\, p(y|x)\, dy$$

$$R(f) = \int \left[E\left((Y-f(X))^2/X=x\right)\right] p(x)\, dx$$

$\because p(x) \geq 0$ so to minimize $R(f)$ we can minimize $E((Y-f(X))^2/X=x)$ for each x.

hence $f^*(x) = \underset{f(x)}{\arg\min}\ E[(Y-f(X))^2/X=x]$

$$\Rightarrow \int (y-f(x))^2\, p(y|x)\, dy$$

~~to find minimum~~ to minimize → $\dfrac{d}{d(f(x))}\left(E[(Y-f(x))^2/X=x]\right) = 0$

$$\frac{d}{d(f(x))}\left[\int (y-f(x))^2\, p(y|x)\, dy\right] = 0$$

$$\Rightarrow \int 2(y-f(x)) \times (-1)\, p(y|x)\, dy = 0$$

$$\rightarrow \frac{d^2}{d(f(x))^2} = \int 2\, p(y|x)\, dy = 2$$

$\cancel{\because p(y|x) \geq 0}$

$$\frac{d^2}{d(f(x))^2} > 0$$

hence it is indeed minimizer

$$\Rightarrow \int (y-f(x))\, p(y|x)\, dy = 0$$

$$\Rightarrow \int y\, p(y|x)\, dy = \int f(x)\, p(y|x)\, dy$$

$$\Rightarrow \int y\, p(y|x)\, dy = f(x) \underbrace{\int p(y|x)\, dy}_{1}$$

$$\Rightarrow f(x) = \int y\, p(y|x)\, dy$$

$$\boxed{f^*(x) = E[y/X=x]}$$

$\therefore$ The optimal function that minimizes true risk $R(f)$

$$\Rightarrow f^*(x) = E[Y/X=x]$$

2) In terms of conditional distribution $p(y|x)$, the optimal function $f^*(x)$ represents conditional expectation of Y for given $X=x$. It represents the expected value of output Y for given input X under conditional distribution $p(y|x)$.

Q2 $f(x) = 2x+3$ $\quad x \sim U[0,5]$

observed response → $y = f(x) + \varepsilon$, $\varepsilon \sim N(0, 0.1)$

Given Training Data → D_1, D_2, D_3

linear Regression model → $\hat{f}(x) = ax+b$ for D_1, D_2, D_3

$\hat{f}_1(x) = 1.9x + 3.5$ $\quad \hat{f}_2(x) = 2.1x+3.2$ $\quad \hat{f}_3(x) = 2x+3.4$

1) Expected Prediction function $= E(\hat{f}(x)) = E(a)\,x + E(b)$

$E(a) = \frac{1}{3}(a_1+a_2+a_3) = \frac{1}{3}(1.9+2.1+2) = \underline{2}$

$E(b) = \frac{1}{3}(b_1+b_2+b_3) = \frac{1}{3}(3.5+3.2+3.4) = \underline{3.367}$

$$\underline{\underline{E[\hat{f}(x)] = 2x + 3.367}}$$

2) Bias $= E(\hat{f}(x)) - f(x)$

for $x=2$ → $E(\hat{f}(x)) = 2\times(2) + 3.367 = 7.367$

$f(2) = 2(2)+3 = 7$

Bias $= E(\hat{f}(x)) - f(x) = 7.367 - 7 = 0.367$

hence Bias $(x=2) = \underline{\underline{0.367}}$

3) Variance $= E\left[\left(\hat{f}(x) - E(\hat{f}(x))\right)^2\right]$

for $x=2$ →

$\hat{f}_1(2) = 1.9\times2 + 3.5 = 7.3$

$\hat{f}_2(2) = 2.1\times2+3.2 = 7.4$

$\hat{f}_3(2) = 2\times2+3.4 = 7.4$

$E(\hat{f}(2)) = 2\times2+3.367 = 7.367$

Variance at $(x=2) = \frac{1}{3}\left[(7.3-7.367)^2 + (7.4-7.367)^2 + (7.4-7.367)^2\right]$

Variance $(x=2) = \frac{1}{3}(0.004489 + 0.001089 + 0.001089)$

$= \frac{1}{3}\times 0.006667 \approx 0.0022$ (approx)

4) Expected Square Error $= E\left[\left(\hat{f}(x) - f(x)\right)^2\right]$
(ESE)

$$f(x) = 2x+3 \Rightarrow f(2) = 2\times2+3 = 7$$

$$\hat{f}_1(x) = 1.9\times x + 3.5 \Rightarrow \hat{f}_1(2) = 1.9\times2+3.5 = 7.3$$

$$\hat{f}_2(x) = 2.1x+3.2 \Rightarrow \hat{f}_2(2) = 2.1\times2+3.2 = 7.4$$

$$\hat{f}_3(x) = 2x+3.4 \Rightarrow \hat{f}_3(2) = 2\times2+3.4 = 7.4$$

$$ESE(2) = \frac{1}{3}\left[\left(\hat{f}_1(x)-f(x)\right)^2 + \left(\hat{f}_2(x)-f(x)\right)^2 + \left(\hat{f}_3(x)-f(x)\right)^2\right]$$

$$= \frac{1}{3}\left[(7.3-7)^2 + (7.4-7)^2 + (7.4-7)^2\right]$$

$$= \frac{1}{3}\left[(0.3)^2 + (0.4)^2 + (0.4)^2\right]$$

$$= \frac{1}{3}\left[0.09+0.16+0.16\right] = \frac{1}{3}\times 0.41 = \underline{0.136}$$

Using Bias-Variance Decomposition →

$$ESE = E\left[\left(\hat{f}(x) - f(x)\right)^2\right] = Bias^2 + Variance + \sigma^2$$

$$= \left(E\hat{f}(x) - f(x)\right)^2 + E\left[\hat{f}(x) - E(\hat{f}(x))\right]^2 + \sigma^2$$

$[\varepsilon \sim N(0, 0.1) \Rightarrow \sigma^2 = 0.1]$ $\quad = (0.367)^2 + 0.0022 + 0.1$

$$= \underline{0.236}$$

$(0.236 - 0.136)$ ~~0.236~~

⇒ Both the values are similar but there is difference of $\boxed{0.1}$
The difference between ESE(2) and what we get from Bias Variance decomposition is due to irreducible noise

difference $= 0.1 = \sigma^2$ (variance of noise)