QUESTION 1]

The learned values of the parameter are:
Degree 1: [0.54789989, -1.70307608]
Degree 2: [0.47690768, -1.04746483, -0.75935132]
Degree 3: [0.45415276, -1.01097588, -0.66036091, -0.11237551]
Degree 4: [0.45927817, -1.06234778, -0.70725285, -0.15171574, 0.2435228]
Degree 5: [0.47272368, -1.10598048, -0.77349855, -0.22425623, 0.1718566 , 0.4115202]
Degree 6: [0.48635647, -1.13086796, -0.82869089, -0.2922155 , 0.10046612, 0.34151806, 0.47451418]
Degree 7: [0.49768493, -1.14183429, -0.86897459, -0.34687064, 0.04034634, 0.28088426, 0.41603021, 0.48256338]
Degree 8: [0.5063993 , -1.1445225 , -0.89695358, -0.38844574, -0.00722334, 0.23176984, 0.36789021, 0.4367719 , 0.46335942]
Degree 9: [0.51287717, -1.14285722, -0.91600046, -0.41938998, -0.04389845, 0.19312394, 0.32948363, 0.39986513, 0.42863347, 0.43175918]

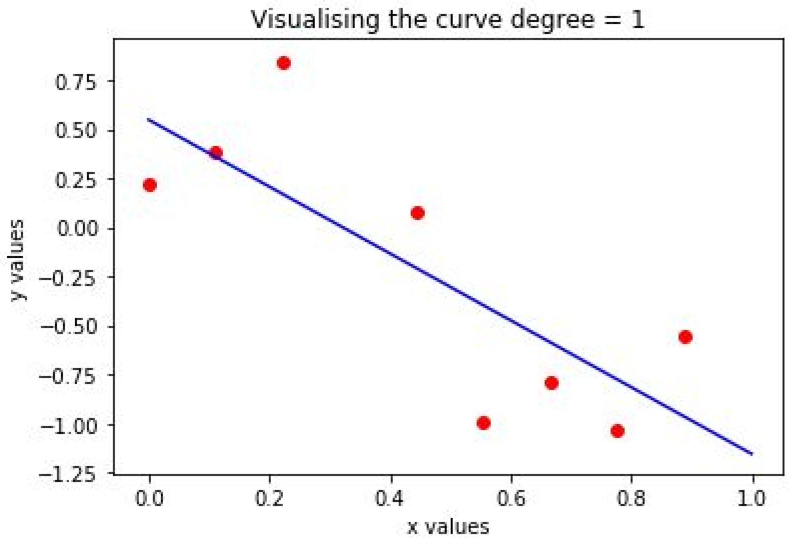
Note that the list contains the coefficients of increasing powers of x (starting from x^0) till x^n (n being degree)

Test error on test data: [0.6698558234784091, 0.726034776438033, 0.7281316135990823, 0.6751387695655584, 0.5971938122659131, 0.5198100069642657, 0.45705225679632755, 0.41473230085382384, 0.39401988622718476]

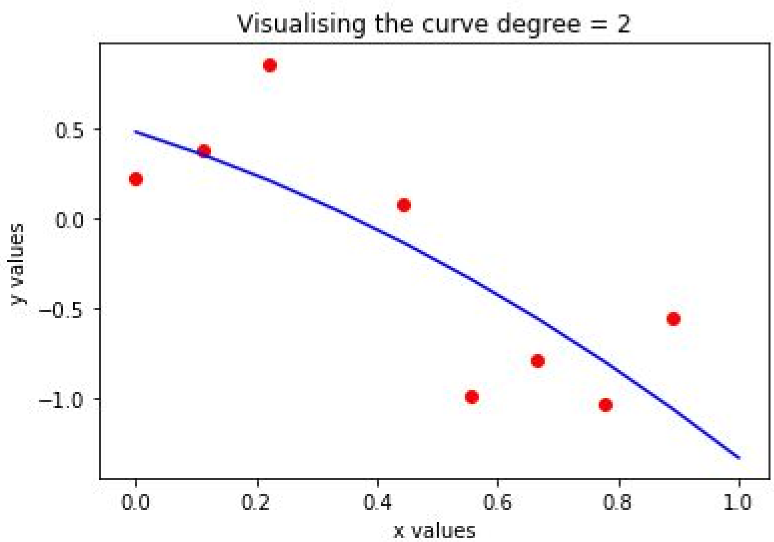
The list contains the test error for test data in increasing degrees of polynomial (from degree 1 to 9)

QUESTION 2]

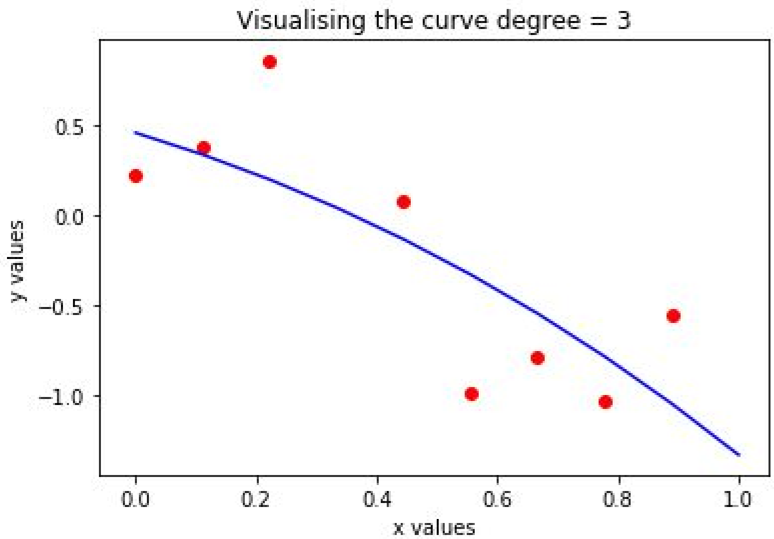
(a)

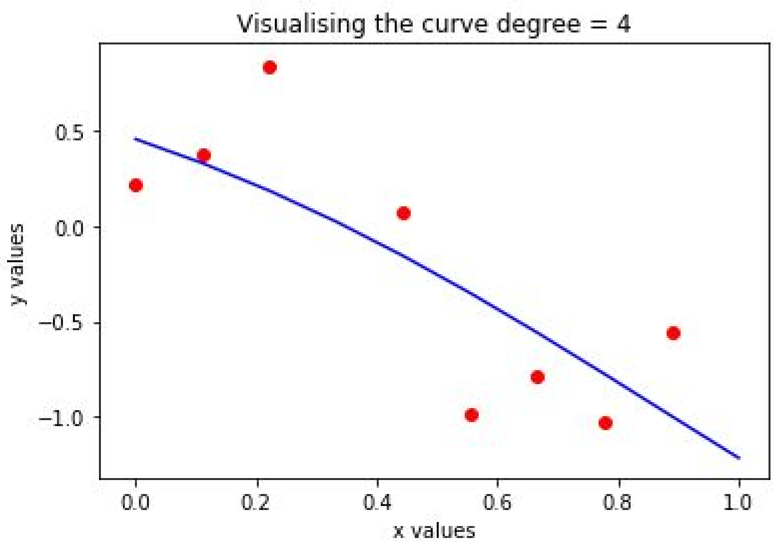


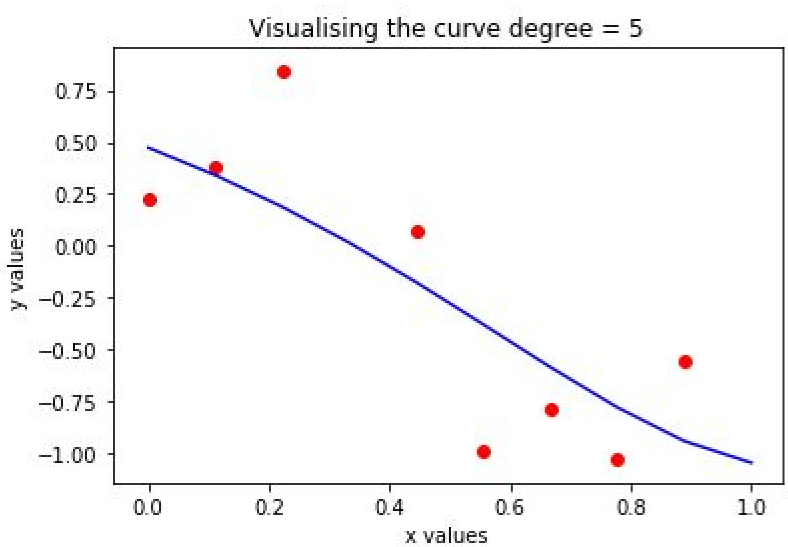
Visualising the curve degree = 1
y values
x values
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
0.0
0.2
0.4
0.6
0.8
1.0

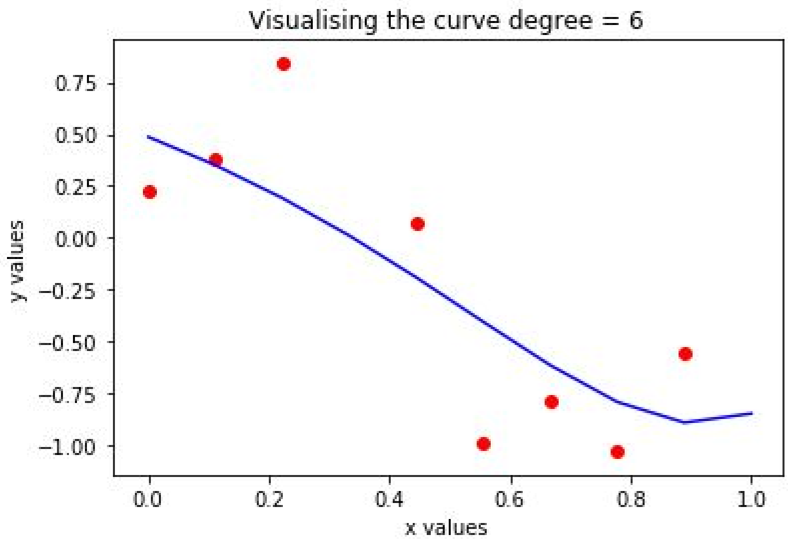


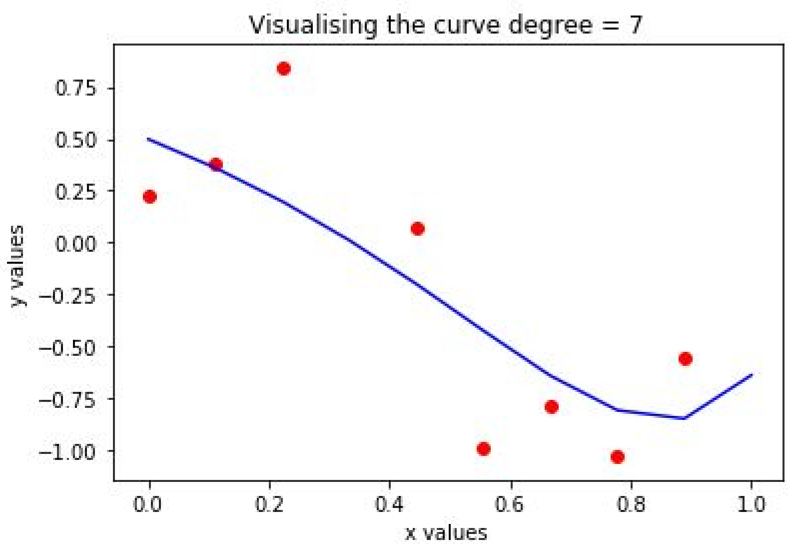
Visualising the curve degree = 2
y values
x values
0.5
0.0
−0.5
−1.0
0.0
0.2
0.4
0.6
0.8
1.0

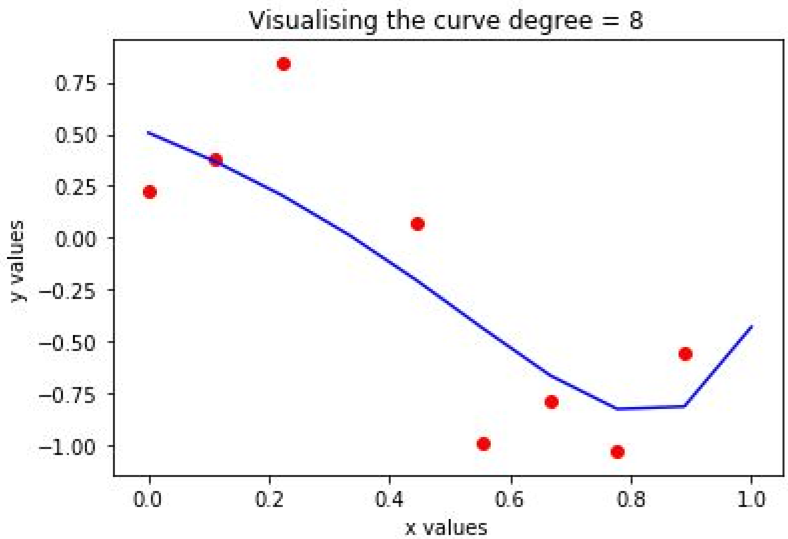

Visualising the curve degree = 3
y values
0.5
0.0
−0.5
−1.0
0.0
0.2
0.4
0.6
0.8
1.0
x values

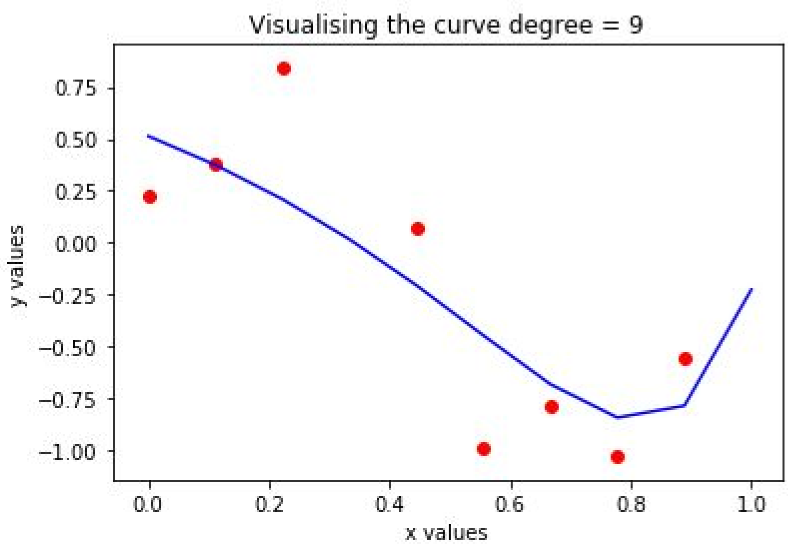

Visualising the curve degree = 4
y values
0.5
0.0
−0.5
−1.0
0.0
0.2
0.4
0.6
0.8
1.0
x values


Visualising the curve degree = 5
y values
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.0
0.2
0.4
0.6
0.8
1.0
x values

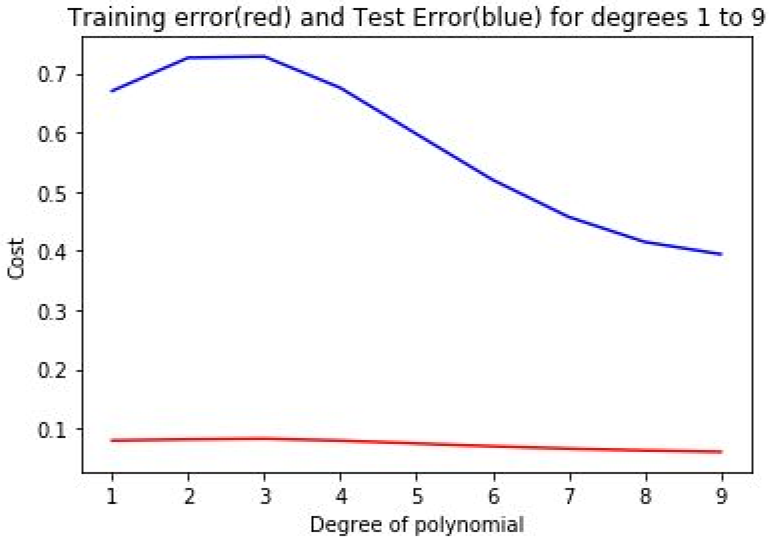

Visualising the curve degree = 6
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
y values
0.0
0.2
0.4
0.6
0.8
1.0
x values


Visualising the curve degree = 7
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
y values
0.0
0.2
0.4
0.6
0.8
1.0
x values


Visualising the curve degree = 8
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
y values
0.0
0.2
0.4
0.6
0.8
1.0
x values


Visualising the curve degree = 9
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
y values
0.0
0.2
0.4
0.6
0.8
1.0
x values

(b)



Training error: [0.0792368361699422, 0.08112281517130852, 0.0822371610443674, 0.07905110922133529, 0.07416347976166085, 0.06941861595398036, 0.06547441477843075, 0.062406076083692086, 0.06008879829921569]

Test error: [0.6698558234784091, 0.726034776438033, 0.7281316135990823, 0.6751387695655584, 0.5971938122659131, 0.5198100069642657, 0.45705225679632755, 0.41473230085382384, 0.39401988622718476]

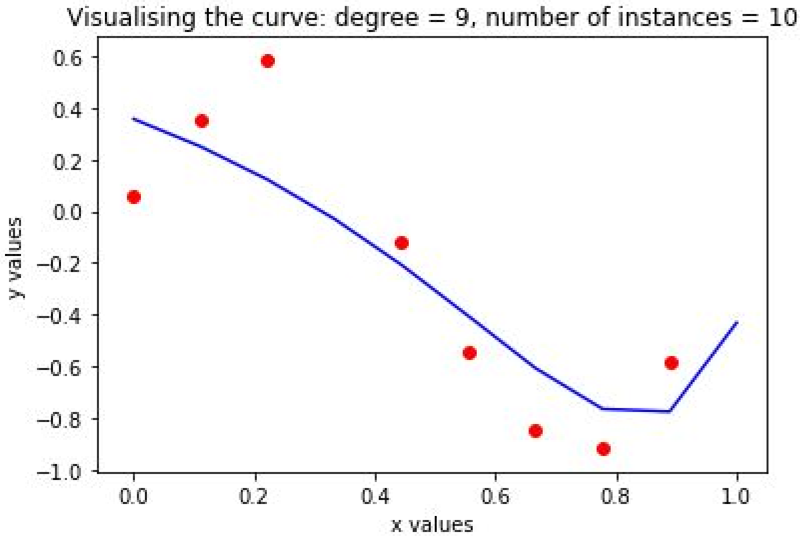
Both lists are arranged for increasing degree of polynomial with first element being for degree 1

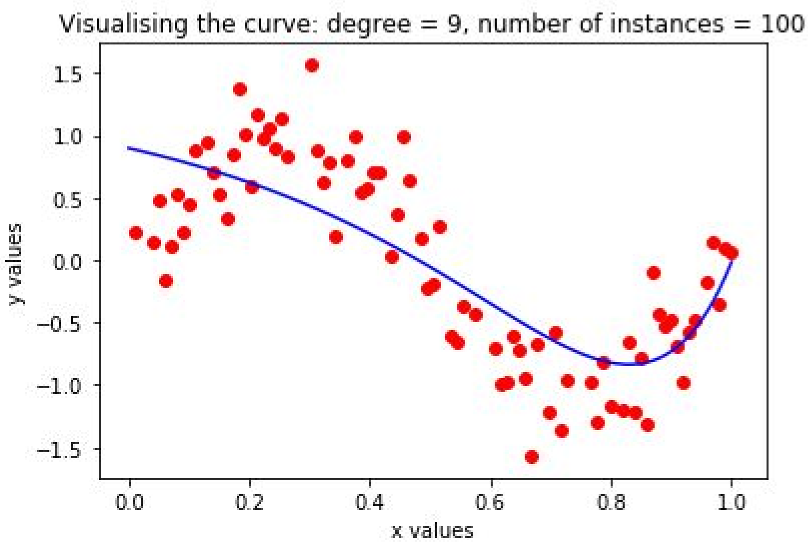
The value of n which is suitable in this case is n=9. This is because for n=9 both training error and test error are minimum which means that it fits the training set well and also generalises to new data (test data) relatively well.
This could be because the synthetic data set so generated would be too complex to be approximated by smaller degree polynomials.

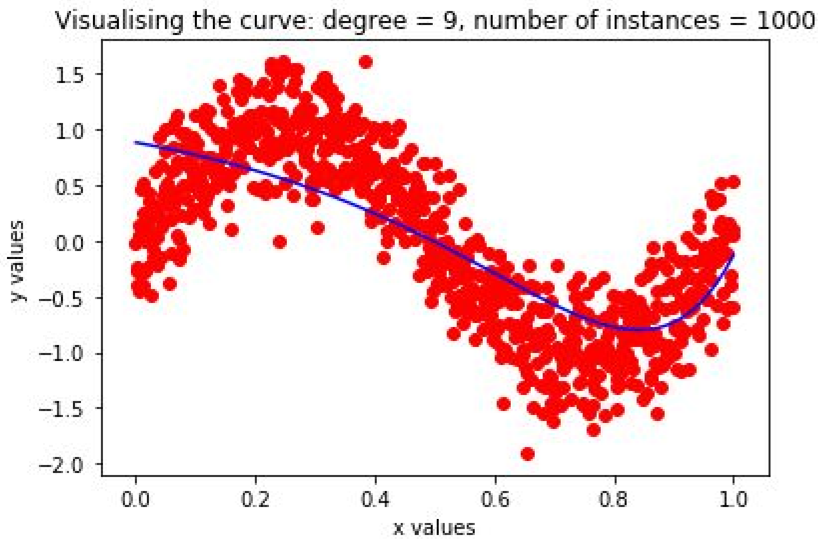
The model parameters for this part would be the same as Part 1 because nothing new has been processed on the data. The data so obtained is only analysed in this part i.e Part 2.

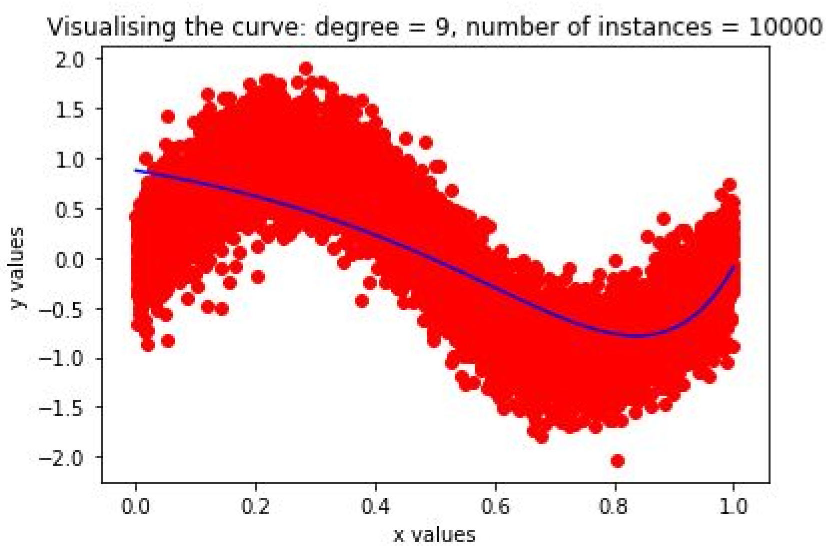
QUESTION 3]

Degree 9 has been chosen to compare between effect of dataset size as degree 9 yielded best results

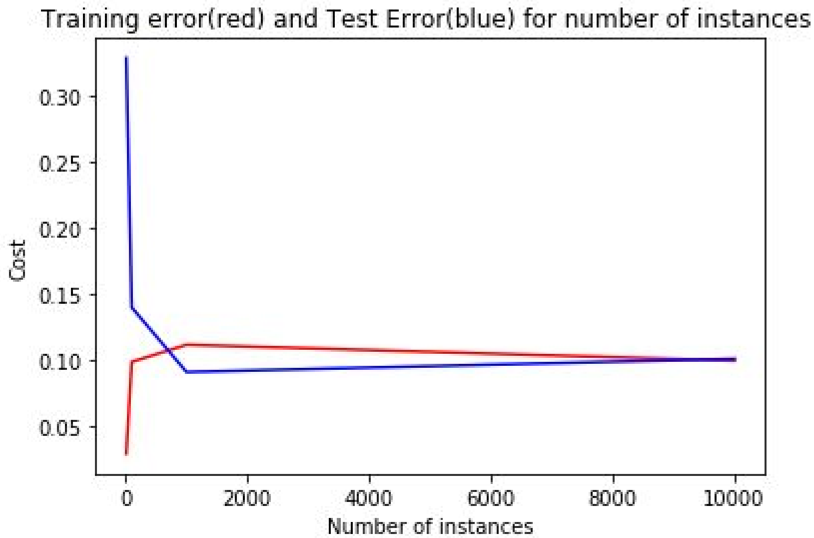





Visualising the curve: degree = 9, number of instances = 1000
y values
x values


Visualising the curve: degree = 9, number of instances = 10000
y values
x values

Required Learning Curve:



These are the parameter values when the dataset size are 10, 100, 1000, 10000 respectively; Note that parameter values are arranged in increasing order of powers of x (starting from x^0 to x^9)

Number of instances = 10 :[0.35667659, -0.87158012, -0.74502146, -0.3697498 , -0.07750751, 0.11275379, 0.22691775, 0.28975138, 0.31912356, 0.3271066]

Number of instances = 100 :[0.89619669, -1.09861538, -1.21914389, -0.77106 , -0.29909476, 0.07827421, 0.35649498, 0.55510372, 0.69453981, 0.7911679]

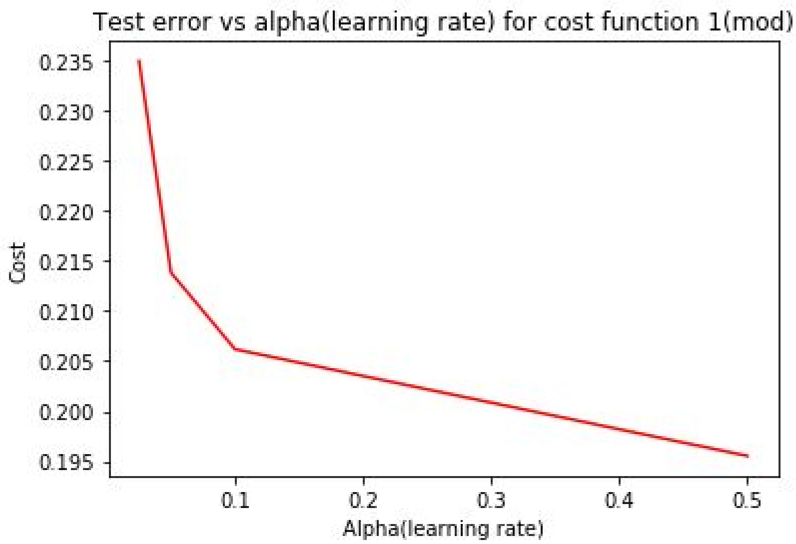
Number of instances = 1000 :[0.88496608, -0.98898993, -1.18535018, -0.7672541 , -0.31329244, 0.05097302, 0.31857793, 0.50798818, 0.63907869, 0.72788765]

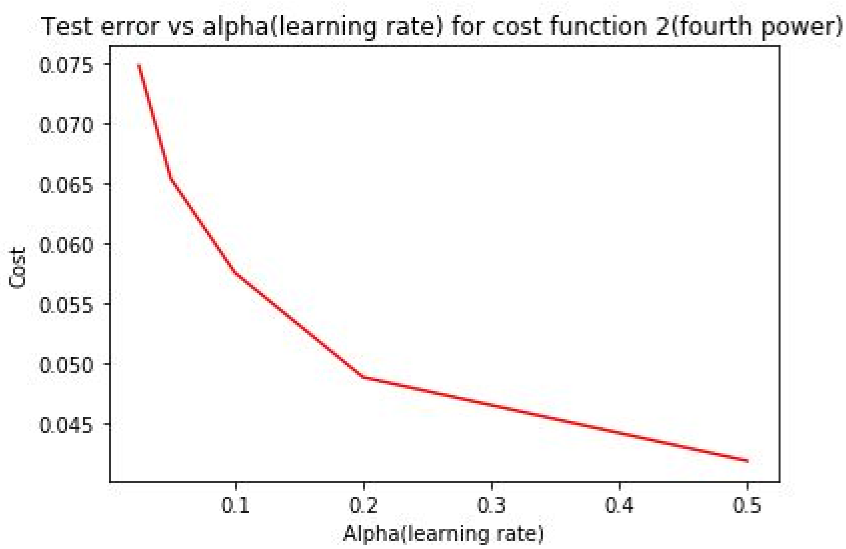
Number of instances = 10000 :[0.87542852, -1.02711968, -1.16140113, -0.73009049, -0.28216541, 0.07010329, 0.32506326, 0.5030755 , 0.62453546, 0.70547973]])]

QUESTION 4]
Degree 9 and number of instances = 100 have been chosen to compare between effect of learning rates.
(a)
The model parameters for cost function 1(mod) are:
Alpha = 0.025
[0.7146875 , -0.72366477, -0.74686451, -0.4963263 , -0.2477525 ,
-0.04971015, 0.09902366, 0.20861223, 0.2888468 , 0.34742804]
Alpha = 0.05
[0.9784375 , -1.11665088, -1.27631935, -0.79816889, -0.30367592,
0.07946301, 0.35114537, 0.53583914, 0.65746144, 0.73459557]
Alpha = 0.1
[0.9746875 , -0.92019886, -1.57609195, -1.01632579, -0.34580206,
0.15974368, 0.48346634, 0.66550801, 0.74854064, 0.76604465]
Alpha = 0.2
[0.9096875 , -0.42817866, -2.00860292, -1.31619906, -0.3627392 ,
0.31301656, 0.67370422, 0.79433034, 0.75483906, 0.61704163]
Alpha = 0.5
[0.7909375 , 0.49454861, -3.02261679, -1.90103472, -0.25976139,
0.78975536, 1.1886586 , 1.10833569, 0.72391516, 0.16757242]])]

The model parameters for cost function 2(fourth power) are:
Alpha = 0.025
[0.65574606, -0.71776733, -0.76119016, -0.50073086, -0.24146282,
-0.04029951, 0.10449835, 0.20552713, 0.27475653, 0.32141245]
Alpha = 0.05
[0.76974637, -0.79679115, -1.10025043, -0.73101757, -0.31002234,
0.02438828, 0.26323882, 0.42561615, 0.53220859, 0.59953095]
Alpha = 0.1
[0.74024778, -0.46434511, -1.42096777, -1.03533266, -0.45724706,
0.02131592, 0.36183591, 0.58695118, 0.72722807, 0.80824273]
Alpha = 0.2
[0.65531585, 0.12435592, -1.86802871, -1.45443447, -0.64609602,
0.02676673, 0.48664468, 0.76733012, 0.91827641, 0.98109563]
Alpha = 0.5
[0.54198608, 0.93192587, -2.63062937, -1.98236093, -0.7291123 ,
0.22349634, 0.77490144, 1.00932511, 1.02790554, 0.91080056]])]

(b)





We would prefer alpha (learning rate) of 0.5 in both the cost functions because it yields the lowest cost relative to other costs obtained from smaller alphas. It means that alpha of 0.5 would be more close to the minimum of cost function than others.

This could possibly be because of alpha (learning rate) being too small implies that it would take longer time to reach the minimum. And in our case when we kept the number of iterations fixed at 1000, smaller alphas were not able to reach the minimum of cost function in the given number of iterations, hence giving greater cost function values.