EE6130 Machine Learning for Computer Vision
Assignment-2
(K-means, EM, Regression Algorithms))

Arulkumar S (CS15S023)

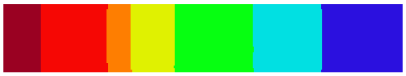
April 11, 2016

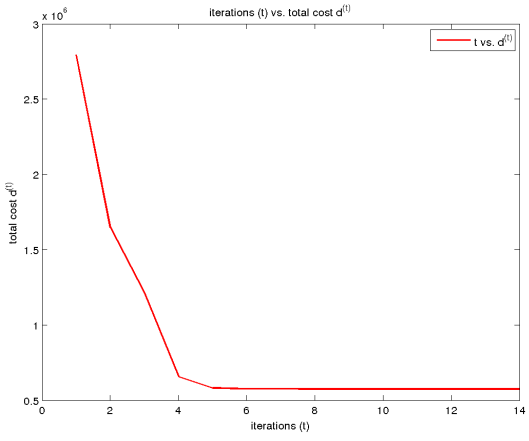
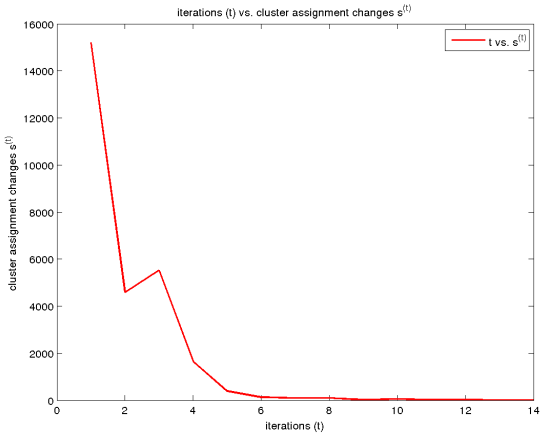
1 Image segmentation using K-Means clustering

1.1 segmentation of spectrum.png ($K = 7$)

1.1.1 Mean values from centre-row

The initial mean values are selected from pixels equally spaced in the horizontal direction in the centre row of the image.

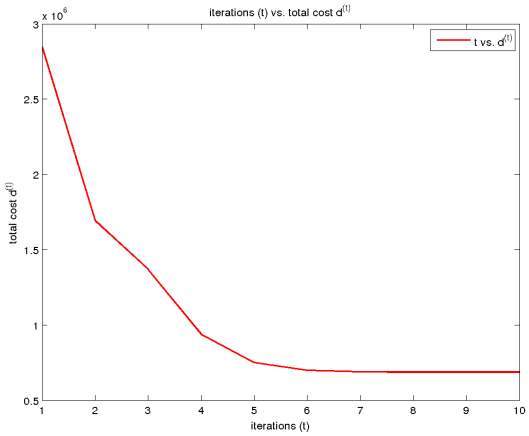
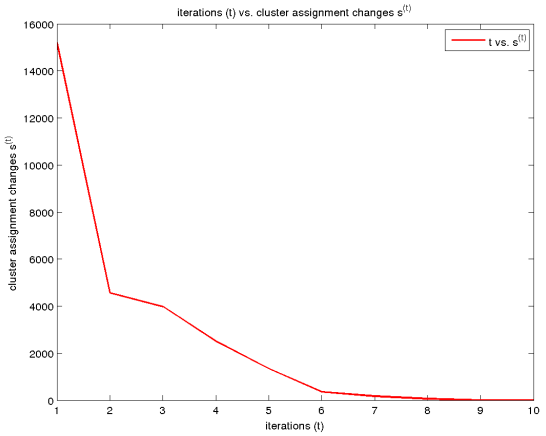


iteration vs. cost	iteration vs. cluster assignment changes
	

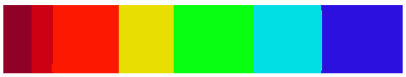
1.1.2 Random mean values

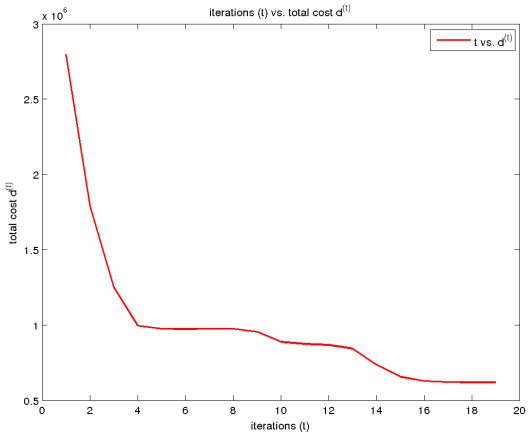
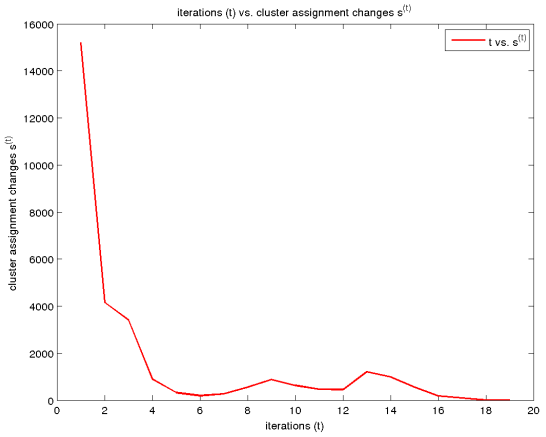
Here, the mean values are chosen randomly from the available pixel values and K-means clustering is carried out until convergence.

1.1.2.1 Random initialization-1

iteration vs. cost	iteration vs. cluster assignment changes
	

1.1.2.2 Random initialization-2



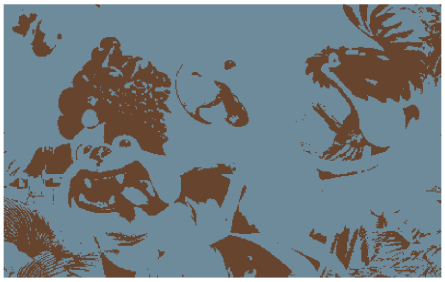
iteration vs. cost	iteration vs. cluster assignment changes
	

As we can see in the segmented images from K-means clustering (randomly initialized means),

- K-means clustering is able to find local minimum of cost function depending on the initialization of means
- Comparing "Random initialization-1" with "Random initialization-2", it can be inferred that the segmented images are different & it is not guaranteed that the segmented images will be same.
- In the K-means clustering process with the means initialized from centre row (equi-distant), since the means are initialized to cover the major colors in the image, the clustering result is always the same.

1.2 segmentation of rio.png

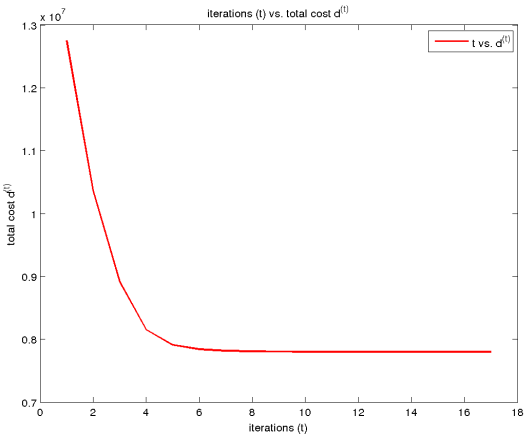
1.2.1 number of clusters = 2

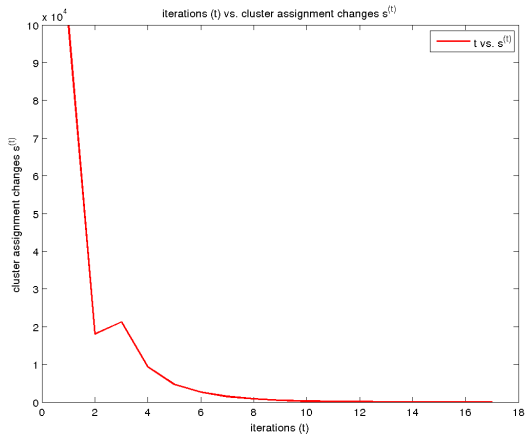


1.2.1.1 final segmented image

iteration vs. cost

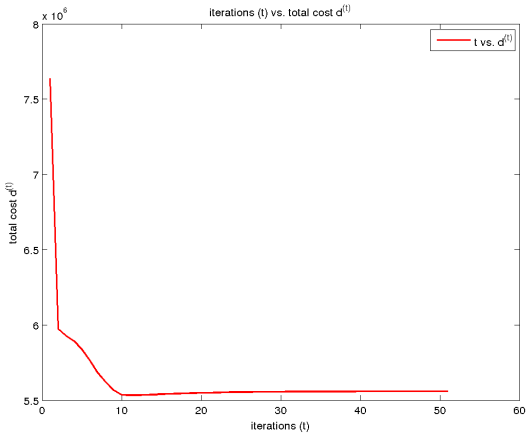
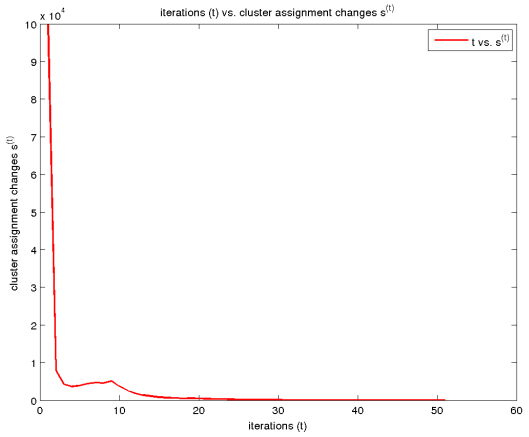
iteration vs. cluster assignment changes





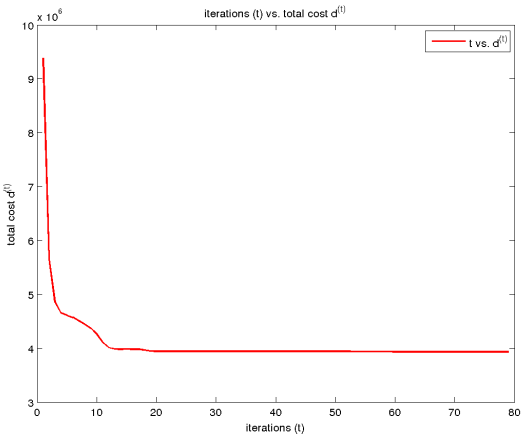
1.2.2 number of clusters = 4

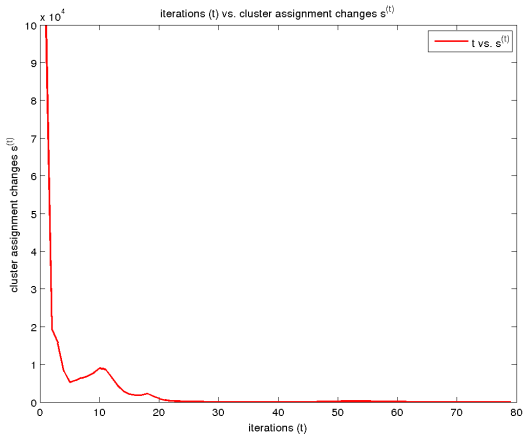
1.2.2.1 final segmented image

iteration vs. cost	iteration vs. cluster assignment changes
	

1.2.3 number of clusters = 8

iteration vs. cost	iteration vs. cluster assignment changes





1.2.4 K-means algorithm - general property

K-means clustering algorithm guarantees only a local minimum (the local minimum is dependent on the random initialization of means).

2 Image segmentation using Gaussian mixture model (EM algorithm)

2.1 spectrum.png with K=7

2.1.1 Random initializations

Random initialization-1	Random initialization-2

2.2 rio.png with K=8

2.2.1 Random initializations

Random initialization-1	Random initialization-2

2.3 Difference between K-means clustering & EM algorithms (e.g., GMM)

- K-means clustering algorithm follows Hard clustering policy in which a particular datapoint is assigned to a cluster completely, where as in EM algorithms such as Gaussian Mixture Model, any data point is assigned to multiple clusters (soft assignment) with a probabilistic weightage.
- K-means clustering algorithm tends to use less number of parameters than EM algorithms. Considering K-means and GMM, K-means algorithm uses K number of means, where as GMM

uses K number of means + K number of weight parameters, K number of dxd covariance matrices.

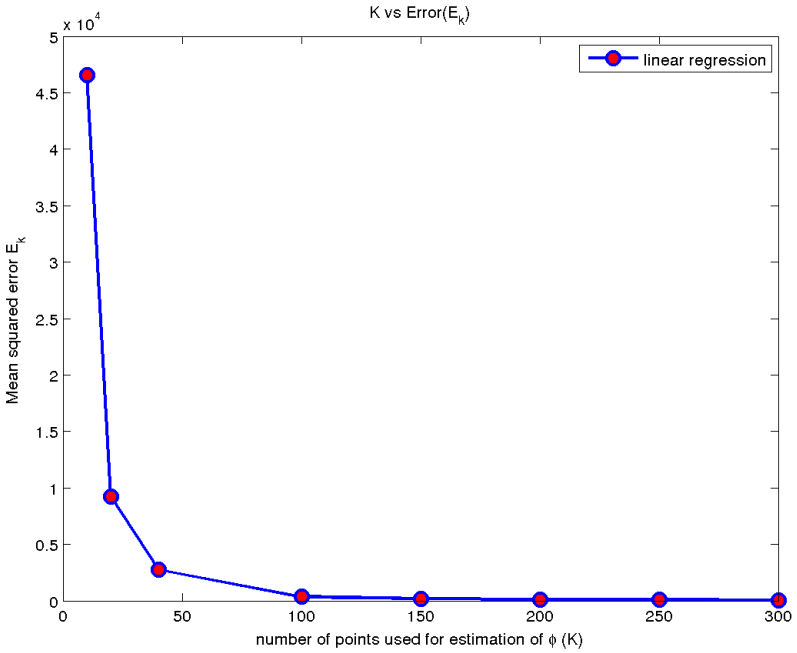
3 Linear regression

3.1 Maximum likelihood method

Let $X = [x_1 x_2 \dots x_I]$ be the data examples given. $(x_i = [1x'])$

we can define that, $\mathrm{P}(w_i|x_i) = N_w(\phi^T x_i, \sigma^2)$.

ML estimate of ϕ is given by, $\phi_{ML} = (XX^T)^{-1}Xw$



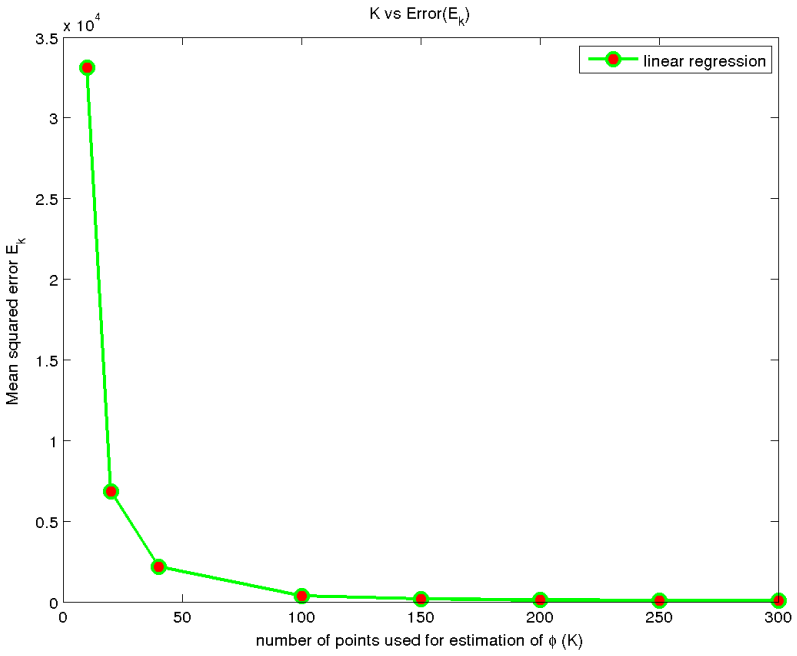
3.2 Bayesian regression method

Posterior distribution $\mathrm{P}(\phi|w, X) = \frac{P(w|X,\phi)P(\phi)}{P(w|X)}$

here, the likelihood is given by $P(w|X, \theta) = N_w(X^T\phi, \sigma^2)$.

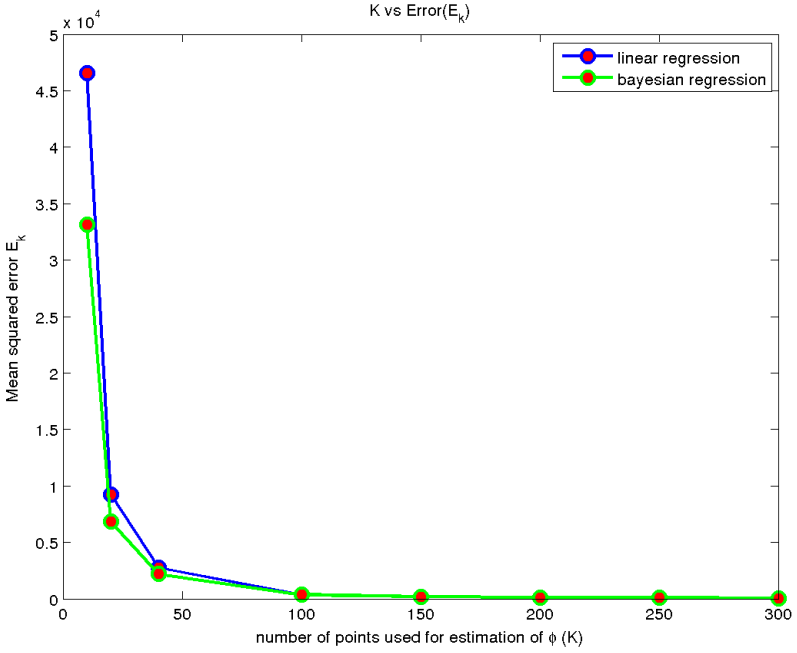
hence, $P(\phi|X, w) = N_\phi(\frac{A^{-1}Xw}{\sigma^2}, A^{-1})$, where A $= \frac{XX^T}{\sigma^2} + \frac{I}{\sigma_p^2}$

Predictive distribution is given by, $P(w^*|x^*, X, w) = N_{w^*}(\frac{x^{*T}A^{-1}Xw}{\sigma^2}, x^{*T}A^{-1}x^*)$



3.3 Comparision

As seen in the image below, the ML estimate is giving high MSE when less number of points are used in Regression. When the number of points are increased, the error of ML estimate converges with Bayesian estimate.

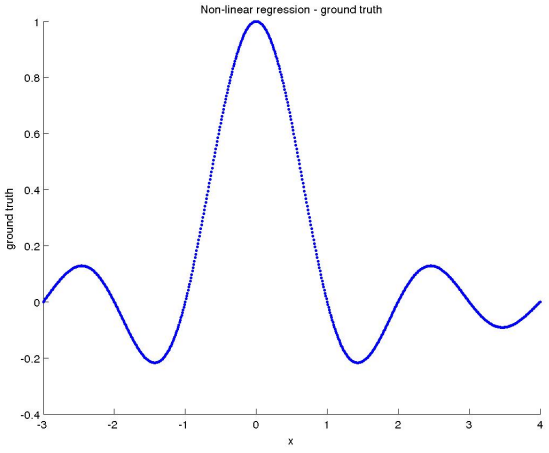
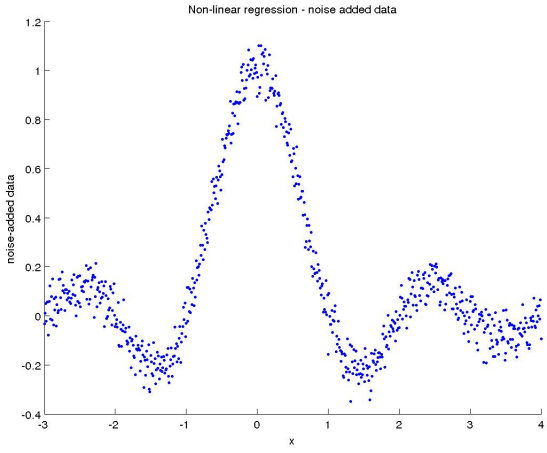


4 Non-linear regression

4.1 Data

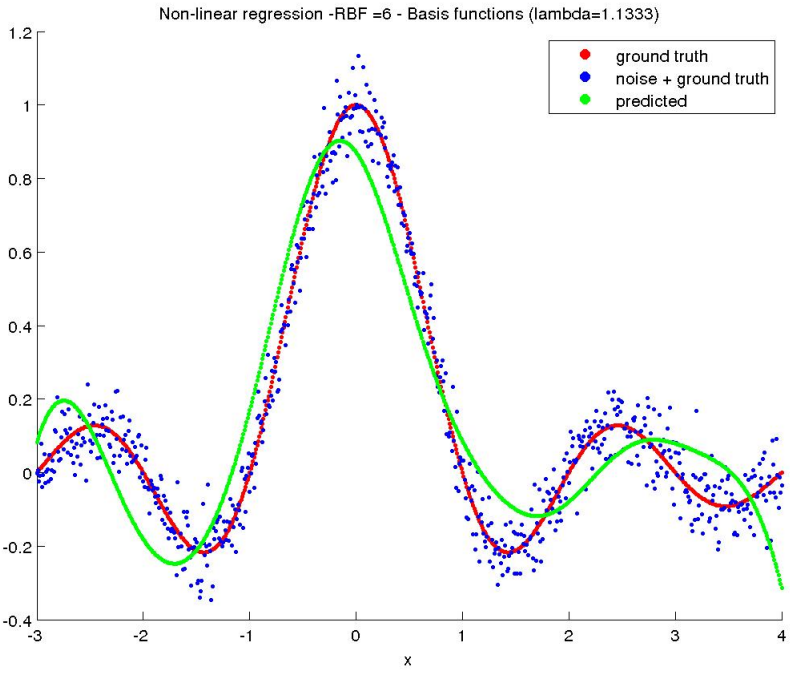
By looking at the data, it shall be inferred that the Linear regression methods (involving linear terms of Data x) will not be sufficient to model the data, as Data seems to involve non-linear characteristics.

To model the data well, we need to go for Non-linear regression where we involve non-linear terms $(x^2, x^3, \ldots, gaussian, arctan)$.

Ground truth	Noise-added data
	

4.2 RBF with 6 kernel functions

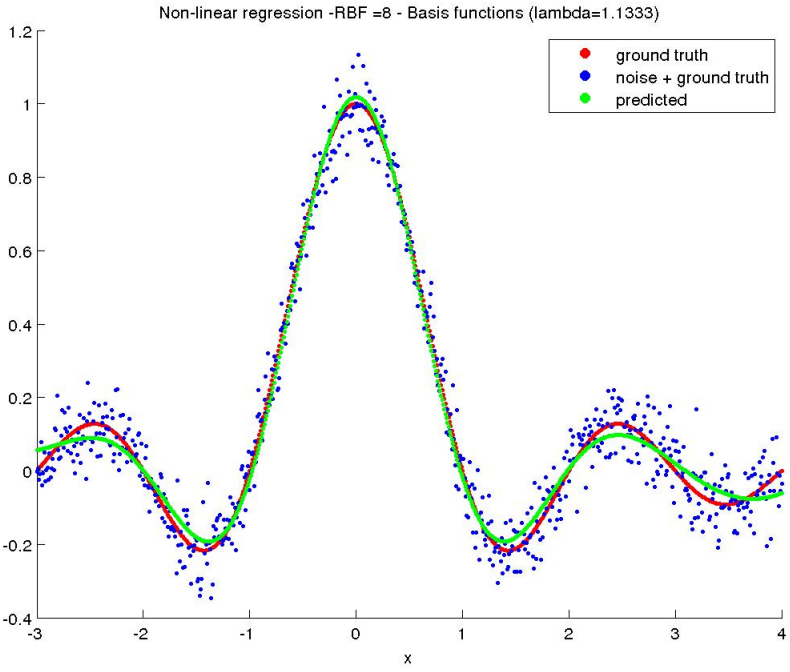
with 6 RBF kernel functions, the predicted curve is given below: (prediction MSE = 0.0133)



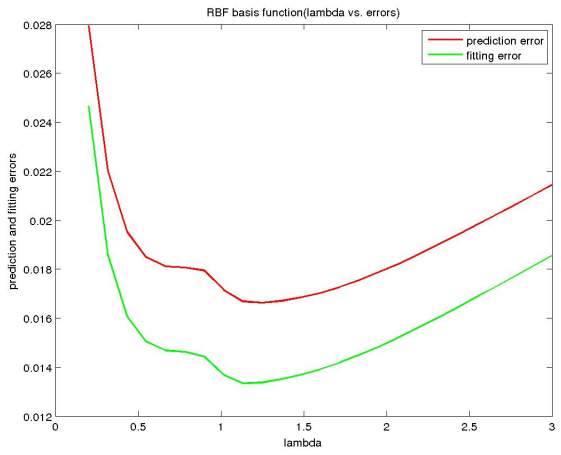
4.2.1 Error analysis

As shown in the error curves, the optimal value of λ is found to be 1.1333 and the optimal number of kernel functions is 8.
When the λ increases, the models tend to underfit the data and hence the prediction error increases.

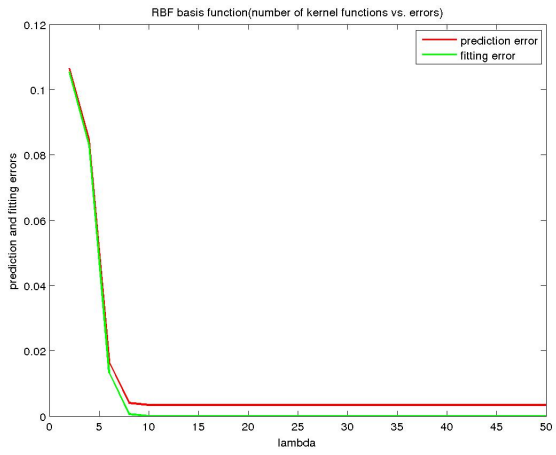
Also, when the number of kernel function increases, normally the model will result in overfitting. However, since the dataset contains less noise, the overfitting behavior is not affecting the models.



λ vs. Error

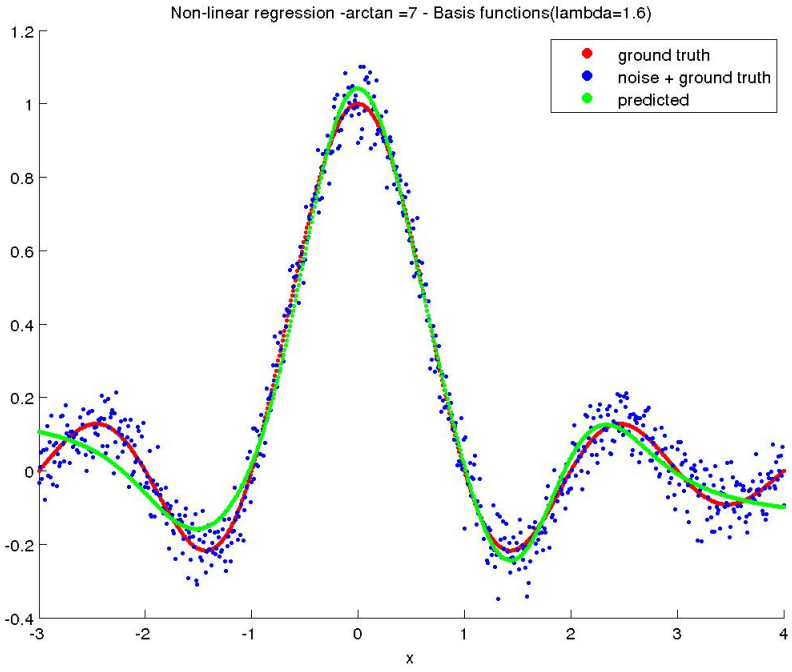


Number of Kernel functions vs. Error



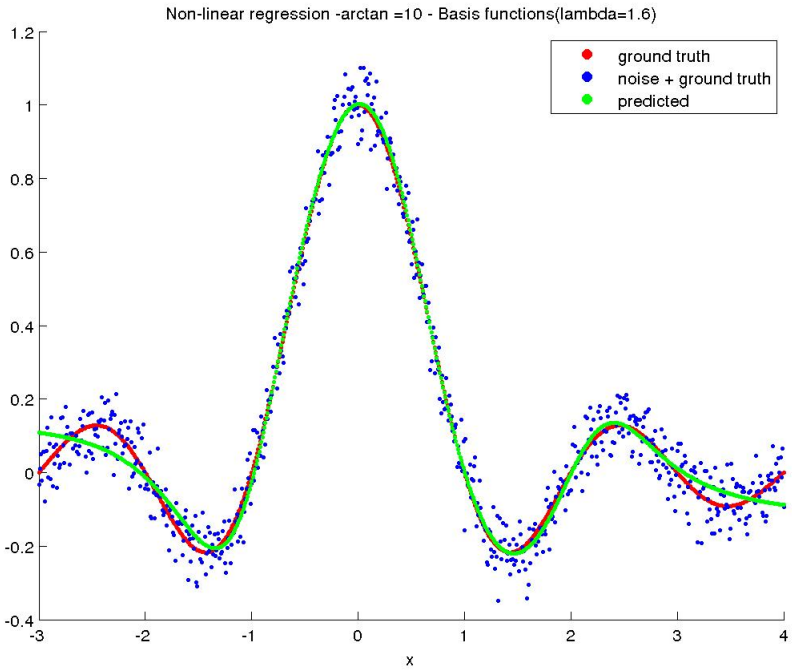
4.3 ARCTAN with 7 basis functions

With 7 arc tangent functions, the predicted curve is given below: (prediction MSE = 0.0016)



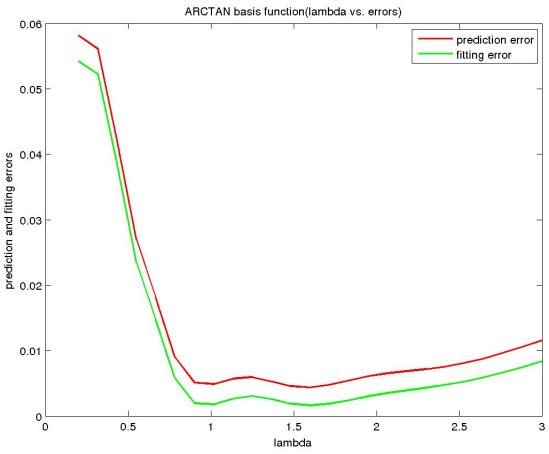
4.3.1 Error analysis

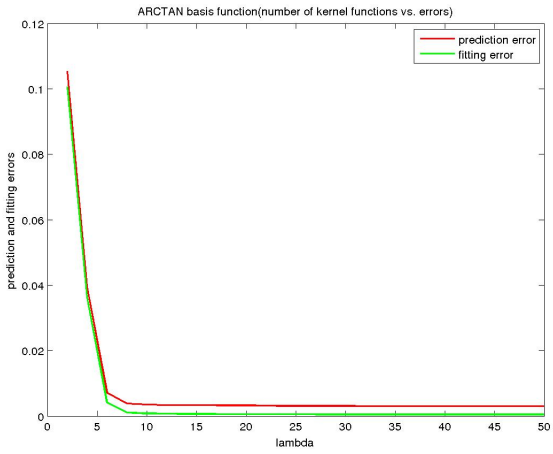
As shown in the error curves, the optimal value of λ is found to be 1.6 and the optimal number of kernel functions is 10.



λ vs. Error

Number of Kernel functions vs. Error

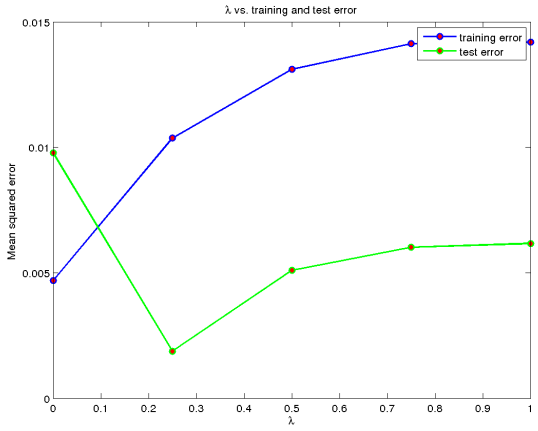

ARCTAN basis function(lambda vs. errors)
prediction error
fitting error
prediction and fitting errors
lambda
0
0.01
0.02
0.03
0.04
0.05
0.06
0.5
1
1.5
2
2.5
3


ARCTAN basis function(number of kernel functions vs. errors)
prediction error
fitting error
prediction and fitting errors
lambda
0
0.02
0.04
0.06
0.08
0.1
0.12
5
10
15
20
25
30
35
40
45
50

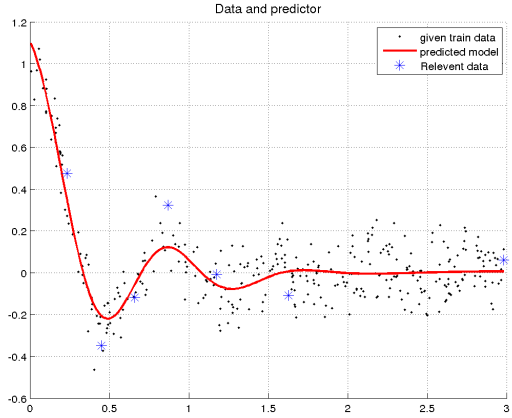
5 Relevance vector regression

5.1 λ vs. MSE analysis

Train and Test errors for λ

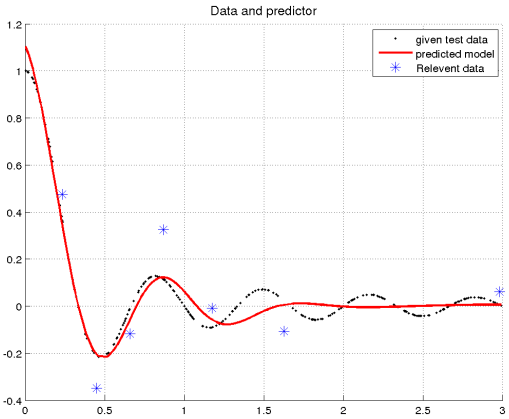


Training data model fit

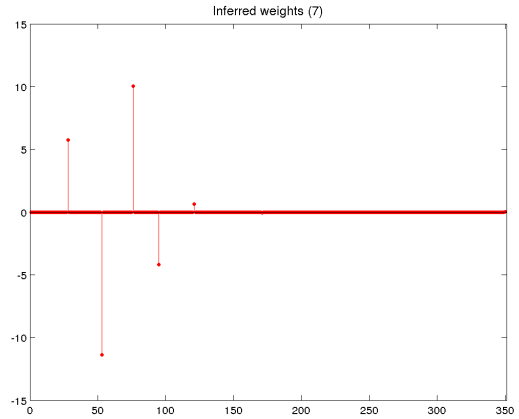


From the error plots for λ, the optimal λ value is found to be 0.25007.

Test data model fit

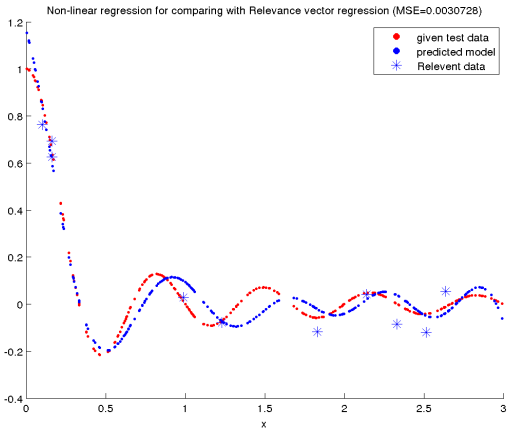


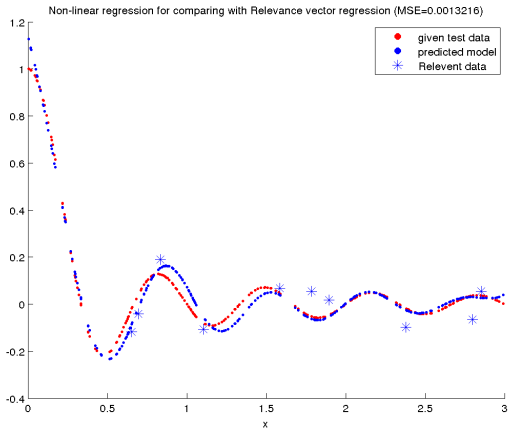
Relevance vector weightage

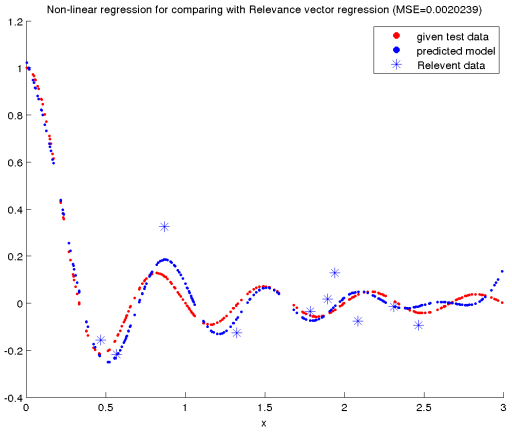


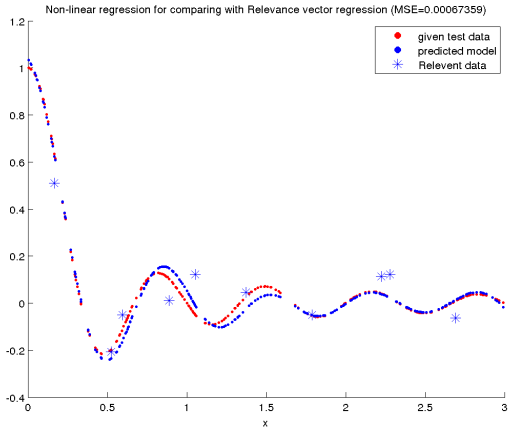
5.2 Non-linear regression to simulate Relevance vector regression

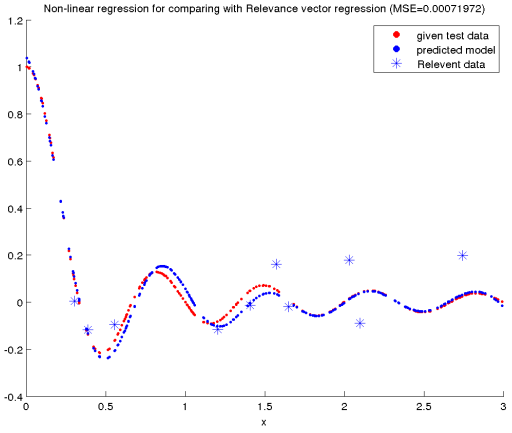
Average MSE = 0.0015


Non-linear regression for comparing with Relevance vector regression (MSE=0.0030728)
given test data
predicted model
Relevent data
x


Non-linear regression for comparing with Relevance vector regression (MSE=0.0013216)
given test data
predicted model
Relevent data
x


Non-linear regression for comparing with Relevance vector regression (MSE=0.0020239)
given test data
predicted model
Relevent data
x


Non-linear regression for comparing with Relevance vector regression (MSE=0.00067359)
given test data
predicted model
Relevent data
x


Non-linear regression for comparing with Relevance vector regression (MSE=0.00071972)
given test data
predicted model
Relevent data
x