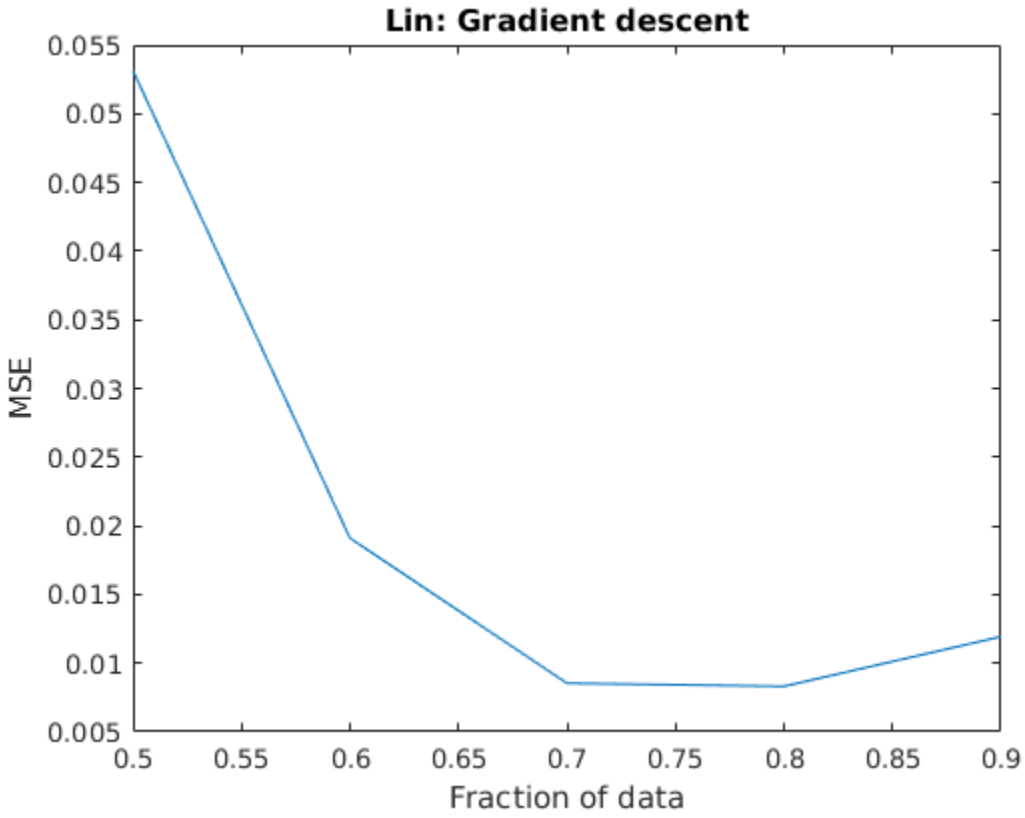
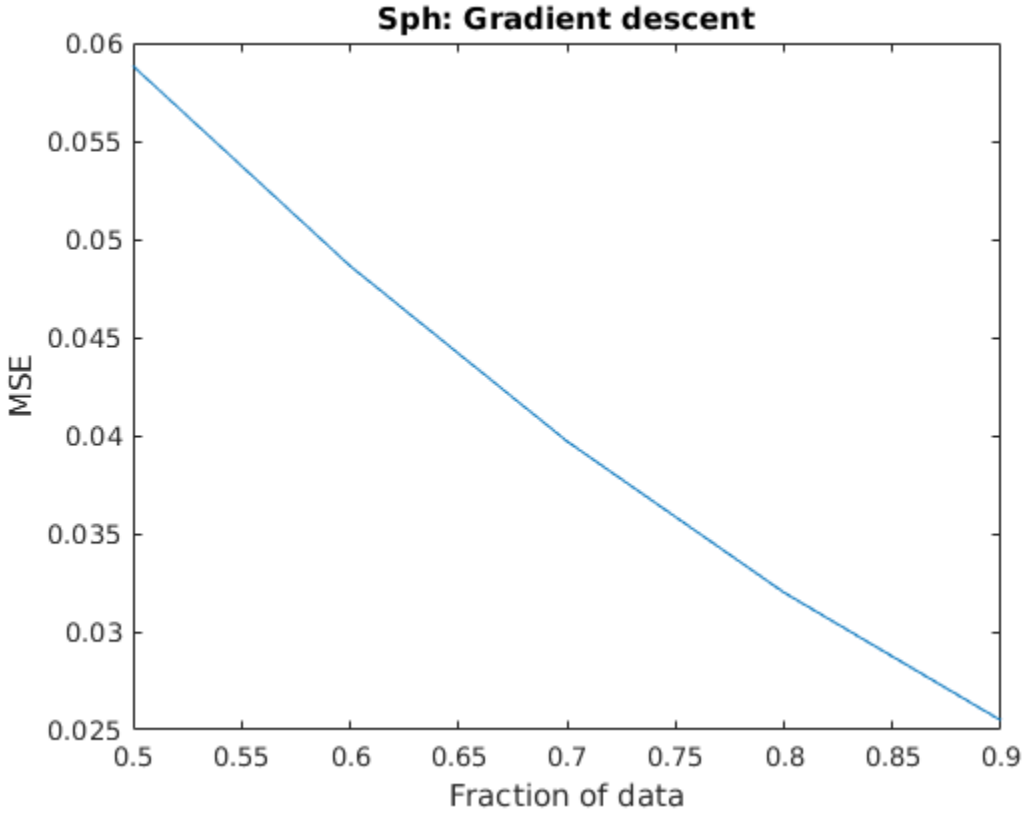
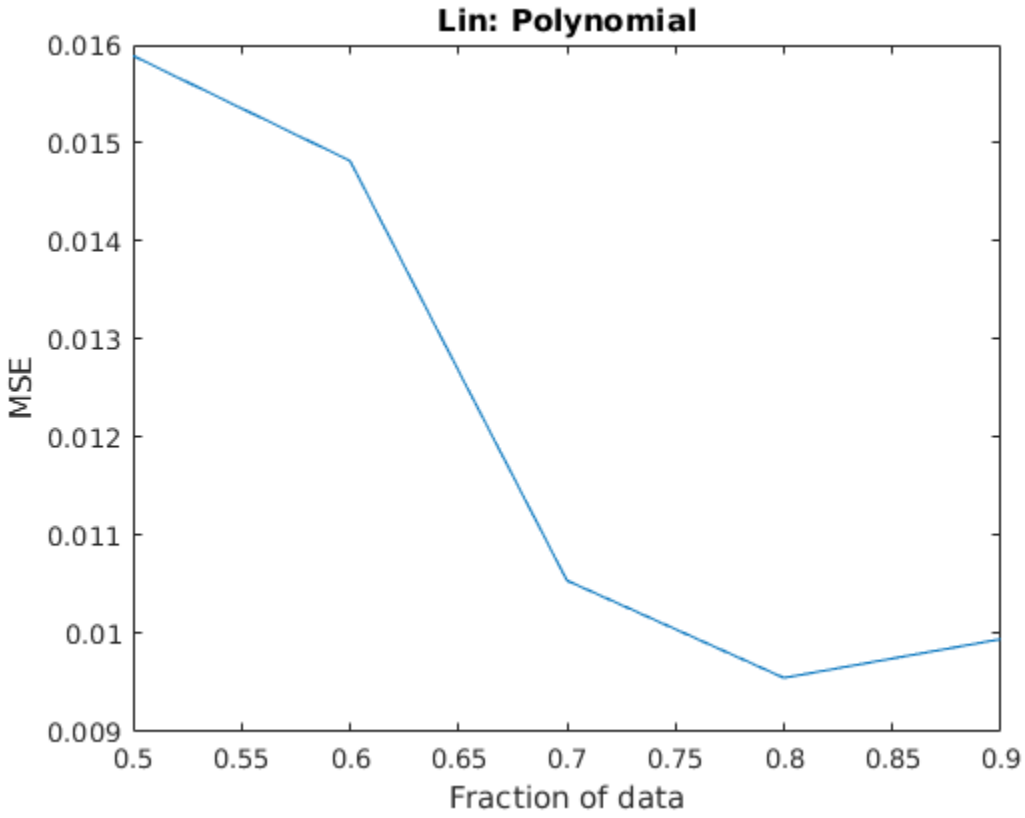
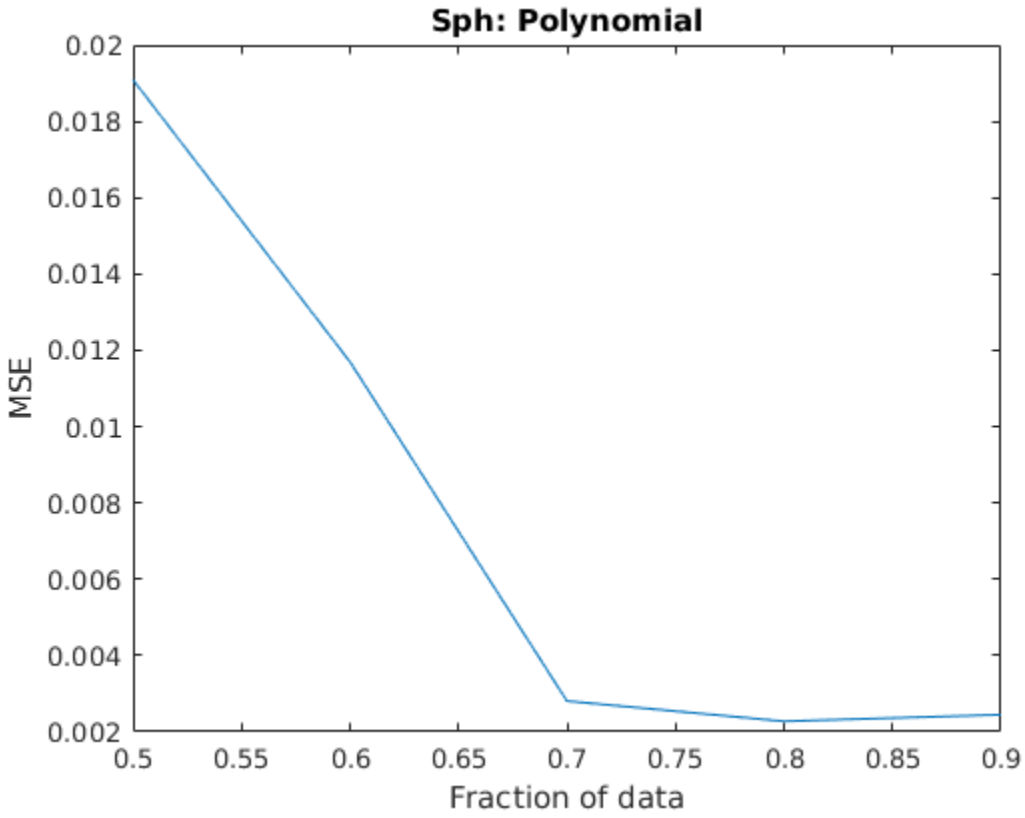
Ans 4. a.


Lin: Gradient descent
MSE
Fraction of data
0.055
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9

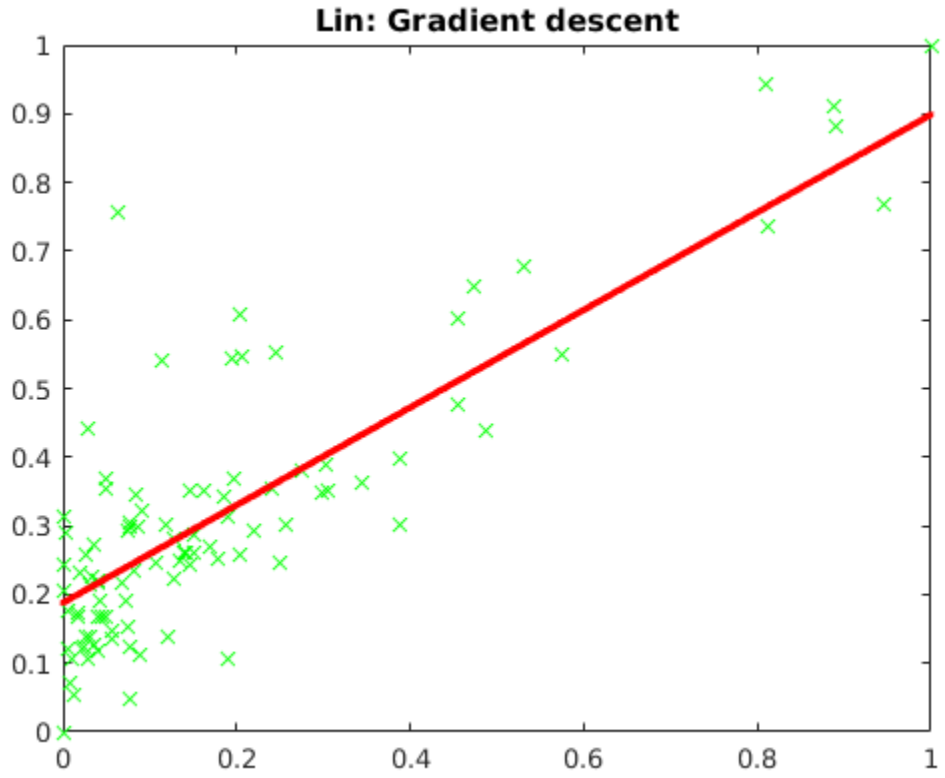

Sph: Gradient descent
MSE
Fraction of data
0.06
0.055
0.05
0.045
0.04
0.035
0.03
0.025
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9

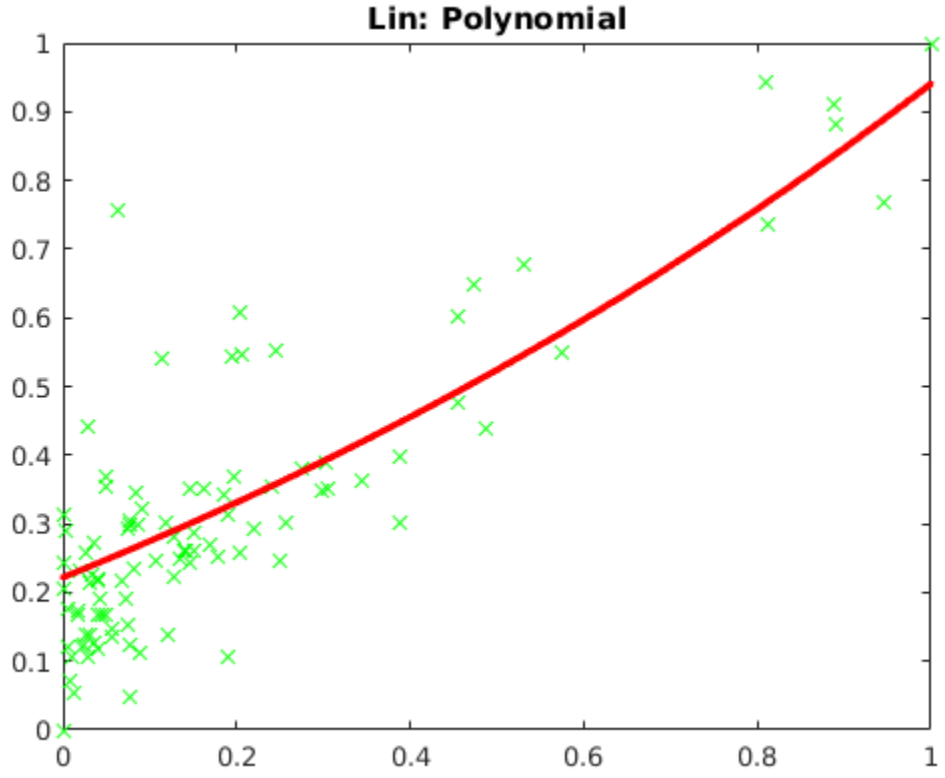
Ans 4. b.

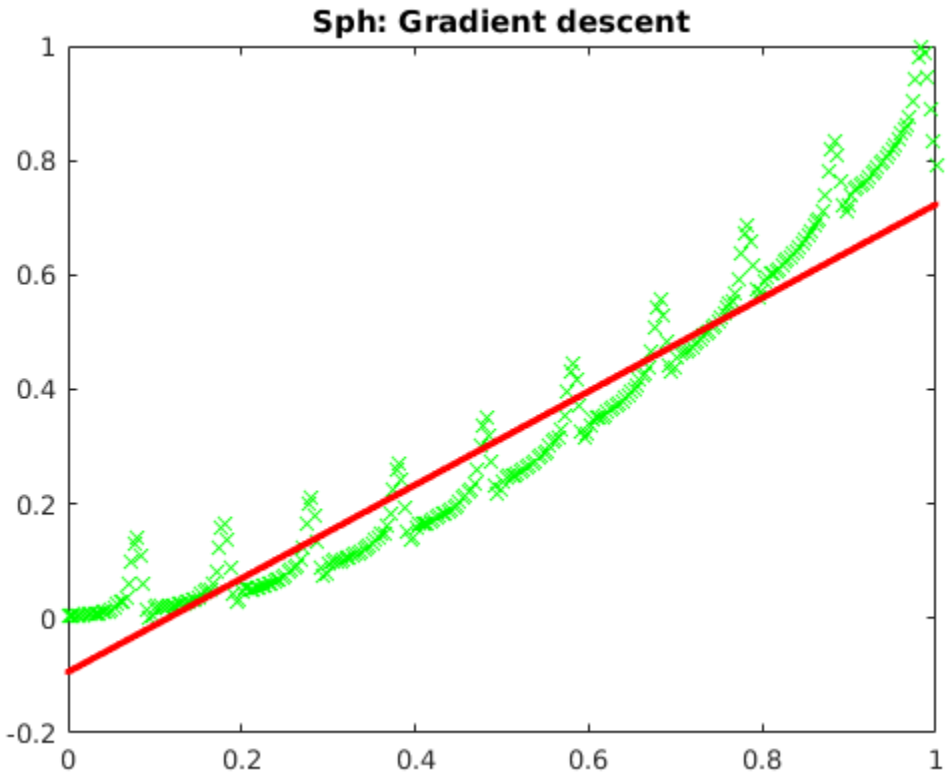

Lin: Polynomial
MSE
Fraction of data
0.016
0.015
0.014
0.013
0.012
0.011
0.01
0.009
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9

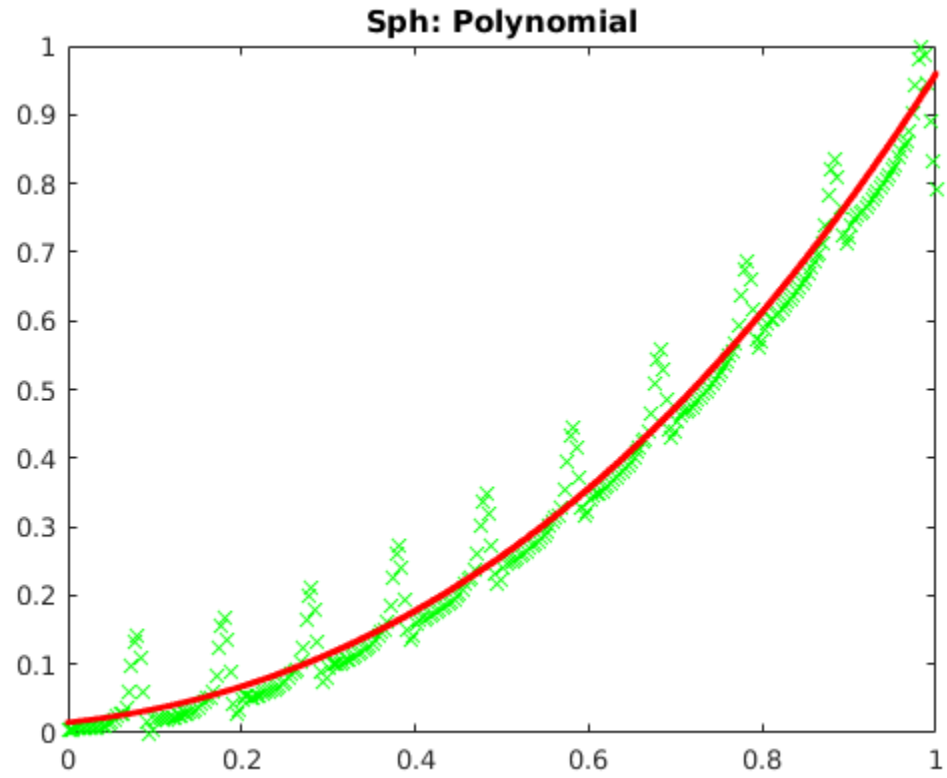

Sph: Polynomial
MSE
Fraction of data
0.02
0.018
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9

Ans 4. c.


Lin: Gradient descent
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1


Lin: Polynomial
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1


Sph: Gradient descent
1
0.8
0.6
0.4
0.2
0
-0.2
0
0.2
0.4
0.6
0.8
1


Sph: Polynomial
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1

Ans 4. d.

Choosing the delta value:

Lin dataset:

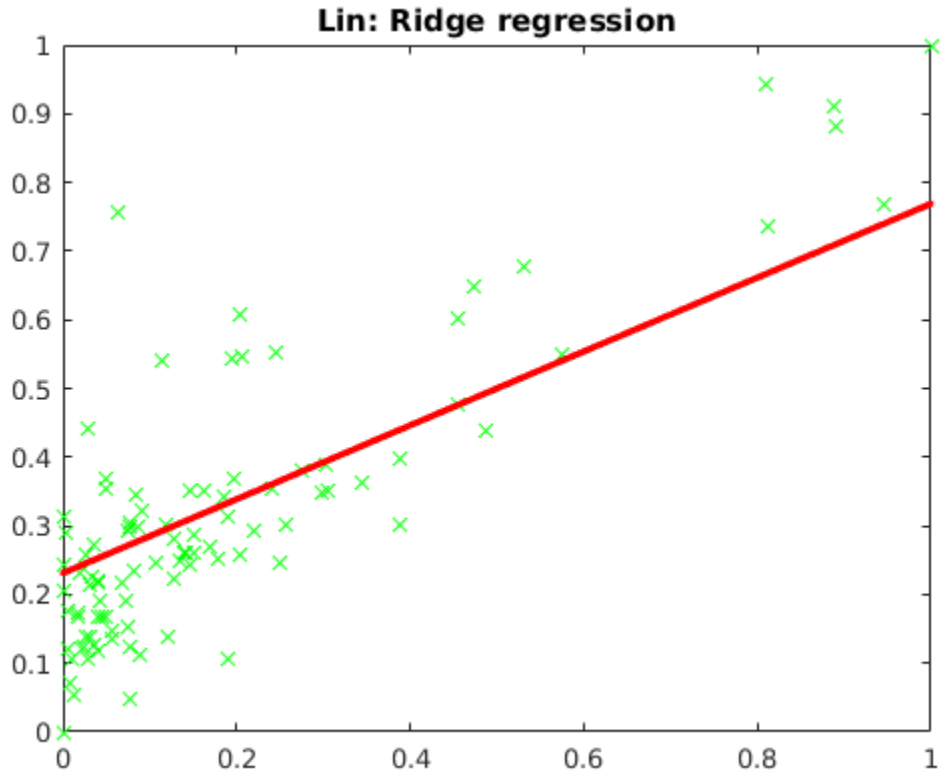
	80 % training data + 20 % test data			90 % training data + 10 % test data		
	Train Error	Test Error	Avg Error	Train Error	Test Error	Avg Error
0.1	0.346640	0.014163	0.180401	0.185215	0.013989	0.099602
0.5	0.050667	0.017297	0.033982	0.149542	0.014900	0.082221
1	0.062826	0.007193	0.035009	0.031174	0.018482	0.024828
10	2.208664	0.008807	1.108736	1.922445	0.009766	0.966106
50	3.498616	0.011004	1.754810	2.168872	0.014136	1.091504

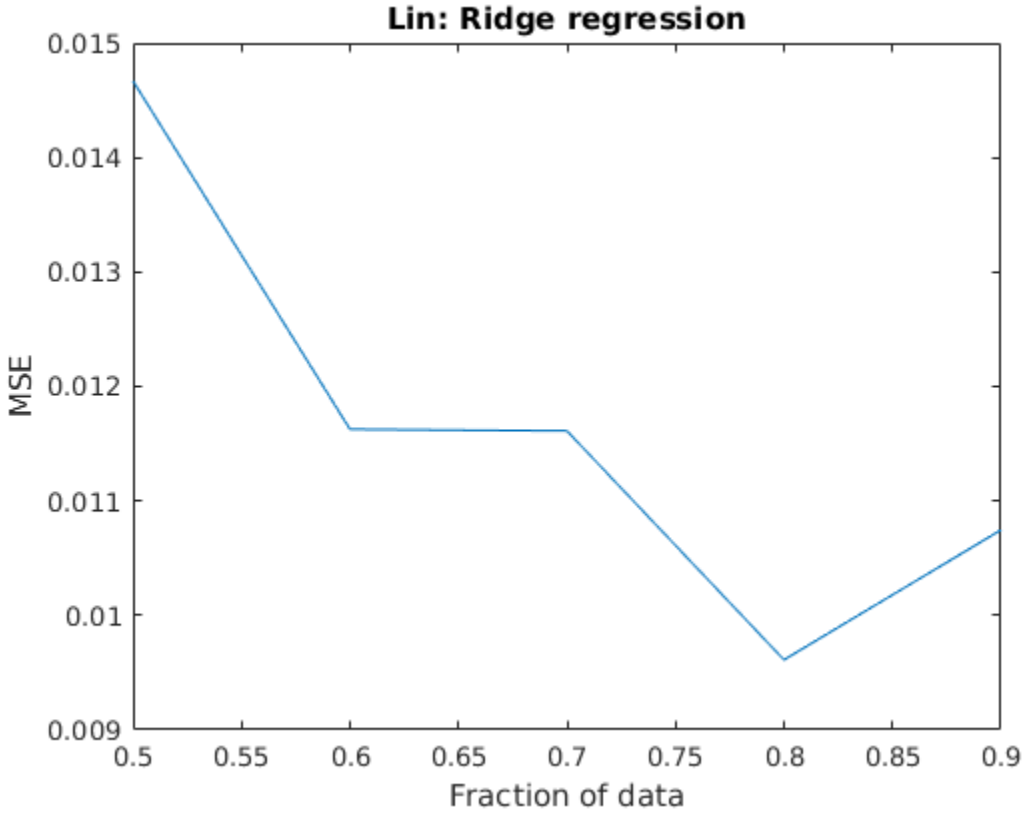
For 90% training example the test error seems to be minimized for delta = 0.1. It performs not so great compared to delta = 1(best avg error) but that is majorly due to poor performance in train error. Therefore, delta = 0.1

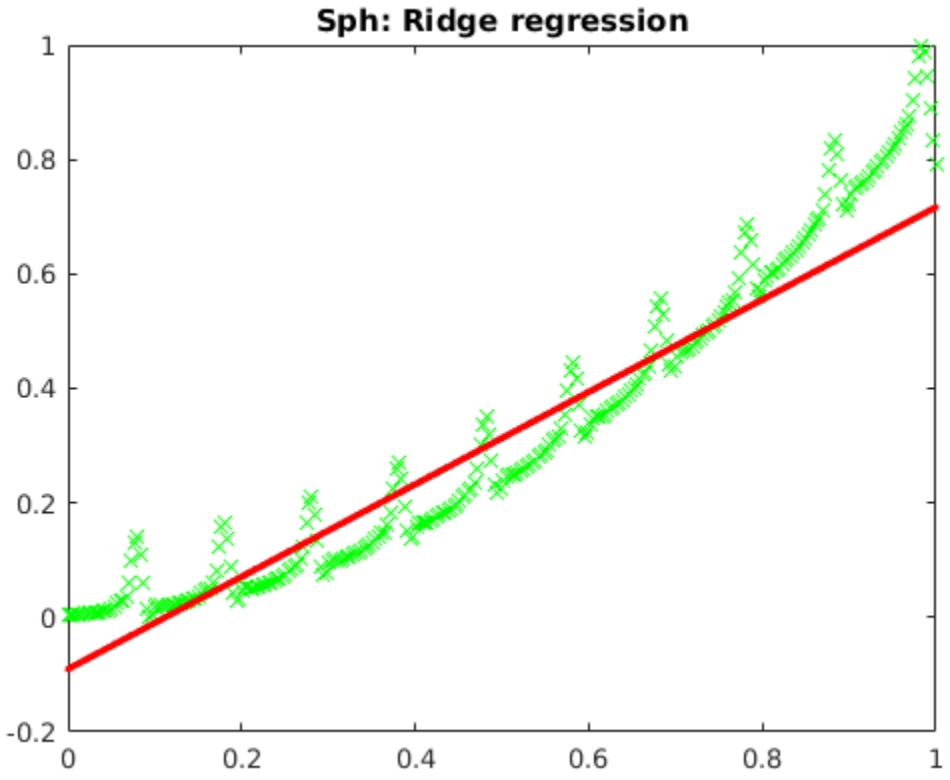
Sph dataset:

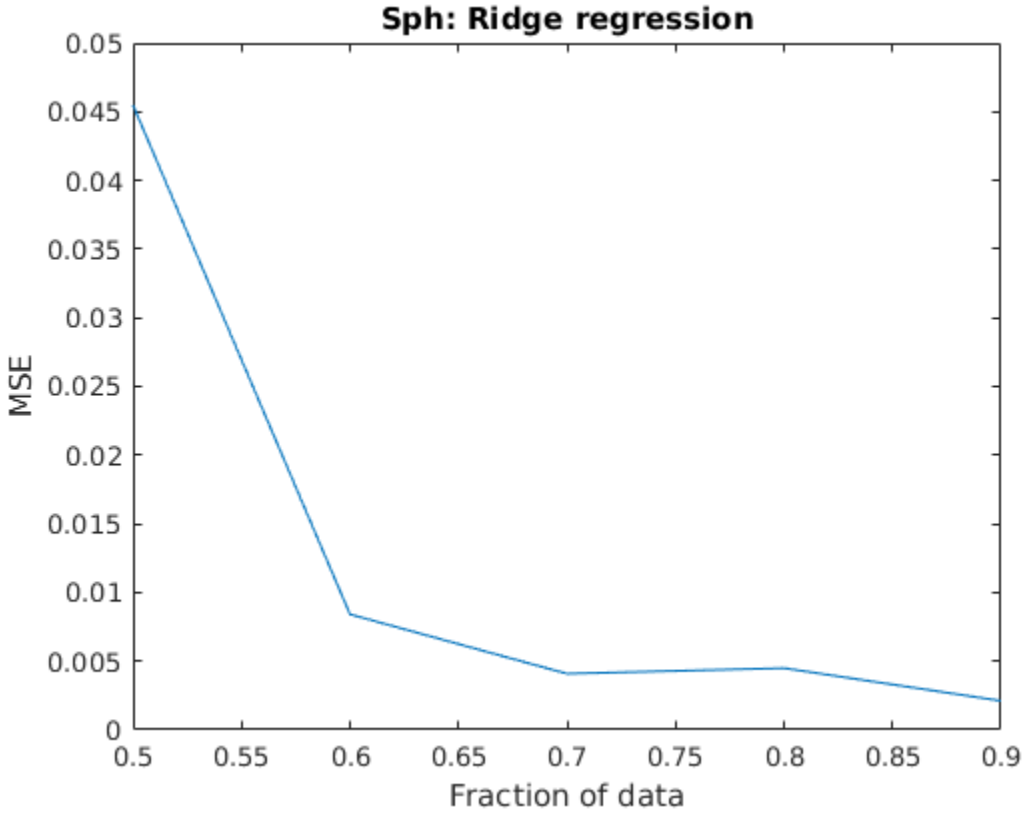
	80 % training data + 20 % test data			90 % training data + 10 % test data		
	Train Error	Test Error	Avg Error	Train Error	Test Error	Avg Error
0.1	0.030096	0.002553	0.016325	0.031361	0.001729	0.016545
0.5	0.145685	0.005339	0.075512	0.153389	0.001715	0.077552
1	0.268837	0.003753	0.136295	0.293244	0.002082	0.147663
10	2.124107	0.024976	1.074542	2.301376	0.006795	1.154086
50	2.589090	0.020900	1.304995	13.872964	0.406893	7.139928

For 90% training example the test error seems to be minimized for delta = 0.5. It performs not so great compared to delta = 0.1(best avg error) but that is majorly due to poor performance in train error. Therefore, delta = 0.5

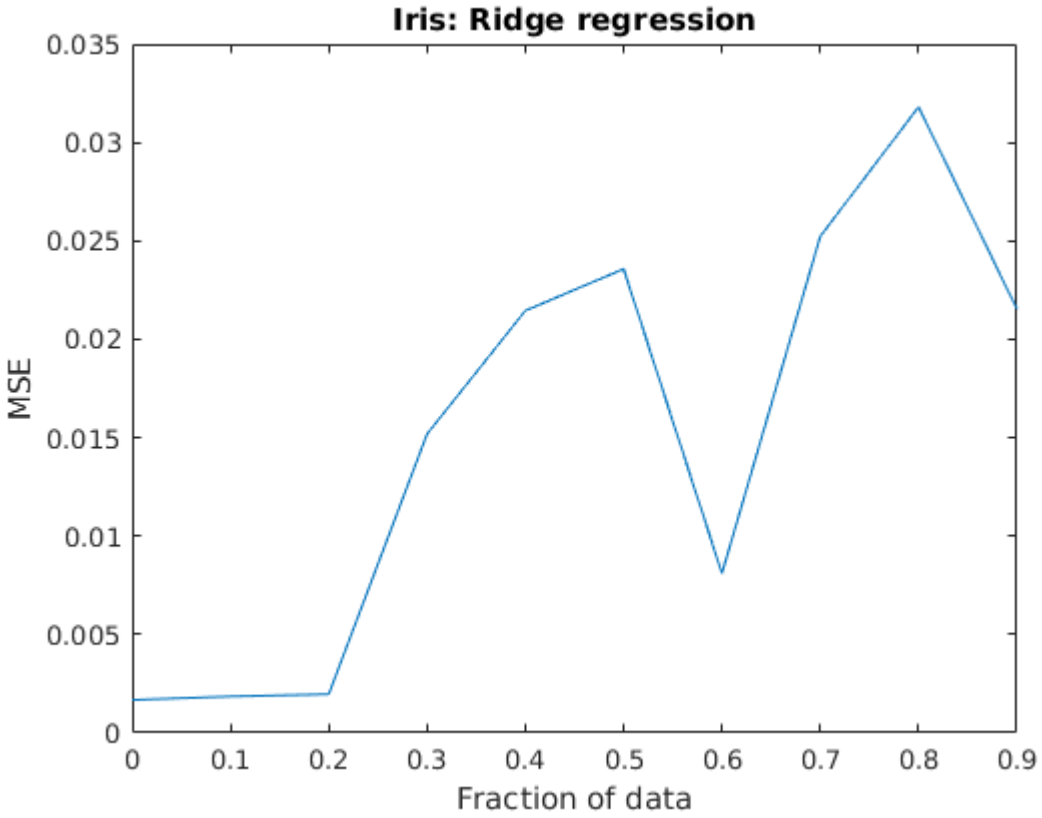

Lin: Ridge regression
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1


Lin: Ridge regression
MSE
0.015
0.014
0.013
0.012
0.011
0.01
0.009
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
Fraction of data

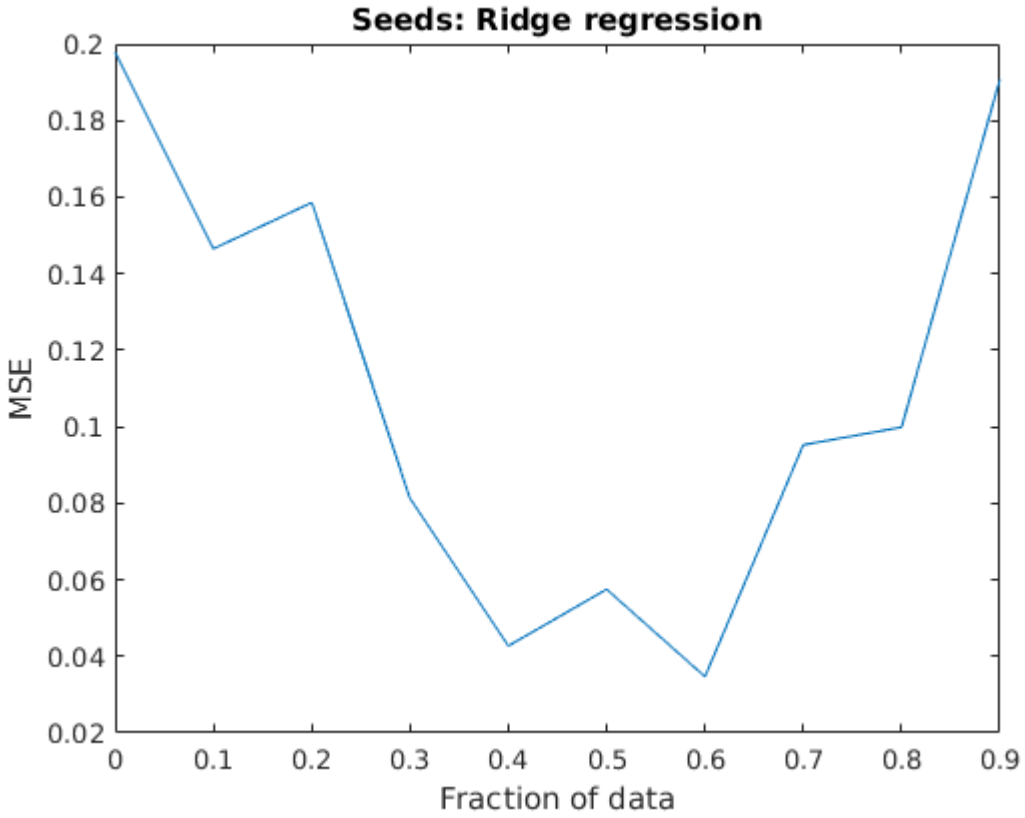

Sph: Ridge regression
1
0.8
0.6
0.4
0.2
0
-0.2
0
0.2
0.4
0.6
0.8
1


Sph: Ridge regression
MSE
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
Fraction of data

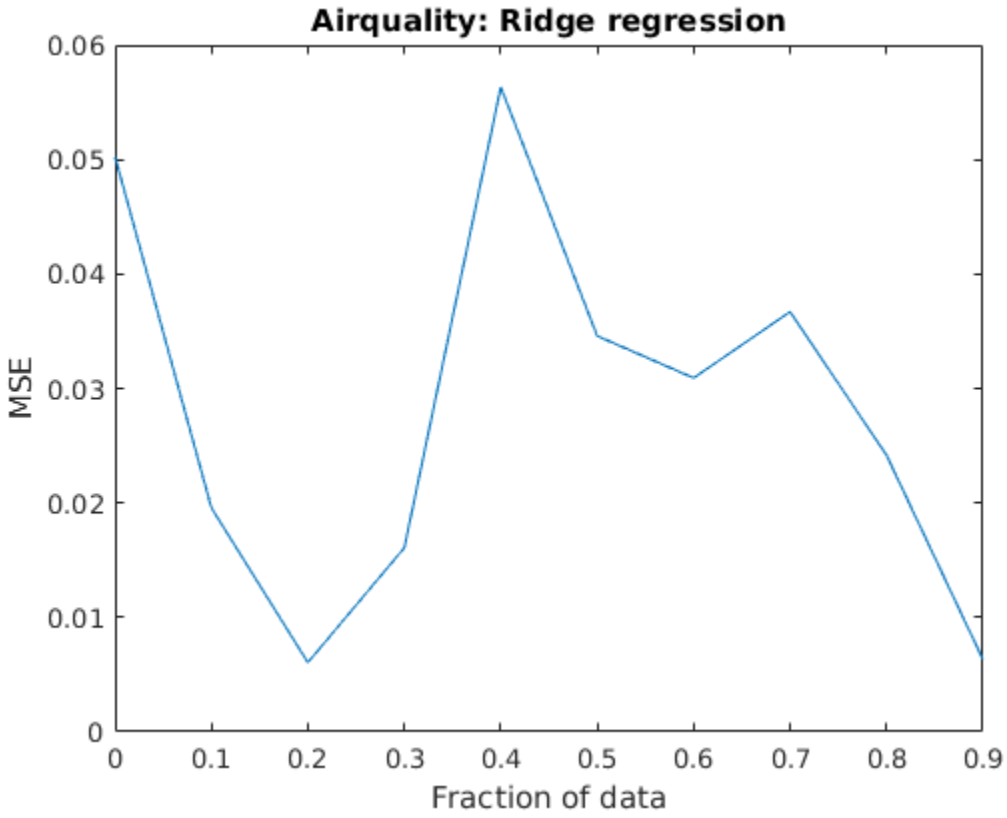
Ans 4. e. Error plots: (Fraction of data represents training data from 'fraction of data' to 'fraction of data'+10% data size)



Mean: 0.015237 Std deviation: 0.011118

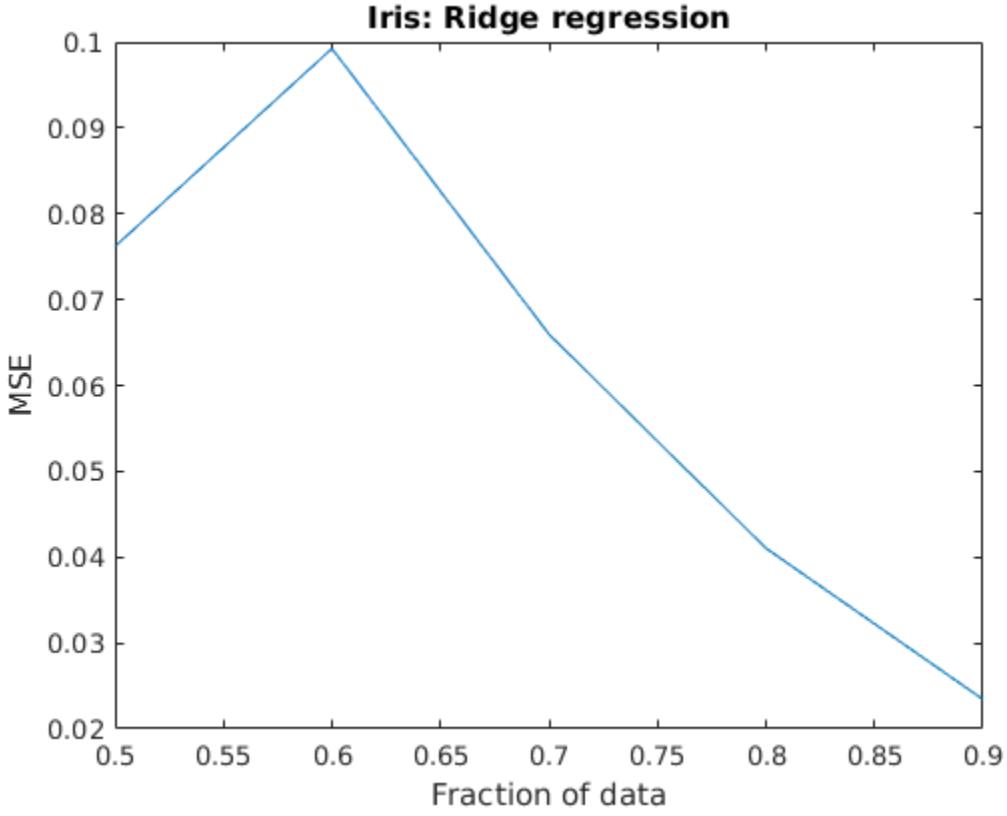


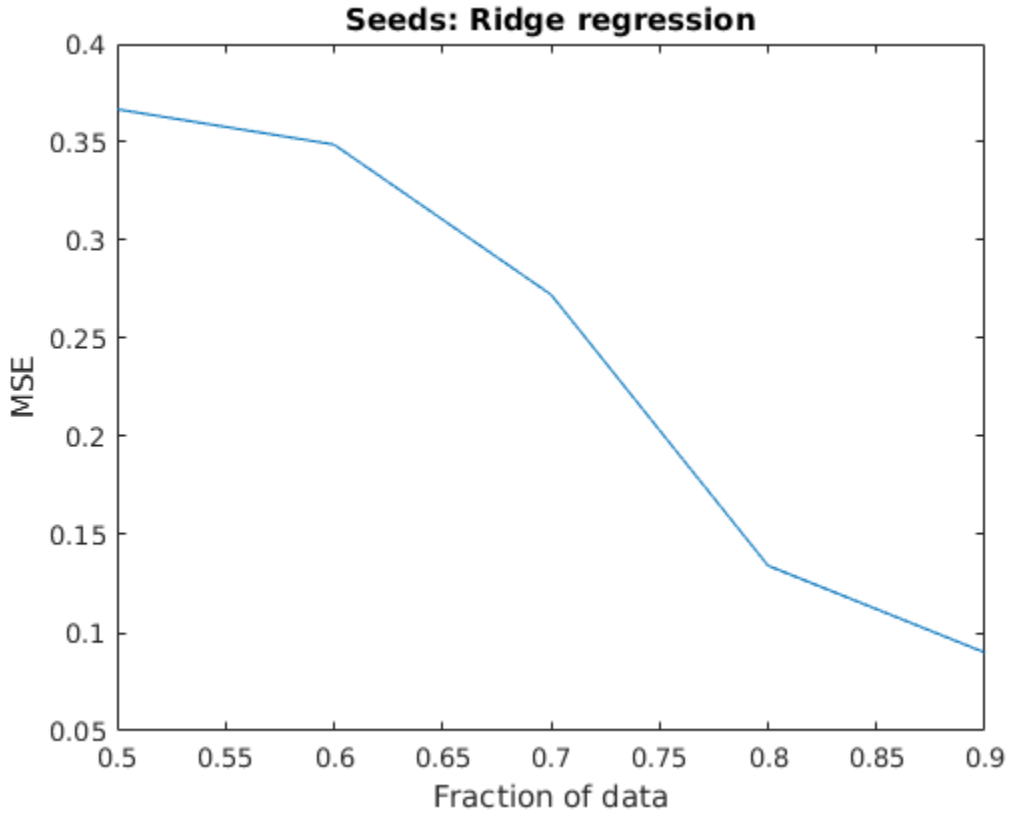
Mean: 0.110497 Std deviation: 0.059651

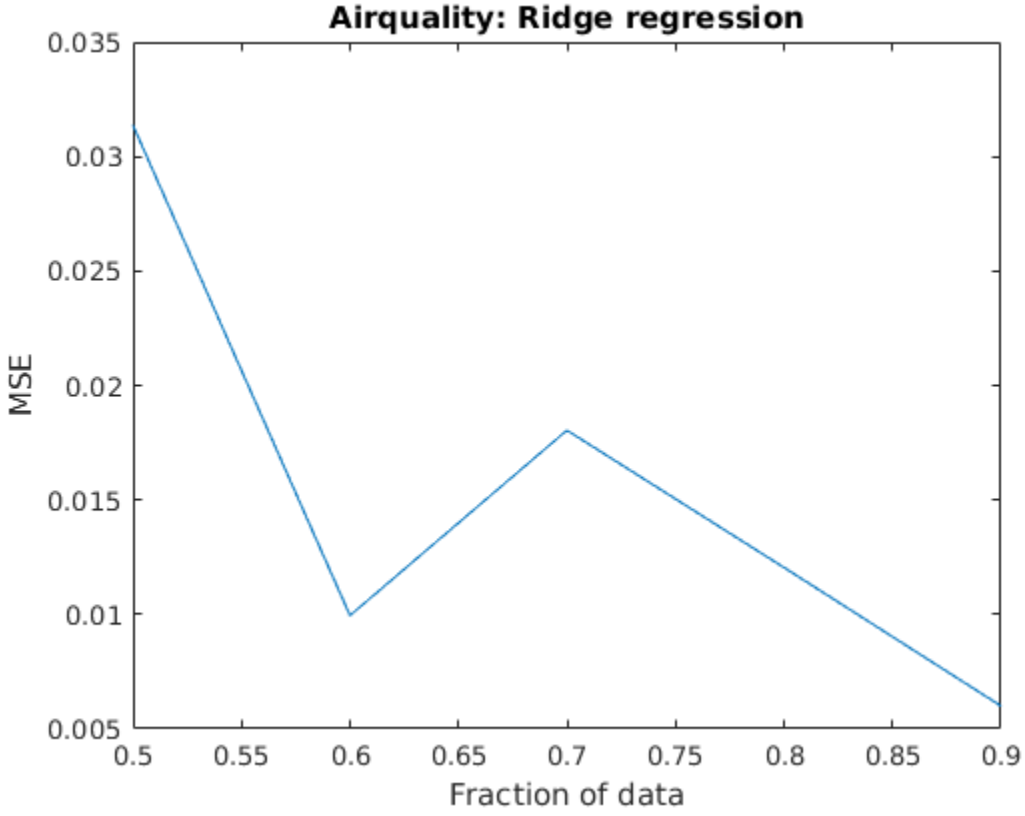


Mean: 0.028098 Std deviation: 0.017013

MSE vs fraction of data plots:




Seeds: Ridge regression
MSE
Fraction of data
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9


Airquality: Ridge regression
MSE
Fraction of data
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9