PROJECT 4 - Problem 2.24

(A) Find $\bar{g}(x)$ analytically

Dataset $D = \{(x_1, x_1^2), (x_2, x_2^2)\}$

$\therefore$ $g^{(D)}(x)$ will be of the form $mx + b$

where $m = \dfrac{x_1^2 - x_2^2}{x_1 - x_2} = x_1 + x_2$

$c = -x_1 x_2$

$\therefore$ $g^{(D)}(x) = (x_1 + x_2)x - x_1 x_2$

$$\bar{g}(x) = E_D\left[g^{(D)}(x)\right]$$

$$= \frac{1}{2}\int_{x_2=-1}^{1} \frac{1}{2}\int_{x_1=-1}^{1} x_1 x + x_2 x - x_1 x_2 \; dx_1\, dx_2$$

$$= \frac{1}{4}\int_{x_2=-1}^{1} x\left[\frac{x_1^2}{2}\right]_{-1}^{1} + x_2 x\,[x]_{-1}^{+1} - x_2\left[\frac{x_1^2}{2}\right]_{-1}^{1} \; dx_2$$

$$= \frac{1}{4}\int_{x_2=-1}^{1} 2x_2 x \; dx_2$$

$$= \frac{1}{4}\, 2x\left[\frac{x_2^2}{2}\right]_{-1}^{1}$$

$$= 0$$

$\therefore$ $\bar{g}(x) = 0$

(D) Compute analytically what Eout, bias and var should be.

$$E_{out} = E_x\left[(g^{(D)}(x) - f(x))^2\right]$$

But $g^{(D)}(x) = ax + b$ where $a = x_1 + x_2$ and $b = -x_1 x_2$

and $f(x) = x^2$

$$E_{out}(g^{(D)}) = E_x\left[(ax + b - x^2)^2\right]$$

$$= E_x\left[a^2x^2 + b^2 + x^4 - 2ax^3 - 2bx^2 + 2abx\right]$$

$$= E_x\left[x^4 - 2ax^3 + x^2(a - 2b) + 2abx + b^2\right]$$

$$\therefore\ E_{out}(g^{(D)}) = \frac{1}{2}\int_{-1}^{1} x^4 - x^3(2a) + x^2(a-2b) + x(2ab) + b^2 \, dx$$

$$= \frac{1}{2}\left[\left[\frac{x^5}{5}\right]_{-1}^{1} - 2a\left[\frac{x^4}{4}\right]_{-1}^{1} + (a-2b)\left[\frac{x^3}{3}\right]_{-1}^{1} + 2ab\left[\frac{x^2}{2}\right]_{-1}^{1} + b\left[x\right]_{-1}^{1}\right]$$

$$= \frac{1}{2}\left[\frac{2}{5} + \frac{2}{3}(a-2b) + 2b^2\right]$$

$$= \frac{1}{5} + \frac{(a-2b)}{3} + b^2$$

$E_{out}(g^{(D)})$ is still dependent on the dataset D. To get the out of sample over all datasets, we need to calculate the Expected value $E_D\left[E_{out}(g^{(D)})\right]$

$$E_D\left[E_{out}(g^{(D)})\right] = E_D\left[\frac{1}{5} + \frac{1}{3}(\overbrace{x_1+x_2}^{a} + \overbrace{2x_1x_2}^{-2b}) + \overbrace{x_1^2x_2^2}^{b^2}\right] \cancel{dx}$$

$$= \frac{1}{2}\int_{x_2=-1}^{1} \frac{1}{2}\int_{x_1=-1}^{1} \frac{1}{5} + \frac{x_1+x_2+2x_1x_2}{3} + x_1^2x_2^2 \, dx_1 \, dx_2$$

$$= \frac{1}{4}\int_{x_2=-1}^{1} \frac{1}{5}\left[x_1\right]_{-1}^{1} + \frac{1}{3}\left[2x_2\right] + 2x_2^2 \, dx_2$$

$$= \frac{1}{4}\int_{x_2=-1}^{1} \frac{2}{5} + \frac{2}{3}x_2 + 2x_2^2 \, dx_2$$

$$= \frac{1}{4}\left[\frac{4}{5} + 0 + \frac{4}{3}\right]$$

$$= \frac{1}{5} + \frac{1}{3} = \frac{8}{15}.$$

$$\therefore\ E_{out} = \frac{8}{15}$$