Training Phase				Validation Phase				Test Phase	
Real Data Set 1 50% of the collcted data		Model 1: Linear Regression	Model 2: Non-Linear Regression	Real Data Set 2 25% of the collcted data		Model 1: Linear Regression	Model 2: Non-Linear Regression	Real Data Set 3 25% of the collcted data	The better model selected from Model 1 and Model 2 depending on the analysis of overfitting
x	y	$\hat{y}=a1 + b1 * x$	$\hat{y}=a2 + b2 * x^2$	x	y	$\hat{y}=a1 + b1 * x$	$\hat{y}=a2 + b2 * x^2$	x	$\hat{y}=a1 + b1 * x$ or $\hat{y}=a2 + b2 * x^2$
1	1.8			1.5	1.7			1.4	
2	2.4			2.9	2.7			2.5	
3.3	2.3			3.7	2.5			3.6	
4.3	3.8			4.7	2.8			4.5	
5.3	5.3			5.1	5.5			5.4	
1.4	1.5			X	X	X	X	X	X
2.5	2.2			X	X	X	X	X	X
2.8	3.8			X	X	X	X	X	X
4.1	4.0			X	X	X	X	X	X
5.1	5.4			X	X	X	X	X	X

X	Y	X*X	X*Y	P	P*y	P*P
1	1.8	1	1.8	1	1.8	1
2	2.4	4	4.8	4	9.6	16
3.3	2.3	10.89	7.59	10.89	25.05	118.59
4.3	3.8	18.49	16.34	18.49	70.26	341.88
5.3	5.3	28.09	28.09	28.09	148.88	789.05
1.4	1.5	1.96	2.1	1.96	2.94	3.84
2.5	2.2	6.25	5.5	6.25	13.75	39.06
2.8	3.8	7.84	10.64	7.84	29.80	61.46
4.1	4	16.81	16.4	16.81	67.24	282.58
5.1	5.4	26.01	27.54	26.01	140.45	676.52

Training Phase

Step 1: Calculate a1, b1, a2, b2

ΣX = 31.8 **ΣP = 121.34**

ΣY = 32.5 **ΣPY = 509.76**

ΣX2 = 121.34 **ΣP2 = 2329.99**

ΣXY = 120.8 **N=10x**

For Model 1

Slope(b) = (NΣXY - (ΣX) (ΣY)) / (NΣX2 - (ΣX)2)

So, b1 = (10*120.8 - (31.8*32.5))/ (10*121.34 – (31.8*31.8))

= 174.5 / 202.16

=0.86

Intercept(a) = (ΣY - b(ΣX)) / N

So, a1 = (32.5- 0.86*31.8)/10

= 0.51

For Model2 (Here P= x*x)

Slope(b) = (NΣPY - (ΣP) (ΣY)) / (NΣP2 - (ΣP)2)

So, b2= (10*509.76 - (121.34*32.5))/ (10*2329.99 – (121.34*121.34))

= 1154.05/8576.5

=0.13

Intercept(a) = (ΣY - b(ΣP)) / N

So, a2 = (32.5 – (0.13*121.34))/10

= 1.67

So Now We have

a1 = 0.51 **b1 = 0.86**

a2 = 1.67 **b2 = 0.13**

Step 2: Calculate Y using Regression Formula

Training Phase				Validation Phase				Test Phase	
Real Data Set 1 50% of the collcted data		Model 1: Linear Regression	Model 2: Non-Linear Regression	Real Data Set 2 25% of the collcted data		Model 1: Linear Regression	Model 2: Non-Linear Regression	Real Data Set 3 25% of the collcted data	The better model selected from Model 1 and Model 2 depending on the analysis of overfitting
x	y	$\hat{y}=a1 + b1 * x$	$\hat{y}=a2 + b2 * x^2$	x	y	$\hat{y}=a1 + b1 * x$	$\hat{y}=a2 + b2 * x^2$	x	$\hat{y}=a1 + b1 * x$ or $\hat{y}=a2 + b2 * x^2$
1	1.8	1.37	1.8	1.5	1.7			1.4	
2	2.4	2.23	2.19	2.9	2.7			2.5	
3.3	2.3	3.35	3.09	3.7	2.5			3.6	
4.3	3.8	4.21	4.07	4.7	2.8			4.5	
5.3	5.3	5.07	5.32	5.1	5.5			5.4	
1.4	1.5	1.71	1.92	X	X	X	X	X	X
2.5	2.2	2.66	2.48	X	X	X	X	X	X
2.8	3.8	2.92	2.69	X	X	X	X	X	X
4.1	4.0	4.04	3.86	X	X	X	X	X	X
5.1	5.4	4.90	5.05	X	X	X	X	X	X

Step 2: Calculate MSE

For Model1 MSE = 0.2819

For Model2 MSE = 0.2370

Validation Phase

Training Phase				Validation Phase				Test Phase	
Real Data Set 1 50% of the collcted data		Model 1: Linear Regression	Model 2: Non-Linear Regression	Real Data Set 2 25% of the collcted data		Model 1: Linear Regression	Model 2: Non-Linear Regression	Real Data Set 3 25% of the collcted data	The better model selected from Model 1 and Model 2 depending on the analysis of overfitting
x	y	$\hat{y}=a1 + b1 * x$	$\hat{y}=a2 + b2 * x^2$	x	y	$\hat{y}=a1 + b1 * x$	$\hat{y}=a2 + b2 * x^2$	x	$\hat{y}=a1 + b1 * x$ or $\hat{y}=a2 + b2 * x^2$
1	1.8	1.37	1.8	1.5	1.7	1.8	1.96	1.4	
2	2.4	2.23	2.19	2.9	2.7	3.00	2.76	2.5	
3.3	2.3	3.35	3.09	3.7	2.5	5.07	3.45	3.6	
4.3	3.8	4.21	4.07	4.7	2.8	4.55	4.54	4.5	
5.3	5.3	5.07	5.32	5.1	5.5	4.90	5.05	5.4	
1.4	1.5	1.71	1.92	X	X	X	X	X	X
2.5	2.2	2.66	2.48	X	X	X	X	X	X
2.8	3.8	2.92	2.69	X	X	X	X	X	X
4.1	4.0	4.04	3.86	X	X	X	X	X	X
5.1	5.4	4.90	5.05	X	X	X	X	X	X

For Model1 MSE = 2.0263

For Model2 MSE = 0.8419

Change in Model 1 = 2.0263/0.2819 = 7.18

Change in Model 2 = 0.8419/0.2370 = 3.55

So, based on the calculation, Model 2 has lower value, so we choose model2 to predict the value of Y.

Test Phase

Final Answer:

Training Phase				Validation Phase				Test Phase	
Real Data Set 1 50% of the collcted data		Model 1: Linear Regression	Model 2: Non-Linear Regression	Real Data Set 2 25% of the collcted data		Model 1: Linear Regression	Model 2: Non-Linear Regression	Real Data Set 3 25% of the collcted data	The better model selected from Model 1 and Model 2 depending on the analysis of overfitting
x	y	$\hat{y}=a1 + b1 * x$	$\hat{y}=a2 + b2 * x^2$	x	y	$\hat{y}=a1 + b1 * x$	$\hat{y}=a2 + b2 * x^2$	x	$\hat{y}=a2 + b2 * x^2$
1	1.8	1.37	1.8	1.5	1.7	1.8	1.96	1.4	1.92
2	2.4	2.23	2.19	2.9	2.7	3.00	2.76	2.5	2.48
3.3	2.3	3.35	3.09	3.7	2.5	5.07	3.45	3.6	3.35
4.3	3.8	4.21	4.07	4.7	2.8	4.55	4.54	4.5	4.30
5.3	5.3	5.07	5.32	5.1	5.5	4.90	5.05	5.4	5.46
1.4	1.5	1.71	1.92	X	X	X	X	X	X
2.5	2.2	2.66	2.48	X	X	X	X	X	X
2.8	3.8	2.92	2.69	X	X	X	X	X	X
4.1	4.0	4.04	3.86	X	X	X	X	X	X
5.1	5.4	4.90	5.05	X	X	X	X	X	X