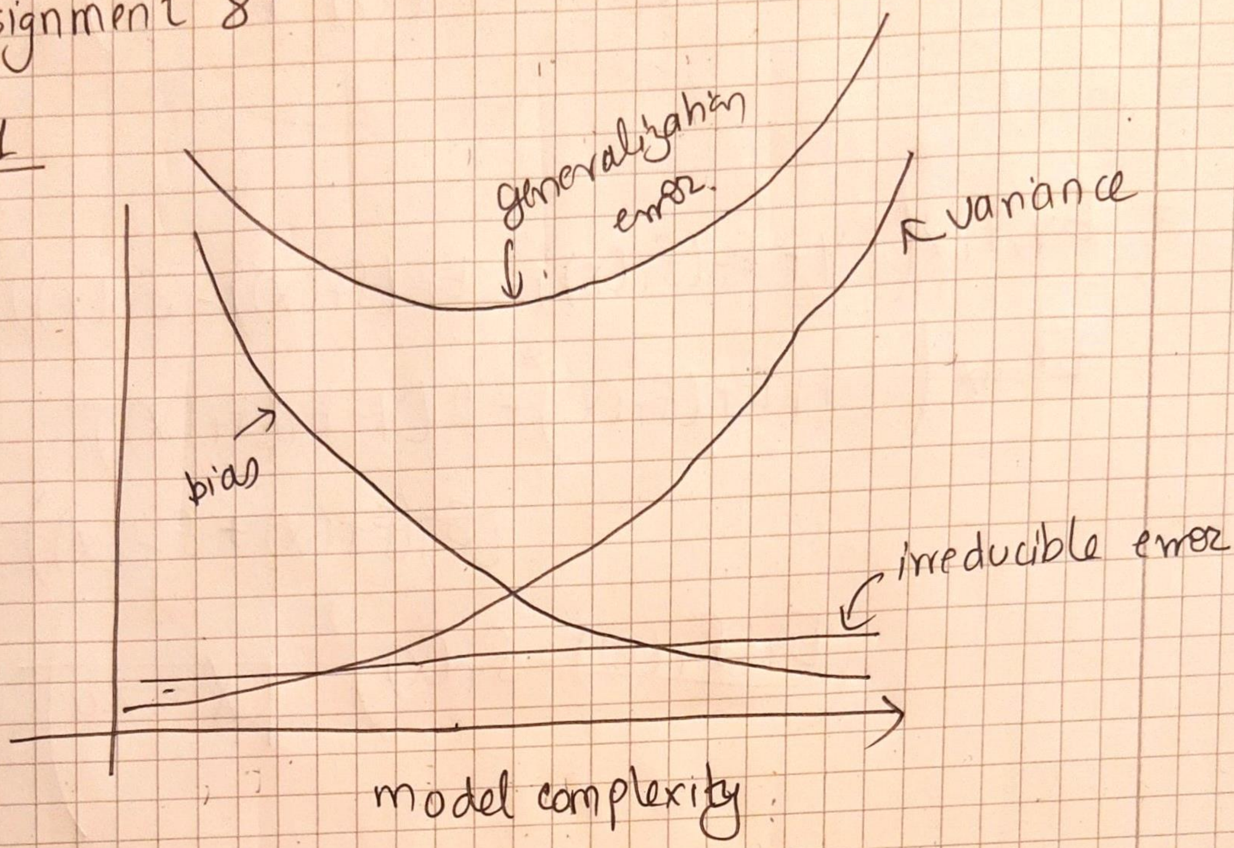
Maulik Chhetri,Subigya Paudel, Mahiem Agrawal

Assignment 8.

Exercise 1



Bias-Variance tradeoff.

From the graph, we can see that for increasing model complexity we have lower bias. This is due to the fact ~~at~~ that at a higher model complexity, the predictor function tries to be as close as to the training sets.

But, as a consequence, it has high variance. because small changes in data points will cause the predictor curve to vary greatly.

Bias is high for lower model complexity because ~~the~~ in this case, the predictor doesn't try to 'adapt' strongly to the datapoints but rather give a general solution. Hence, changing the data points for low model complexity won't affect its nature much which implies lower variance.

The irreducible error ~~ub a caure~~ comes from the noise in the data set. Hence, this is an intrinsic property and remains the same.

The sum of bias, variance and irreducible error is the generalization error.

Exercise 2

Given, $f_{exact}(x) = x^3$

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	-1.1
1	2	-0.5	0.1	-0.025
1	3	-1	0	-1
2	1	-0.5	0.1	-0.025
2	2	0.5	0.2	0.325
2	3	1	0	1
3	1	0.1	0	0.001
3	2	0.5	-0.1	0.025
3	3	1.0	0.1	1.1
4	1	-0.1	-0.1	-0.101
4	2	-0.5	0	-0.125
4	3	-1.0	0.2	-0.8

For the first sample,

$$X = \begin{bmatrix} 1 & -1.0 \\ 1 & -0.5 \\ 1 & -1 \end{bmatrix}, \quad y = \begin{bmatrix} -1.1 \\ -0.025 \\ -1 \end{bmatrix},$$

solving the least square $\beta = (X^TX)^{-1}X^Ty$, we get,

$$\beta = \begin{bmatrix} 1.0 \\ 2.05 \end{bmatrix}$$

$$f(x) = 1 + 2.05\,x$$

for, second sample

$$X = \begin{bmatrix} 1 & -0.5 \\ 1 & 0.25 \\ 1 & 1 \end{bmatrix}, \quad y = \begin{bmatrix} -0.025 \\ 0.325 \\ 1 \end{bmatrix}.$$

$$\beta = \begin{bmatrix} 0.2214 \\ 0.6357 \end{bmatrix}, \quad f(y) = 0.2214 + 0.6357x$$

for, 3rd sample,

$$X = \begin{bmatrix} 1 & 0.1 \\ 1 & 0.5 \\ 1 & 1 \end{bmatrix}, \quad y = \begin{bmatrix} 0.001 \\ 0.025 \\ 1.1 \end{bmatrix}, \quad \beta = \begin{bmatrix} -0.2962 \\ 1.2592 \end{bmatrix}$$

$$f(y) = -0.2962 + 1.2592x$$

for, 4th sample

$$X = \begin{bmatrix} 1 & -0.1 \\ 1 & -0.5 \\ 1 & -1 \end{bmatrix}, \quad y = \begin{bmatrix} -0.101 \\ -0.125 \\ -0.8 \end{bmatrix}, \quad \beta = \begin{bmatrix} 0.0848 \\ 0.8002 \end{bmatrix}$$

$$f(y) = 0.0848 + 0.8002x$$

(b) Given, $x_0 = 0$.

for $f(x) = 1+2.05x$.

$f_{T_1}(x_0) = 1$, $f_{T_2}(x_0) = 0.2214$, $f_{T_3}(x_0) = -0.2962$

$f_{T_4}(x_0) = 0.0848$.

$f_{exact}(x_0) = 0$

$\cancel{(1-0)^2 +}$

$$E_T(f_T(x_0) - f_{exact}(x_0)) = ((1-0) + (0.2214-0) + (-0.2962-0) + (0.0848-0)) \times 1/4$$

$$= 0.2525$$

$$(E_T(f_T(x_0) - f_{exact}(x_0)))^2 = 0.06375$$

(c) $f_{T_1}(x_0) = 1$, $f_{T_2}(x_0) = 0.2214$, $f_{T_3}(x_0) = -0.2962$,

$f_{T_4}(x_0) = 0.084$. , Mean $= (1+0.2214-0.2962+0.084) \times 1/4$

$= 0.2523$

$$\text{Variance} = \frac{1}{3}\left[(1-0.2523)^2 + (0.2214-0.2523)^2 + (-0.2962-0.2523)^2 + (0.084-0.2523)^2\right]$$

$$= \frac{1}{3}(0.88918)$$

$$= 0.2963$$