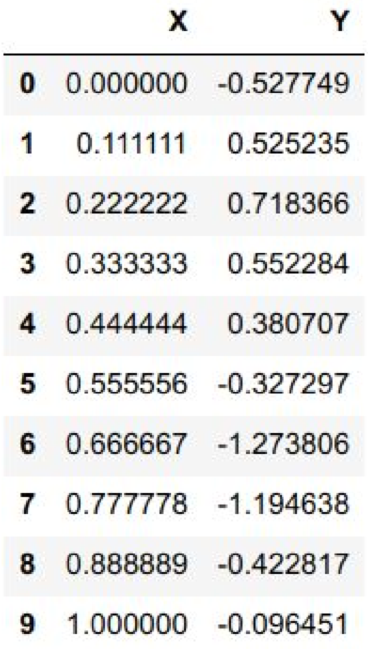
MACHINE LEARNING
Linear Regression

VAIBHAV PODDAR (16CS10051)

Part 1: Synthetic data generation and simple curve fitting

a) Generation of a synthetic dataset with dataset size =10

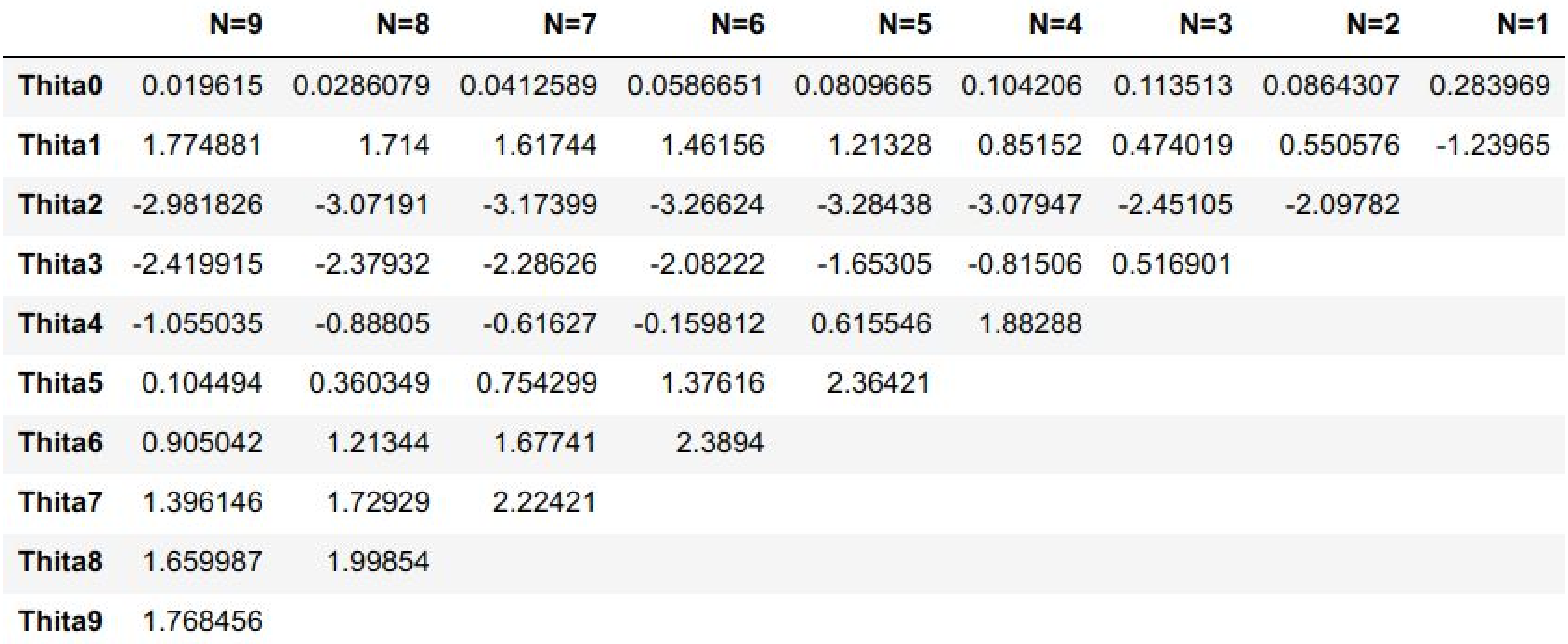


	X	Y
0	0.000000	-0.527749
1	0.111111	0.525235
2	0.222222	0.718366
3	0.333333	0.552284
4	0.444444	0.380707
5	0.555556	-0.327297
6	0.666667	-1.273806
7	0.777778	-1.194638
8	0.888889	-0.422817
9	1.000000	-0.096451

b) Splitting was done in code.

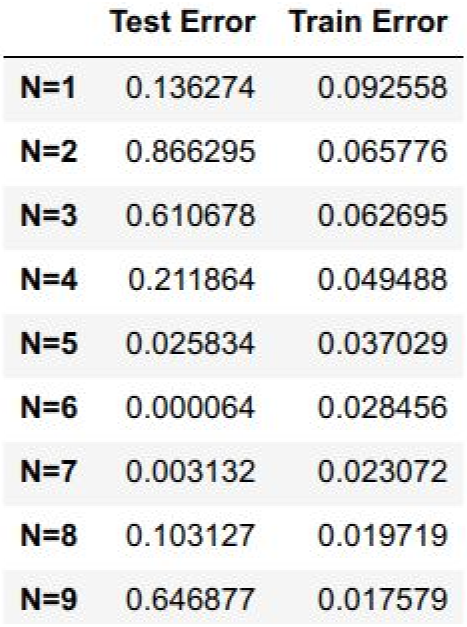
c) Learned parameters:

After fitting 9 different curves parameters obtained are as follows:



	N=9	N=8	N=7	N=6	N=5	N=4	N=3	N=2	N=1
Thita0	0.019615	0.0286079	0.0412589	0.0586651	0.0809665	0.104206	0.113513	0.0864307	0.283969
Thita1	1.774881	1.714	1.61744	1.46156	1.21328	0.85152	0.474019	0.550576	-1.23965
Thita2	-2.981826	-3.07191	-3.17399	-3.26624	-3.28438	-3.07947	-2.45105	-2.09782	
Thita3	-2.419915	-2.37932	-2.28626	-2.08222	-1.65305	-0.81506	0.516901		
Thita4	-1.055035	-0.88805	-0.61627	-0.159812	0.615546	1.88288			
Thita5	0.104494	0.360349	0.754299	1.37616	2.36421				
Thita6	0.905042	1.21344	1.67741	2.3894					
Thita7	1.396146	1.72929	2.22421						
Thita8	1.659987	1.99854							
Thita9	1.768456								

Squared Error on the test set and train set obtained are as follows:



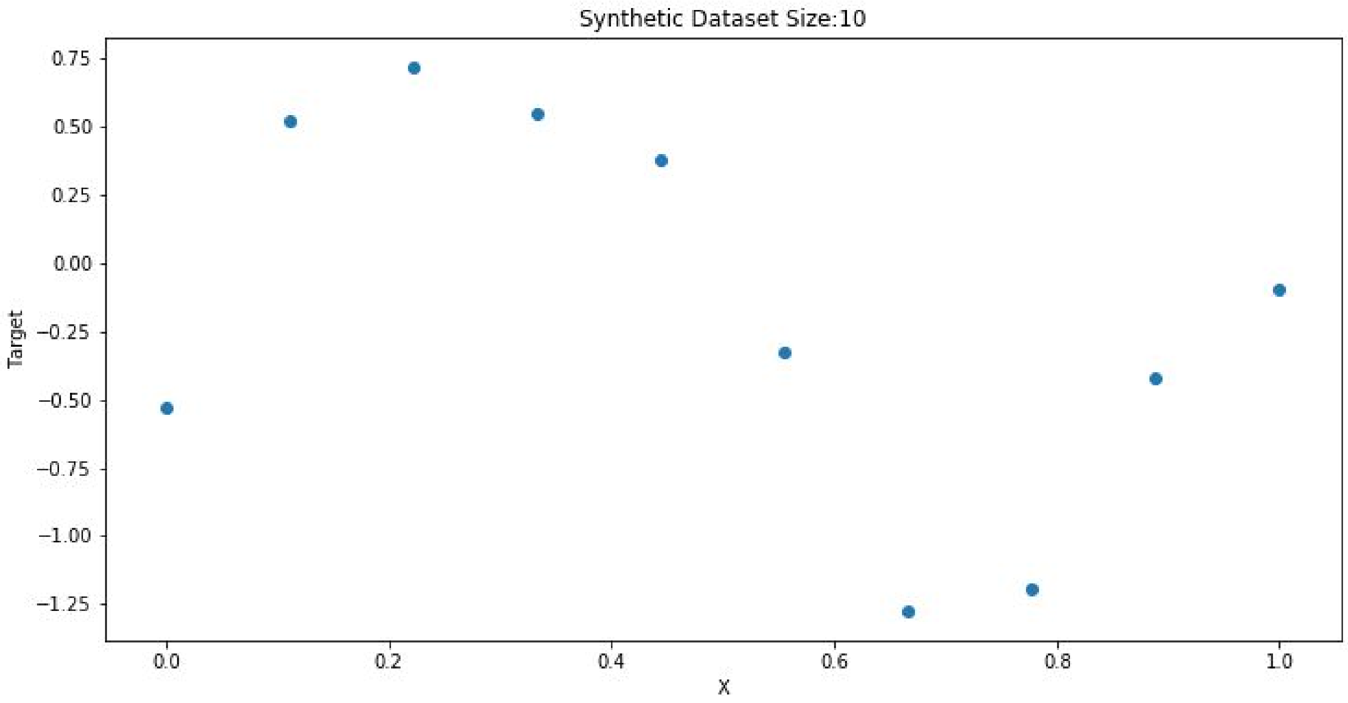
	Test Error	Train Error
N=1	0.136274	0.092558
N=2	0.866295	0.065776
N=3	0.610678	0.062695
N=4	0.211864	0.049488
N=5	0.025834	0.037029
N=6	0.000064	0.028456
N=7	0.003132	0.023072
N=8	0.103127	0.019719
N=9	0.646877	0.017579

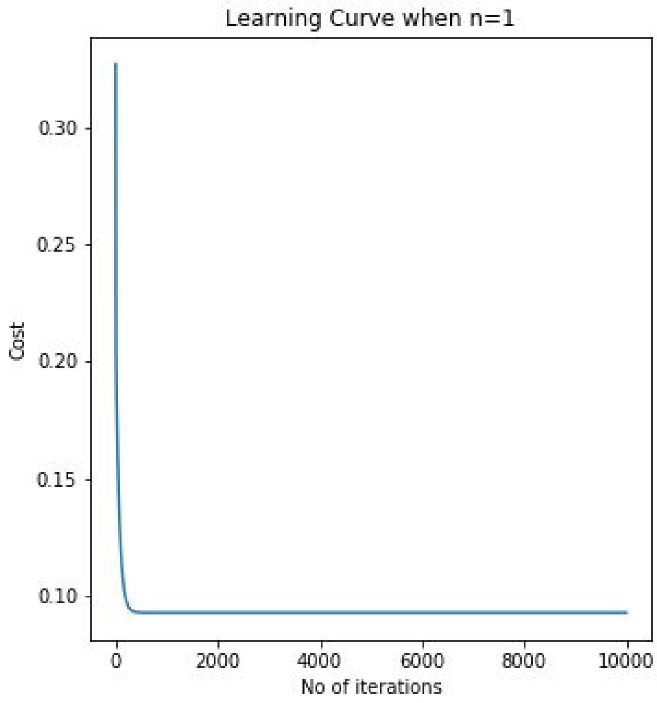
==

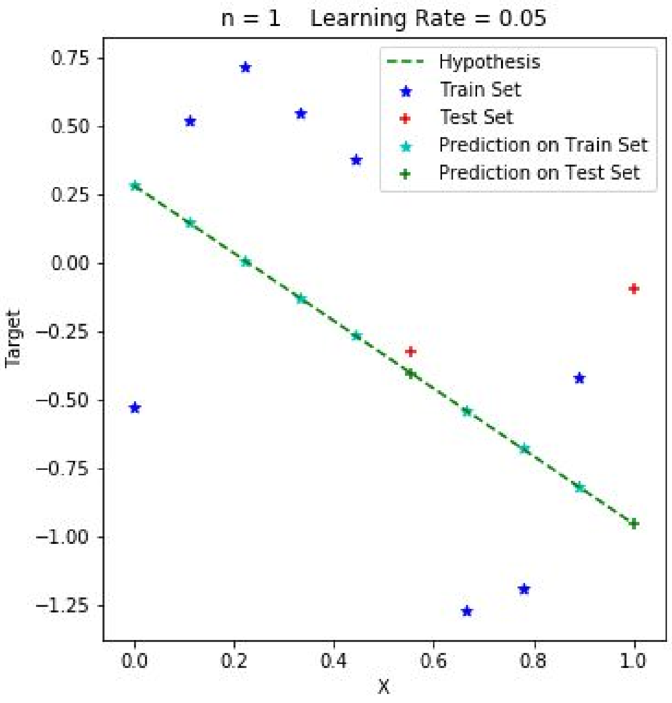
Part 2: Visualization of the dataset and the fitted curves

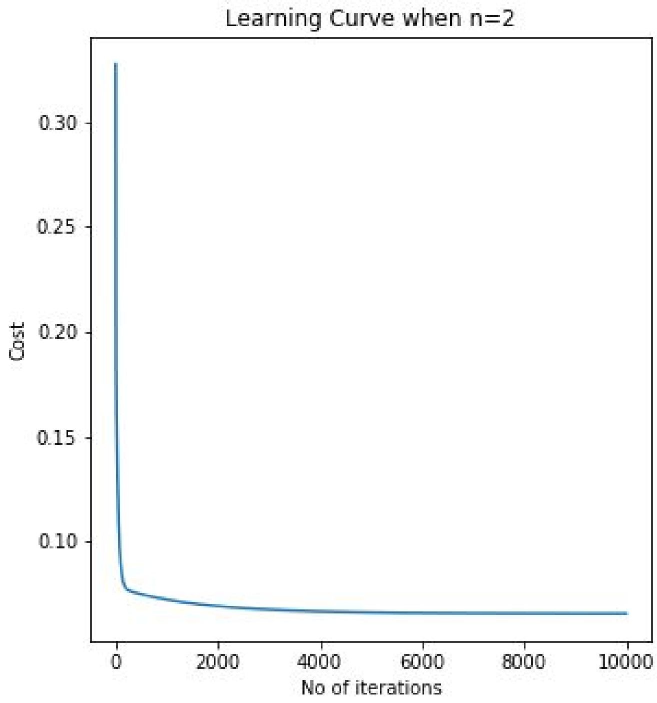
a) Plots of synthetic points

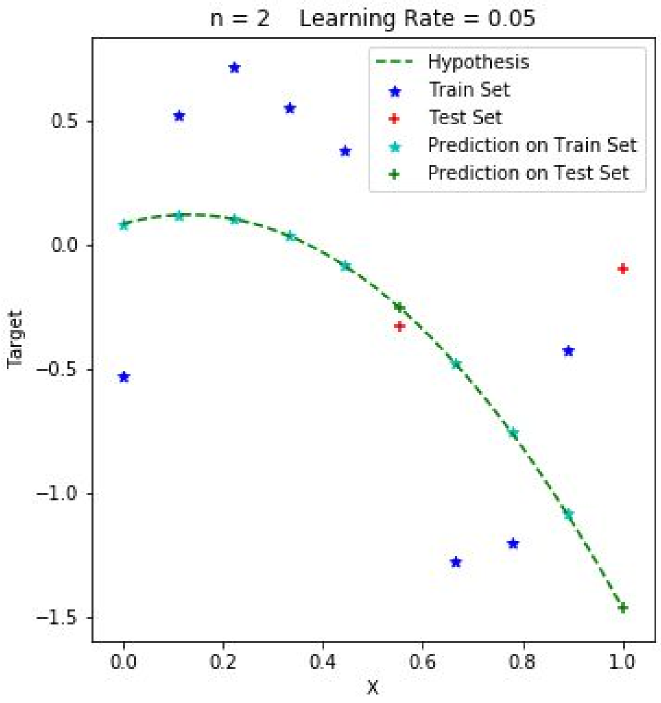
WHEN DATASIZE=10

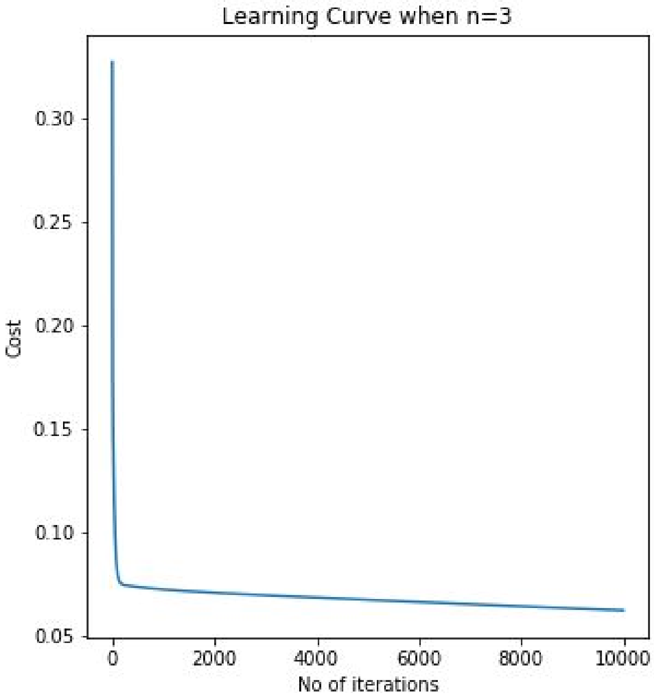


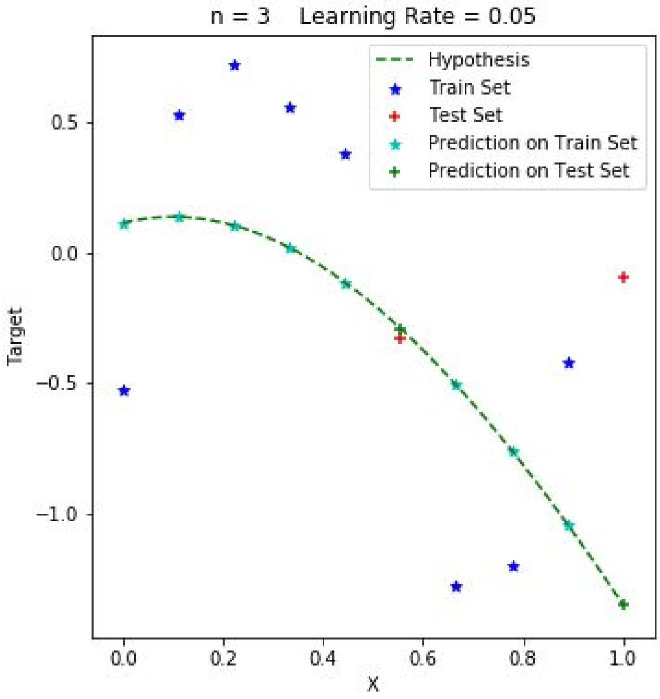


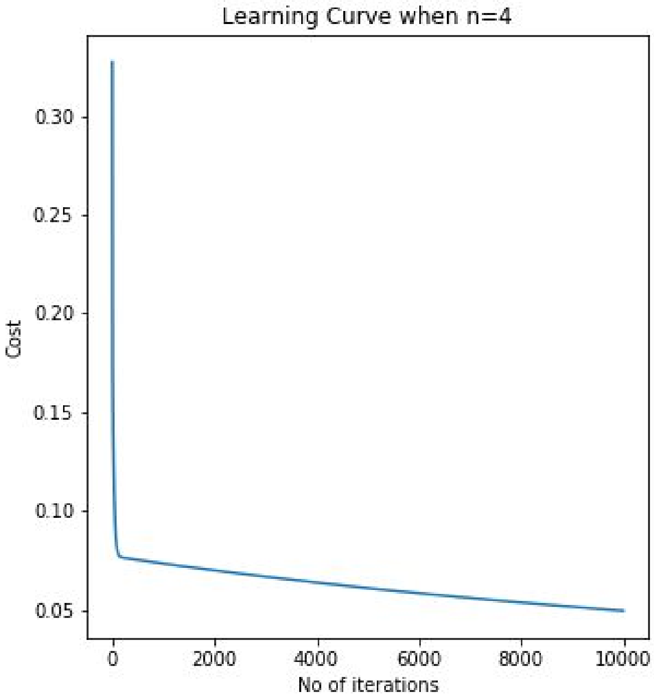


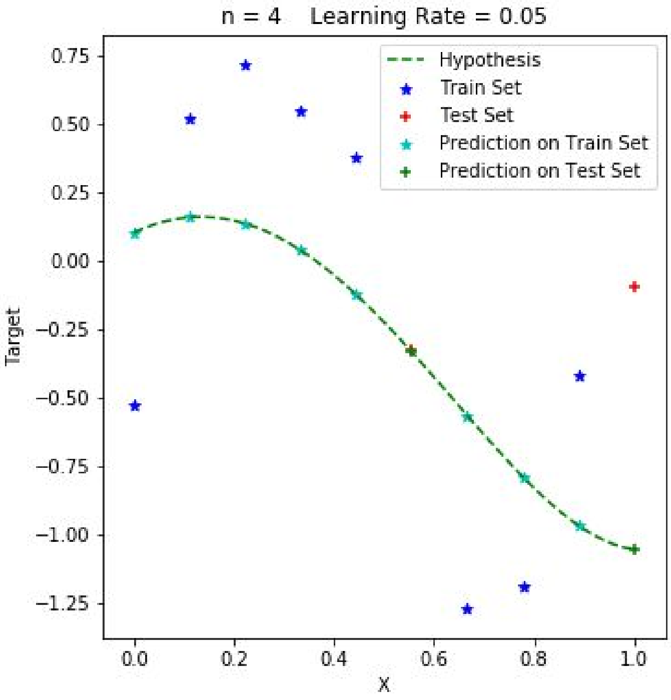


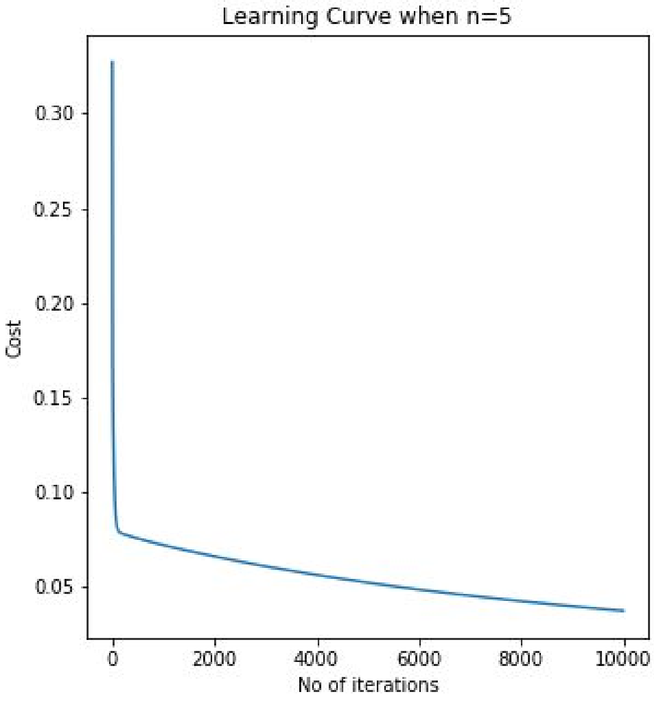


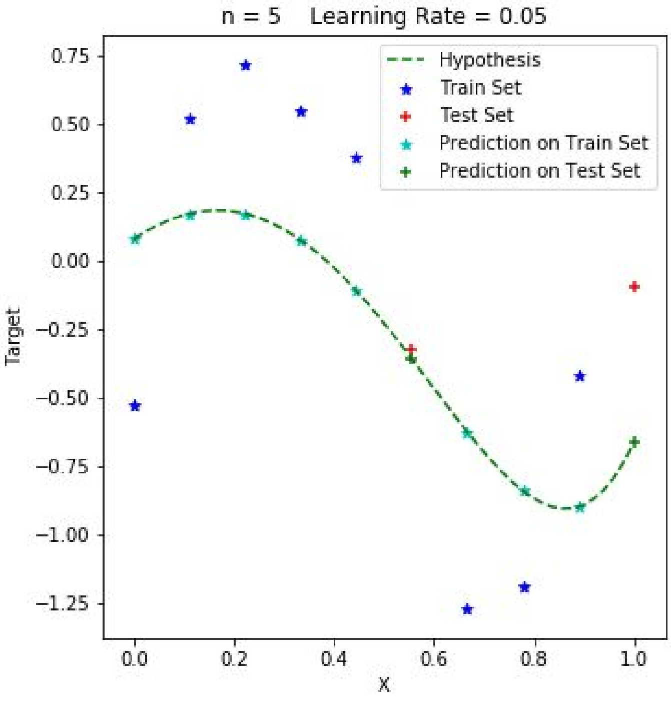

Learning Curve when n=3

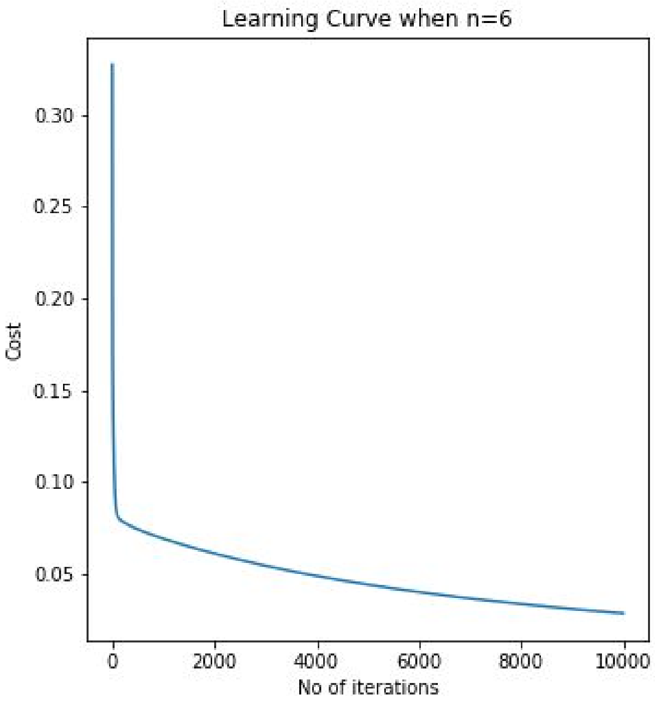

n = 3   Learning Rate = 0.05

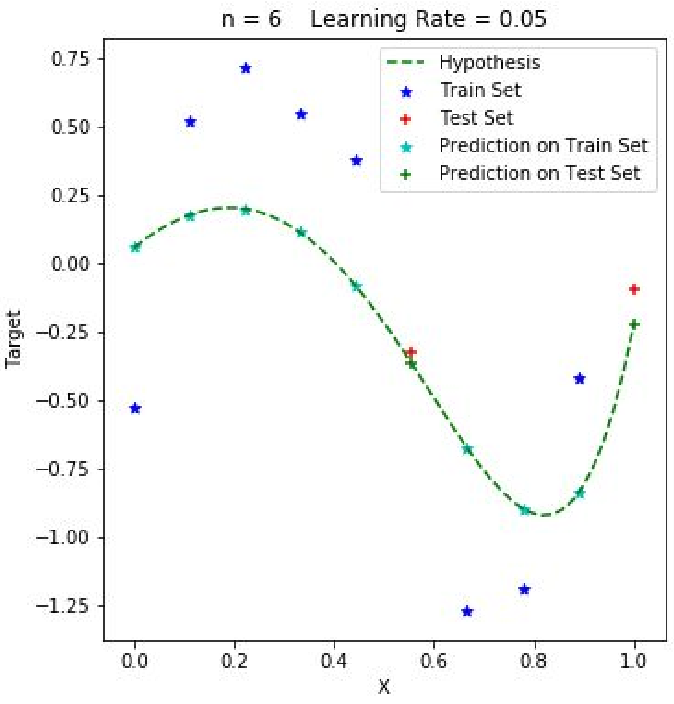

Learning Curve when n=4

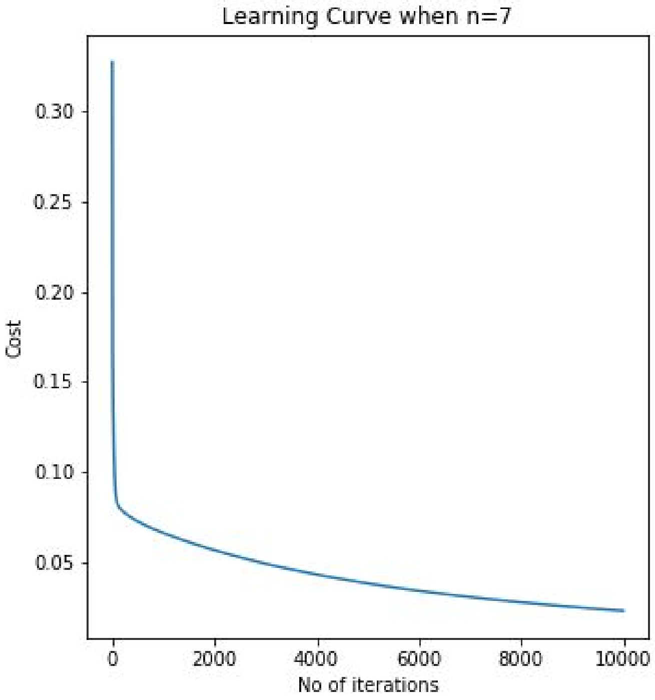

n = 4   Learning Rate = 0.05

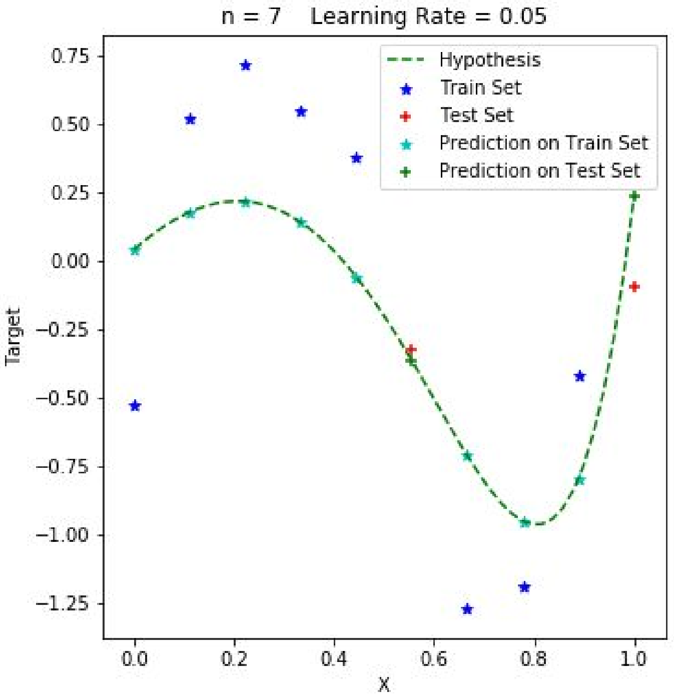

Learning Curve when n=5

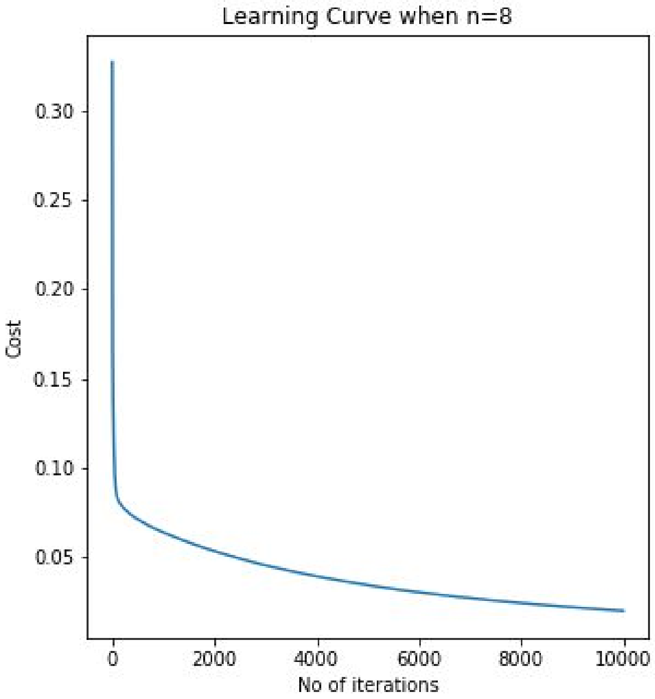

n = 5   Learning Rate = 0.05

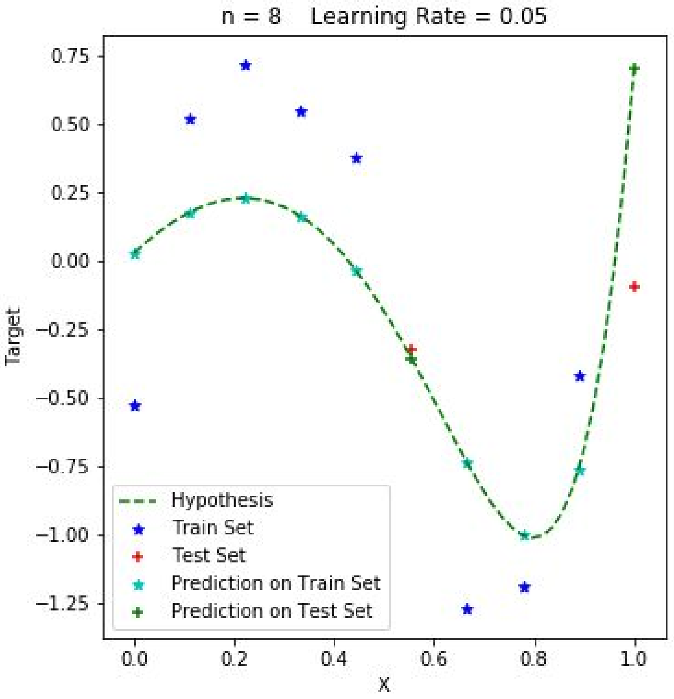

Learning Curve when n=6
Cost
0.30
0.25
0.20
0.15
0.10
0.05
0
2000
4000
6000
8000
10000
No of iterations

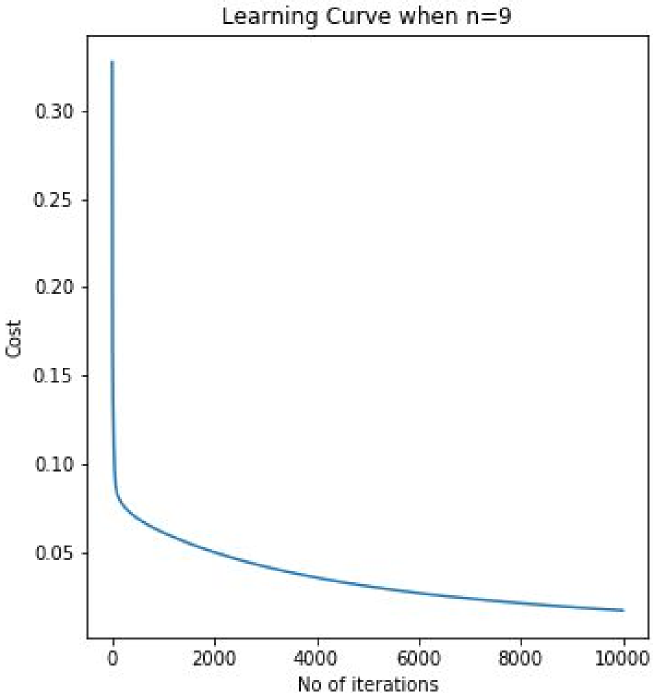

n = 6   Learning Rate = 0.05
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set
Target
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
0.0
0.2
0.4
0.6
0.8
1.0
X

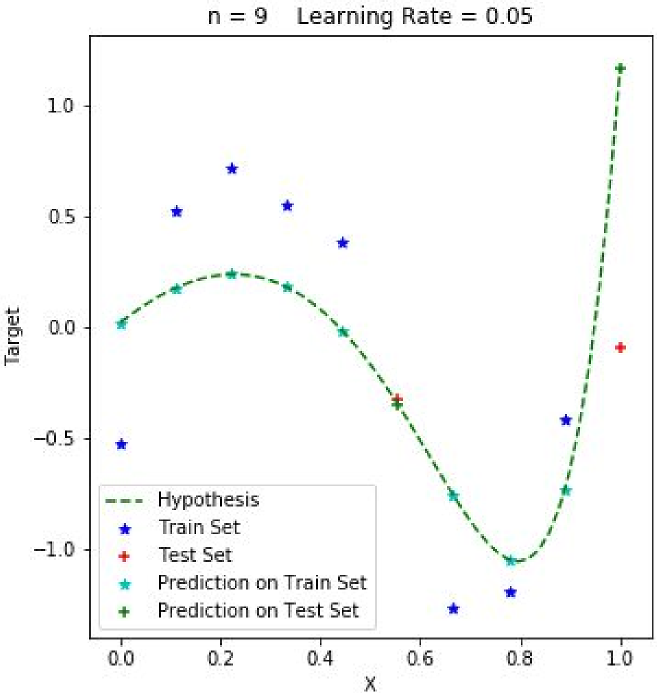

Learning Curve when n=7
Cost
0.30
0.25
0.20
0.15
0.10
0.05
0
2000
4000
6000
8000
10000
No of iterations


n = 7   Learning Rate = 0.05
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set
Target
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
0.0
0.2
0.4
0.6
0.8
1.0
X

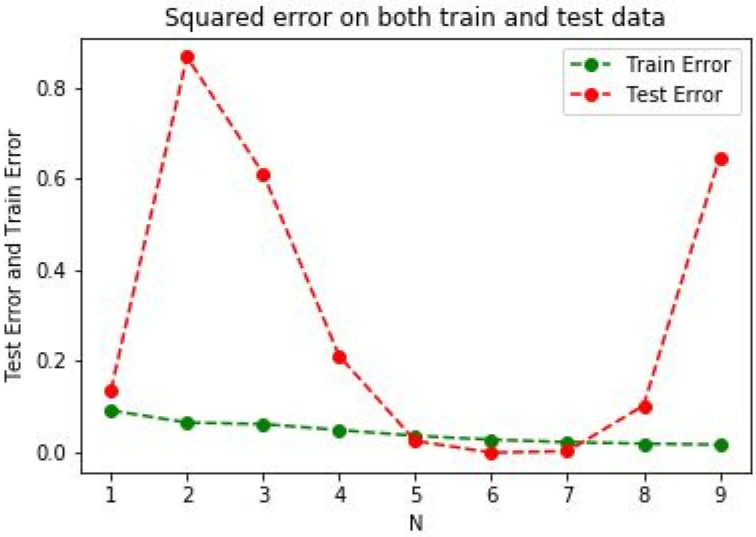

Learning Curve when n=8
Cost
0.30
0.25
0.20
0.15
0.10
0.05
0
2000
4000
6000
8000
10000
No of iterations

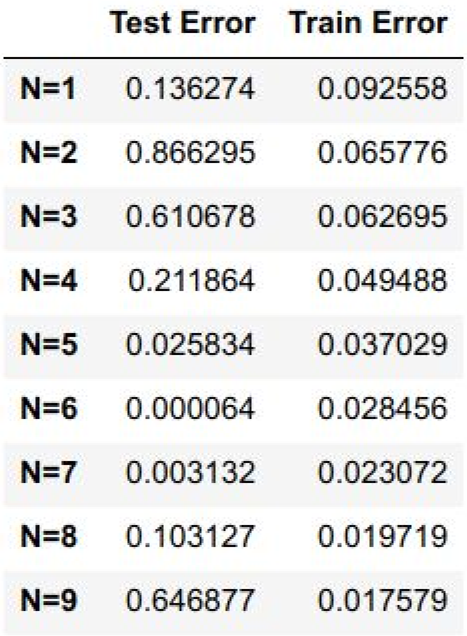

n = 8   Learning Rate = 0.05
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set
Target
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
0.0
0.2
0.4
0.6
0.8
1.0
X





b) Squared error on both train and test data:





	Test Error	Train Error
N=1	0.136274	0.092558
N=2	0.866295	0.065776
N=3	0.610678	0.062695
N=4	0.211864	0.049488
N=5	0.025834	0.037029
N=6	0.000064	0.028456
N=7	0.003132	0.023072
N=8	0.103127	0.019719
N=9	0.646877	0.017579

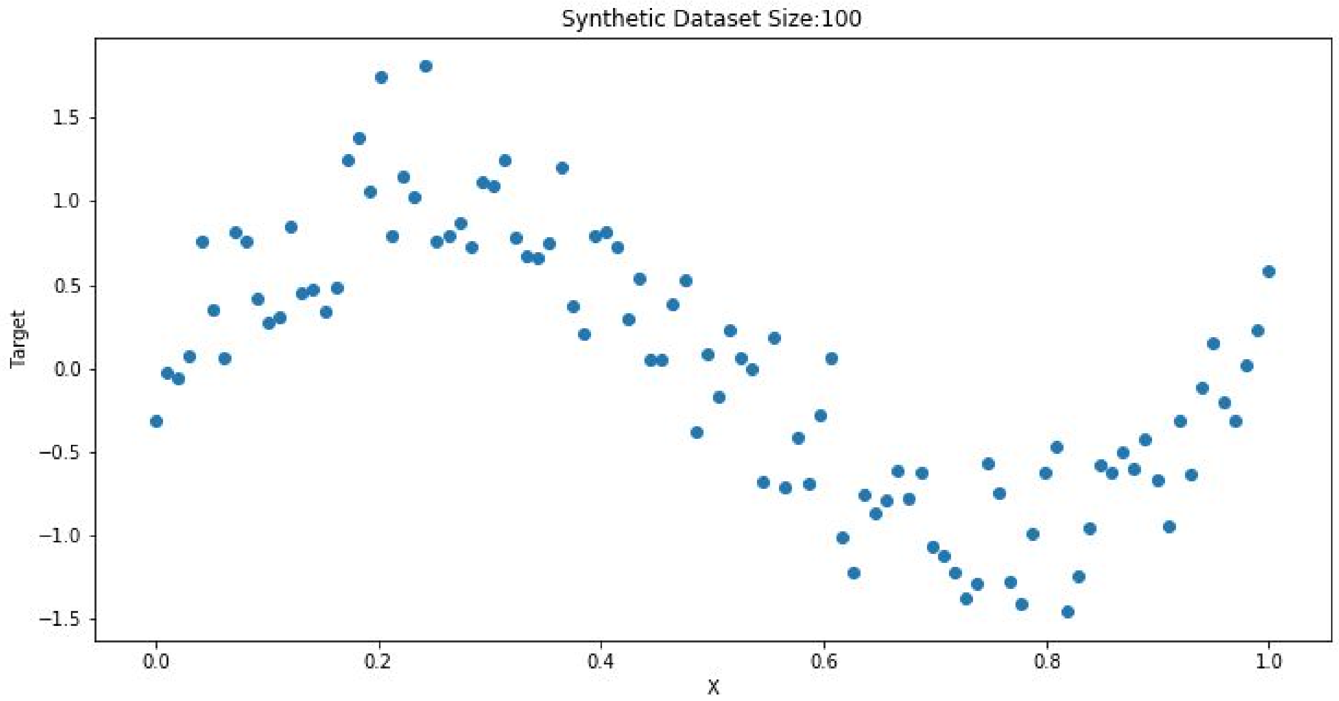
N = 6 seems to be suitable for this synthetic dataset because at N=6 test error is minimum and after N=6 as N increases test error also increases.

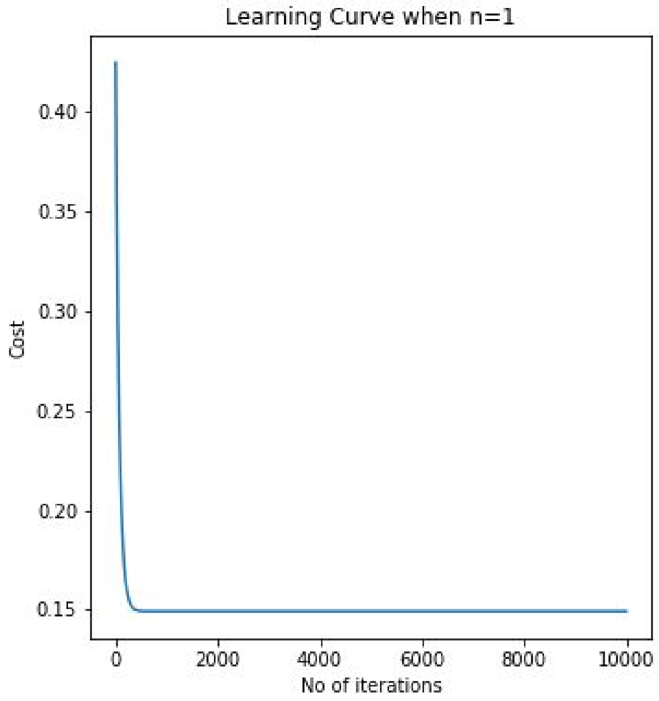
So, after N=6 even though our training error decreases our model will perform poor on the test set.

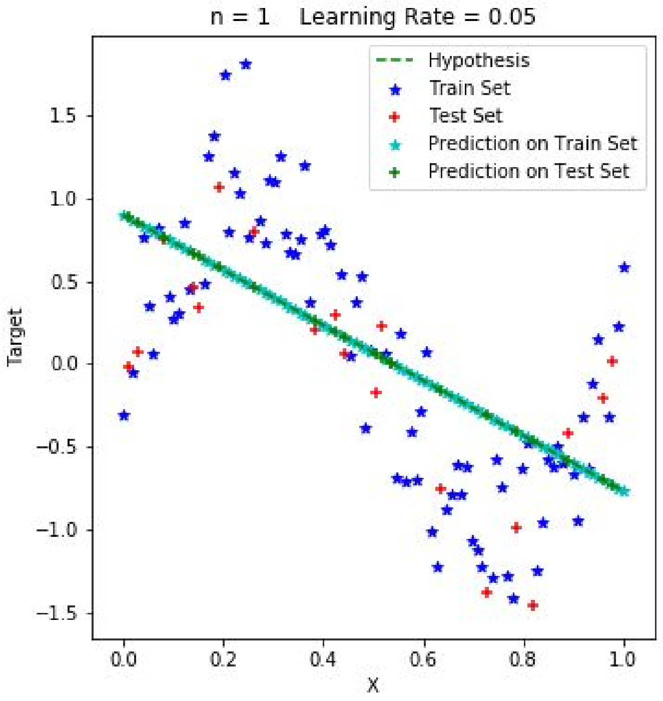
==

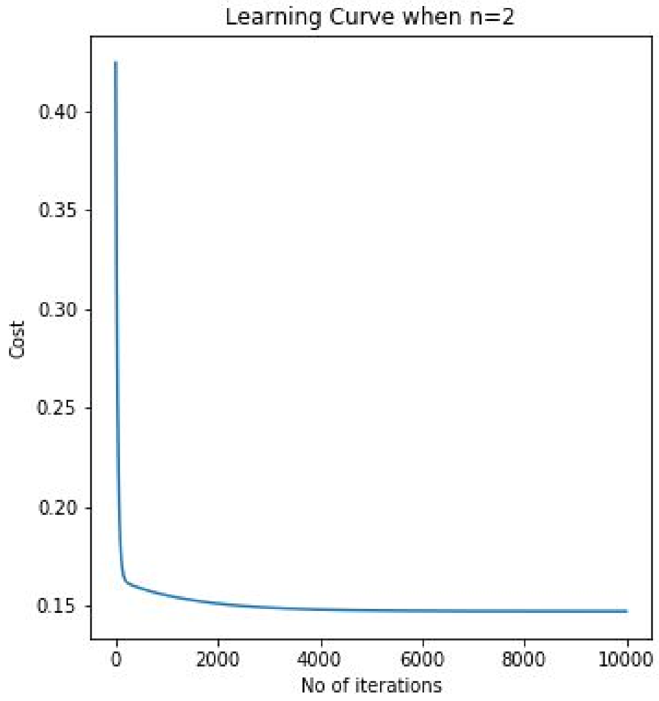
Part 3: Experimenting with the larger training set

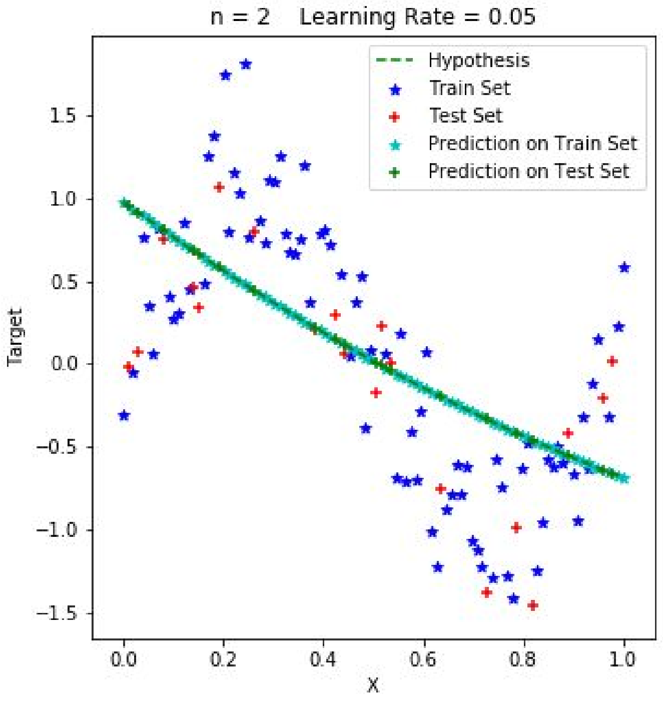
WHEN DATASIZE=100

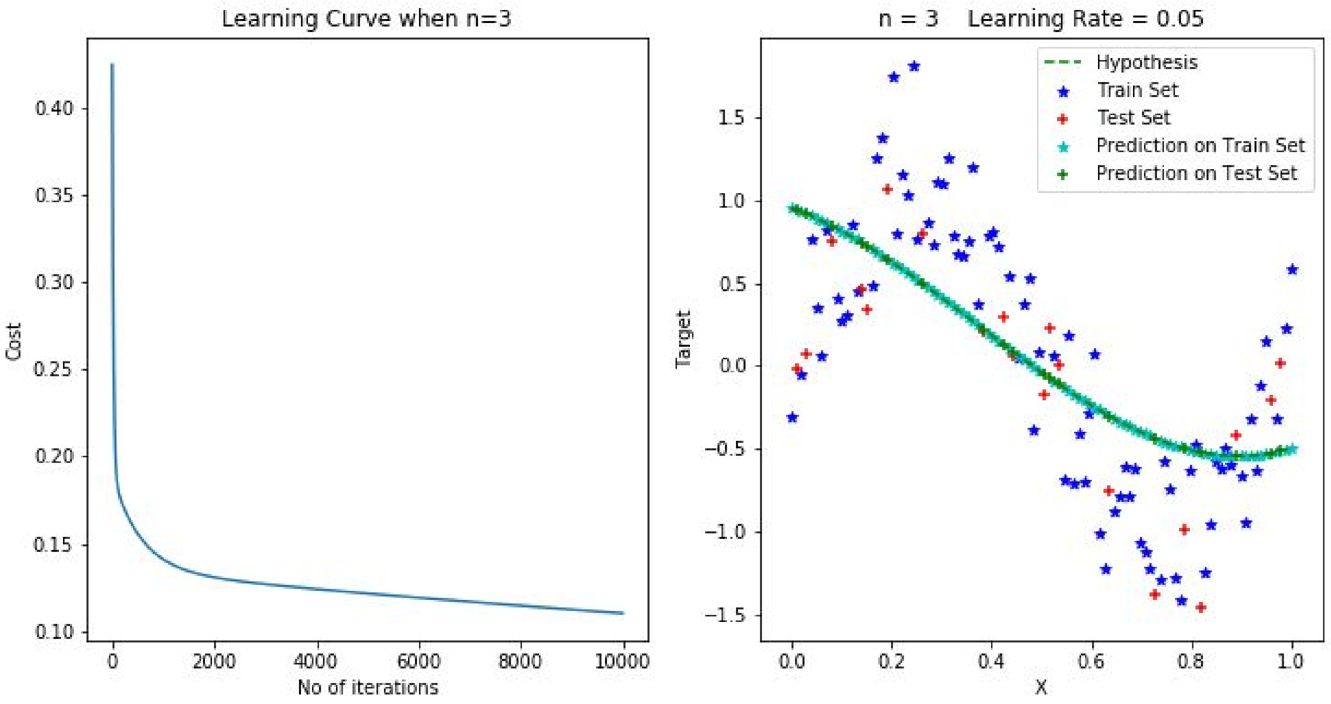


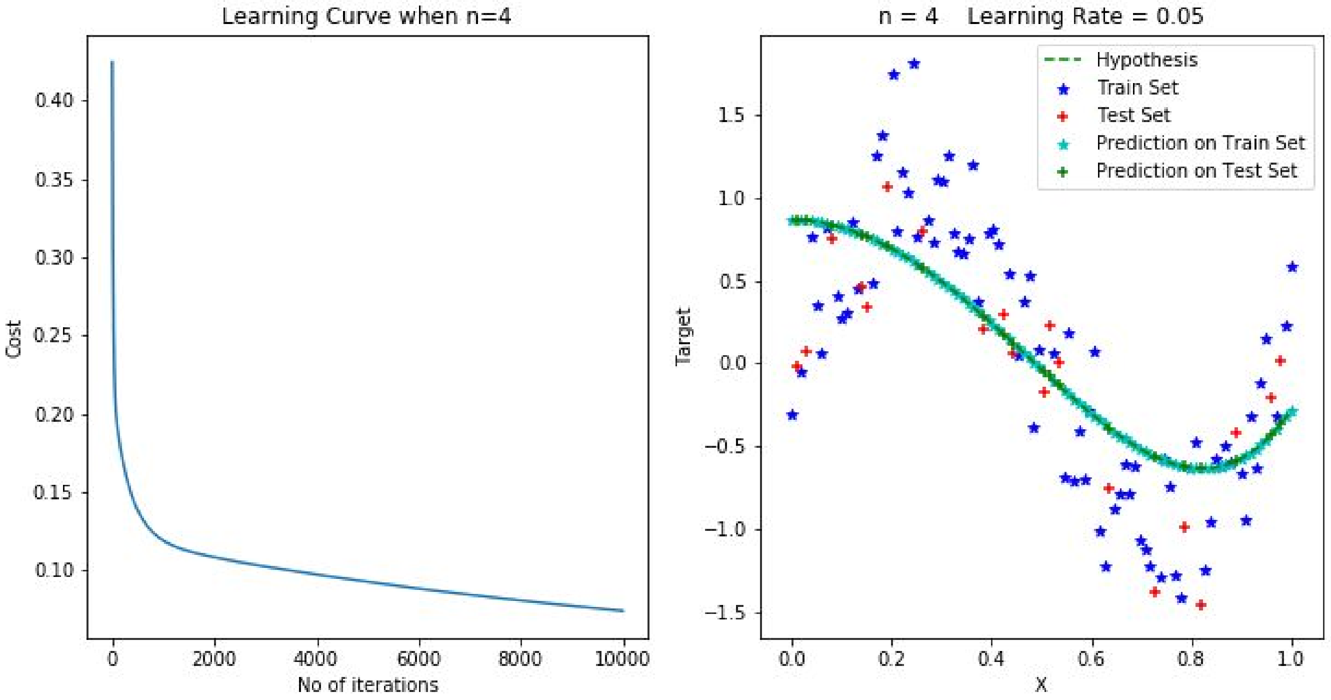


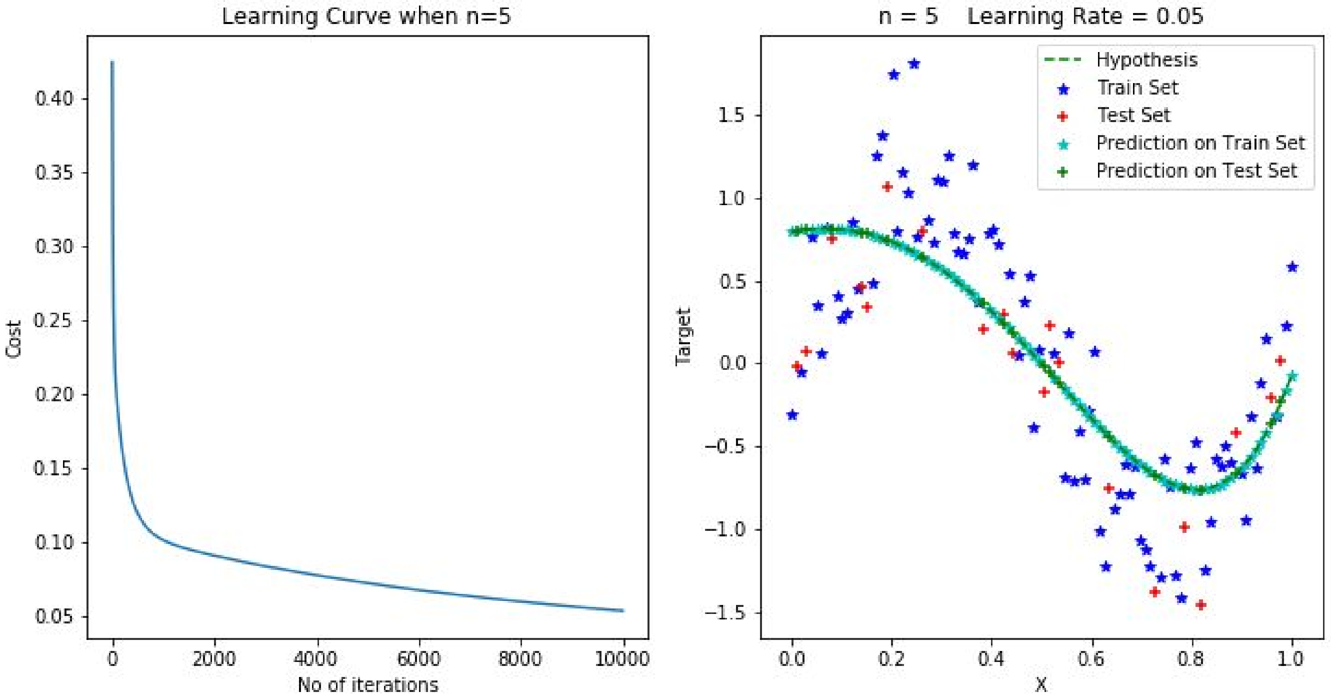


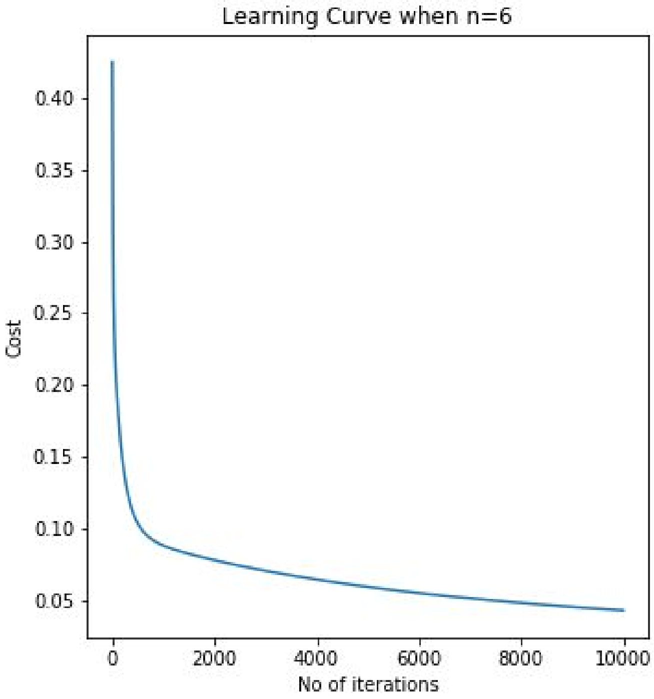


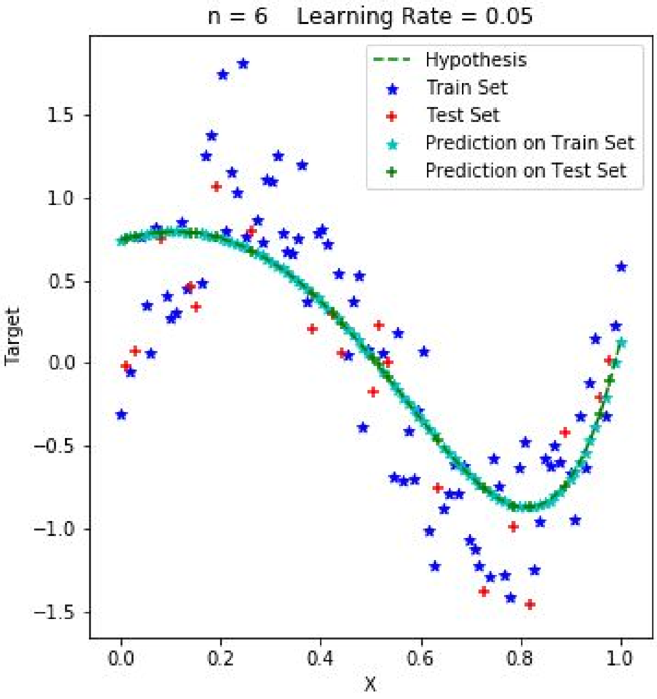


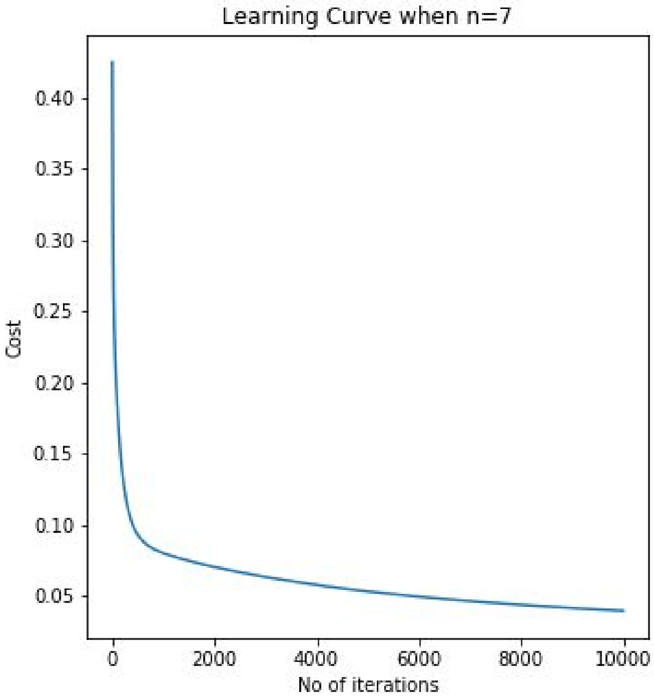

Learning Curve when n=3
Cost
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0
2000
4000
6000
8000
10000
No of iterations
n = 3 Learning Rate = 0.05
Target
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0.0
0.2
0.4
0.6
0.8
1.0
X
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set

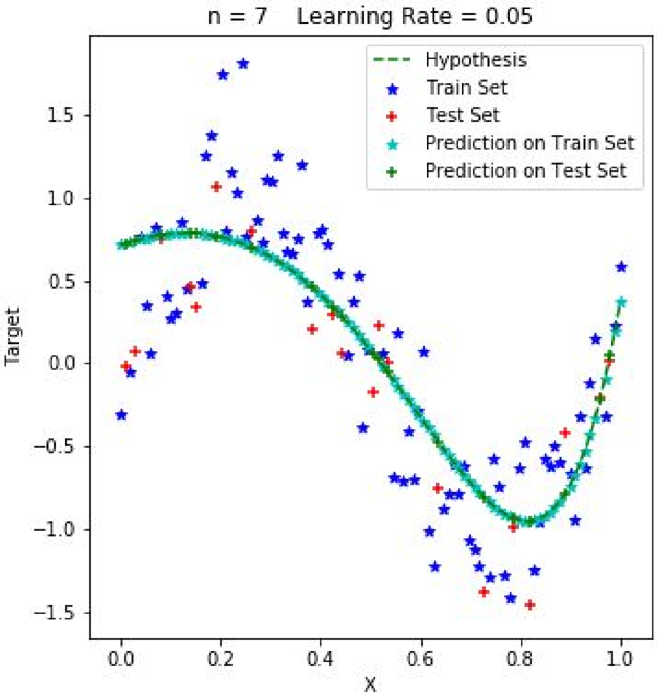

Learning Curve when n=4
Cost
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0
2000
4000
6000
8000
10000
No of iterations
n = 4 Learning Rate = 0.05
Target
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0.0
0.2
0.4
0.6
0.8
1.0
X
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set

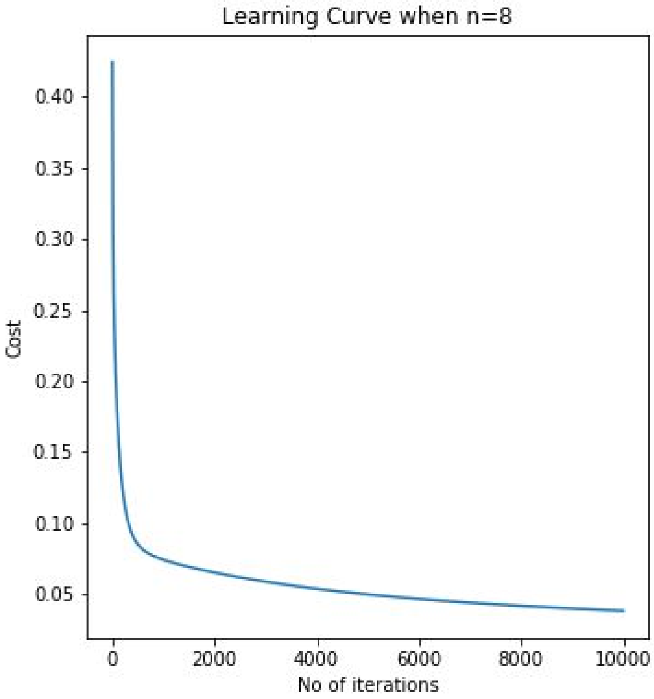

Learning Curve when n=5
Cost
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0
2000
4000
6000
8000
10000
No of iterations
n = 5 Learning Rate = 0.05
Target
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0.0
0.2
0.4
0.6
0.8
1.0
X
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set

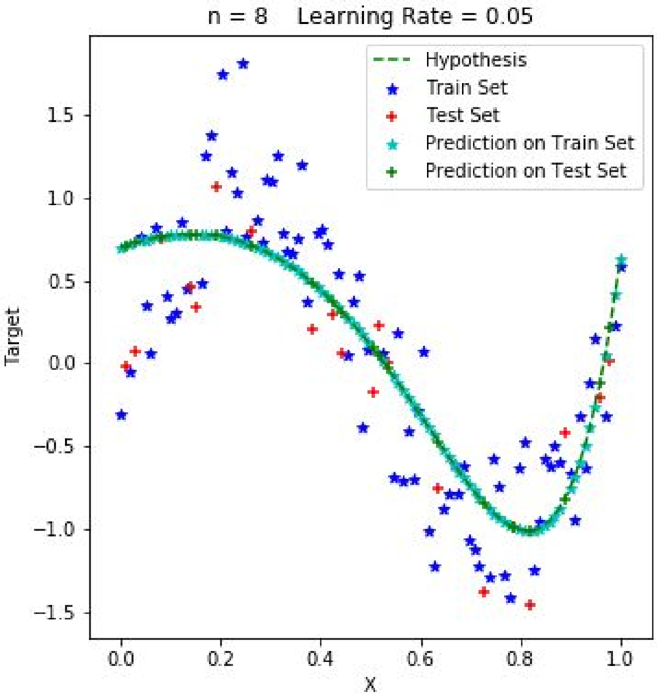

Learning Curve when n=6

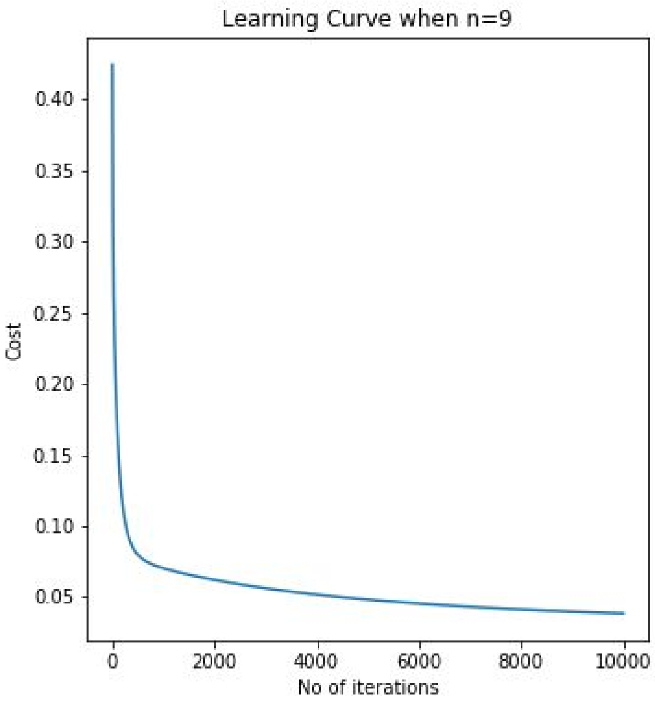

n = 6   Learning Rate = 0.05

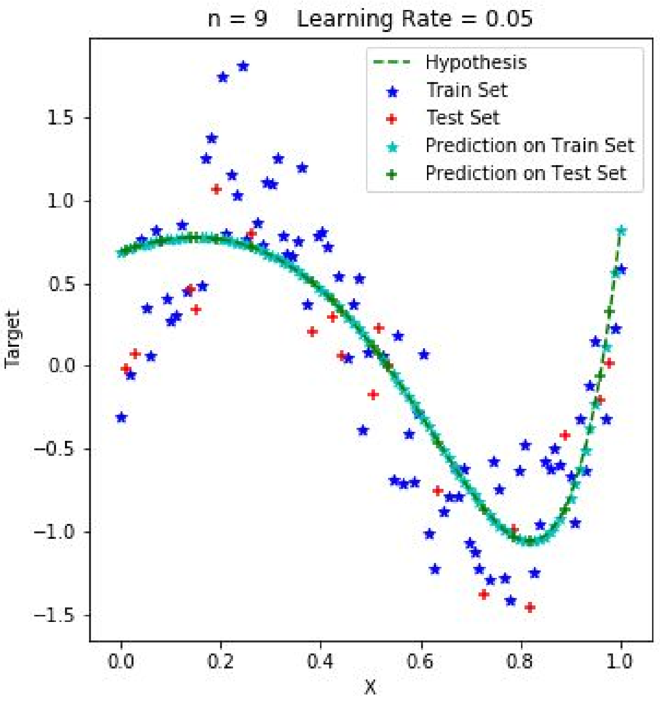

Learning Curve when n=7


n = 7   Learning Rate = 0.05


Learning Curve when n=8

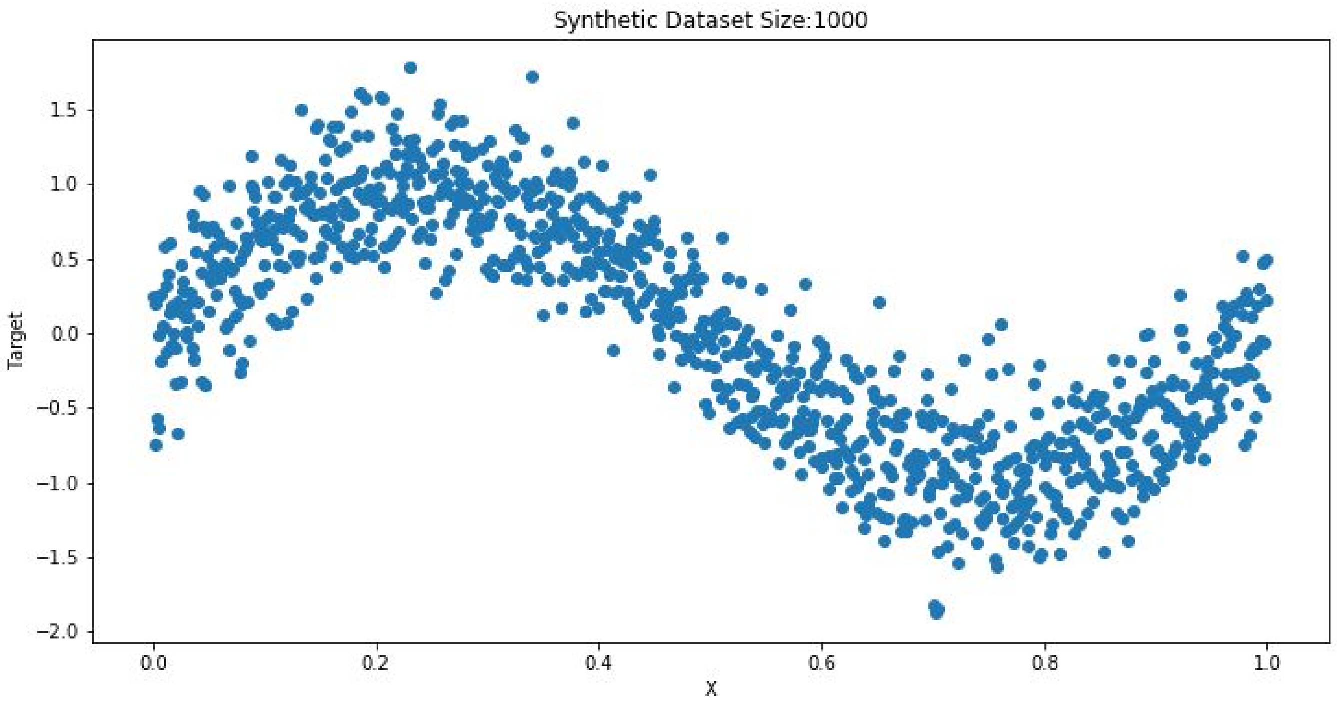

n = 8   Learning Rate = 0.05

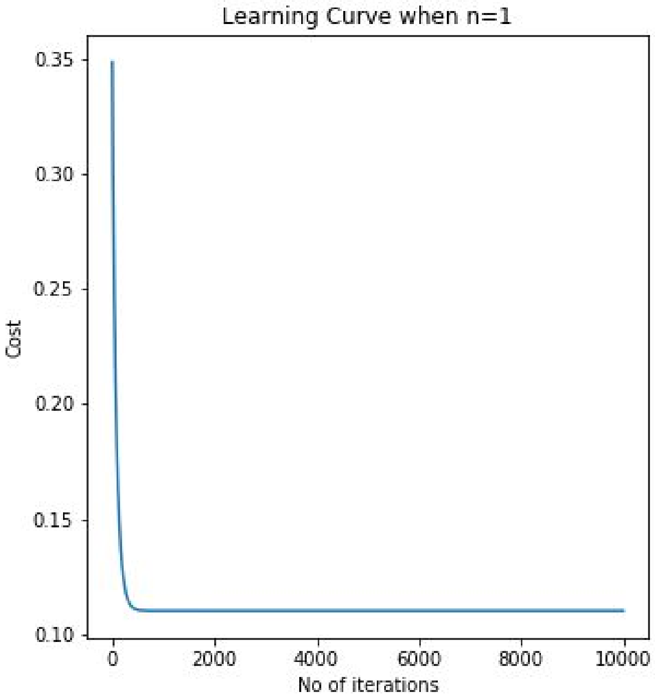


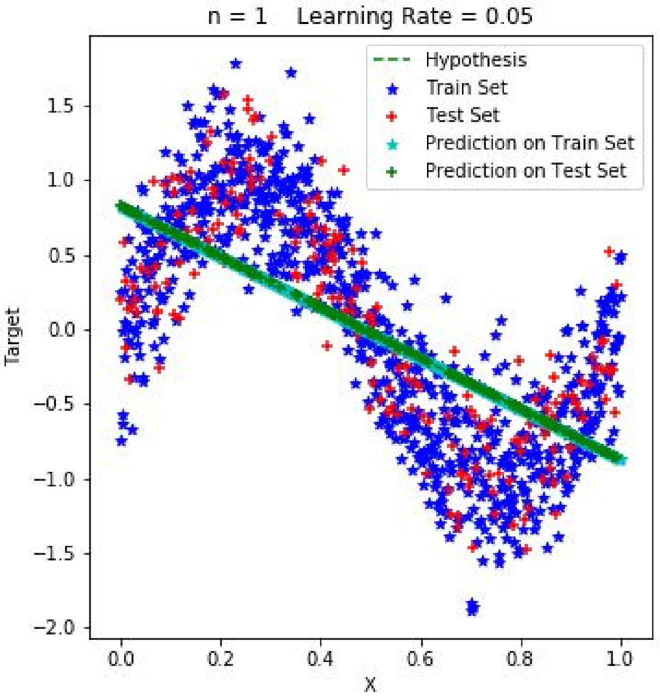


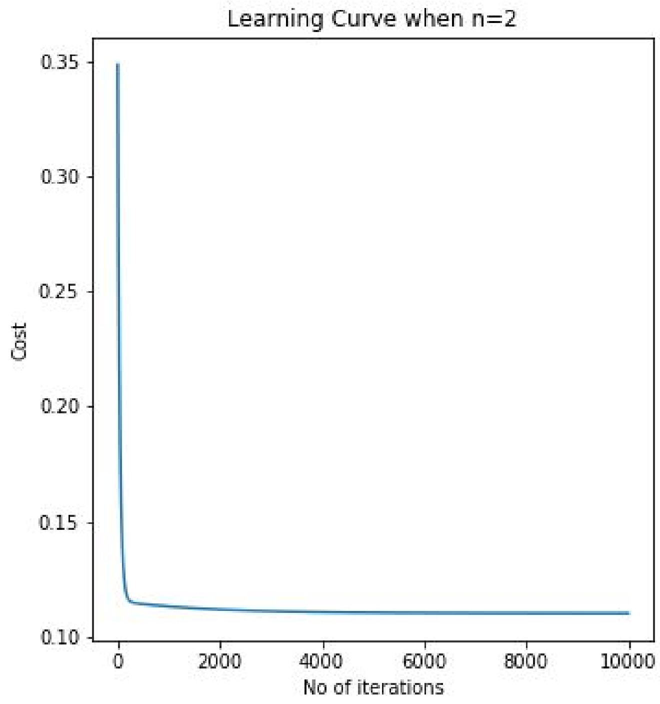
****

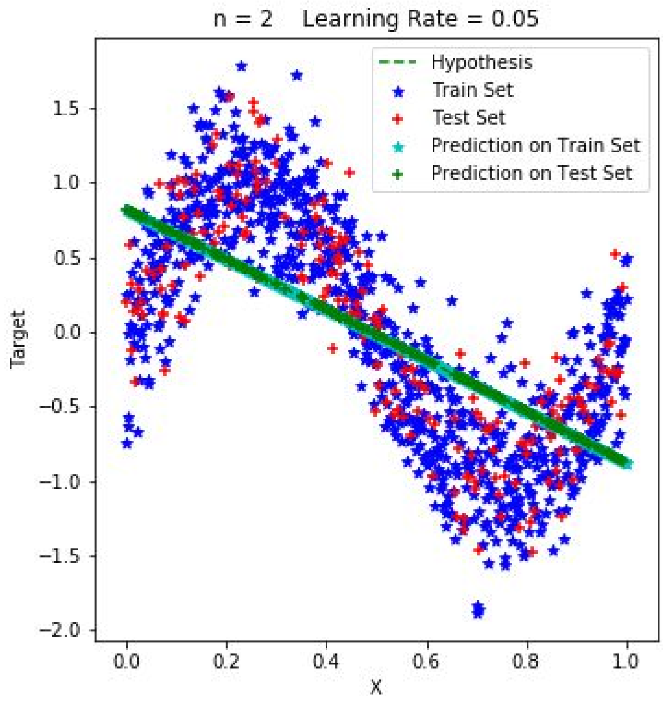
WHEN DATASIZE=1000

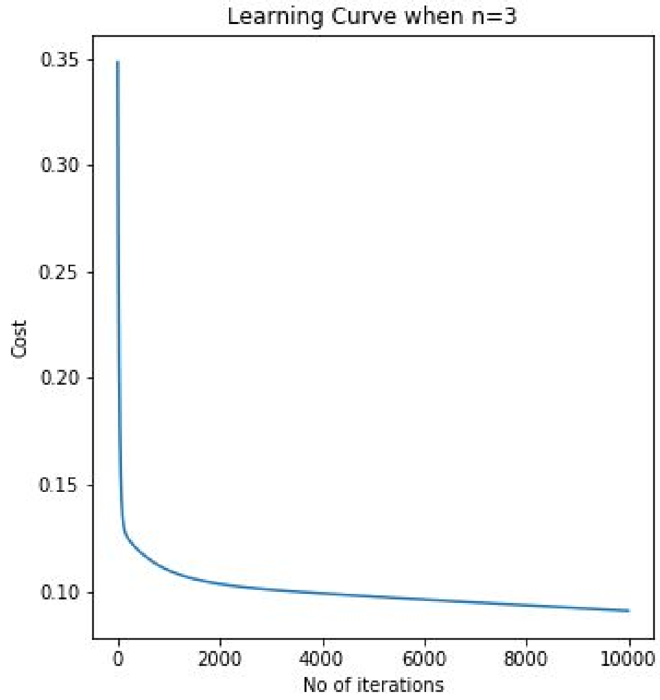


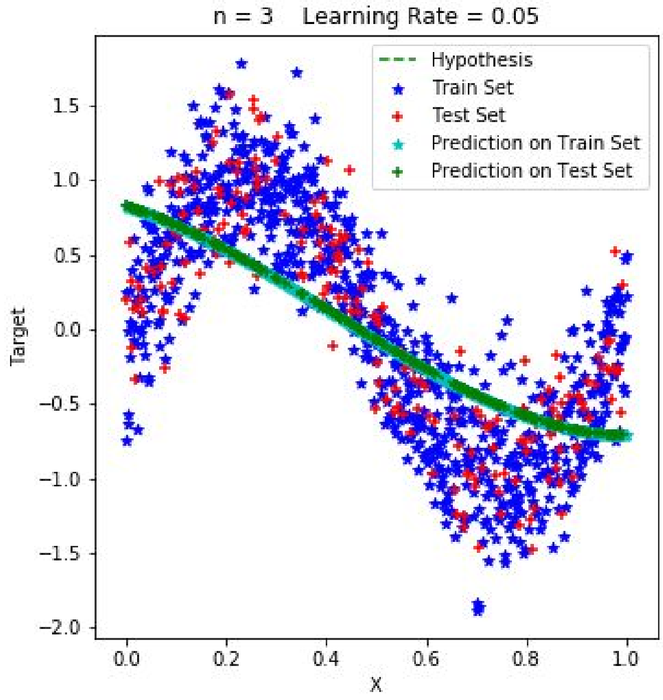


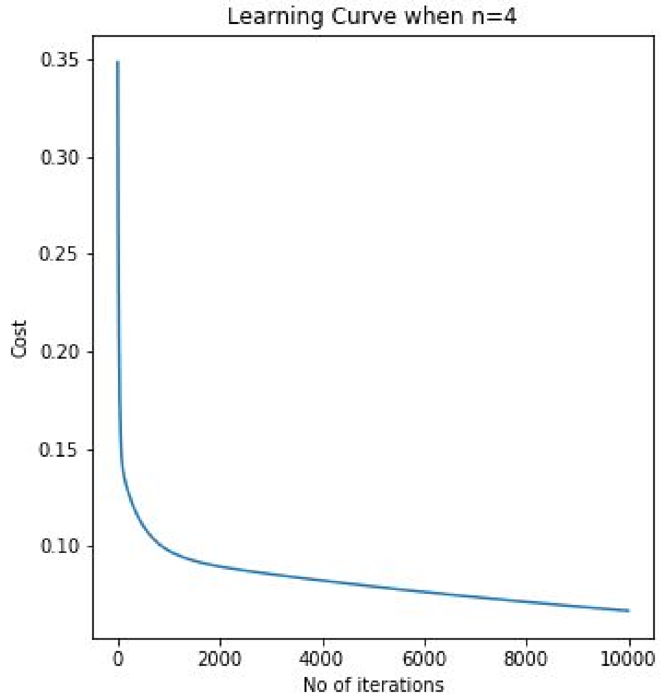


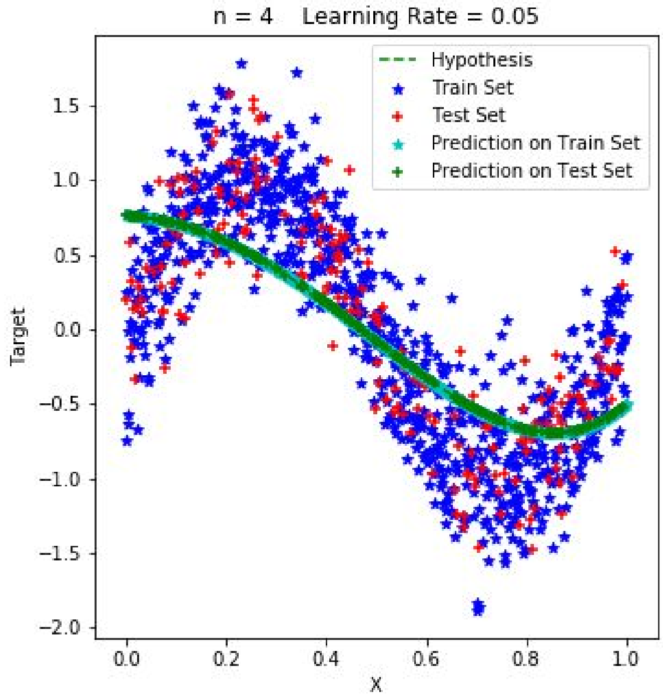

Learning Curve when n=2
Cost
0.35
0.30
0.25
0.20
0.15
0.10
0
2000
4000
6000
8000
10000
No of iterations

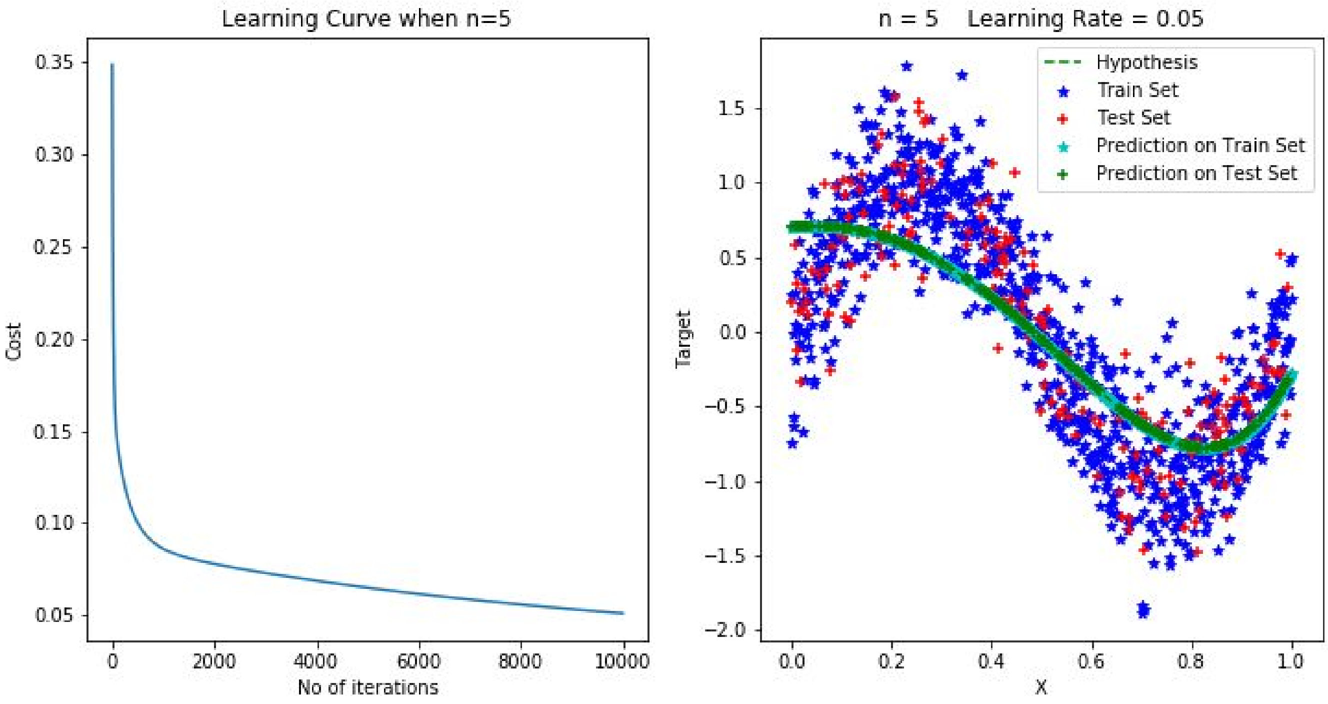

n = 2   Learning Rate = 0.05
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set
Target
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0
X

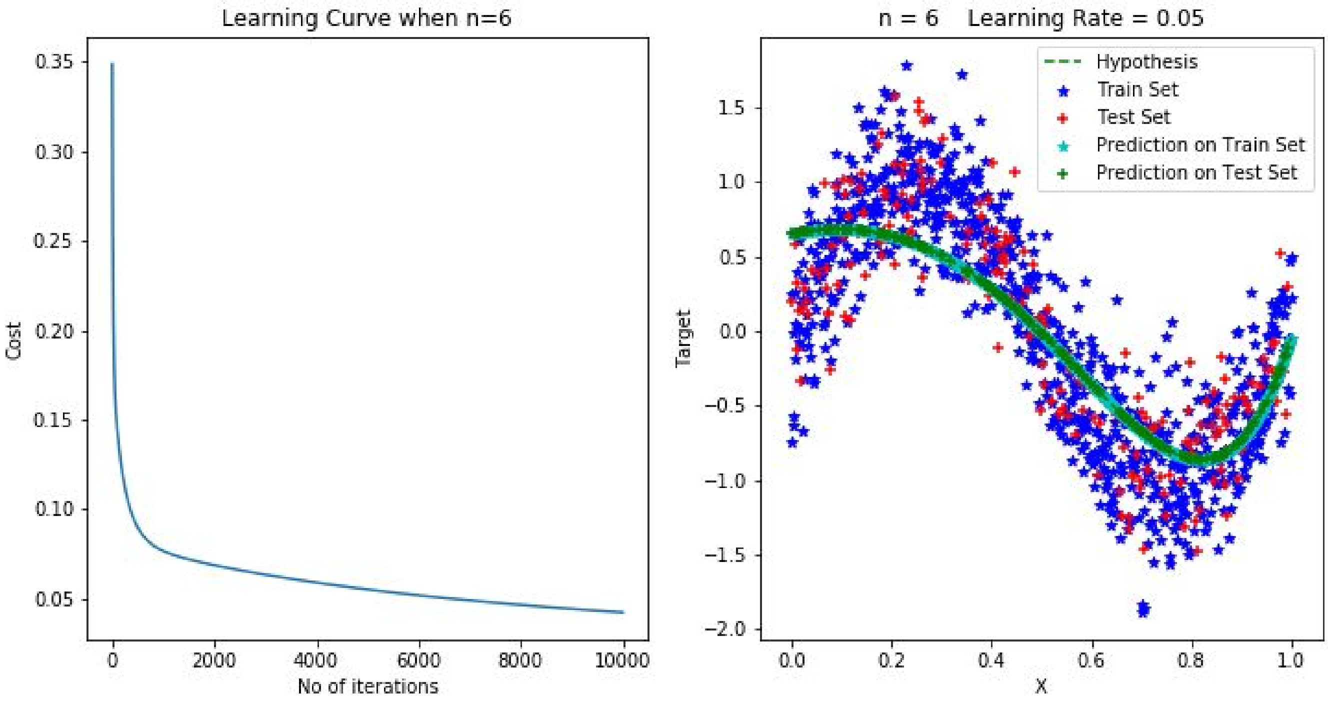

Learning Curve when n=3
Cost
0.35
0.30
0.25
0.20
0.15
0.10
0
2000
4000
6000
8000
10000
No of iterations

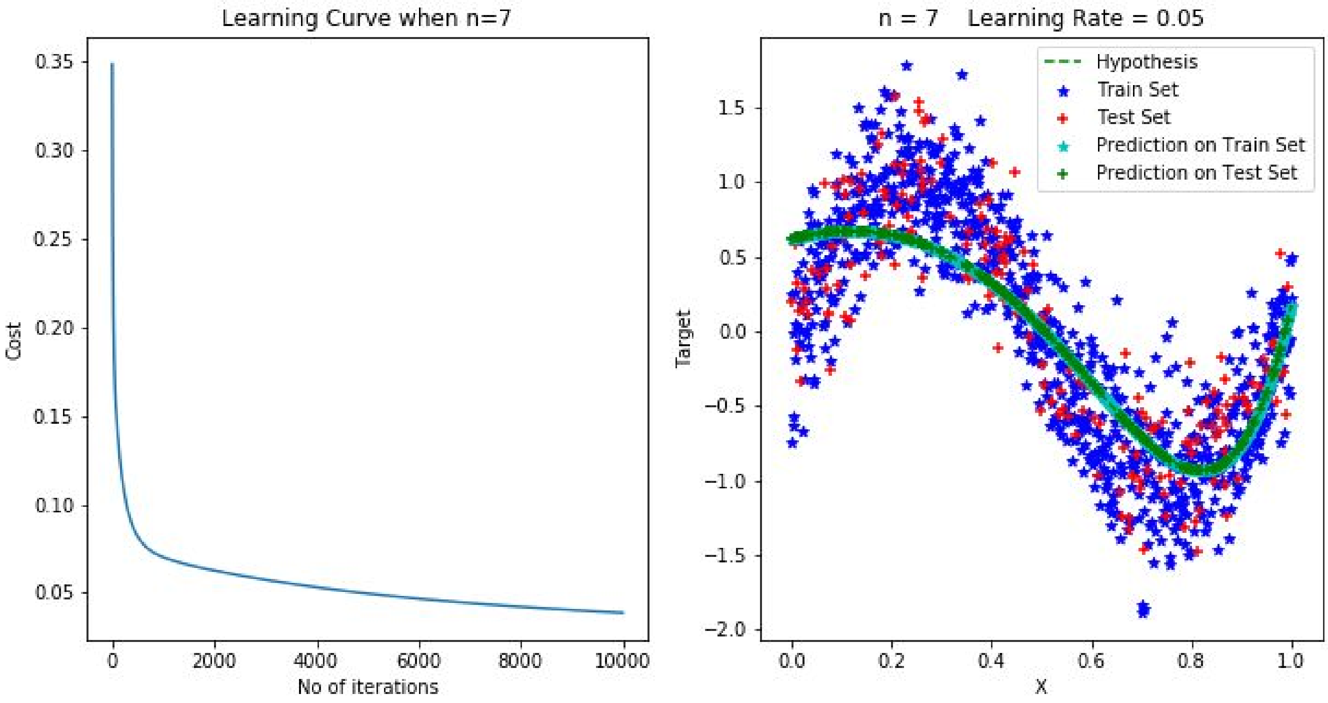

n = 3   Learning Rate = 0.05
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set
Target
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0
X

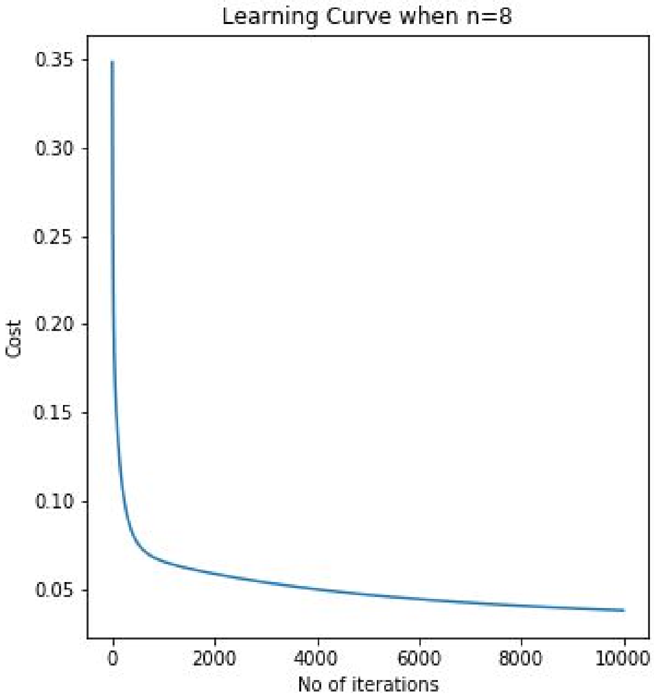

Learning Curve when n=4
Cost
0.35
0.30
0.25
0.20
0.15
0.10
0
2000
4000
6000
8000
10000
No of iterations

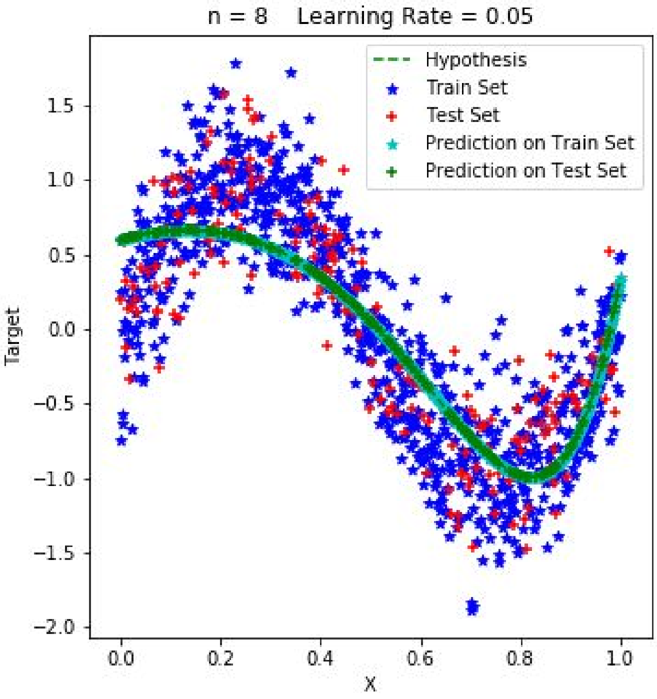

n = 4   Learning Rate = 0.05
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set
Target
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0
X

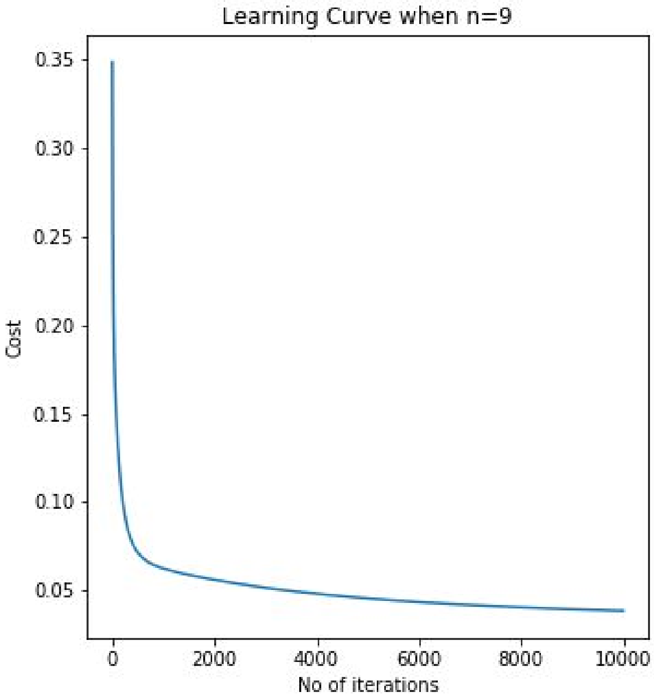

Learning Curve when n=5
Cost
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0
2000
4000
6000
8000
10000
No of iterations
n = 5   Learning Rate = 0.05
Target
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0
X
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set

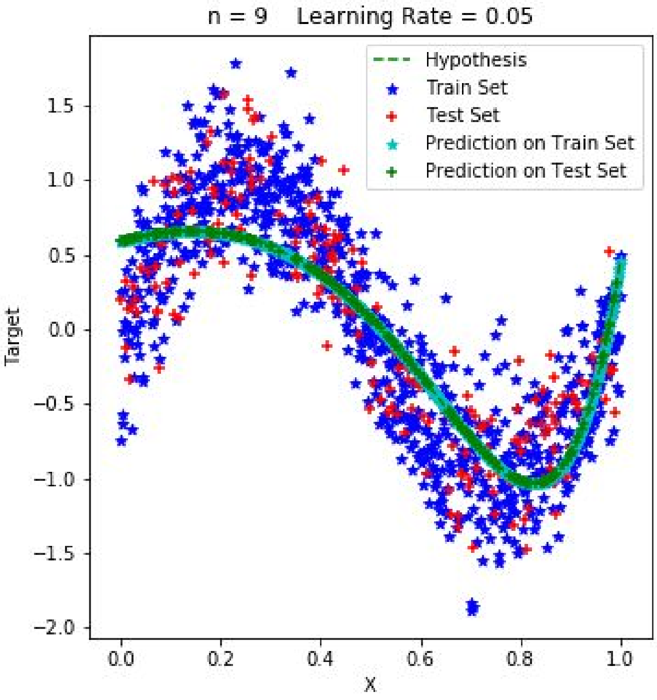

Learning Curve when n=6
Cost
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0
2000
4000
6000
8000
10000
No of iterations
n = 6   Learning Rate = 0.05
Target
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0
X
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set


Learning Curve when n=7
Cost
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0
2000
4000
6000
8000
10000
No of iterations
n = 7   Learning Rate = 0.05
Target
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0
X
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set



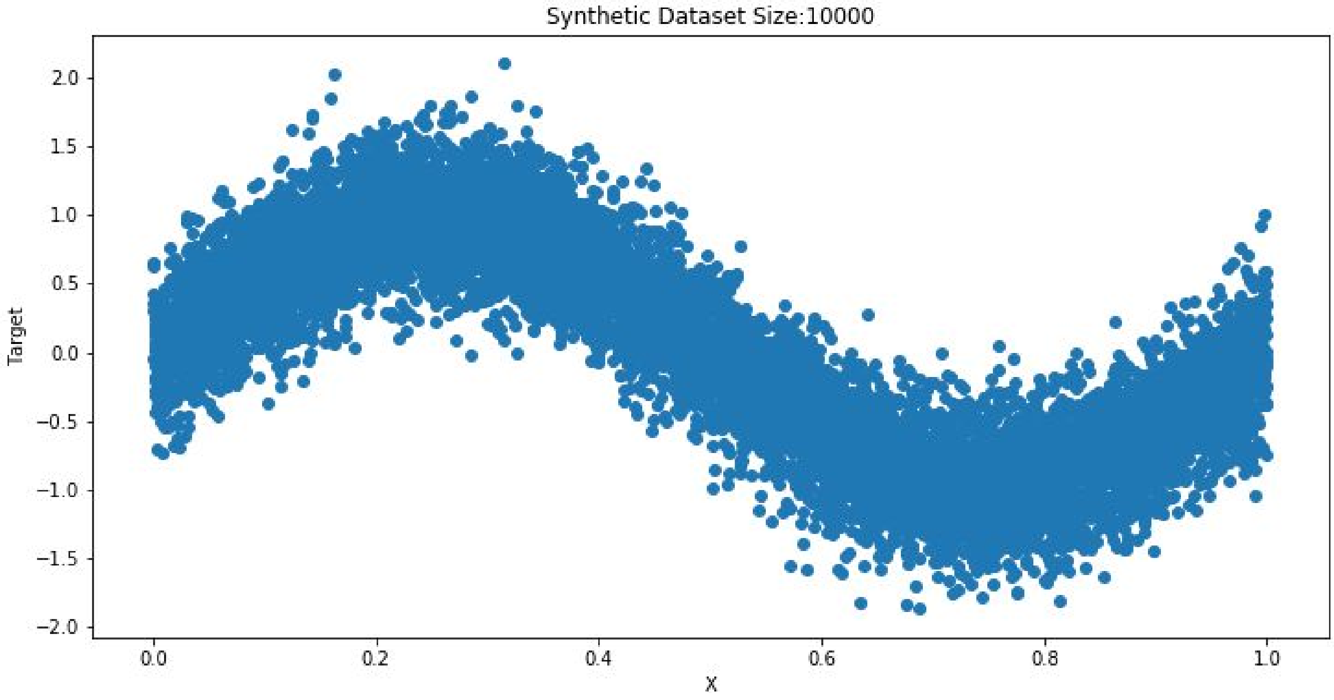


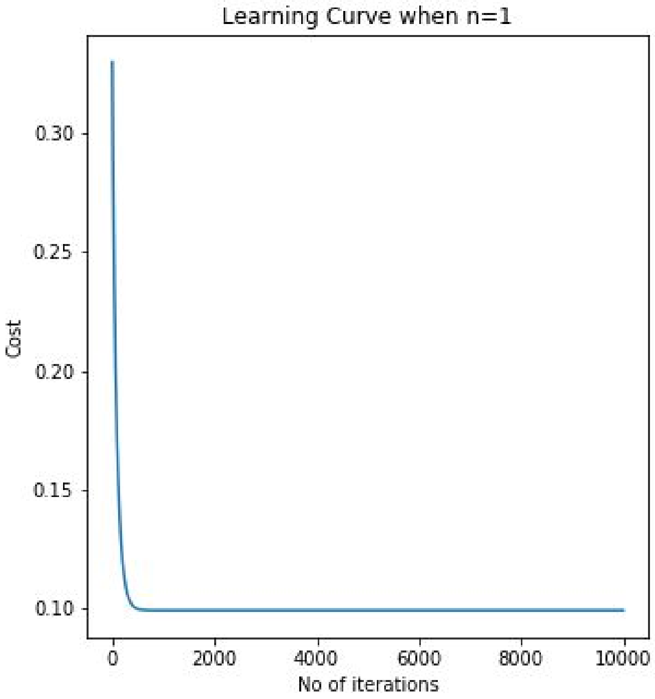


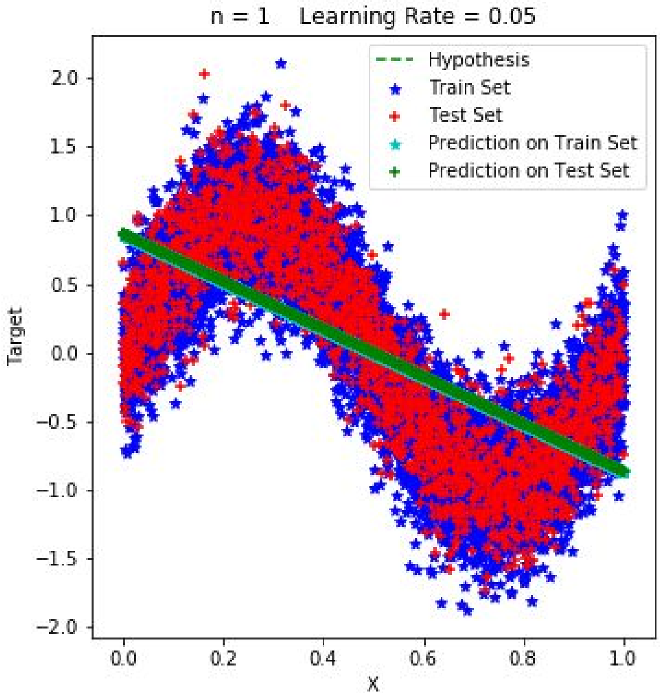


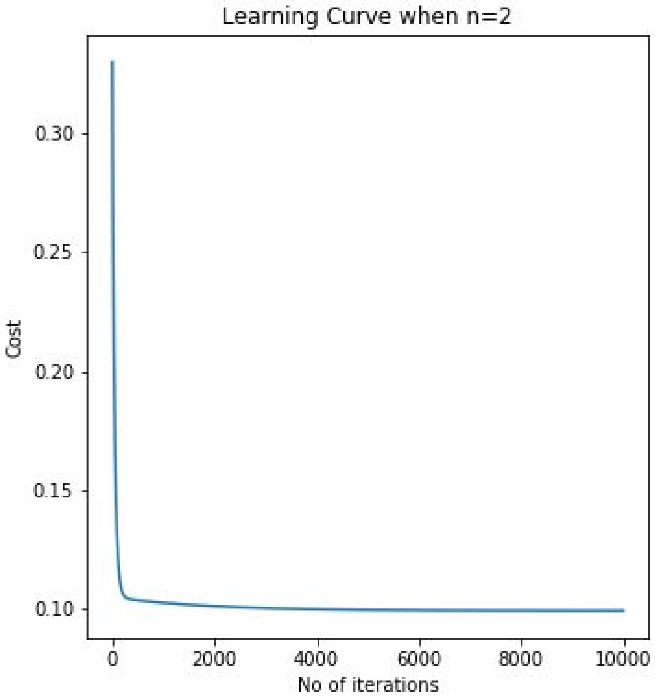
****

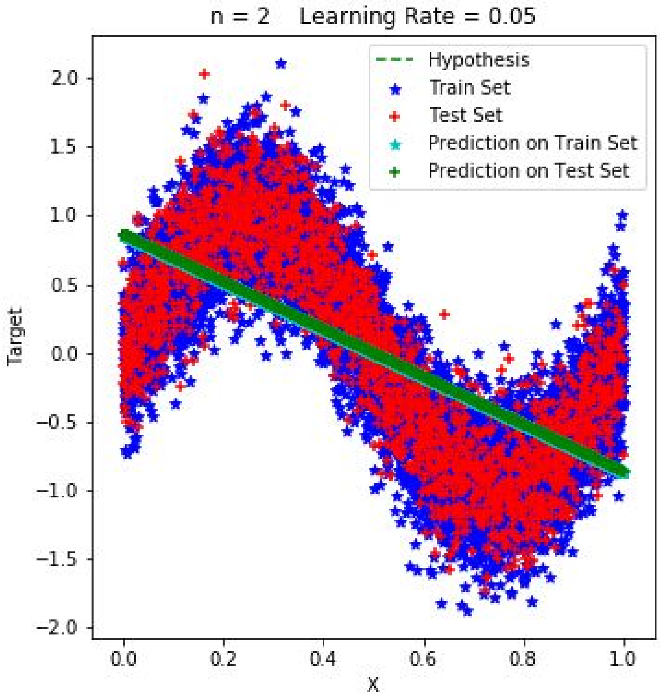
WHEN DATASIZE = 10000

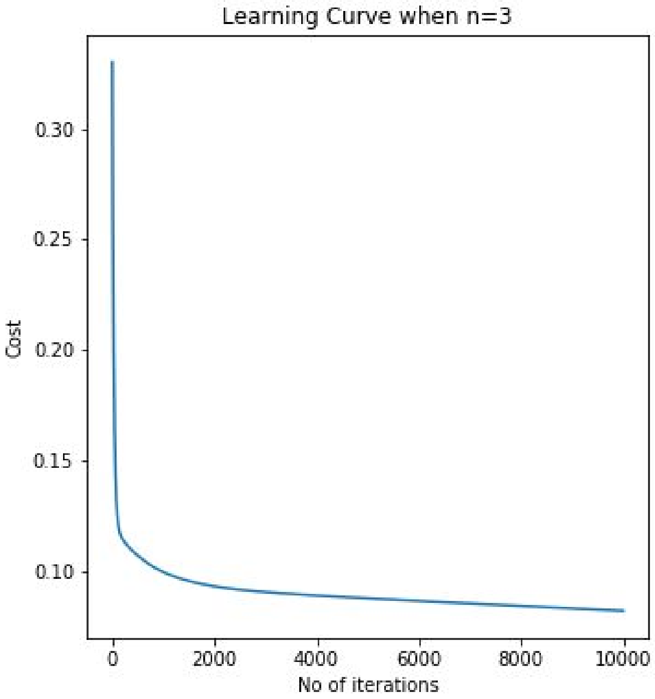


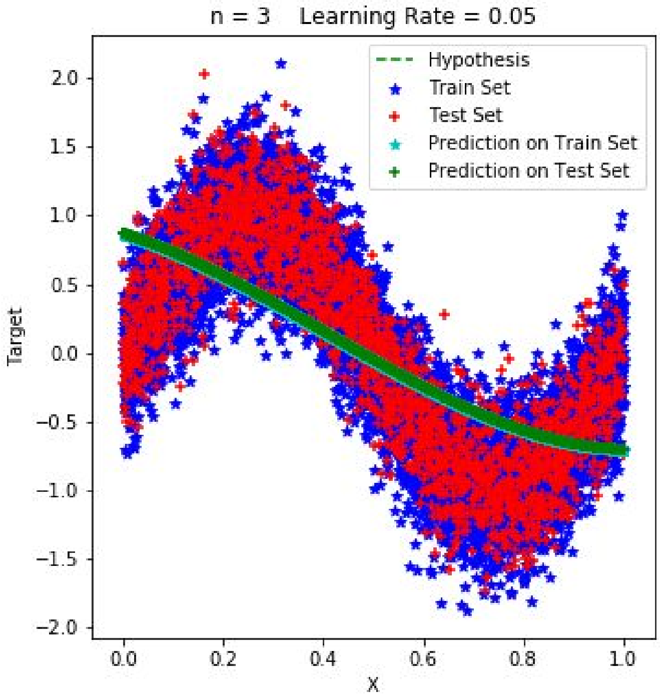

Learning Curve when n=1
Cost
0.30
0.25
0.20
0.15
0.10
0
2000
4000
6000
8000
10000
No of iterations

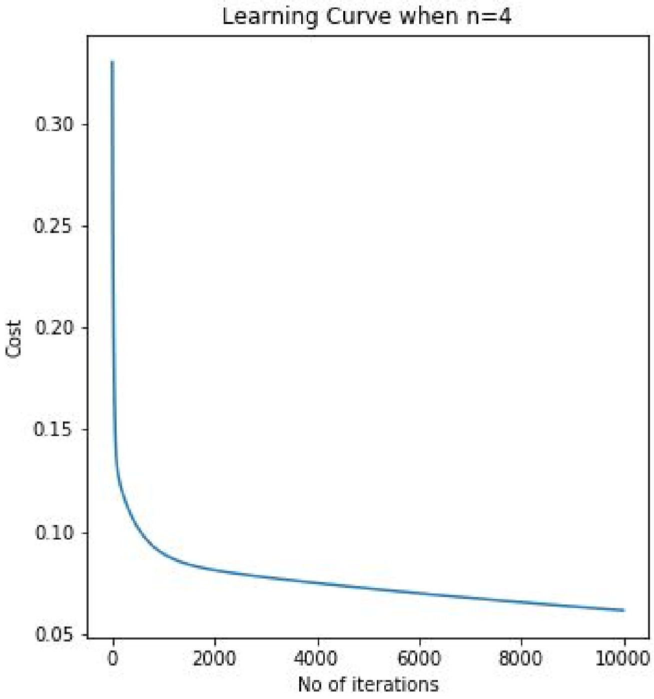

n = 1 Learning Rate = 0.05
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set
Target
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0
X

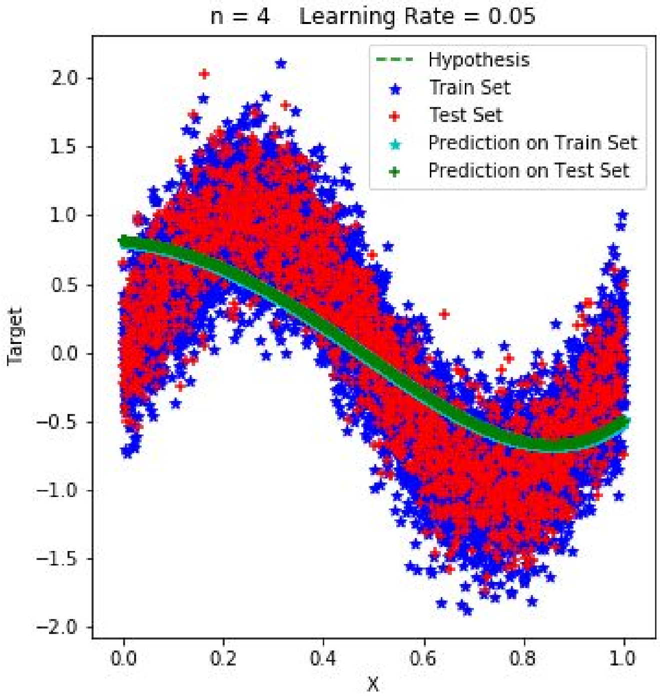

Learning Curve when n=2
Cost
0.30
0.25
0.20
0.15
0.10
0
2000
4000
6000
8000
10000
No of iterations

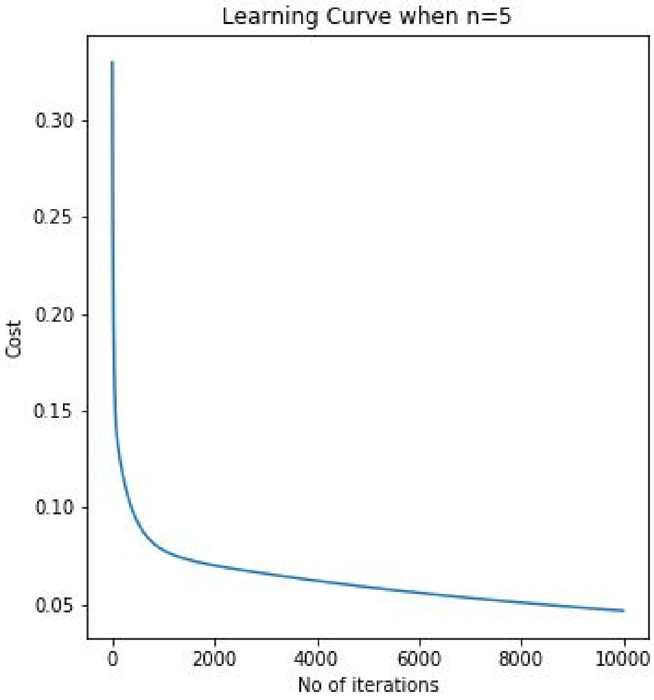

n = 2 Learning Rate = 0.05
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set
Target
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0
X

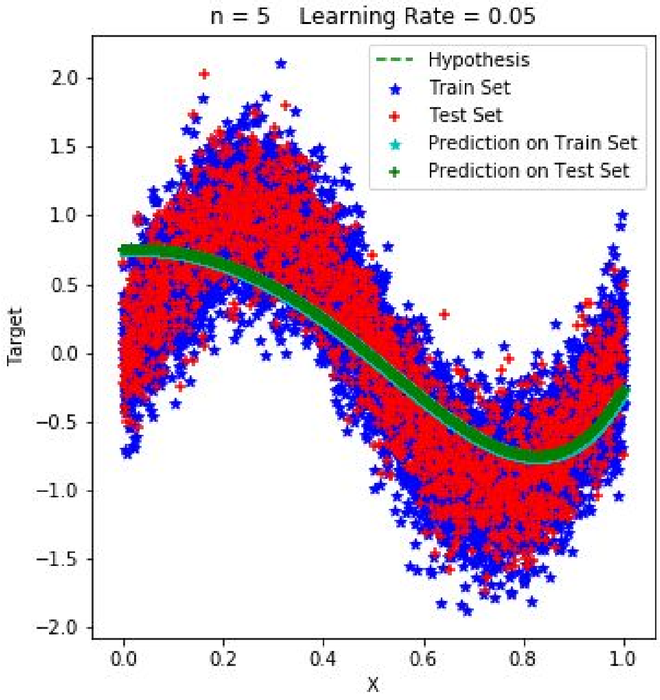

Learning Curve when n=3
Cost
0.30
0.25
0.20
0.15
0.10
0
2000
4000
6000
8000
10000
No of iterations

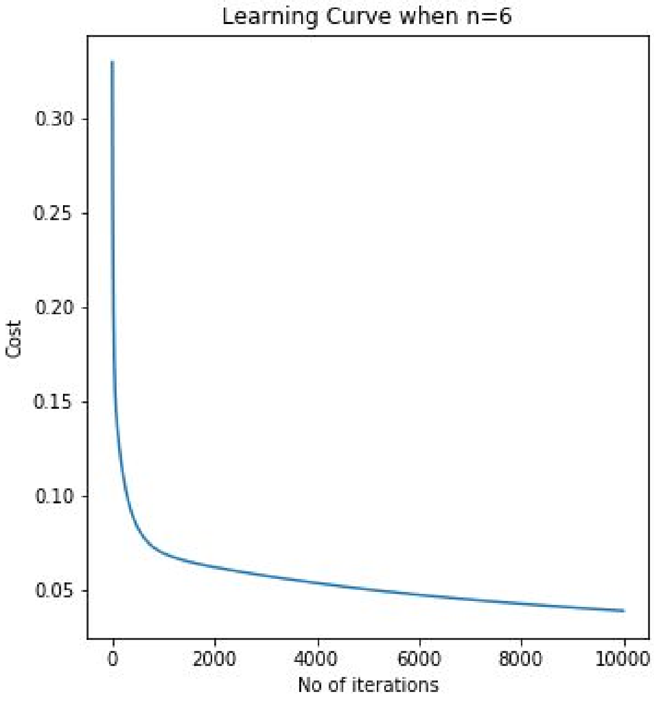

n = 3 Learning Rate = 0.05
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set
Target
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0
X

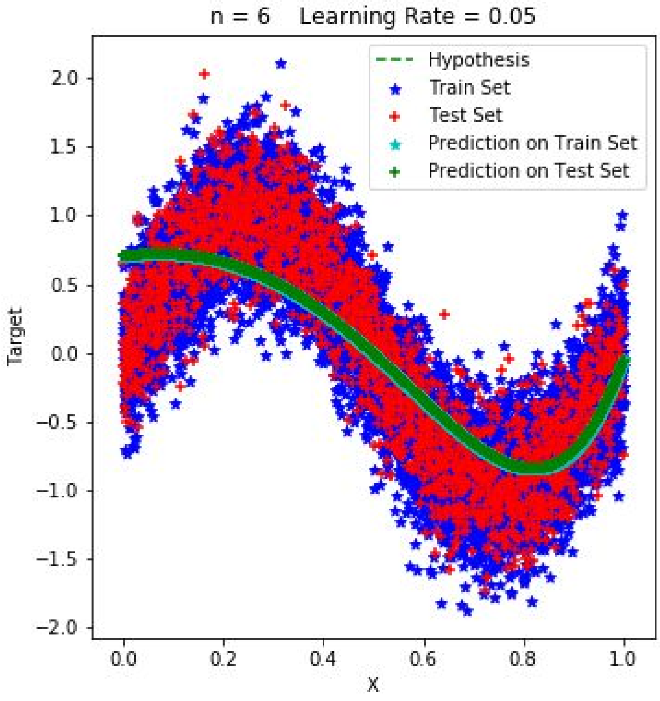

Learning Curve when n=4
Cost
0.30
0.25
0.20
0.15
0.10
0.05
0
2000
4000
6000
8000
10000
No of iterations

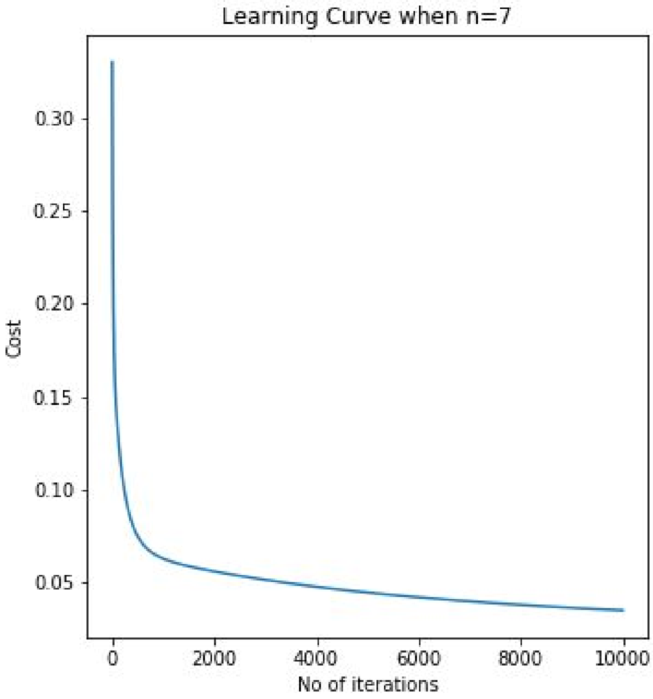

n = 4   Learning Rate = 0.05
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set
Target
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0
X

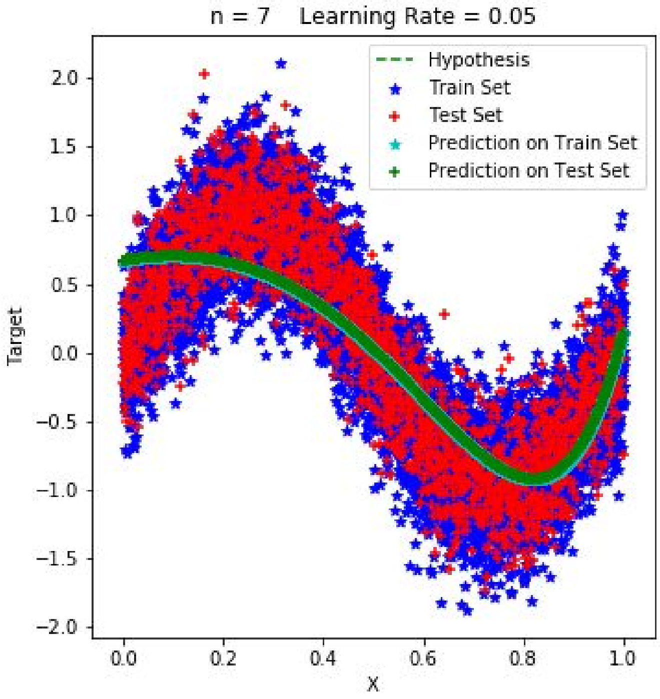

Learning Curve when n=5
Cost
0.30
0.25
0.20
0.15
0.10
0.05
0
2000
4000
6000
8000
10000
No of iterations

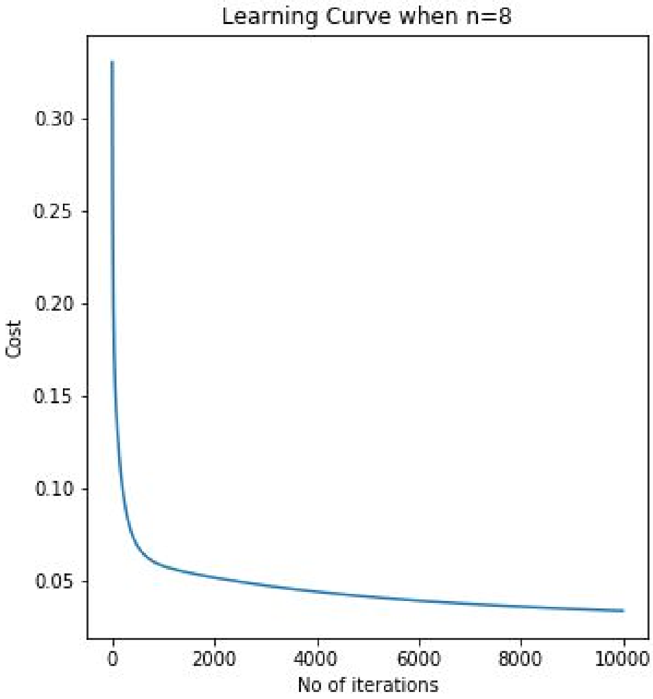

n = 5   Learning Rate = 0.05
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set
Target
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0
X

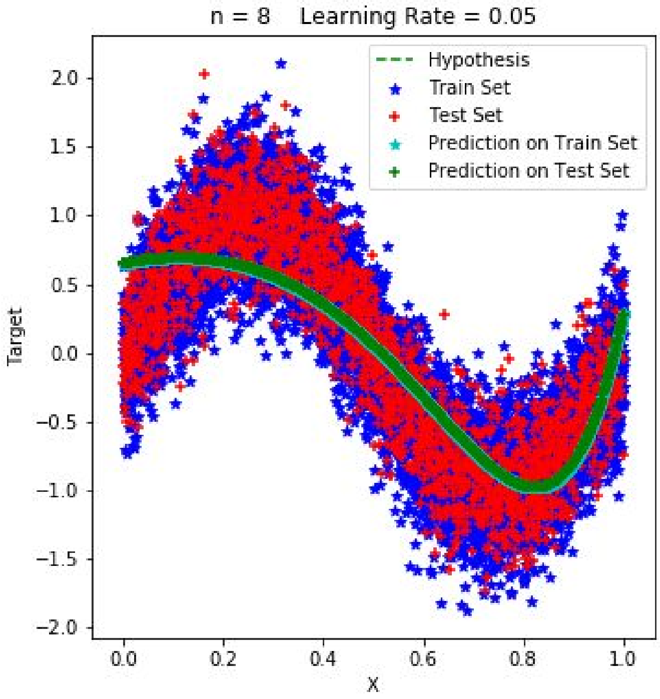

Learning Curve when n=6
Cost
0.30
0.25
0.20
0.15
0.10
0.05
0
2000
4000
6000
8000
10000
No of iterations

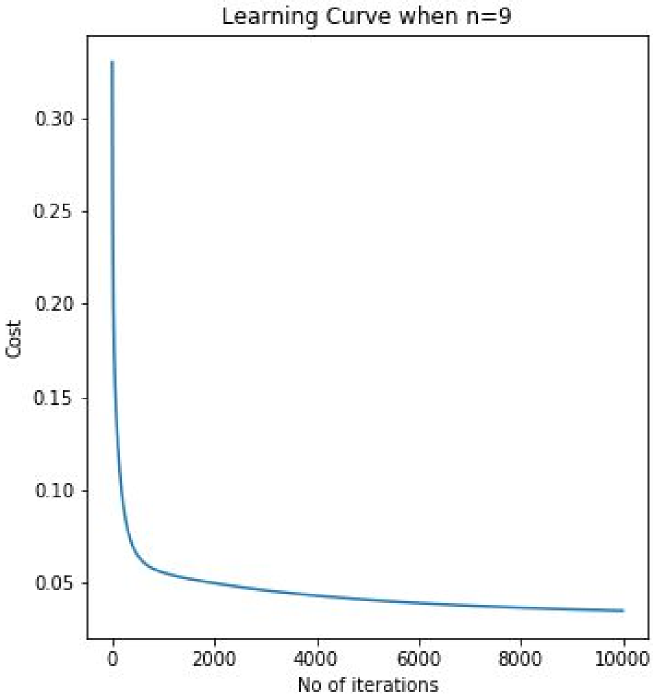

n = 6   Learning Rate = 0.05
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set
Target
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0
X

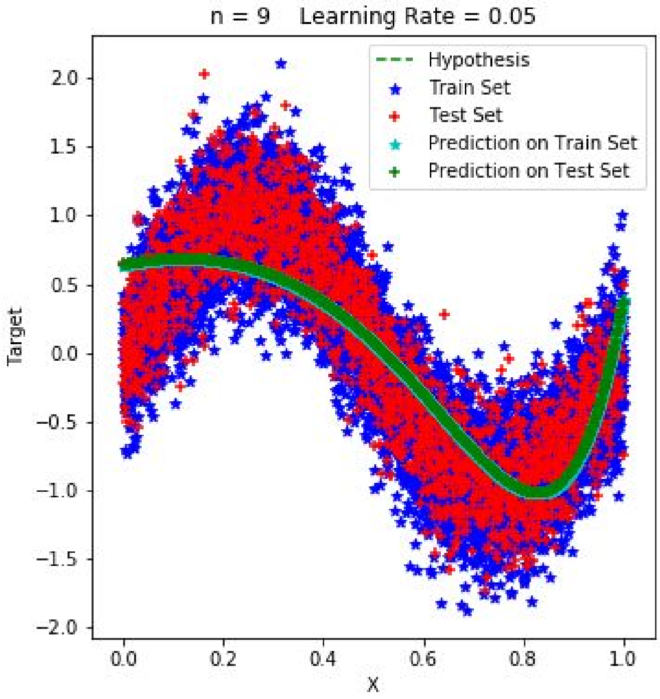

Learning Curve when n=7
Cost
0.30
0.25
0.20
0.15
0.10
0.05
0
2000
4000
6000
8000
10000
No of iterations


n = 7   Learning Rate = 0.05
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set
Target
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0
X


Learning Curve when n=8
Cost
0.30
0.25
0.20
0.15
0.10
0.05
0
2000
4000
6000
8000
10000
No of iterations

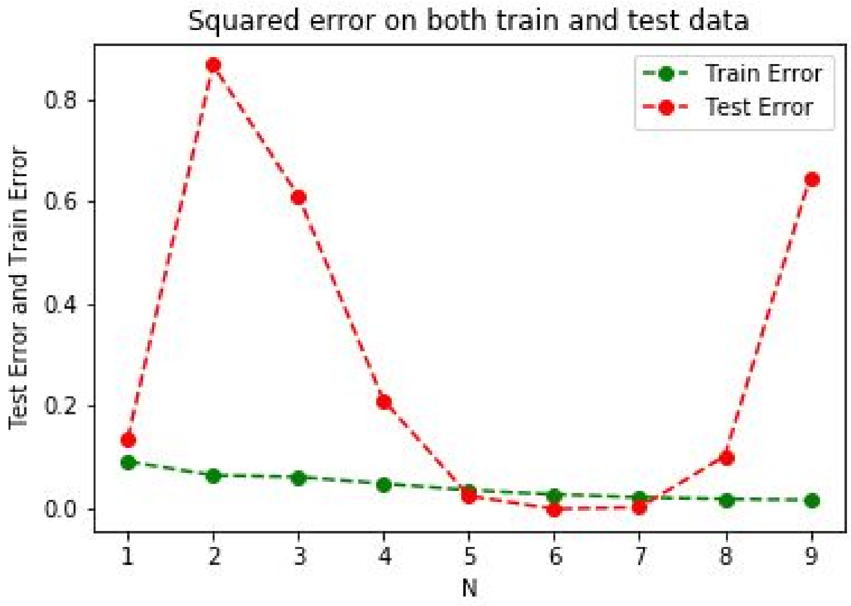

n = 8   Learning Rate = 0.05
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set
Target
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0
X

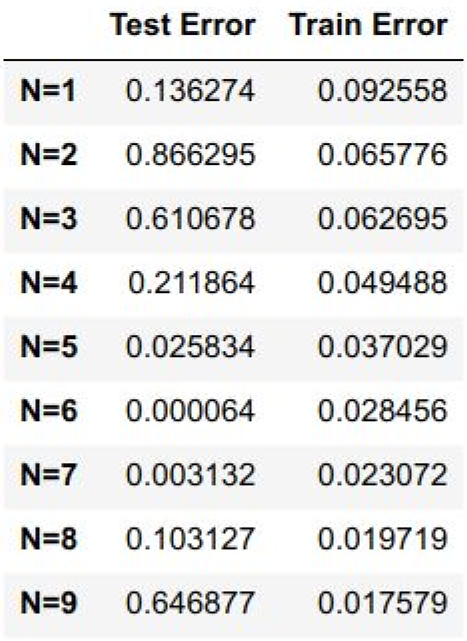

Learning Curve when n=9
Cost
0.30
0.25
0.20
0.15
0.10
0.05
0
2000
4000
6000
8000
10000
No of iterations


n = 9   Learning Rate = 0.05
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set
Target
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0
X

Squared error on both train and test data:

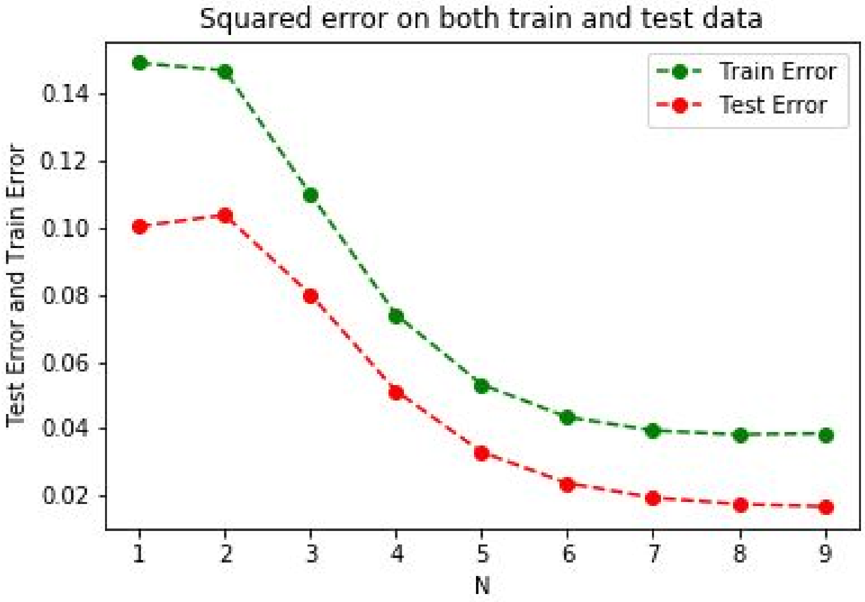
Data size = 10

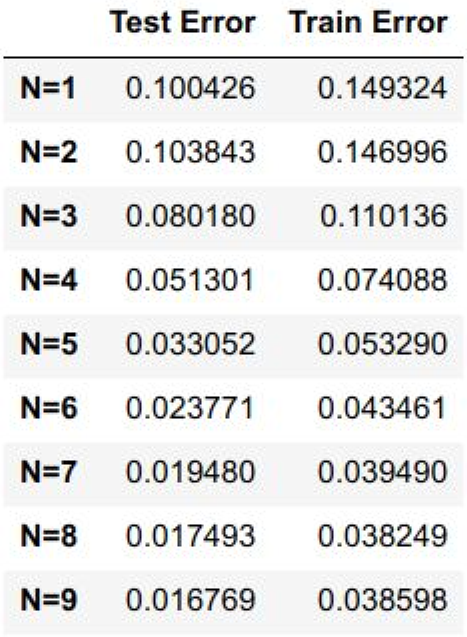




	Test Error	Train Error
N=1	0.136274	0.092558
N=2	0.866295	0.065776
N=3	0.610678	0.062695
N=4	0.211864	0.049488
N=5	0.025834	0.037029
N=6	0.000064	0.028456
N=7	0.003132	0.023072
N=8	0.103127	0.019719
N=9	0.646877	0.017579

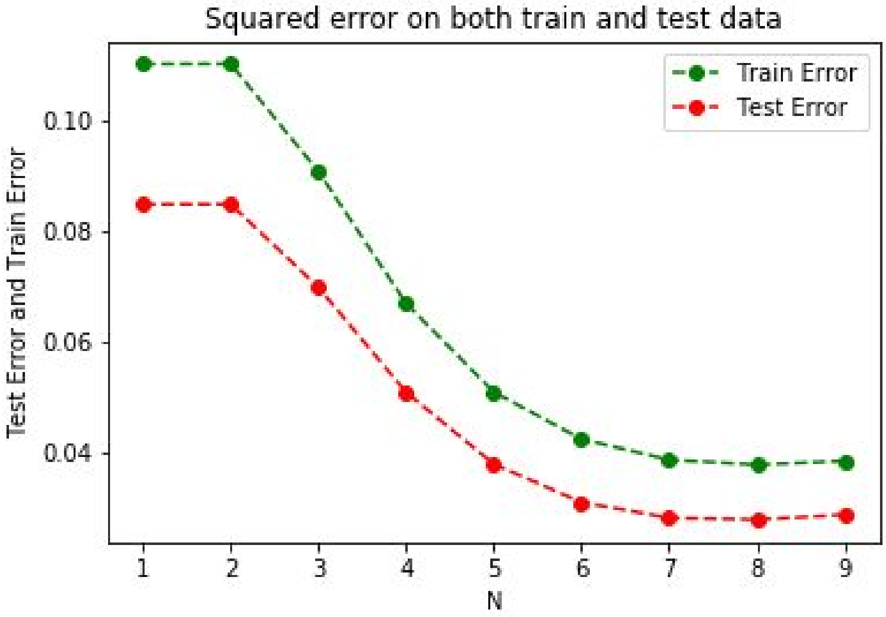
Data size = 100

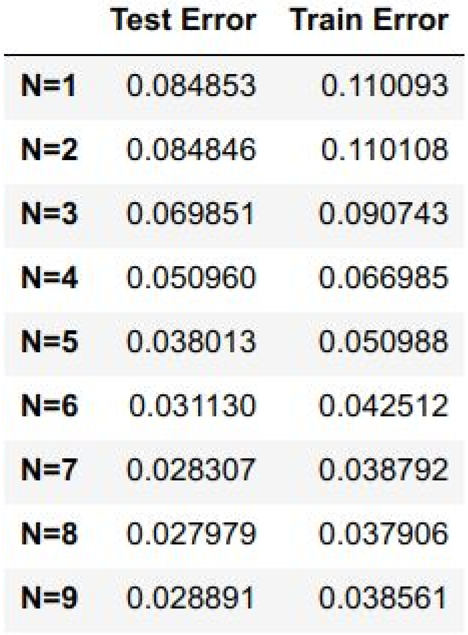




	Test Error	Train Error
N=1	0.100426	0.149324
N=2	0.103843	0.146996
N=3	0.080180	0.110136
N=4	0.051301	0.074088
N=5	0.033052	0.053290
N=6	0.023771	0.043461
N=7	0.019480	0.039490
N=8	0.017493	0.038249
N=9	0.016769	0.038598

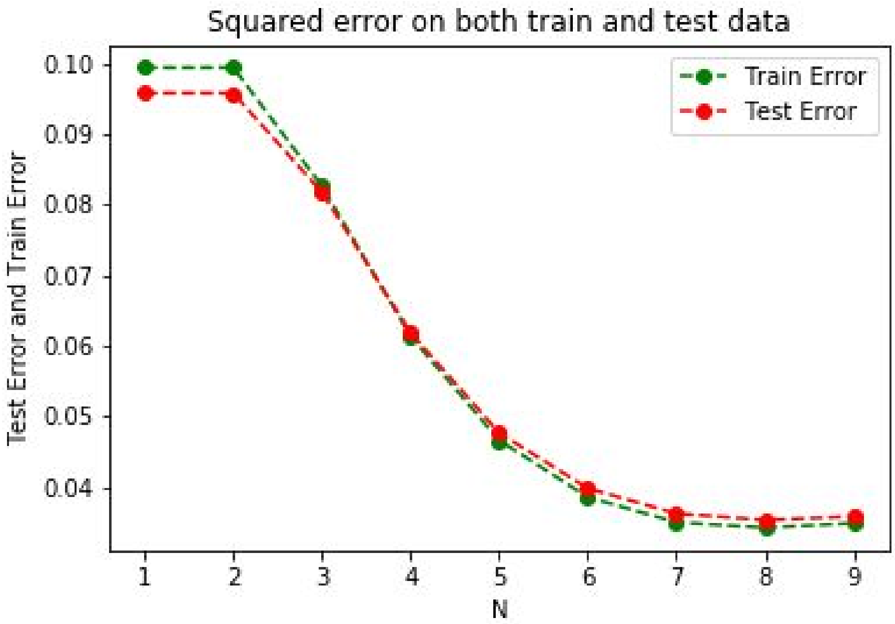
Data size = 1000

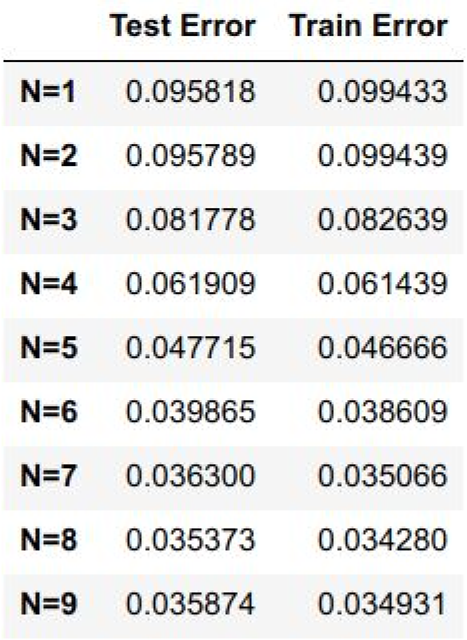




	Test Error	Train Error
N=1	0.084853	0.110093
N=2	0.084846	0.110108
N=3	0.069851	0.090743
N=4	0.050960	0.066985
N=5	0.038013	0.050988
N=6	0.031130	0.042512
N=7	0.028307	0.038792
N=8	0.027979	0.037906
N=9	0.028891	0.038561

Data size = 10000

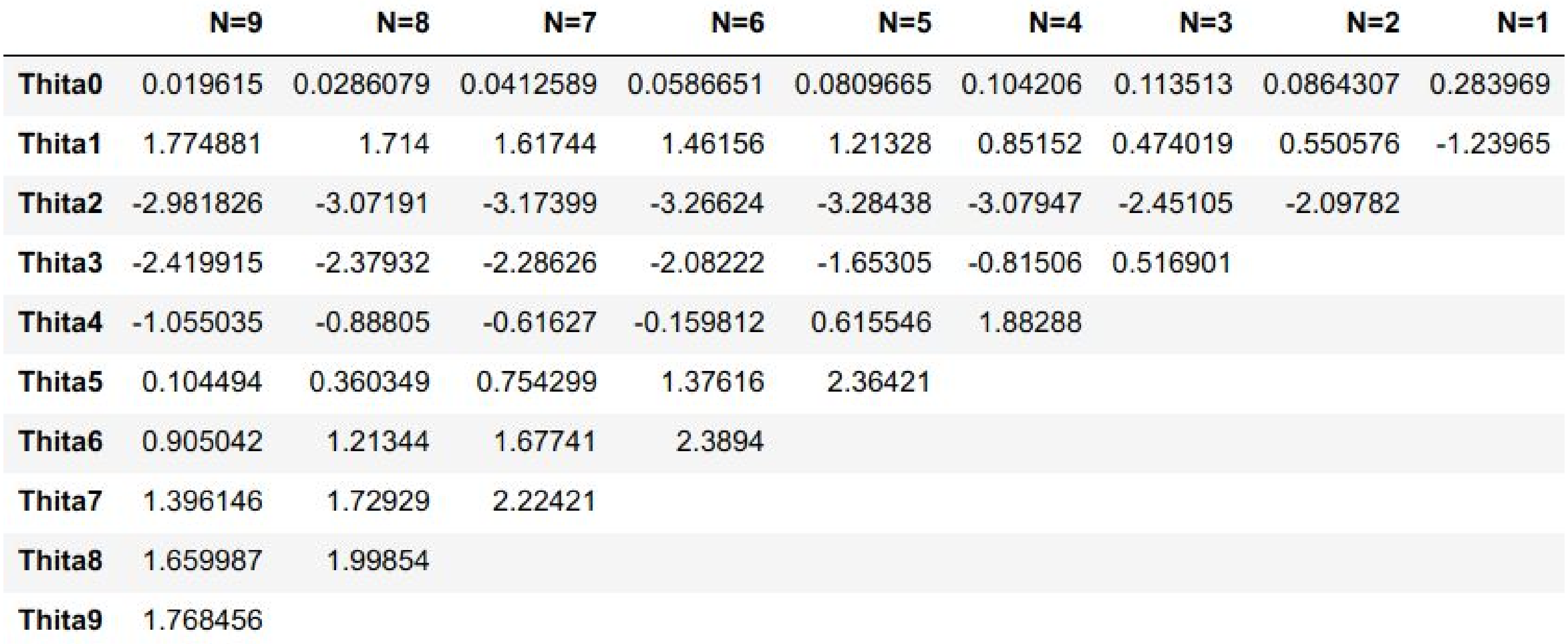




	Test Error	Train Error
N=1	0.095818	0.099433
N=2	0.095789	0.099439
N=3	0.081778	0.082639
N=4	0.061909	0.061439
N=5	0.047715	0.046666
N=6	0.039865	0.038609
N=7	0.036300	0.035066
N=8	0.035373	0.034280
N=9	0.035874	0.034931

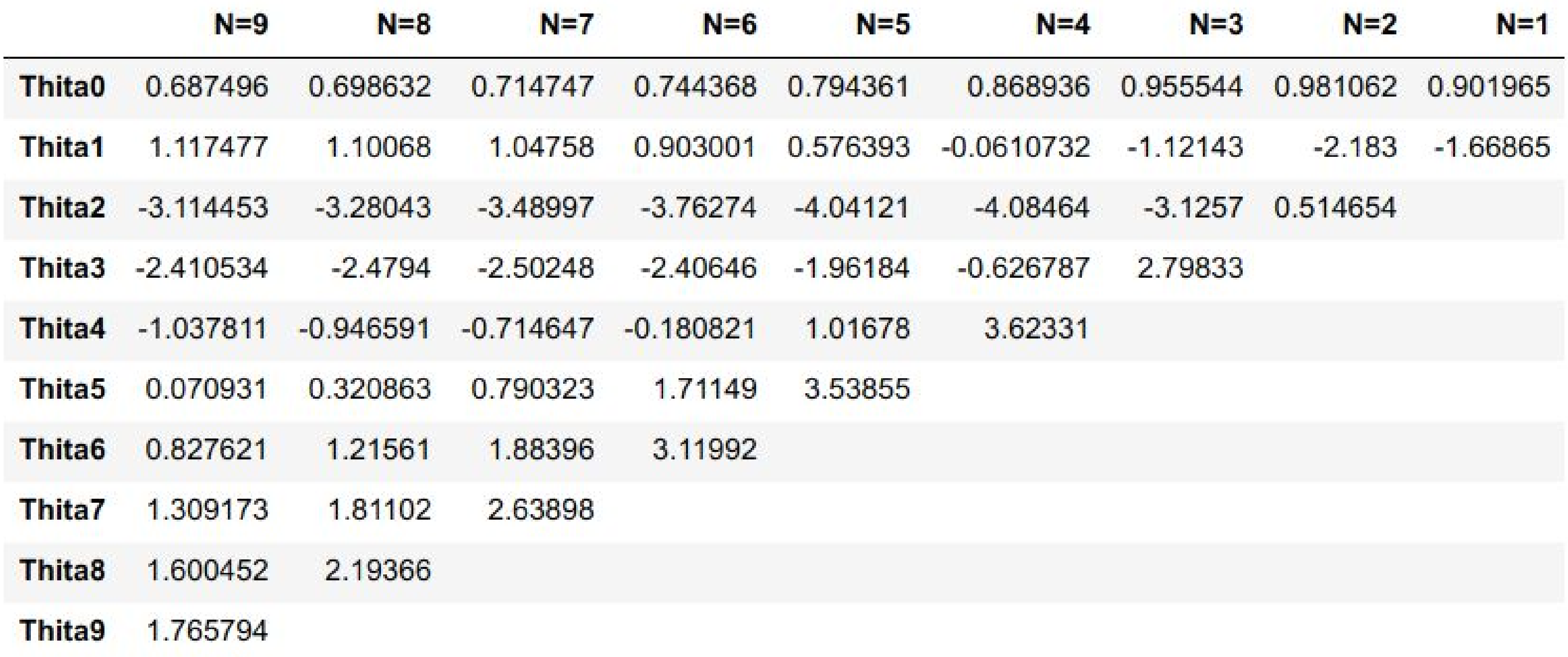
Parameters values for different data size :

Data size = 10



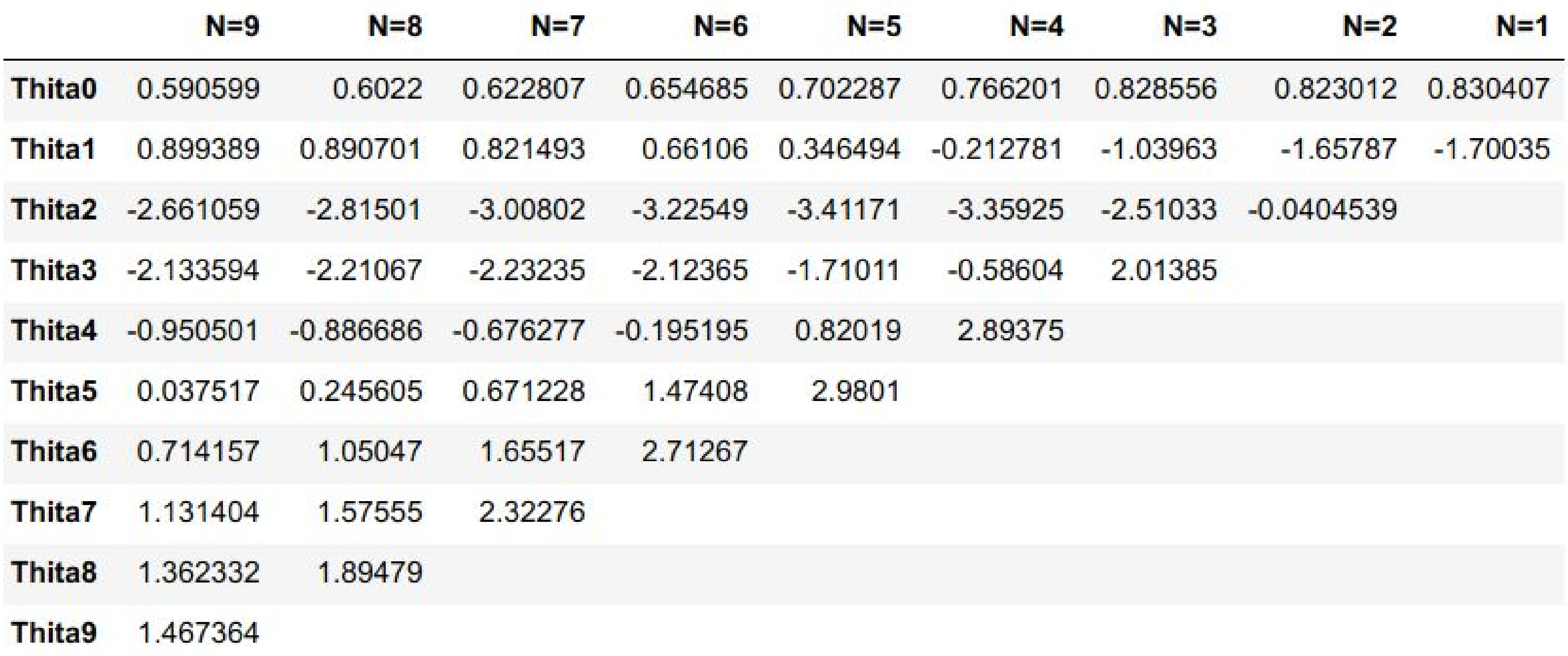
	N=9	N=8	N=7	N=6	N=5	N=4	N=3	N=2	N=1
Thita0	0.019615	0.0286079	0.0412589	0.0586651	0.0809665	0.104206	0.113513	0.0864307	0.283969
Thita1	1.774881	1.714	1.61744	1.46156	1.21328	0.85152	0.474019	0.550576	-1.23965
Thita2	-2.981826	-3.07191	-3.17399	-3.26624	-3.28438	-3.07947	-2.45105	-2.09782	
Thita3	-2.419915	-2.37932	-2.28626	-2.08222	-1.65305	-0.81506	0.516901		
Thita4	-1.055035	-0.88805	-0.61627	-0.159812	0.615546	1.88288			
Thita5	0.104494	0.360349	0.754299	1.37616	2.36421				
Thita6	0.905042	1.21344	1.67741	2.3894					
Thita7	1.396146	1.72929	2.22421						
Thita8	1.659987	1.99854							
Thita9	1.768456								

Data size = 100



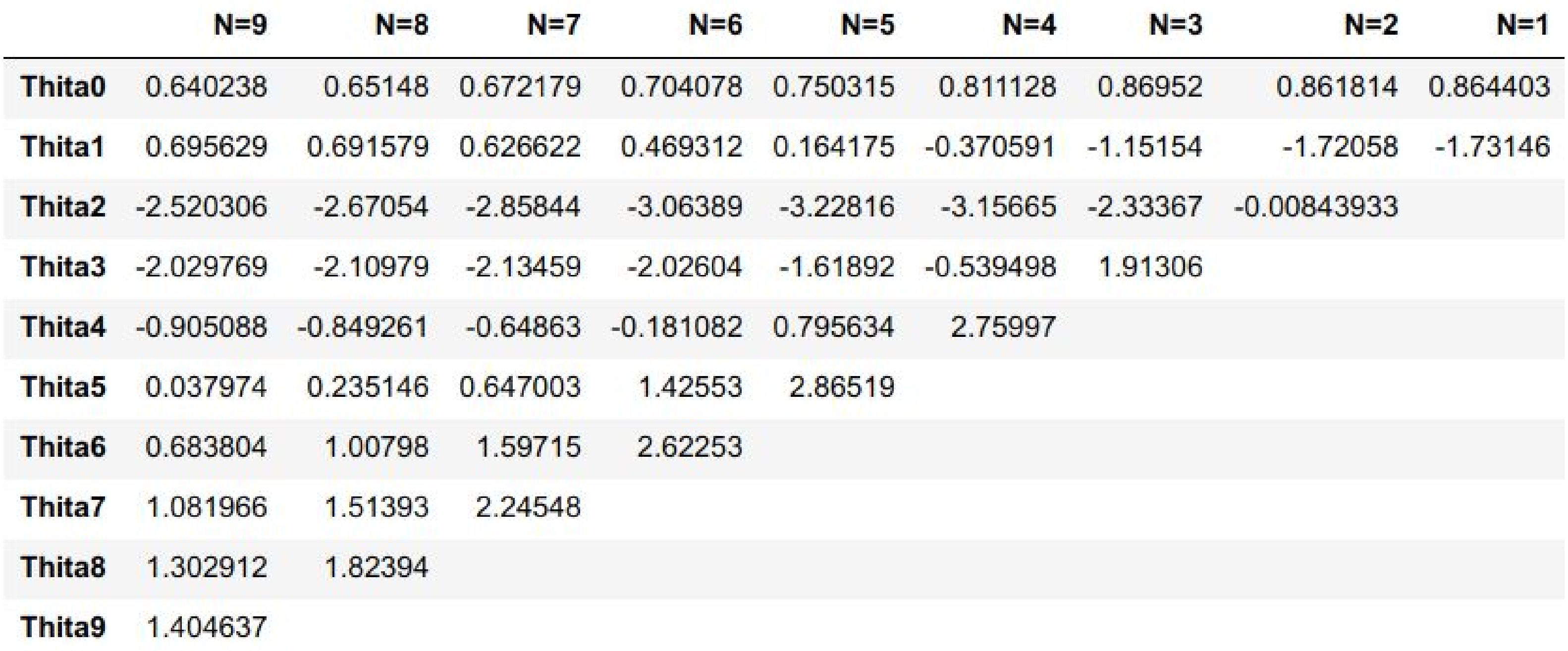
	N=9	N=8	N=7	N=6	N=5	N=4	N=3	N=2	N=1
Thita0	0.687496	0.698632	0.714747	0.744368	0.794361	0.868936	0.955544	0.981062	0.901965
Thita1	1.117477	1.10068	1.04758	0.903001	0.576393	-0.0610732	-1.12143	-2.183	-1.66865
Thita2	-3.114453	-3.28043	-3.48997	-3.76274	-4.04121	-4.08464	-3.1257	0.514654	
Thita3	-2.410534	-2.4794	-2.50248	-2.40646	-1.96184	-0.626787	2.79833		
Thita4	-1.037811	-0.946591	-0.714647	-0.180821	1.01678	3.62331			
Thita5	0.070931	0.320863	0.790323	1.71149	3.53855				
Thita6	0.827621	1.21561	1.88396	3.11992					
Thita7	1.309173	1.81102	2.63898						
Thita8	1.600452	2.19366							
Thita9	1.765794								

Data size = 1000



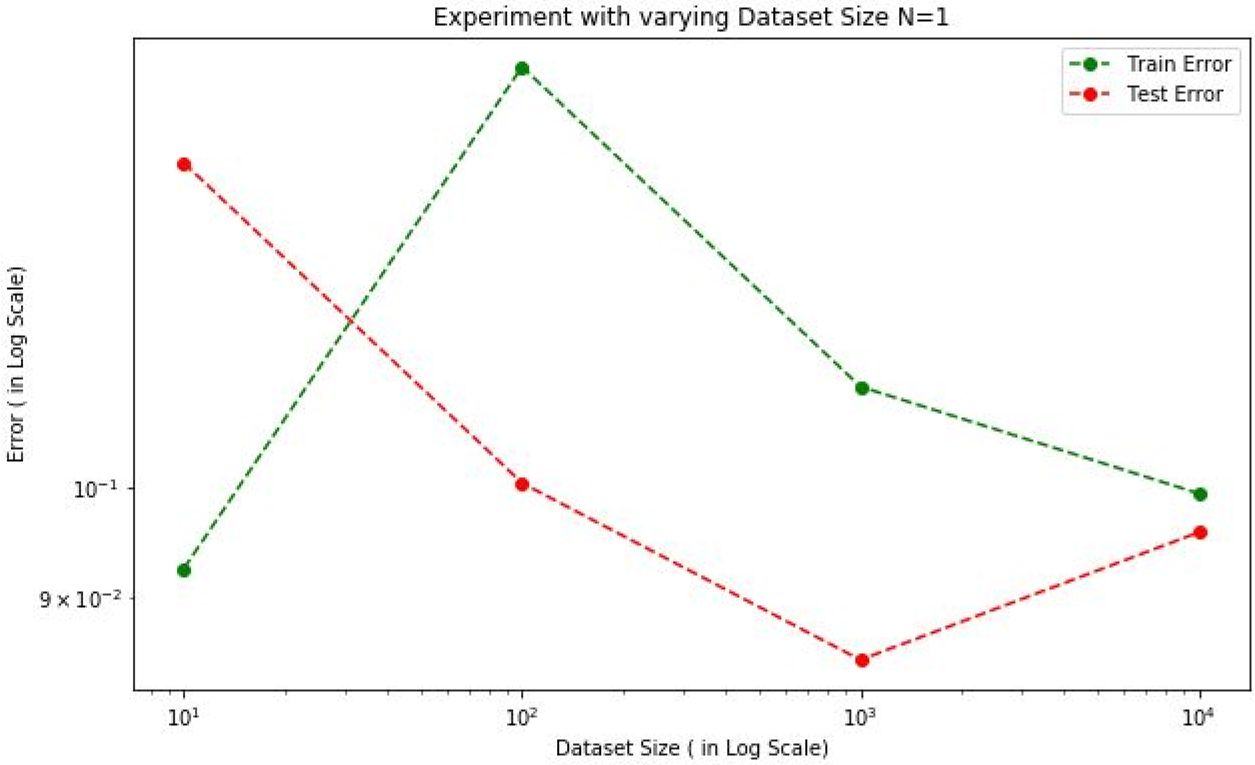
	N=9	N=8	N=7	N=6	N=5	N=4	N=3	N=2	N=1
Thita0	0.590599	0.6022	0.622807	0.654685	0.702287	0.766201	0.828556	0.823012	0.830407
Thita1	0.899389	0.890701	0.821493	0.66106	0.346494	-0.212781	-1.03963	-1.65787	-1.70035
Thita2	-2.661059	-2.81501	-3.00802	-3.22549	-3.41171	-3.35925	-2.51033	-0.0404539	
Thita3	-2.133594	-2.21067	-2.23235	-2.12365	-1.71011	-0.58604	2.01385		
Thita4	-0.950501	-0.886686	-0.676277	-0.195195	0.82019	2.89375			
Thita5	0.037517	0.245605	0.671228	1.47408	2.9801				
Thita6	0.714157	1.05047	1.65517	2.71267					
Thita7	1.131404	1.57555	2.32276						
Thita8	1.362332	1.89479							
Thita9	1.467364								

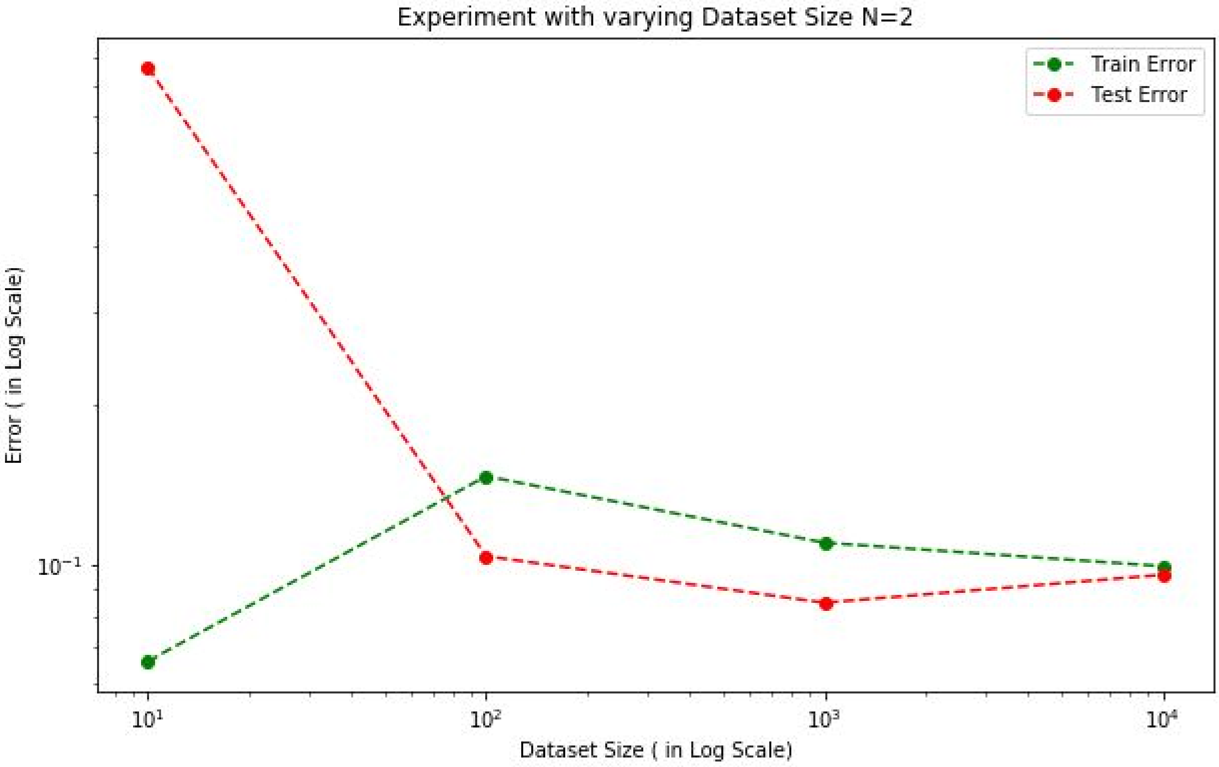
Data size = 10000

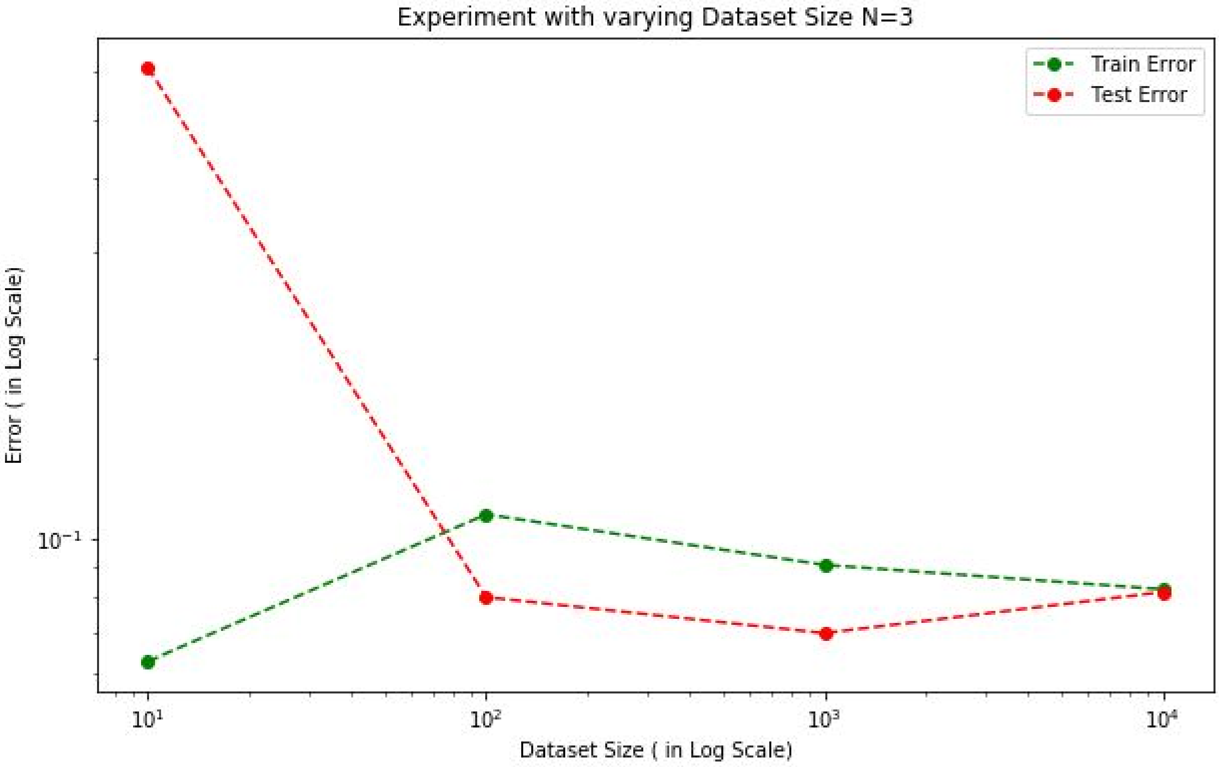


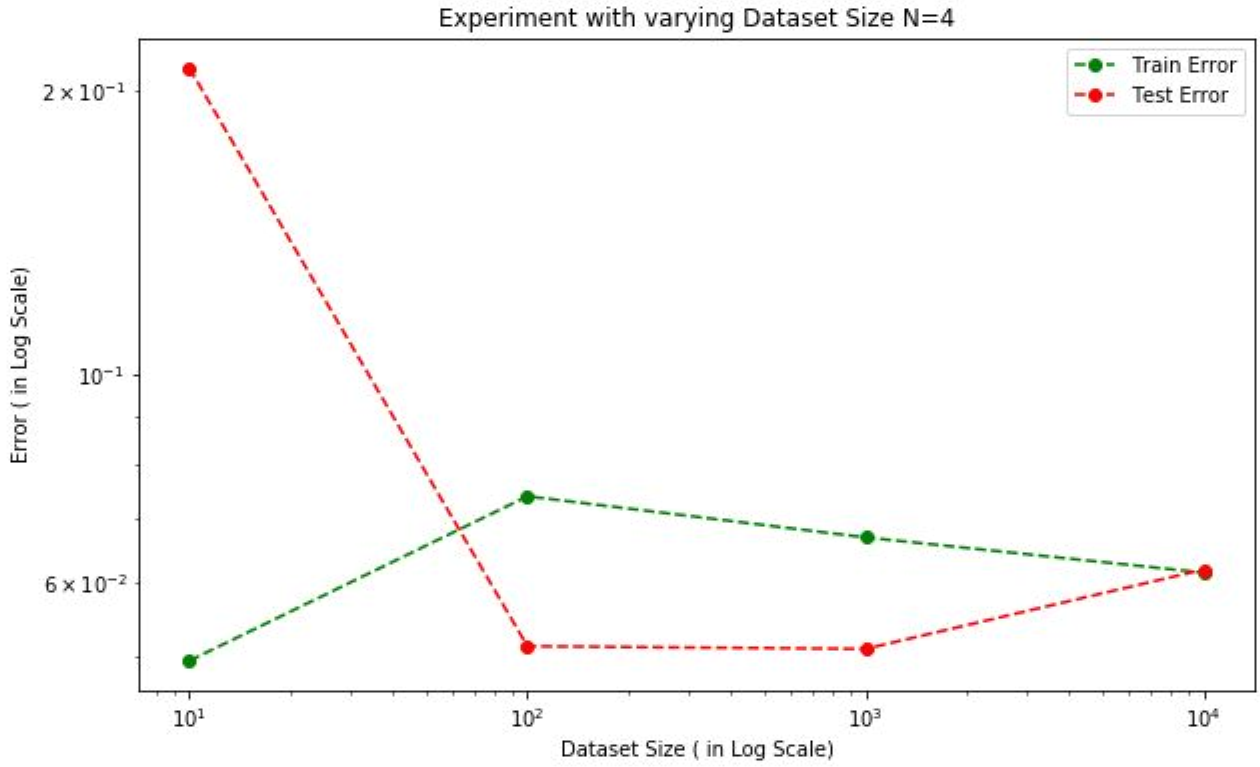
	N=9	N=8	N=7	N=6	N=5	N=4	N=3	N=2	N=1
Thita0	0.640238	0.65148	0.672179	0.704078	0.750315	0.811128	0.86952	0.861814	0.864403
Thita1	0.695629	0.691579	0.626622	0.469312	0.164175	-0.370591	-1.15154	-1.72058	-1.73146
Thita2	-2.520306	-2.67054	-2.85844	-3.06389	-3.22816	-3.15665	-2.33367	-0.00843933	
Thita3	-2.029769	-2.10979	-2.13459	-2.02604	-1.61892	-0.539498	1.91306		
Thita4	-0.905088	-0.849261	-0.64863	-0.181082	0.795634	2.75997			
Thita5	0.037974	0.235146	0.647003	1.42553	2.86519				
Thita6	0.683804	1.00798	1.59715	2.62253					
Thita7	1.081966	1.51393	2.24548						
Thita8	1.302912	1.82394							
Thita9	1.404637								

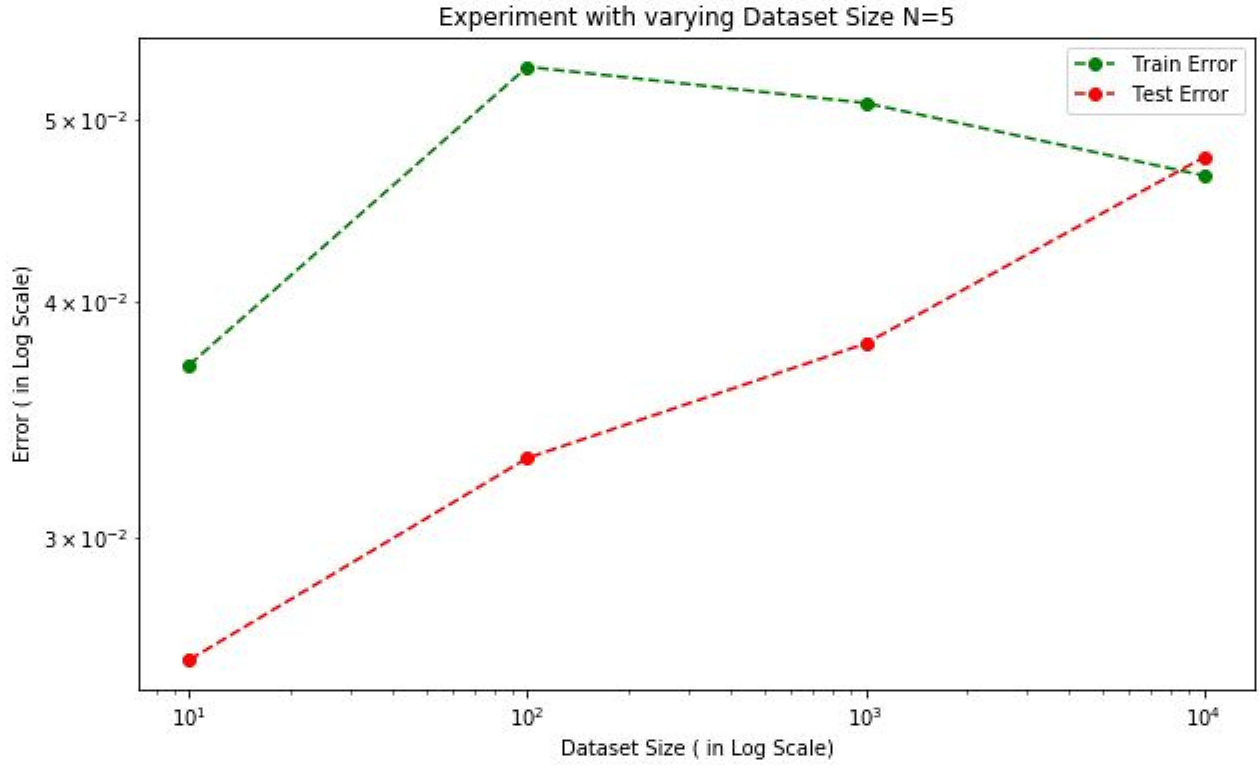
Learning Curve of How train and test error varies with an increase in the size of datasets:

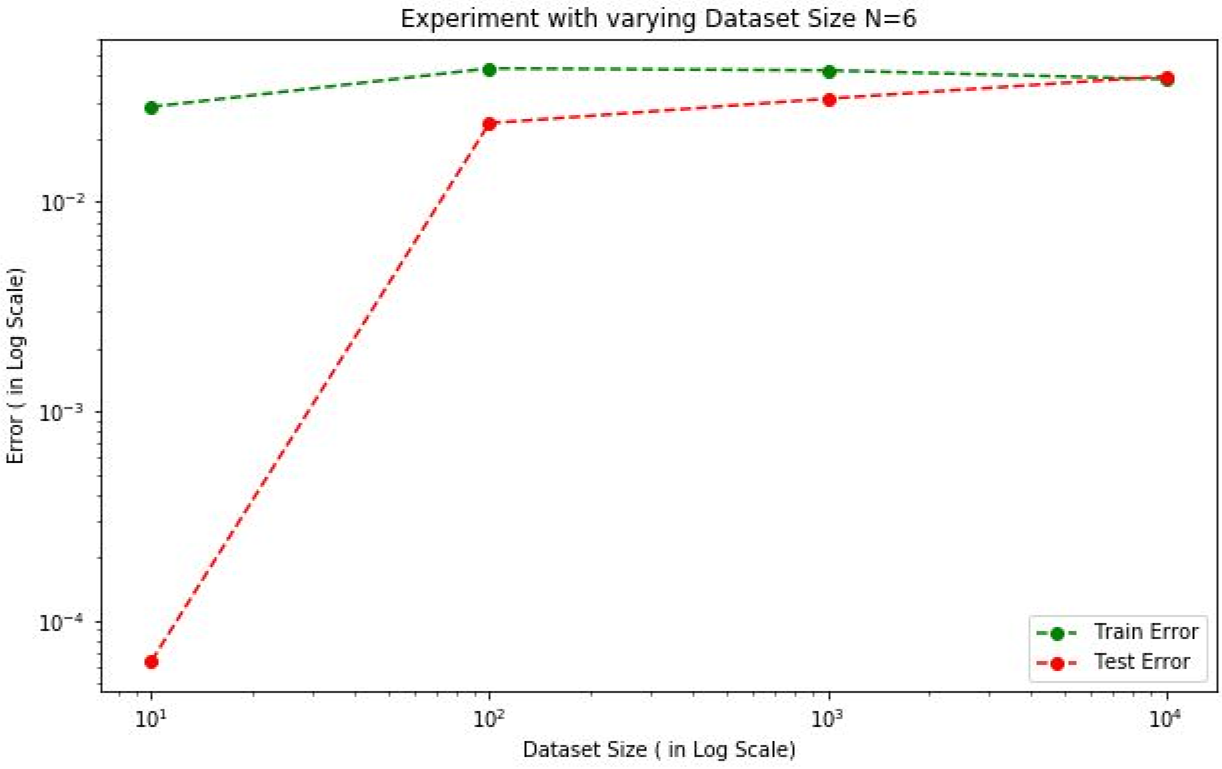


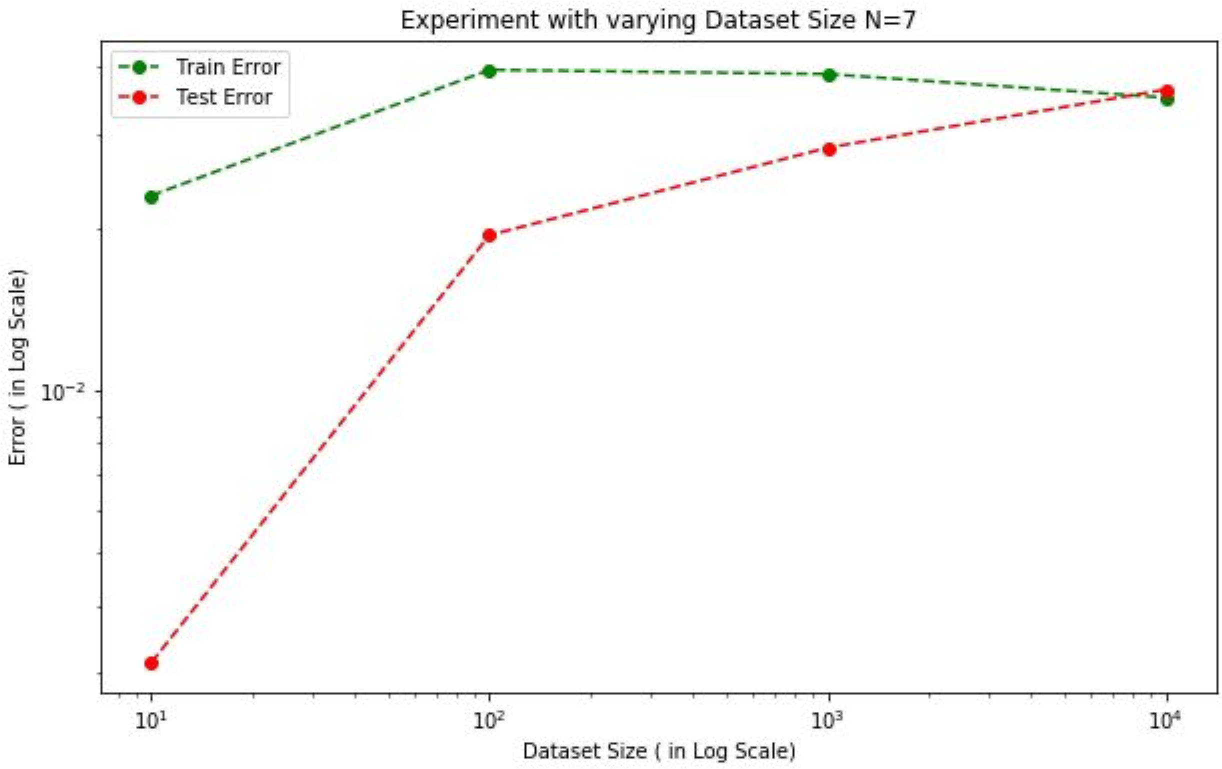


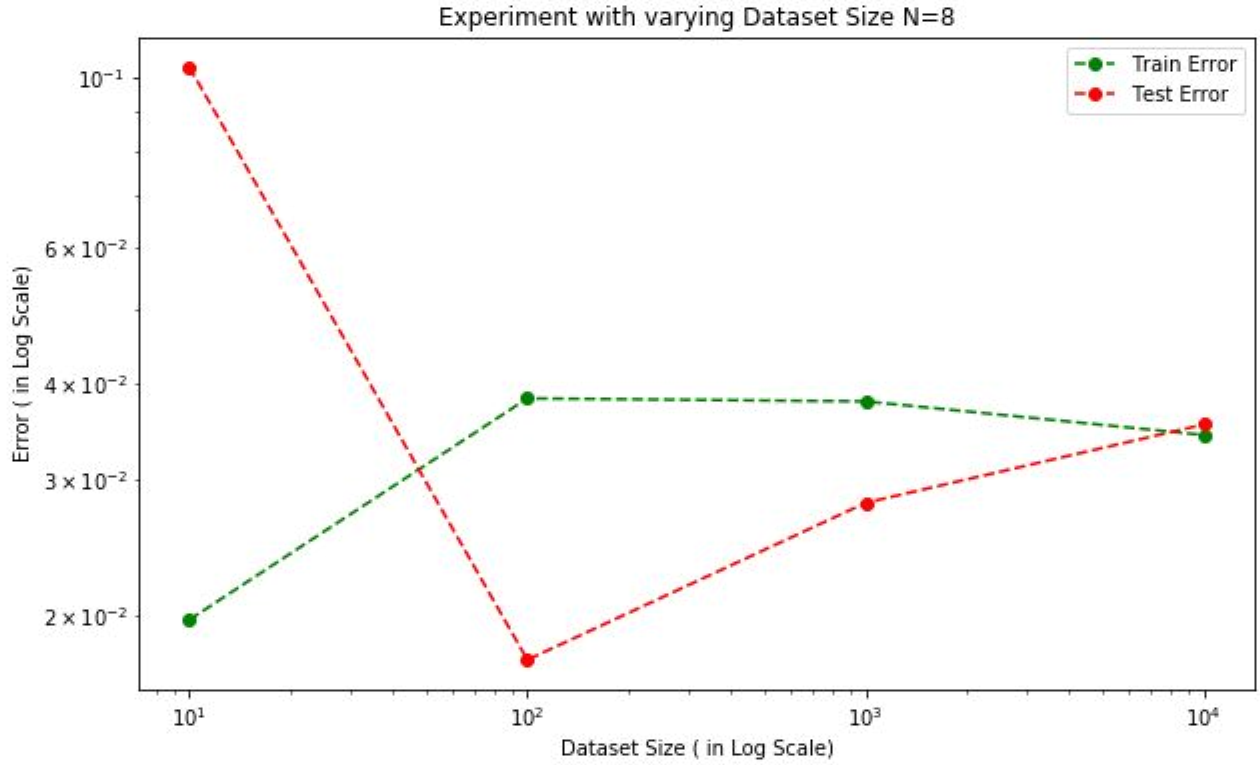


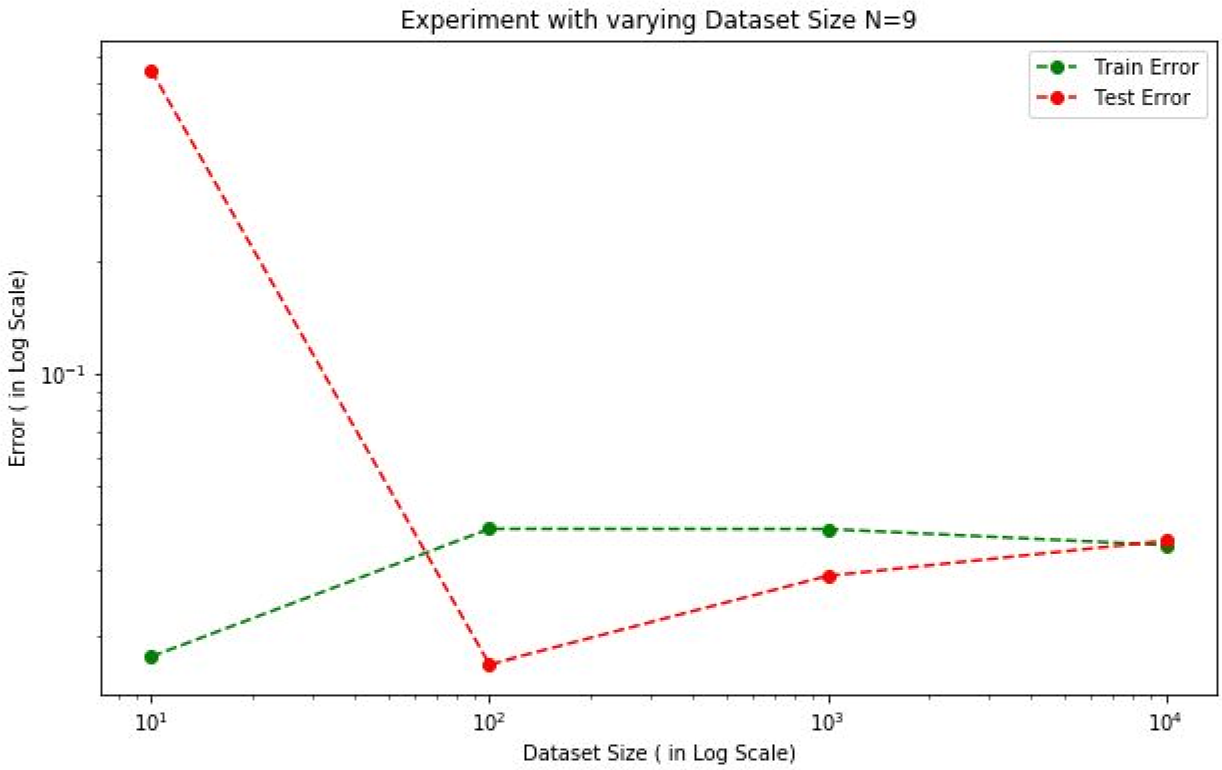

Experiment with varying Dataset Size N=4
Train Error
Test Error
2×10^{-1}
10^{-1}
6×10^{-2}
Error (in Log Scale)
10^1
10^2
10^3
10^4
Dataset Size (in Log Scale)


Experiment with varying Dataset Size N=5
Train Error
Test Error
5×10^{-2}
4×10^{-2}
3×10^{-2}
Error (in Log Scale)
10^1
10^2
10^3
10^4
Dataset Size (in Log Scale)


Experiment with varying Dataset Size N=6
10^{-2}
10^{-3}
10^{-4}
Error (in Log Scale)
Train Error
Test Error
10^1
10^2
10^3
10^4
Dataset Size (in Log Scale)


Experiment with varying Dataset Size N=7
Train Error
Test Error
Error (in Log Scale)
10^{-2}
10^1
10^2
10^3
10^4
Dataset Size (in Log Scale)


Experiment with varying Dataset Size N=8
Train Error
Test Error
Error (in Log Scale)
10^{-1}
6×10^{-2}
4×10^{-2}
3×10^{-2}
2×10^{-2}
10^1
10^2
10^3
10^4
Dataset Size (in Log Scale)


Experiment with varying Dataset Size N=9
Train Error
Test Error
Error (in Log Scale)
10^{-1}
10^1
10^2
10^3
10^4
Dataset Size (in Log Scale)

==

Part 4: Experimenting with cost functions

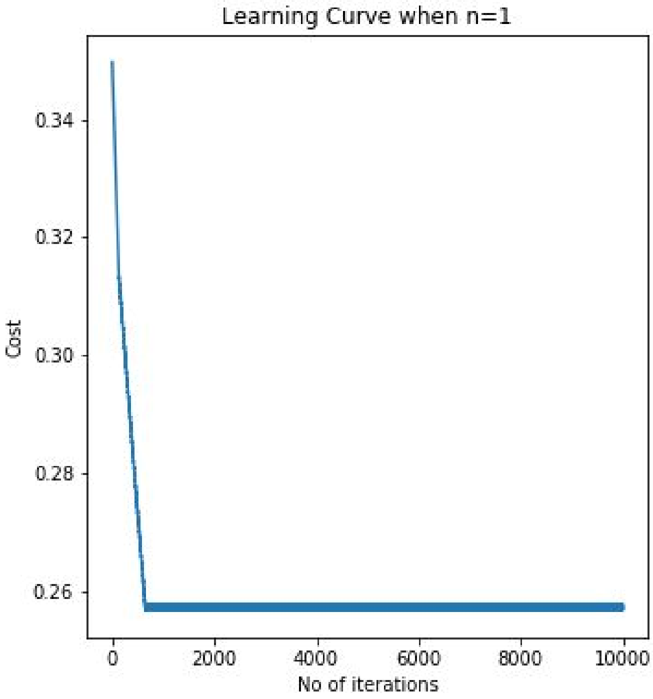
a) Mean absolute error

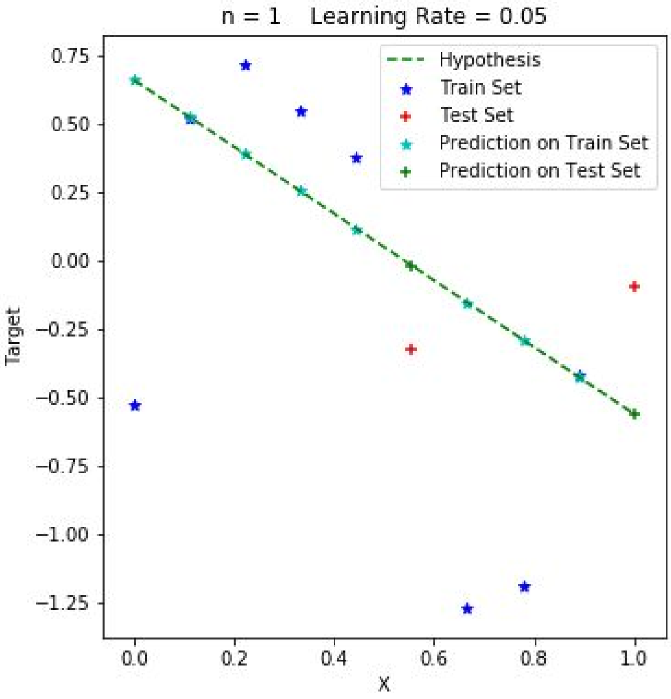
$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left|W^T\Phi_n(x)-y\right|$$

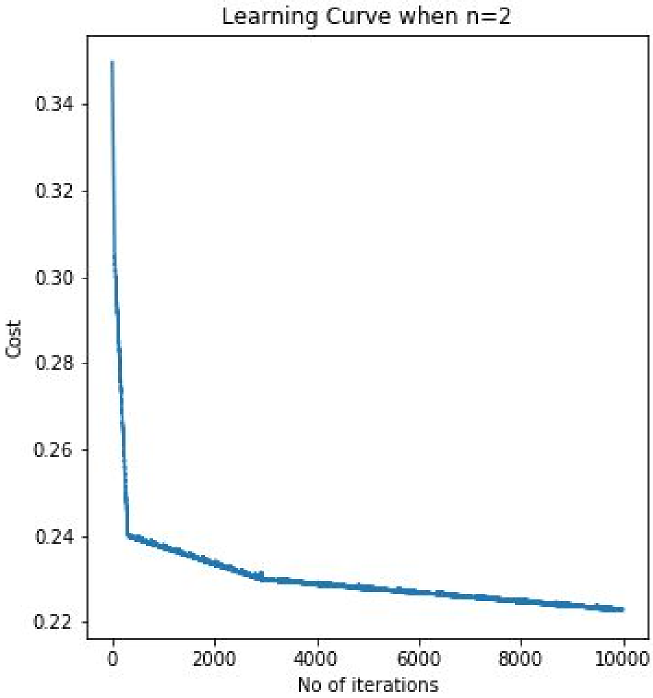
As this Cost function is not differentiable, So we can't find the partial derivative of this J(Θ).

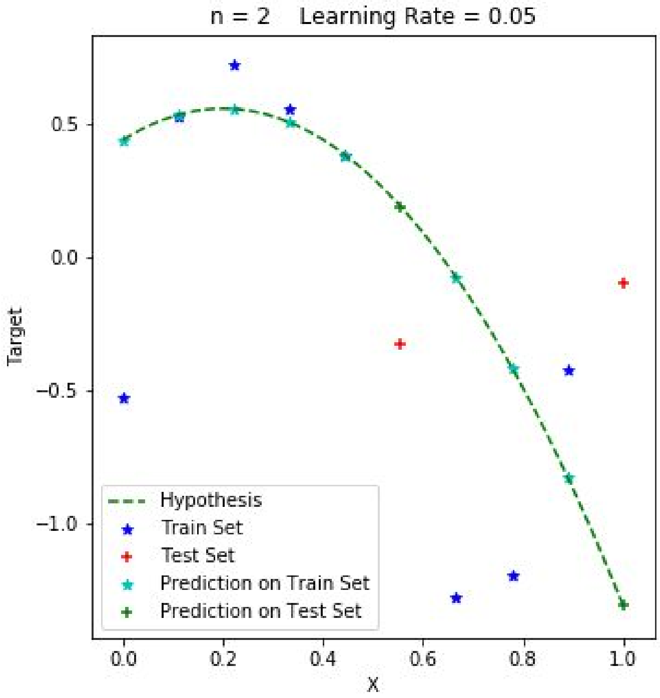
Ulimatelty we can’t find the equations for gradient descent since it involves finding of the partial derivative of the Cost function which in this case is not differentiable.

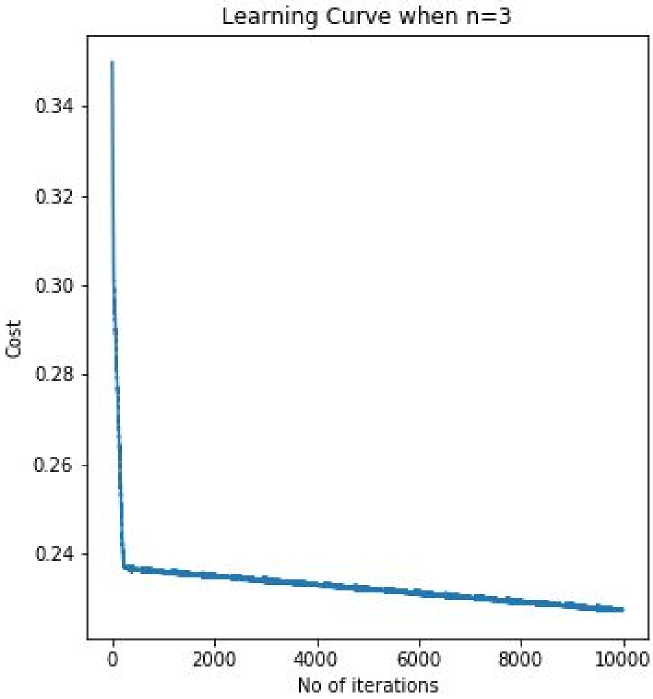
But if we create 2 cases first when inner part is positive and other when it is negative, then we can find the derivative:

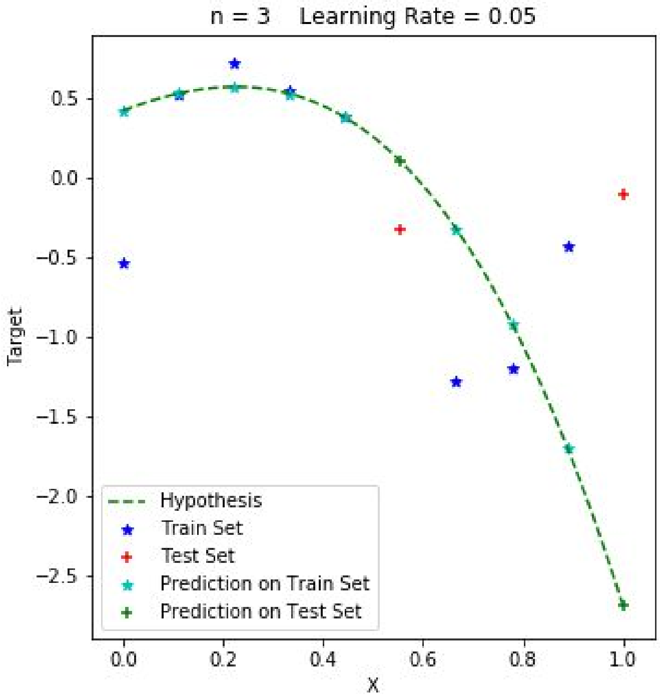


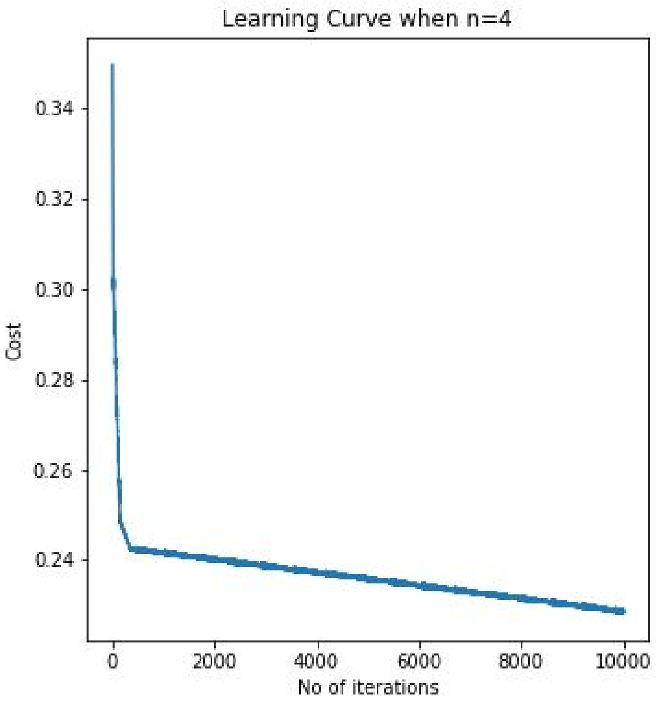


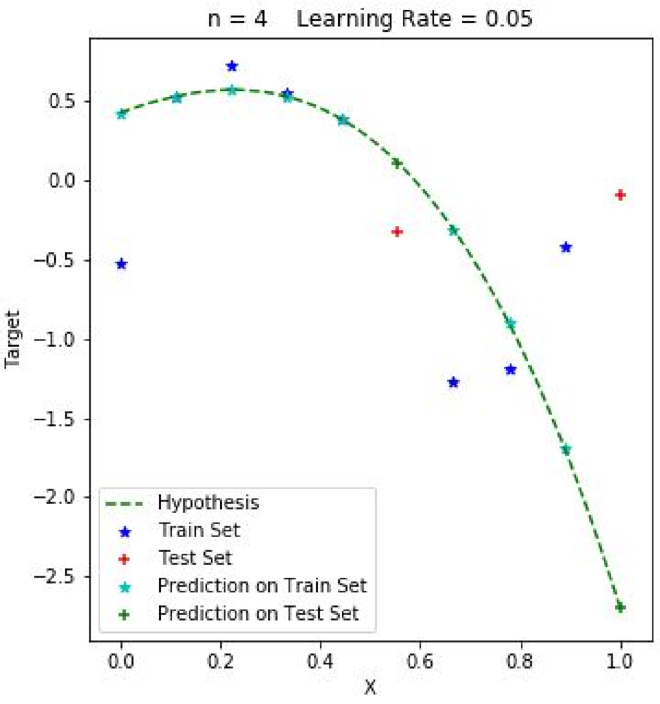


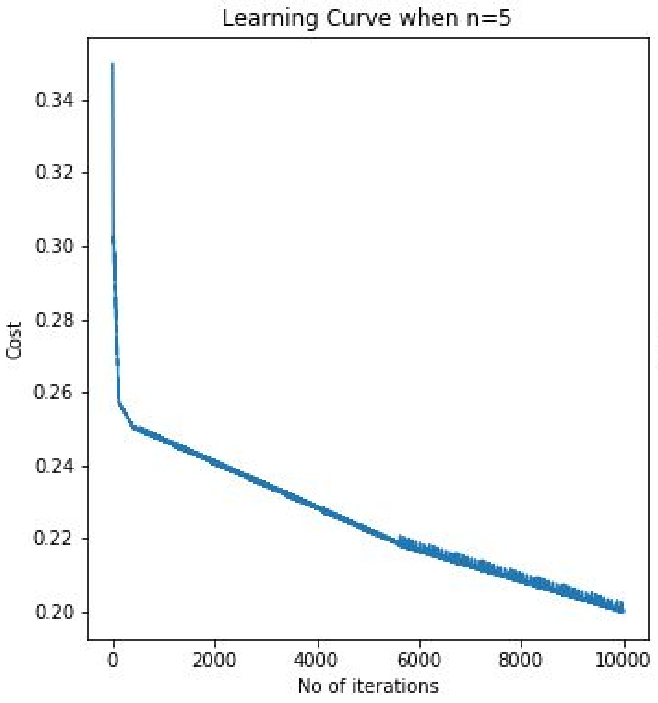


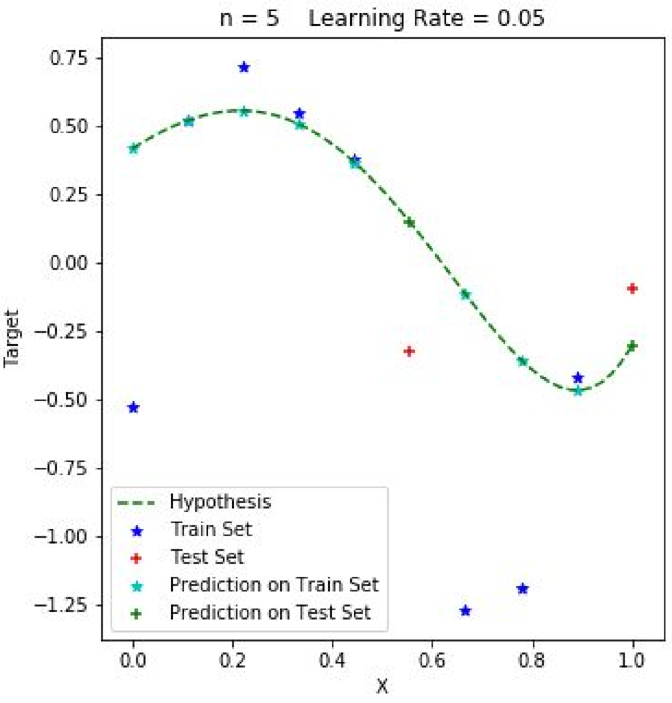


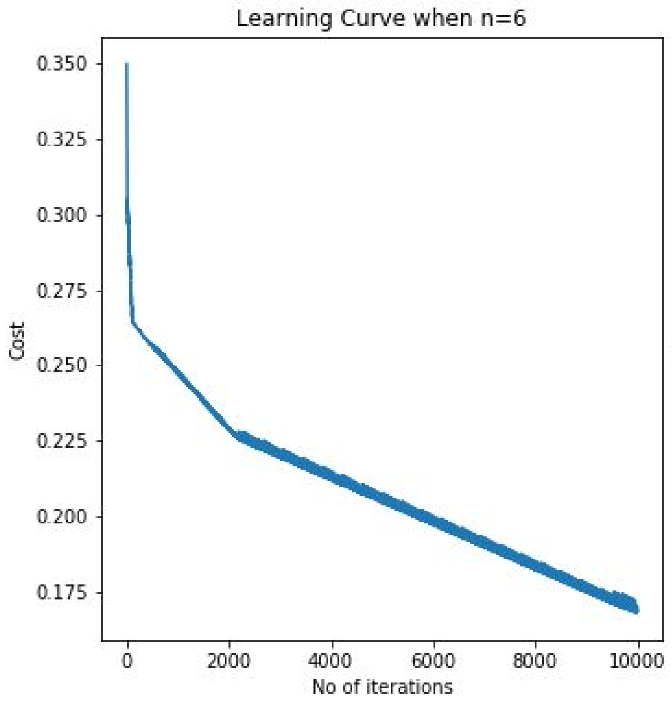


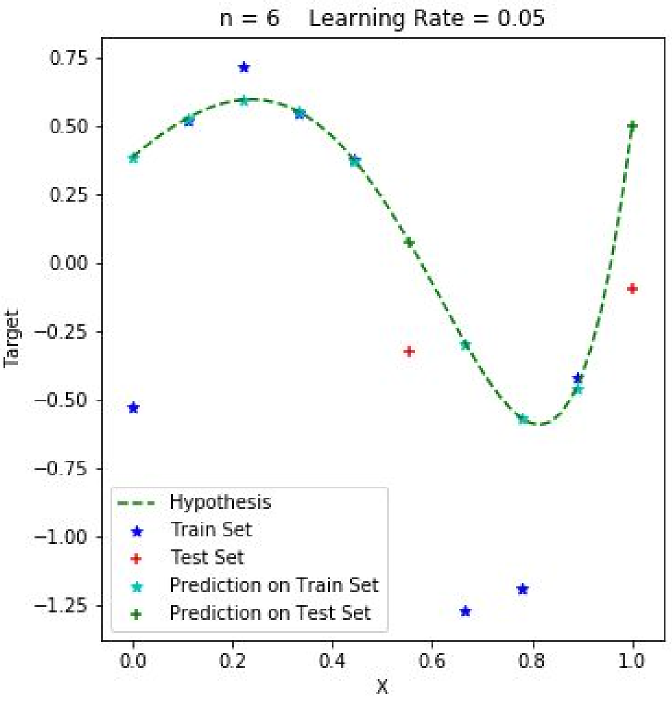

Learning Curve when n=4
Cost
0.34
0.32
0.30
0.28
0.26
0.24
0
2000
4000
6000
8000
10000
No of iterations

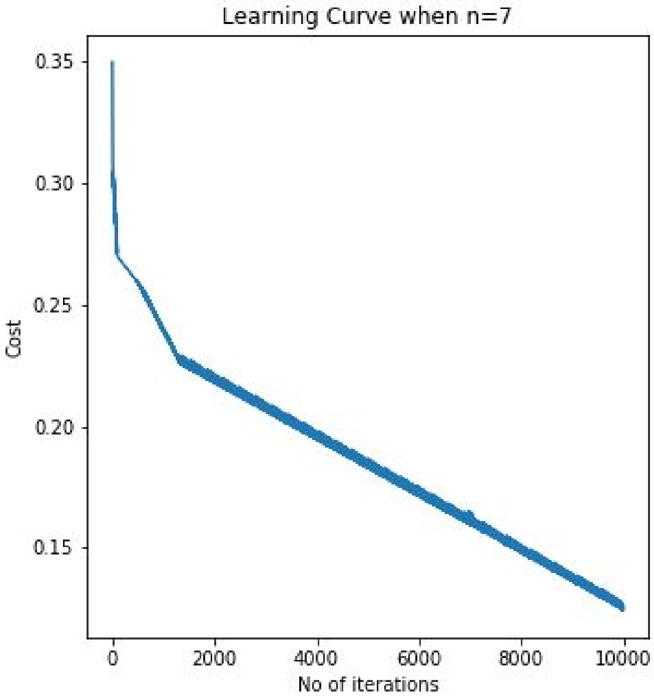

n = 4   Learning Rate = 0.05
Target
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
0.0
0.2
0.4
0.6
0.8
1.0
X
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set

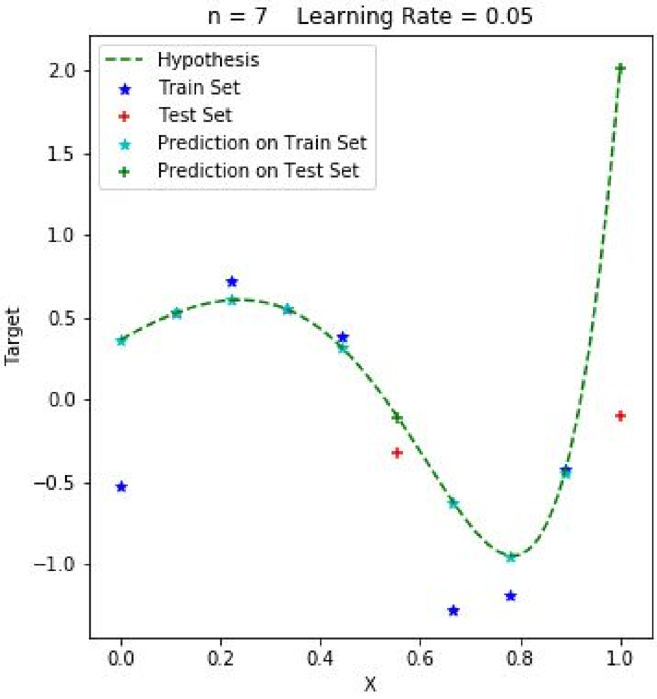

Learning Curve when n=5
Cost
0.34
0.32
0.30
0.28
0.26
0.24
0.22
0.20
0
2000
4000
6000
8000
10000
No of iterations

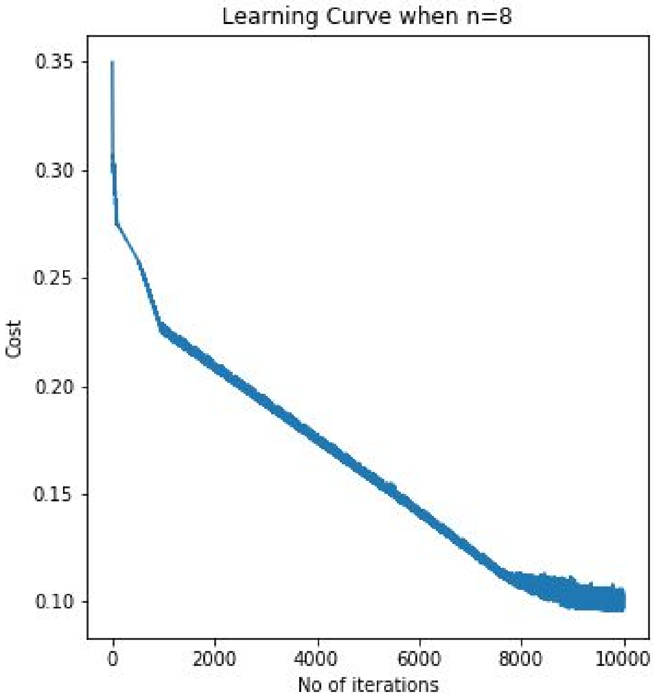

n = 5   Learning Rate = 0.05
Target
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
0.0
0.2
0.4
0.6
0.8
1.0
X
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set

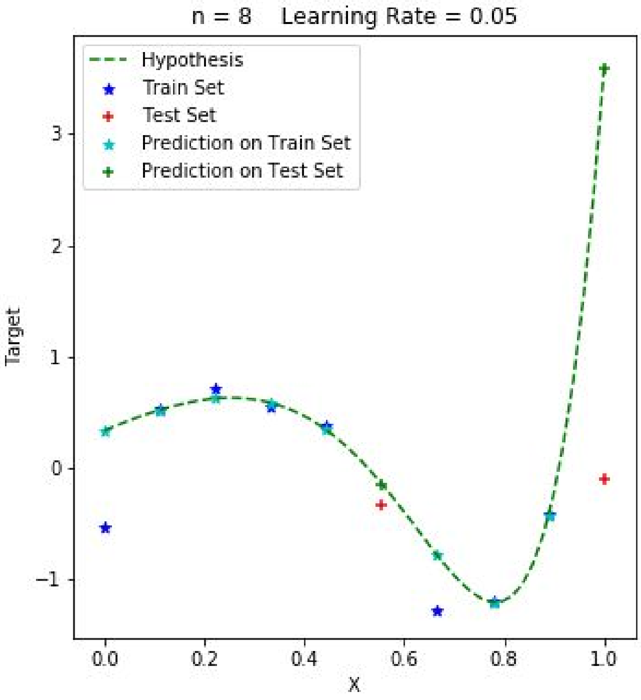

Learning Curve when n=6
Cost
0.350
0.325
0.300
0.275
0.250
0.225
0.200
0.175
0
2000
4000
6000
8000
10000
No of iterations

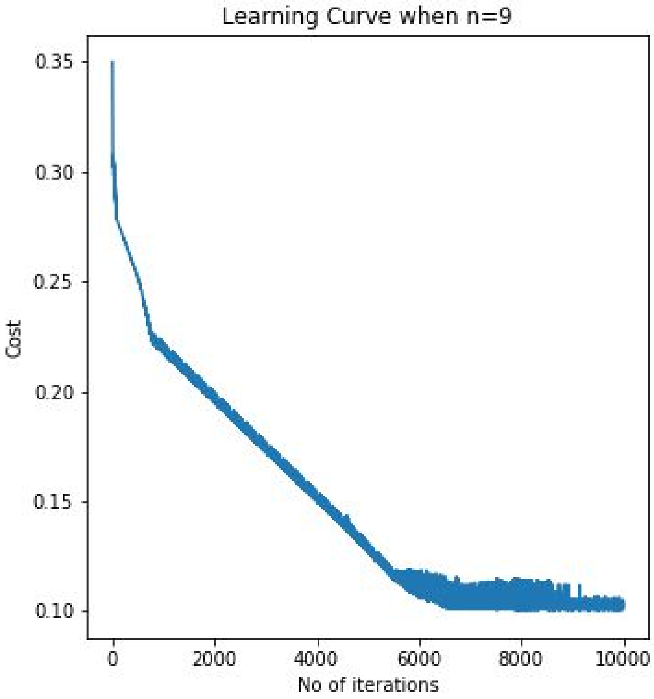

n = 6   Learning Rate = 0.05
Target
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
0.0
0.2
0.4
0.6
0.8
1.0
X
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set

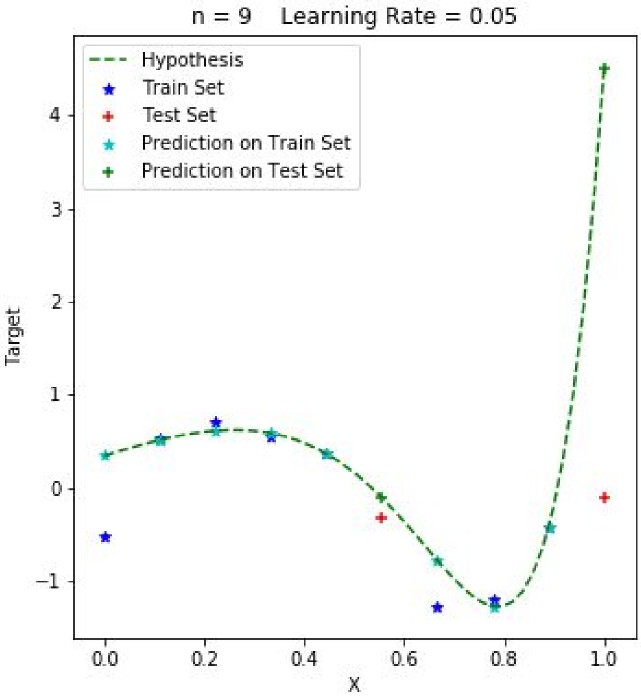

Learning Curve when n=7
0.35
0.30
0.25
0.20
0.15
Cost
0
2000
4000
6000
8000
10000
No of iterations


n = 7 Learning Rate = 0.05
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
Target
0.0
0.2
0.4
0.6
0.8
1.0
X

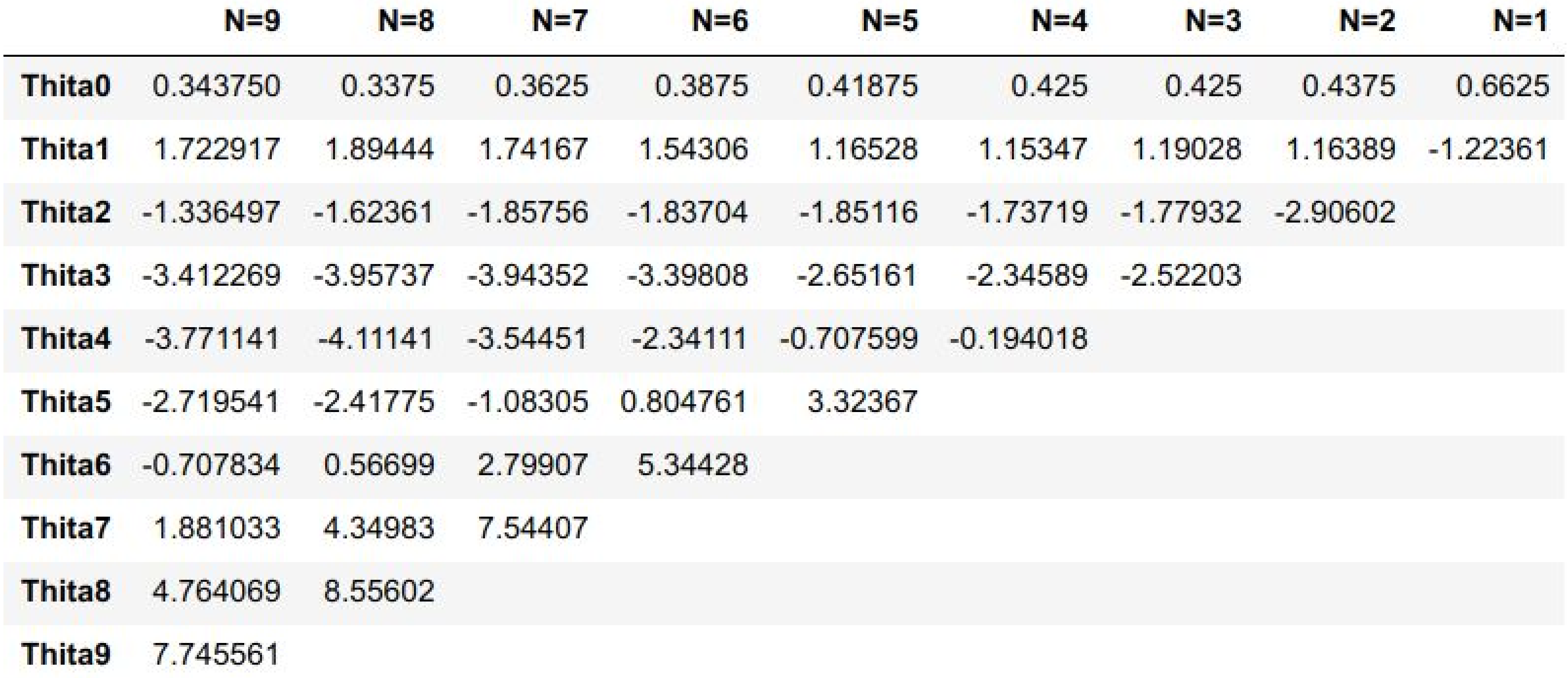

Learning Curve when n=8
0.35
0.30
0.25
0.20
0.15
0.10
Cost
0
2000
4000
6000
8000
10000
No of iterations


n = 8 Learning Rate = 0.05
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set
3
2
1
0
−1
Target
0.0
0.2
0.4
0.6
0.8
1.0
X


Learning Curve when n=9
0.35
0.30
0.25
0.20
0.15
0.10
Cost
0
2000
4000
6000
8000
10000
No of iterations

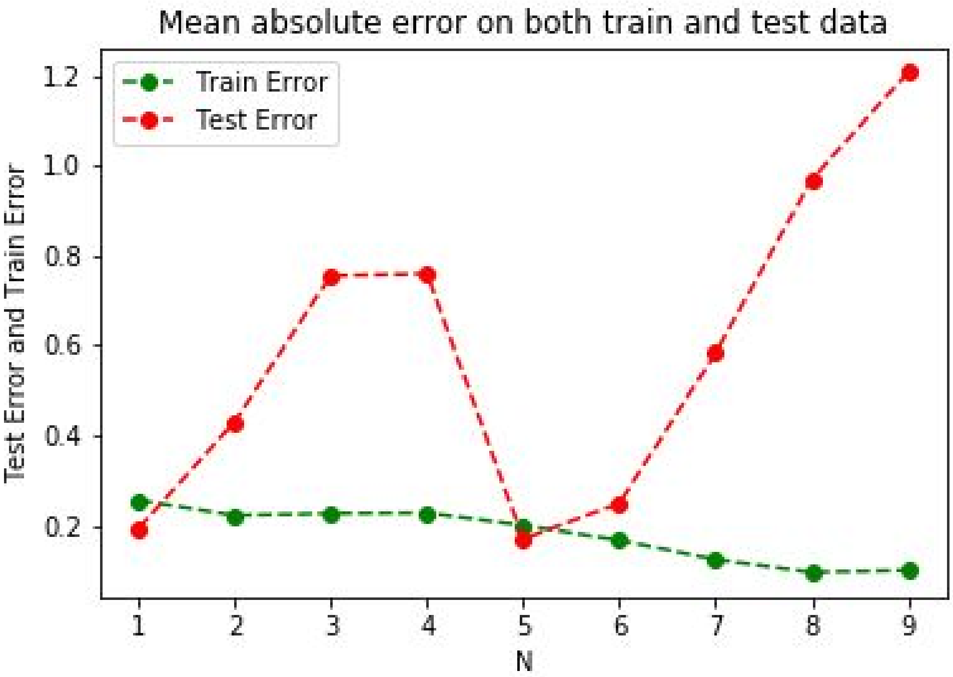

n = 9 Learning Rate = 0.05
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set
4
3
2
1
0
−1
Target
0.0
0.2
0.4
0.6
0.8
1.0
X

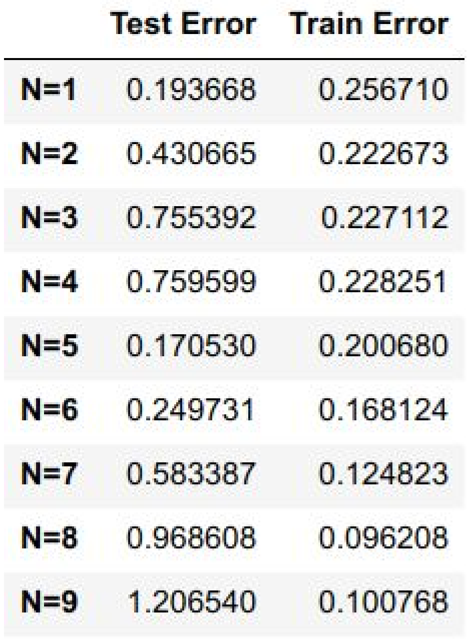
Parameters learned are as follows:



	N=9	N=8	N=7	N=6	N=5	N=4	N=3	N=2	N=1
Thita0	0.343750	0.3375	0.3625	0.3875	0.41875	0.425	0.425	0.4375	0.6625
Thita1	1.722917	1.89444	1.74167	1.54306	1.16528	1.15347	1.19028	1.16389	-1.22361
Thita2	-1.336497	-1.62361	-1.85756	-1.83704	-1.85116	-1.73719	-1.77932	-2.90602	
Thita3	-3.412269	-3.95737	-3.94352	-3.39808	-2.65161	-2.34589	-2.52203		
Thita4	-3.771141	-4.11141	-3.54451	-2.34111	-0.707599	-0.194018			
Thita5	-2.719541	-2.41775	-1.08305	0.804761	3.32367				
Thita6	-0.707834	0.56699	2.79907	5.34428					
Thita7	1.881033	4.34983	7.54407						
Thita8	4.764069	8.55602							
Thita9	7.745561								

Test and Train errors are as follows:

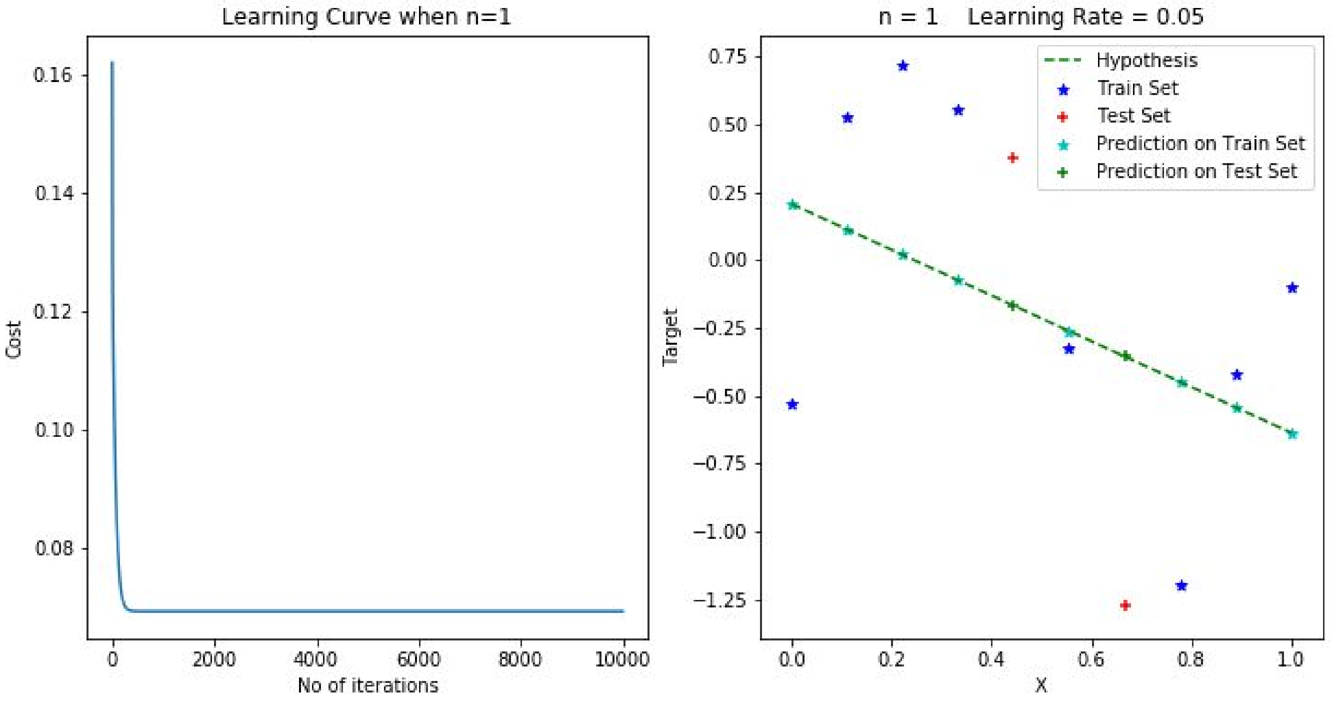


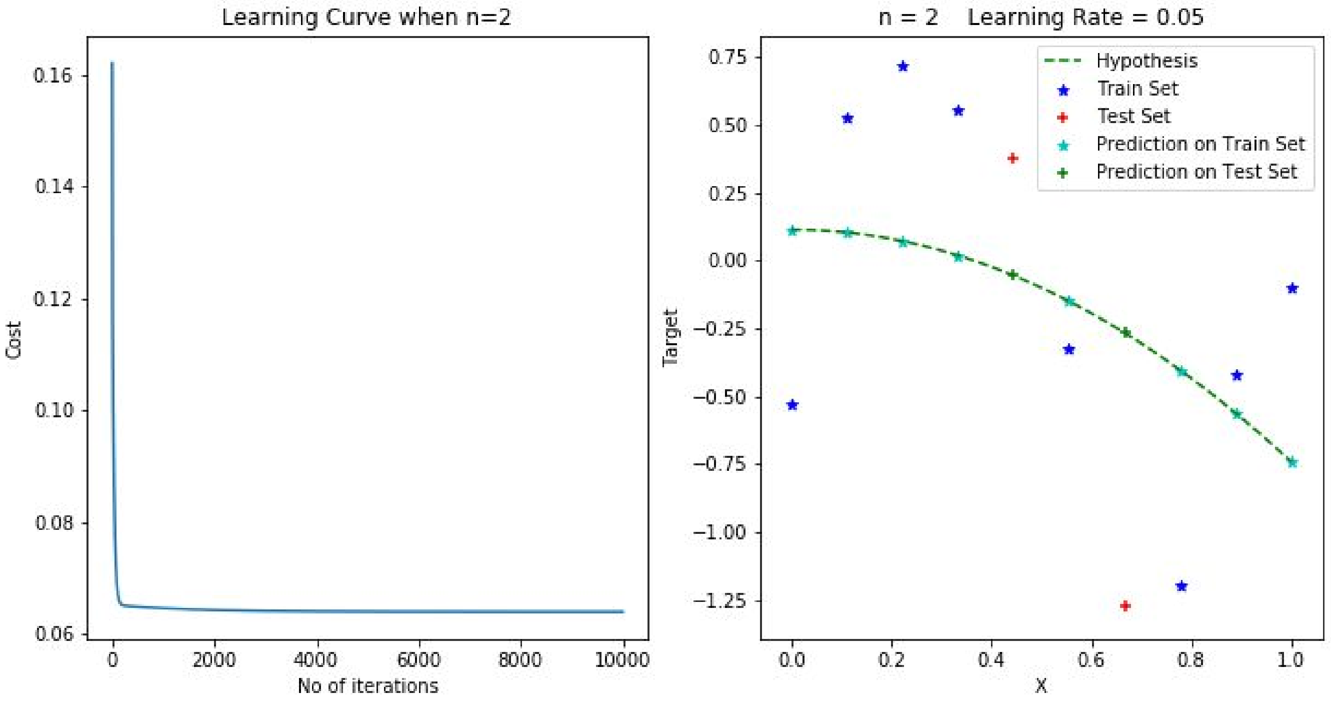


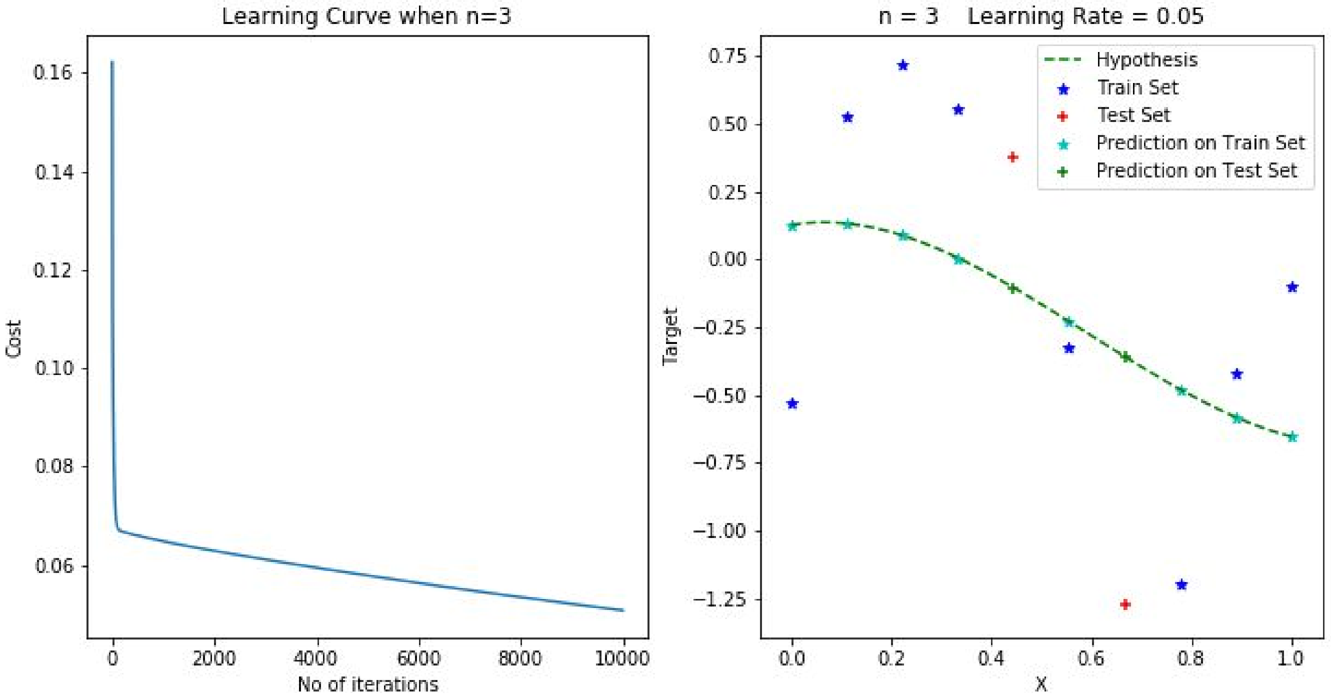
	Test Error	Train Error
N=1	0.193668	0.256710
N=2	0.430665	0.222673
N=3	0.755392	0.227112
N=4	0.759599	0.228251
N=5	0.170530	0.200680
N=6	0.249731	0.168124
N=7	0.583387	0.124823
N=8	0.968608	0.096208
N=9	1.206540	0.100768

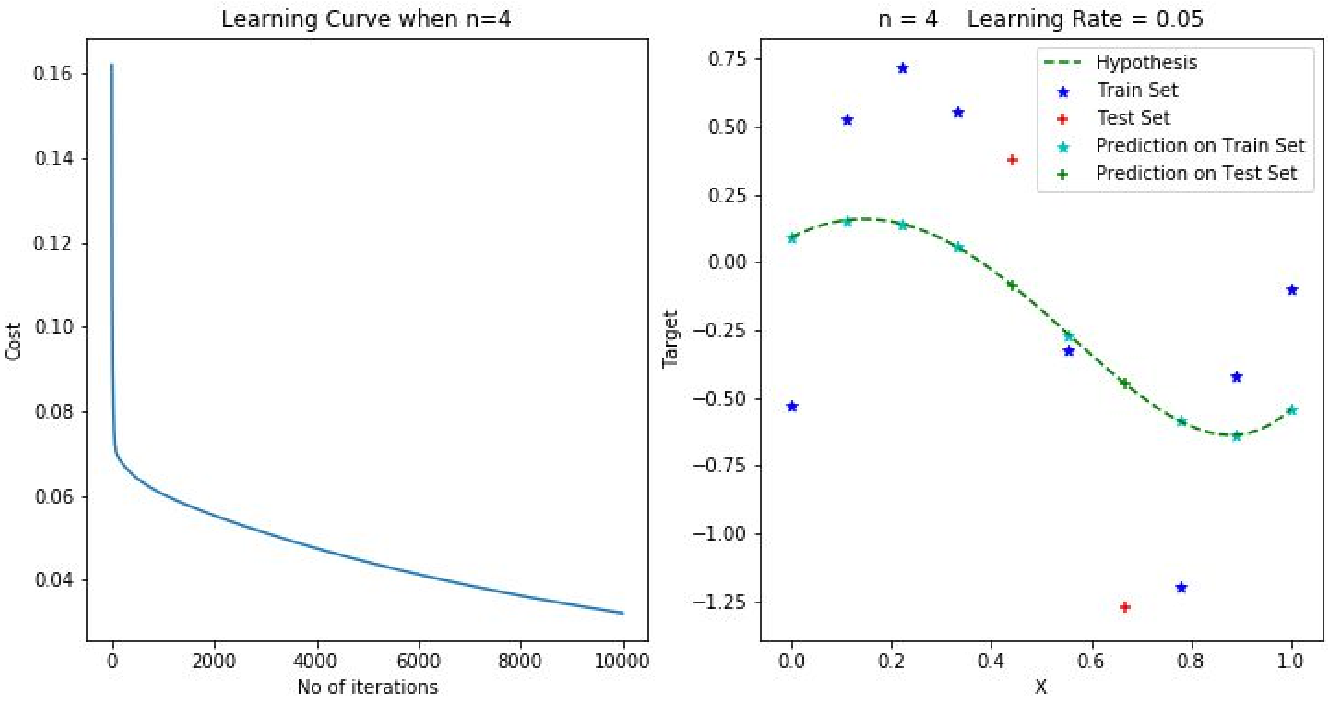
b) Fourth power error

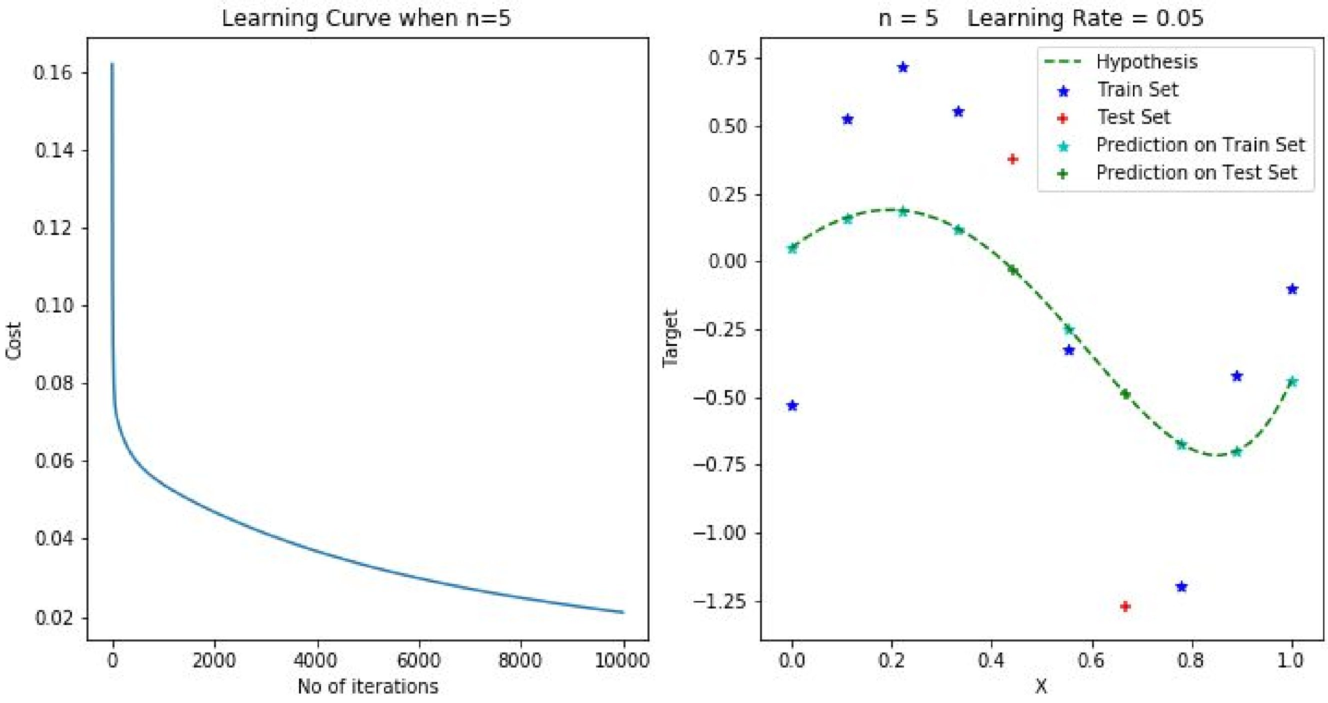
$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(W^T\Phi_n(x)-y\right)^4$$

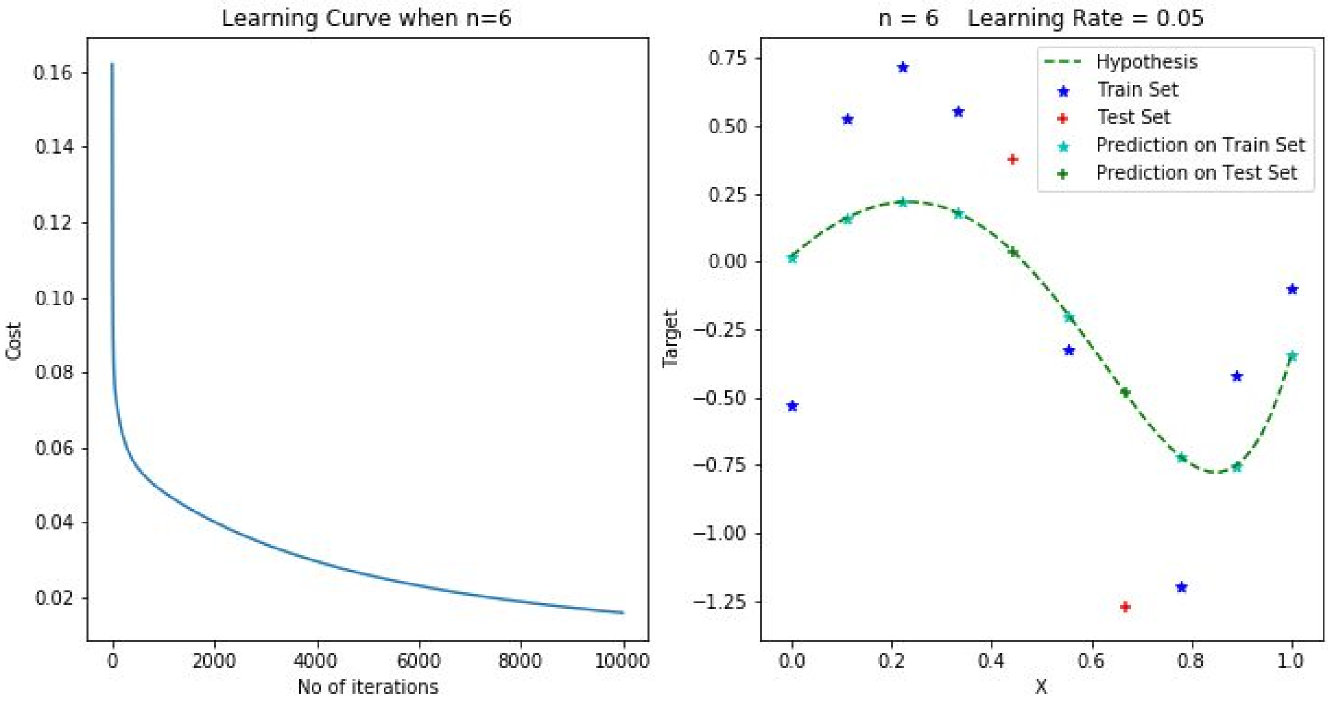


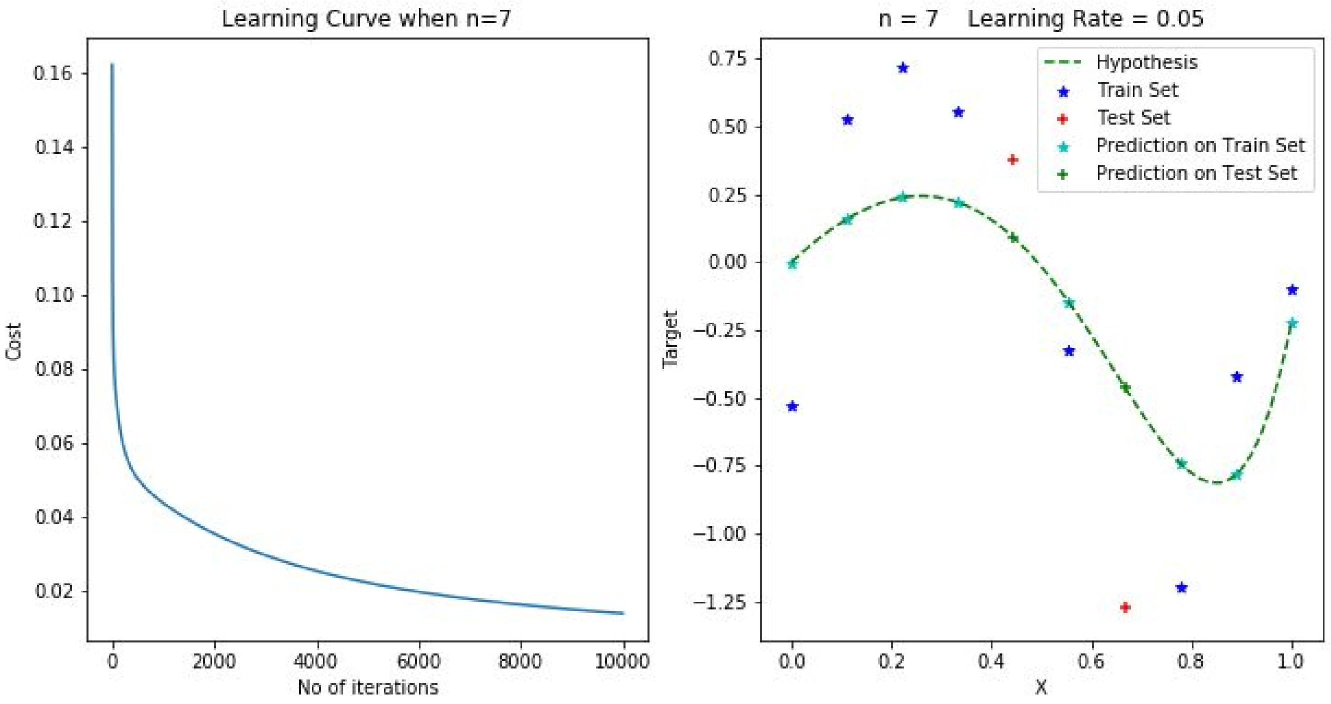


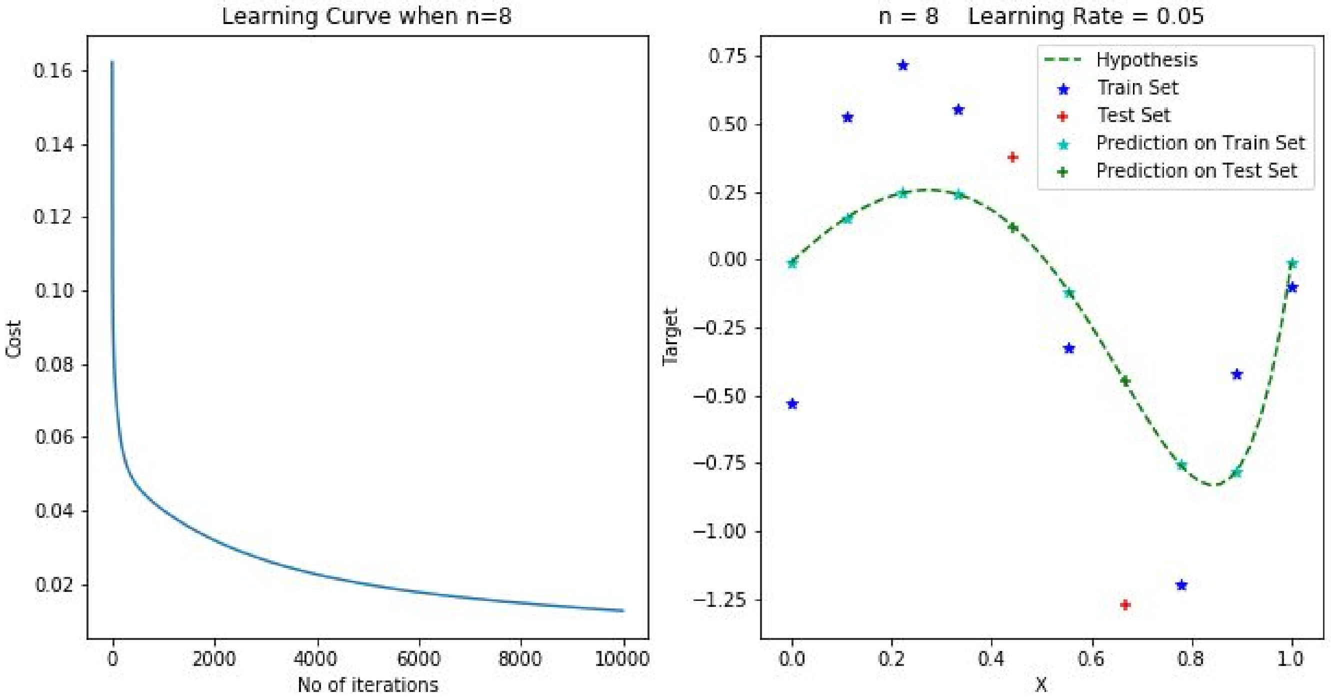


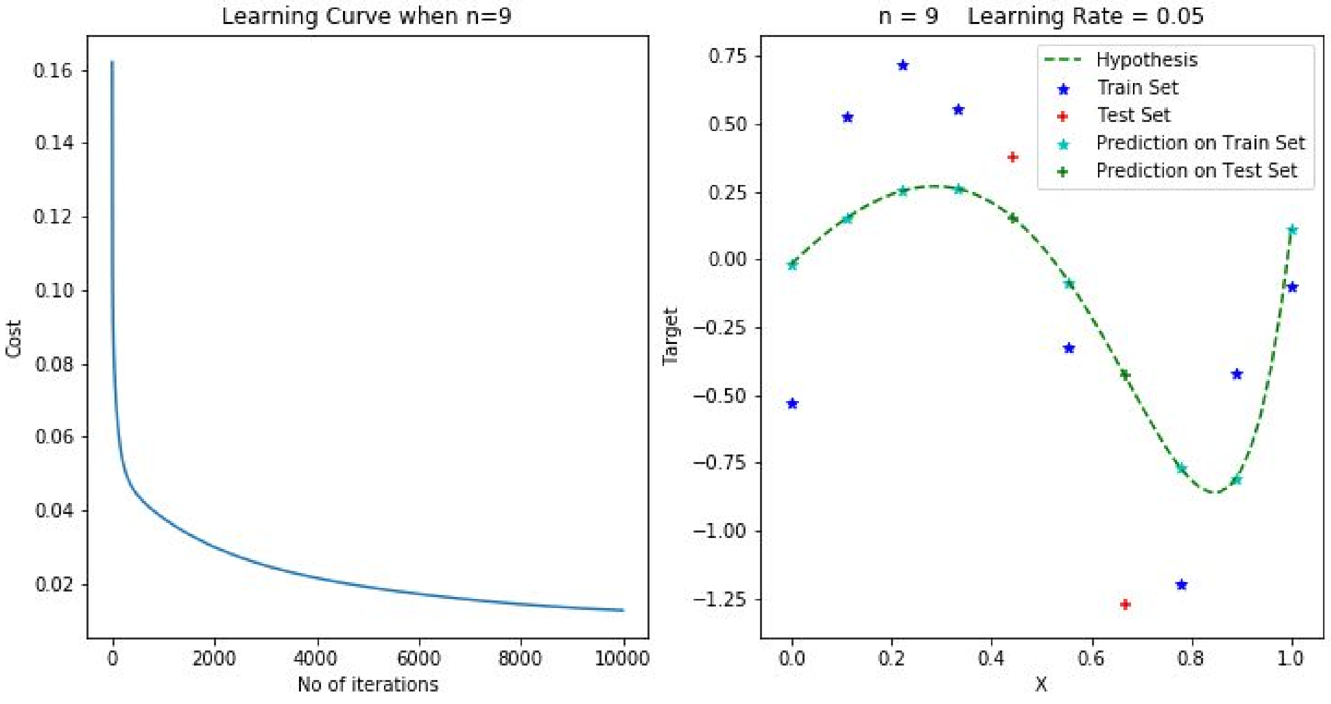

Learning Curve when n=4
Cost
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0
2000
4000
6000
8000
10000
No of iterations
n = 4   Learning Rate = 0.05
Target
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
0.0
0.2
0.4
0.6
0.8
1.0
X
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set


Learning Curve when n=5
Cost
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0
2000
4000
6000
8000
10000
No of iterations
n = 5   Learning Rate = 0.05
Target
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
0.0
0.2
0.4
0.6
0.8
1.0
X
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set

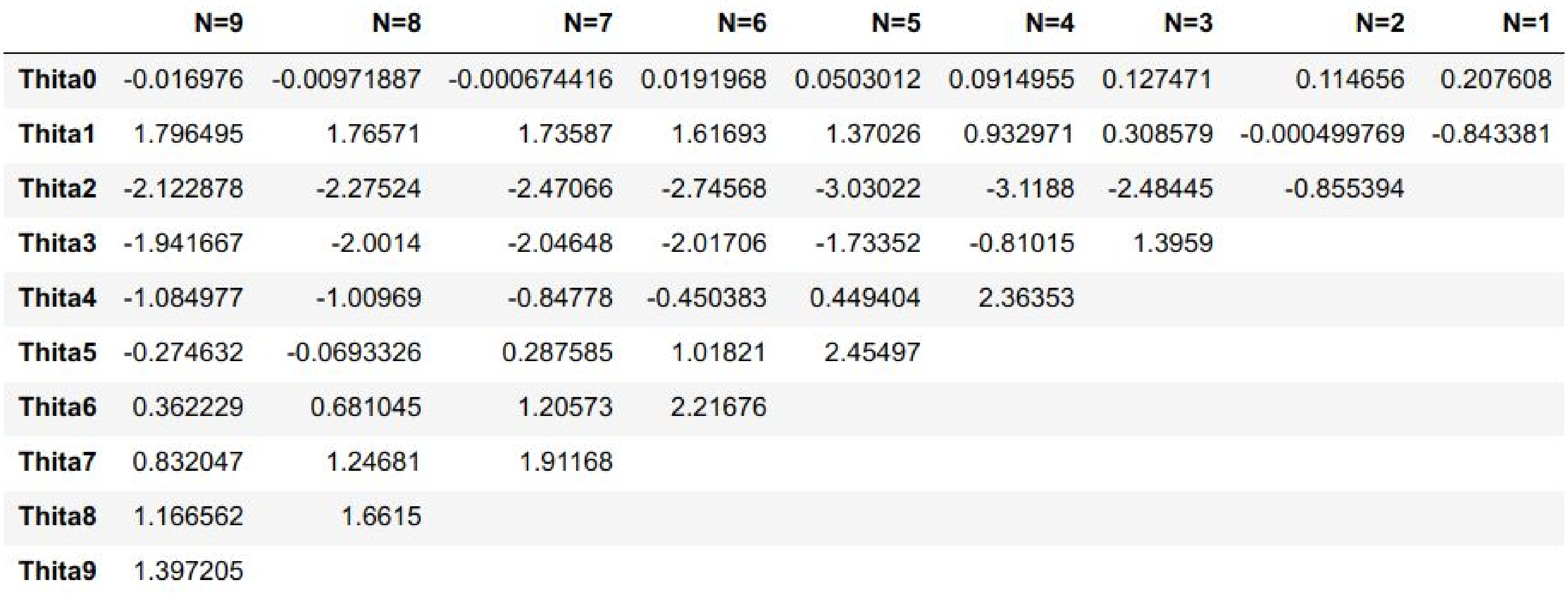

Learning Curve when n=6
Cost
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0
2000
4000
6000
8000
10000
No of iterations
n = 6   Learning Rate = 0.05
Target
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
0.0
0.2
0.4
0.6
0.8
1.0
X
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set


Learning Curve when n=7
Cost
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0
2000
4000
6000
8000
10000
No of iterations
n = 7   Learning Rate = 0.05
Target
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
0.0
0.2
0.4
0.6
0.8
1.0
X
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set


Learning Curve when n=8
Cost
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0
2000
4000
6000
8000
10000
No of iterations
n = 8   Learning Rate = 0.05
Target
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
0.0
0.2
0.4
0.6
0.8
1.0
X
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set

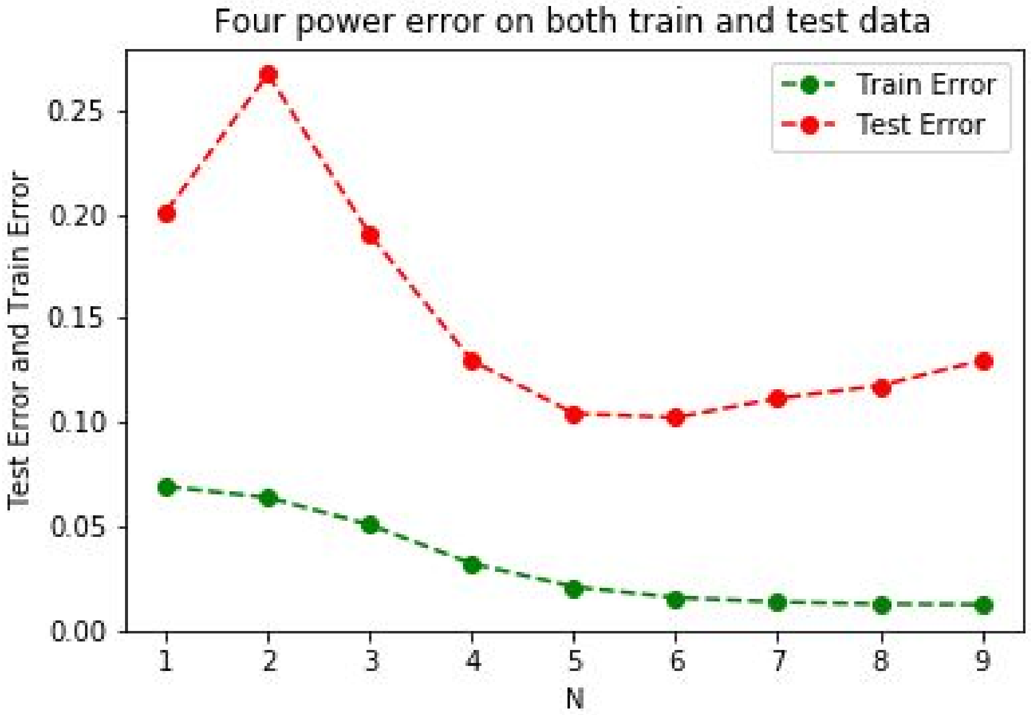

Learning Curve when n=9
Cost
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0
2000
4000
6000
8000
10000
No of iterations
n = 9   Learning Rate = 0.05
Target
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
0.0
0.2
0.4
0.6
0.8
1.0
X
Hypothesis
Train Set
Test Set
Prediction on Train Set
Prediction on Test Set

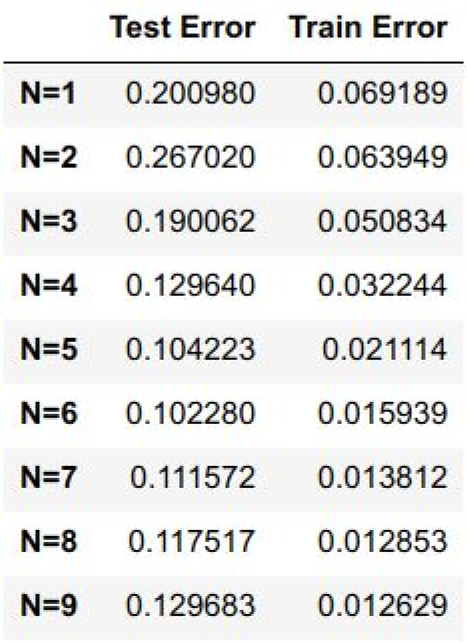
Parameters learned are as follows:



	N=9	N=8	N=7	N=6	N=5	N=4	N=3	N=2	N=1
Thita0	-0.016976	-0.00971887	-0.000674416	0.0191968	0.0503012	0.0914955	0.127471	0.114656	0.207608
Thita1	1.796495	1.76571	1.73587	1.61693	1.37026	0.932971	0.308579	-0.000499769	-0.843381
Thita2	-2.122878	-2.27524	-2.47066	-2.74568	-3.03022	-3.1188	-2.48445	-0.855394	
Thita3	-1.941667	-2.0014	-2.04648	-2.01706	-1.73352	-0.81015	1.3959		
Thita4	-1.084977	-1.00969	-0.84778	-0.450383	0.449404	2.36353			
Thita5	-0.274632	-0.0693326	0.287585	1.01821	2.45497				
Thita6	0.362229	0.681045	1.20573	2.21676					
Thita7	0.832047	1.24681	1.91168						
Thita8	1.166562	1.6615							
Thita9	1.397205								

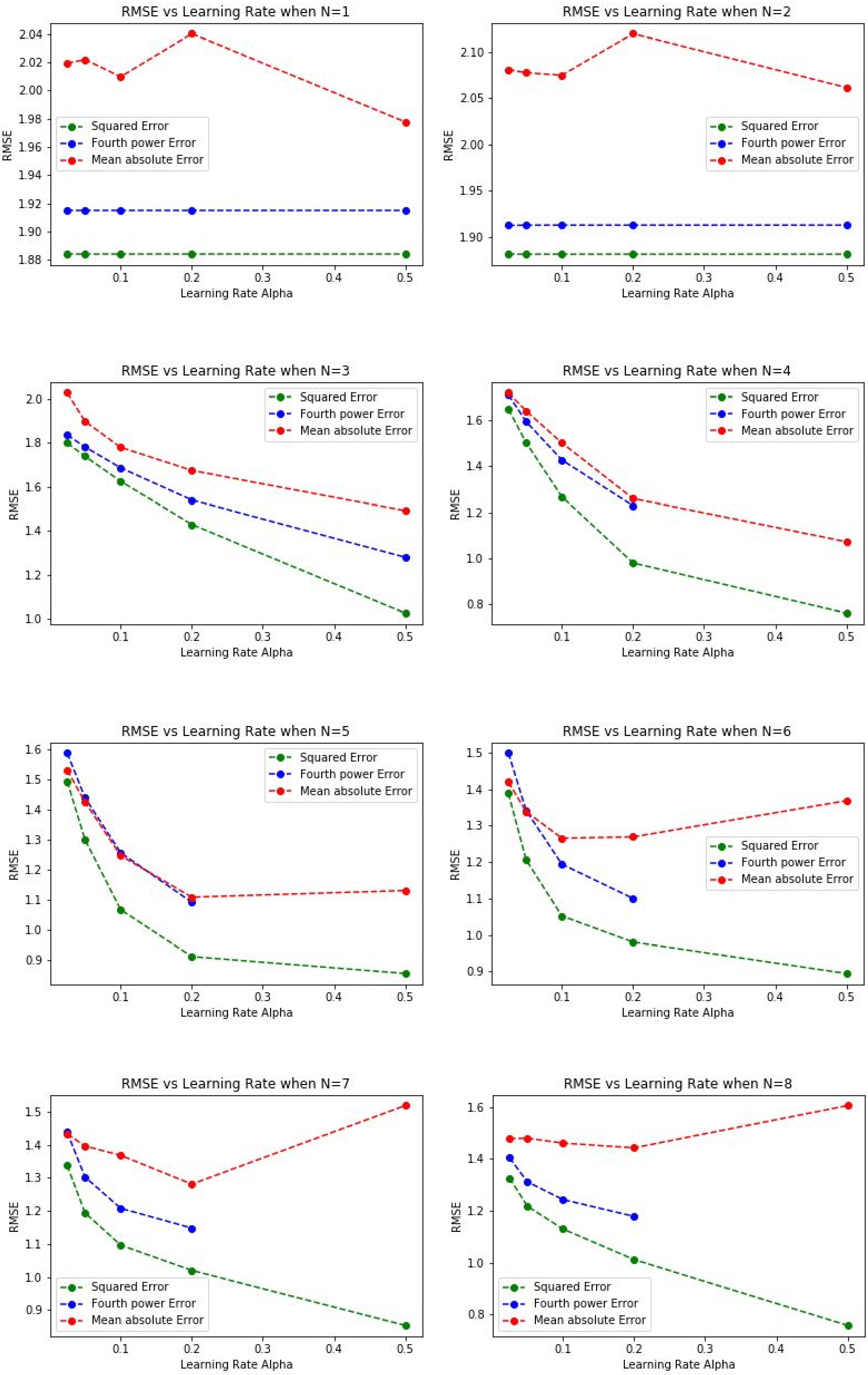
Train and Test error by Fourth power error as Cost function:

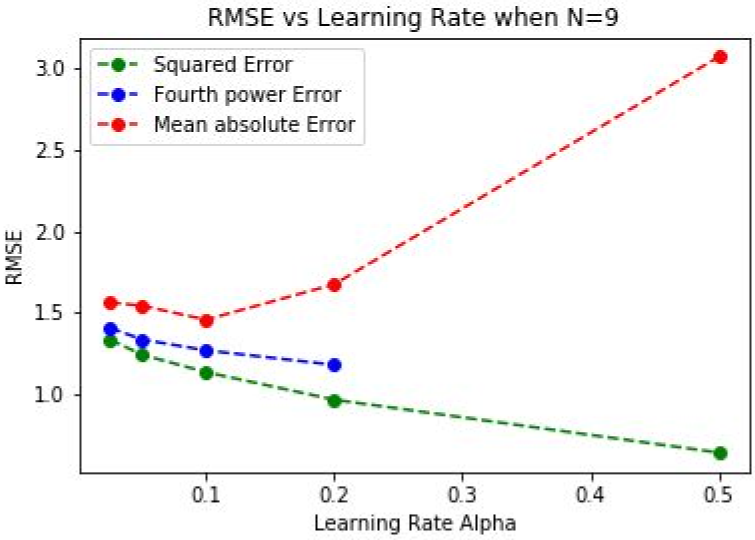




	Test Error	Train Error
N=1	0.200980	0.069189
N=2	0.267020	0.063949
N=3	0.190062	0.050834
N=4	0.129640	0.032244
N=5	0.104223	0.021114
N=6	0.102280	0.015939
N=7	0.111572	0.013812
N=8	0.117517	0.012853
N=9	0.129683	0.012629

b) Plot of RMSE vs learning rate for each cost function





For all those plots which don't have a point at Learning rate Alpha = 0.5, there Model doesn't converge to a minima for that value of alpha rather cost keeps on increasing as the number of iterations increases(i.e. Cost diverges) and RMSE value tends to infinity.

So the optimal learning rate Alpha which should be used as per deduction from these plots is

0.2 if Cost function used is Fourth Power Error.

0.5 if Cost function used is Squared Error.

For Mean Absolute Error, the optimal value of alpha varies as N changes.

- By Vaibhav Poddar
16CS10051

** **THE END** **