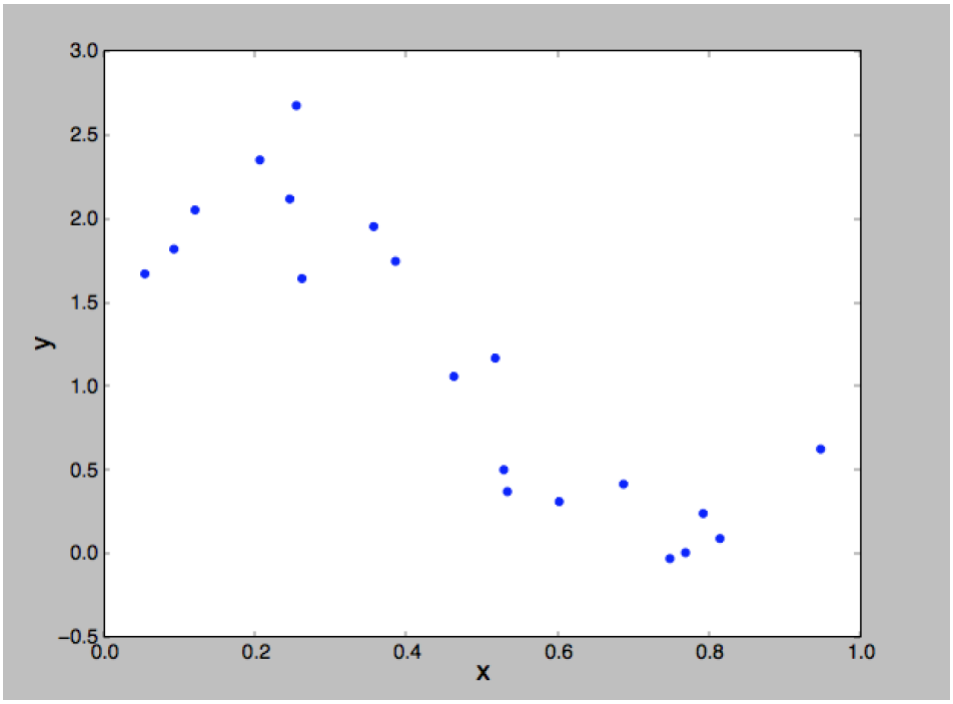
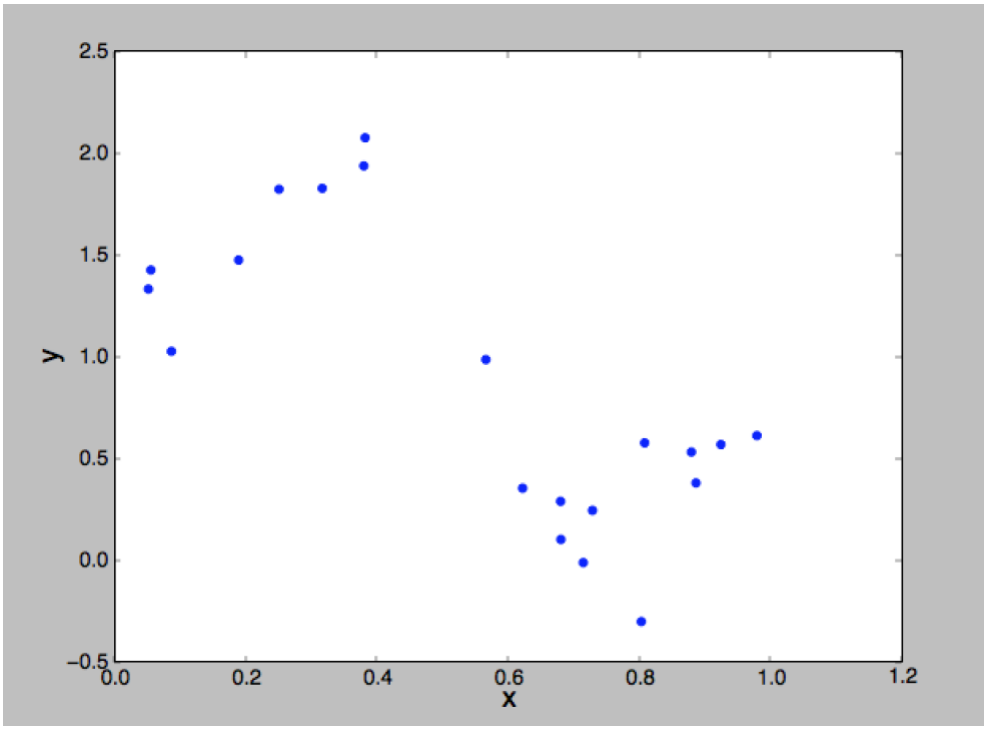
Problem 5:

a).

Train set visualized data:



Test set visualized data:



From the two plots, we can see that the data set could be linear separate easily. Although there are two "noisy" data in the middle of Train set plot, we could still separate the data set into positive and negative groups. Therefore, linear regression would be effective in predicting the data.

d).

η	Coefficient	Cost	Iterations	Time(s)
0.0001	[1.91573585 -1.74358989]	5. 4935655887	10000	0. 670198
0.001	[2.4463815 -2.81630184]	3.91257640947	10000	0.566816
0.01	[2.44640699 -2.81635338]	3.91257640579	1490	0.075016
0.0407	[2.44640706 -2.81635352]	3.91257640579	383	0.020842

Besides when η = 0.0001 which is not converge, the coefficient of rests is almost the same. η = 0.0001 and η = 0.001 reached the limit of 10000 iterations, and they took about the same time. η = 0.01 and η = 0.0407 took much less iterations and time.

e)

For closed-form, the resulting coefficient is: [2.44640709 -2.81635359], and the cost is: 3.91257640579, which is about same as those get from GD. The time spent is: 0.00278s, which is much faster than GD. However, the lower time cost on closed-form solution is because of the small size of the data set. If we run both methods on a very large data set, eventually GD would be faster than closed-form solution.

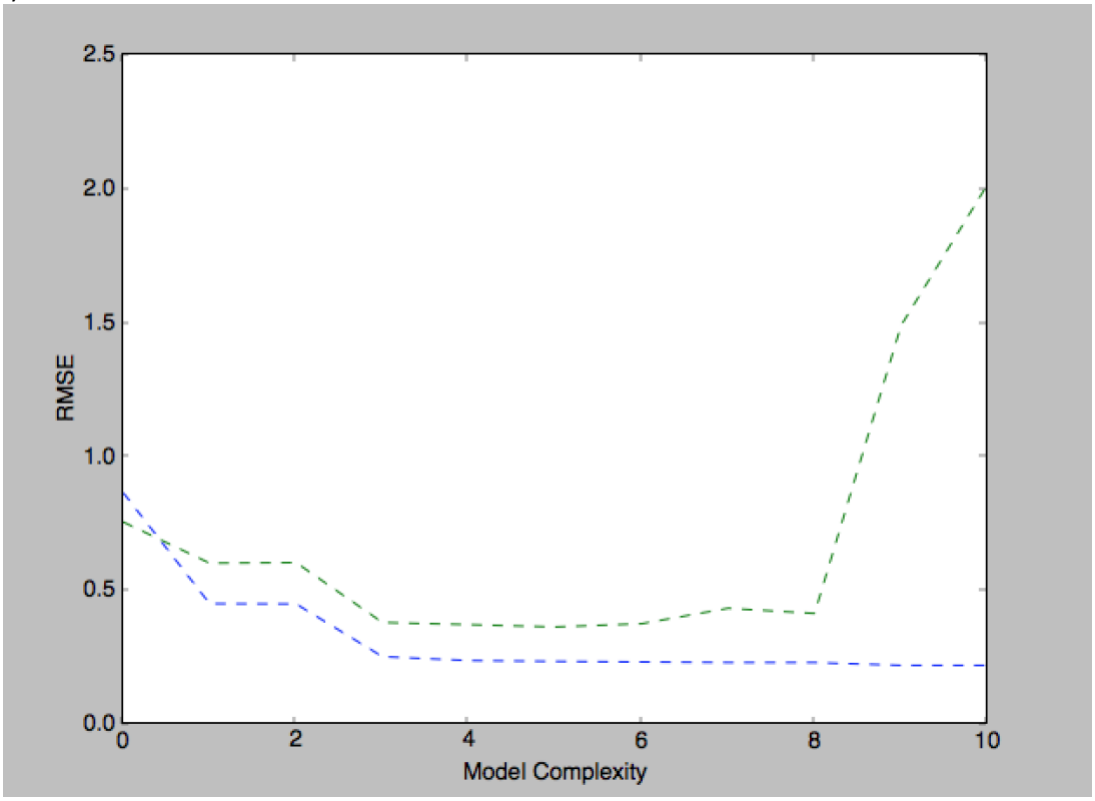
f)

The iteration is: 10000, time spent is: 0.6054s, and the cost is: 3.91257642432, which used up the limit of iterations. The coefficient I got is: [2.44634965 -2.81623746].

h).

Root-Mean-Square error (RMSE) represents the sample standard deviation of the differences between real data and predicted data, RMSE would normalizes the errors. J(θ) only measures the magnitude which doesn't match. Therefore, RMSE is preferable to measure overfitting problem since it doesn't rely on not-matched data only.

i).



*Green line indicates test error; Blue line indicates train error.

From the graph, we could know that degree 4 and 5 best fit the data since the error rate is the lowest and most steady. Test error after degree 6 is increasing very fast since overfitting happens.