Exercise 2 - Solution

a) Compute the output samples of the given training set.

t	i	$x_i^{(t)}$	$\varepsilon_i^{(t)}$	$y_i^{(t)} = f_{exact}(x_i^{(t)}) + \varepsilon_i^{(t)}$
1	1	1.0	-0.1	0.9
1	2	-0.5	0.0	0.25
1	3	3.0	0.2	9.2
2	1	-2.0	0.3	4.3
2	2	-1.5	-0.2	2.05
2	3	0.5	0.1	0.35
3	1	2.0	0.3	4.3
3	2	1.0	0.1	1.1
3	3	-3.0	0.2	9.2
4	1	-1.0	0.3	1.3
4	2	-1.5	0.0	2.25
4	3	2.5	-0.2	6.05

$T_1 = \{(1.0,\ 0.9),\ (-0.5,\ 0.25),\ (3.0,\ 9.2)\}$

$T_2 = \{(-2.0,\ 4.3),\ (-1.5,\ 2.05),\ (0.5,\ 0.35)\}$

$T_3 = \{(2.0,\ 4.3),\ (1.0,\ 1.1),\ (-3.0,\ 9.2)\}$

$T_4 = \{(-1.0,\ 1.3),\ (-1.5,\ 2.25),\ (2.5,\ 6.05)\}$

b) Compute an estimator for the bias term

Recall for a random variable Z, an estimator for its mean is given by:

$$E(Z) \approx \bar{Z} = \frac{1}{M} \sum_{i=1}^{M} z_i,$$

where $z_1, \ldots, z_M$ are M samples drawn from that random variable.

Firstly compute the linear model for each training set:

Using socscistatistics.com LR Calculator:

Linear regression on T_1, T_2, T_3, T_4

$f_{T_1}(x) = 0.3662 + 2.6432x$

$f_{T_2}(x) = 0.8619 - 1.3714x$

$f_{T_3}(x) = 4.8667 - 1.2786x$

$f_{T_4}(x) = 3.2 + 1.1x$

$\text{Bias} = E_T(f_T(x_0)) - f_{exact}(x_0)$ $\quad x_0 = 0$

$\to f_{exact}(x_0) = f_{exact}(0) = 0^2 = 0$

$$E_T(f_T(x_0)) \approx \frac{1}{4}\sum_{i=1}^{4} E_{T_i}(f_{T_i}(x_0)) = \frac{1}{4}\sum_{i=1}^{4} f_{T_i}(x_0) =$$

$$= \frac{1}{4}(f_{T_1}(x_0) + f_{T_2}(x_0) + f_{T_3}(x_0) + f_{T_4}(x_0)) =$$

$$= \frac{1}{4}(0.3662 + 0.8619 + 4.8667 + 3.2) =$$

$$= 2.3237$$

c) $E_T((f_T(x_0) - E_T(f_T(x_0)))^2) = Var_T(f_T(x_0))$

Biased estimator:

$$Var_T(f_T(x_0)) = \frac{1}{M}\sum_{i=1}^{M}(f_{T_i}(x_0) - E_T(f_T(x_0)))^2$$

$$= \frac{1}{4}\sum_{i=1}^{4}(f_{T_i}(x_0) - E_T(f_T(x_0)))^2$$

$$= \frac{1}{4}((0.3662 - 2.3237)^2 + (0.8619 - 2.3237)^2 + (4.8667 - 2.3237)^2 + (3.2 - 2.3237)^2)$$

$$\approx 3.3$$

Small sample → Variance is underestimated

→ use unbiased estimator

$$Var_T(f_T(x_0)) = \frac{1}{M-1} \sum_{i=1}^{n} \left(f_{T_i}(x_0) - E_T(f_T(x_0))\right)^2$$

$$= \frac{1}{3} \sum_{i=1}^{4} \Big((0.3662 - 2{,}3237)^2 + (0.8619 - 2{,}3237)^2 + (4{,}8667 - 2{,}3237)^2 + (3.2 - 2.3237)^2\Big)$$

$$\approx 4{,}4$$