CS6011: Kernel Methods for Pattern Analysis Programming Assignment 1

Aravind Sankar CS11B033
Ramnandan SK CS11B061
Adit Krishnan CS11B063

Group 2

February 5, 2015

Contents

1 Objective of the assignment

The basic goal of this assignment is to use linear models of regression to do 2 Function Approximation tasks, namely Polynomial Curve fitting on Dataset 1 (univariate) and Surface Fitting using Gaussian Basis functions for Dataset 2 (bivariate). Here, the means of the Gaussians used are the centers of clusters obtained using K-means clustering of the training data. The experimental results obtained are to be summarized using plots obtained on varying different model parameters.

2 Dataset 1

2.1 Generation

The given function to be approximated is shown below :

$$f(x) = e^{\cos(2\pi x)}$$

$$x \in (0, 1)$$

To generate points from this function, we randomly generated points $x \in (0, 1)$ and found the function value $f(x)$ at each point, and added to it Gaussian Noise with mean $\mu = 0$ and standard deviation $\sigma = 0.1$. 3 different training sets were generated, of sizes 20, 100 and 1000. The size of the validation and test set generated were 50 each.

2.2 Model creation

As the goal is to fit a polynomial function to the given data, we need to decide on the degree M of the polynomial which decides the model complexity.
Through the plots, we show the variation of model outputs and RMS errors on varying different parameters and our analysis of the results to choose the best model.
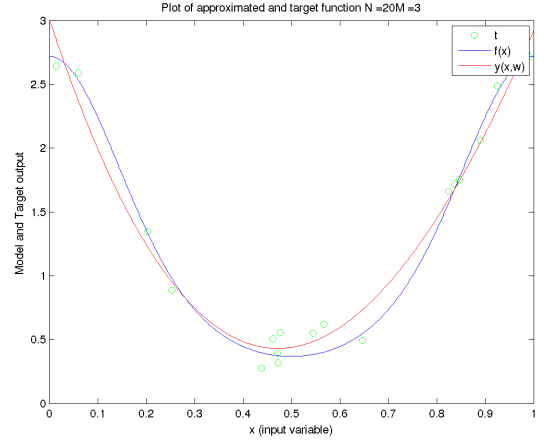
2.3   Results

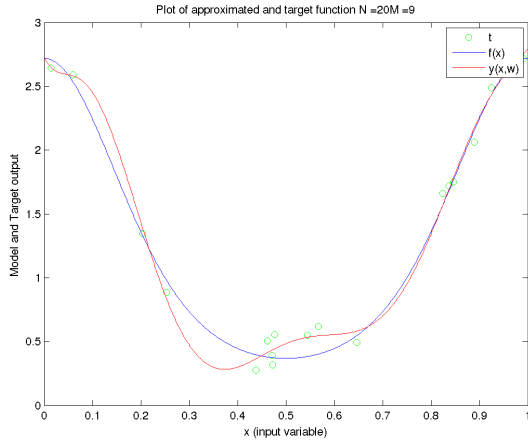### 2.3.1   Plots of approximated functions - Question 1

The following plots showing variation of the polynomial w.r.t degree of polynomial M, for training dataset of size $N = 20$. Each plot shows the data points x, the target function $f(x)$ and the approximated polynomial $y(x, \bar{w})$
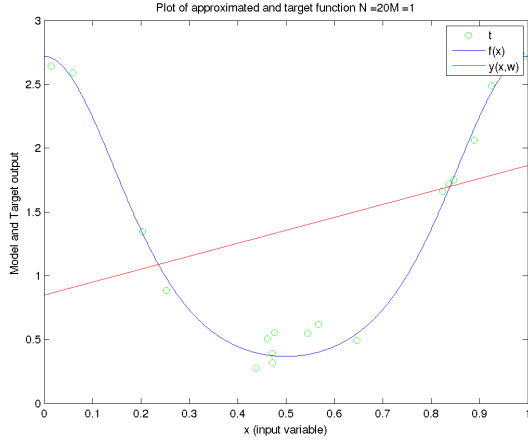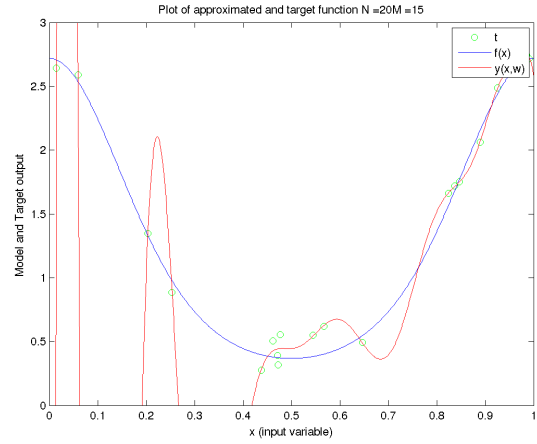

(a) N = 20 M = 1

(b) N = 20 M = 3

(c) N = 20 M = 9

(d) N = 20 M = 15

Observations :

- We chose the training dataset with 20 points to clearly show the overfitting problem (for large model complexities) in the absence of regularization.

- Notice that the model with complexity 3 is very close to the true function.

- For complexities 9 and 15, due to the small number of training data points, we observe overfitting which leads to significant deviation from the underlying target function.
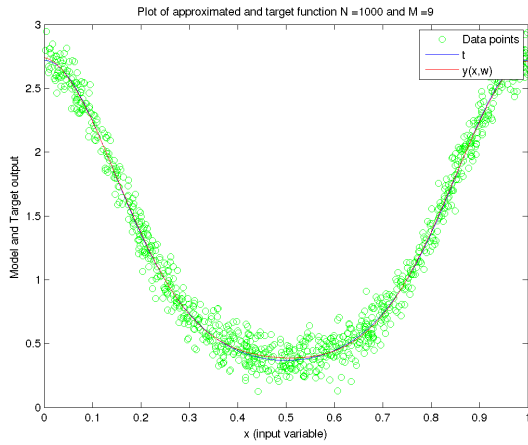
The following plots show the different polynomials obtained on varying the training dataset size N. Each of these plots are shown for 2 values of M, 9 and 15. The plots for N = 20 and these values of M has already been plotted above. So, the following plots correspond to training dataset sizes of 100 and 1000.
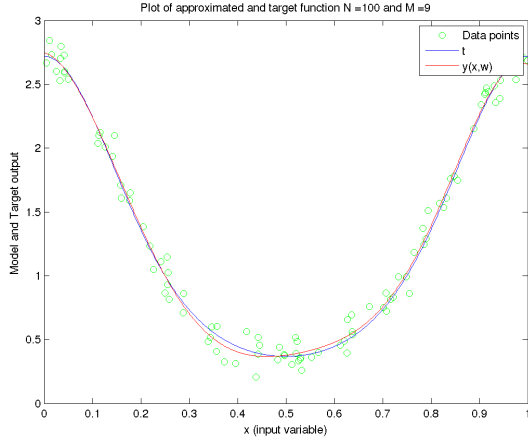
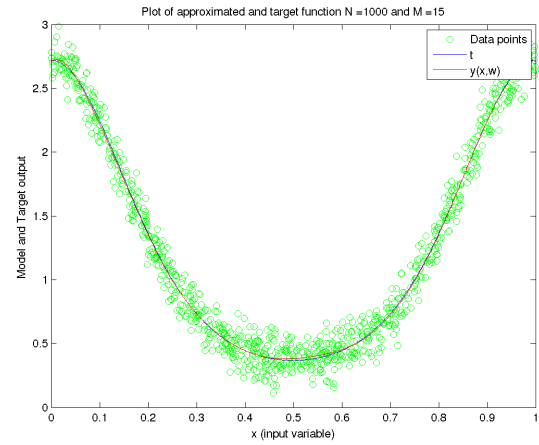(a) N = 100 M = 9
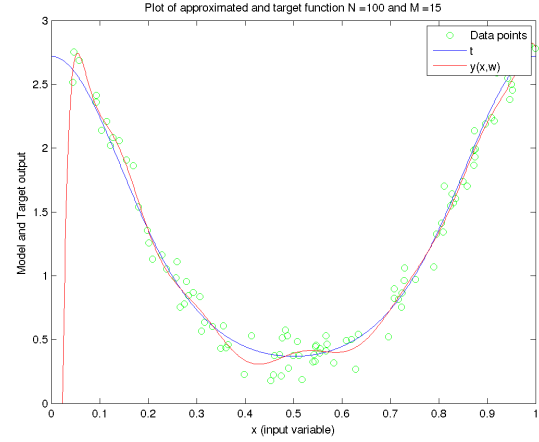
(b) N = 100 M = 15

(c) N = 1000 M = 9

(d) N = 1000 M = 15

Observations:

- We observe that Model complexity $M = 9$ almost matches the target function for both datasets since there are sufficient samples.

- However, Model complexity $M = 15$ leads to slight overfitting and hence some deviation from $f(x)$ in case of $N = 100$.

- However, for the 1000 sample dataset($N = 1000$) both models are almost overlapping with the target function due to the sufficiency of training points.
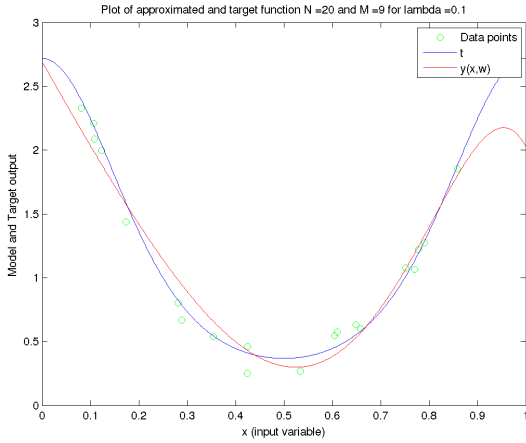
- Next we look at the effect of regularization.

The following plots show the effects of varying the regularization parameter λ on the curve fit. Since the overfitting effect is pronounced only for a small training set, these plots correspond to the 20 point training set and model complexity $M = 9$.
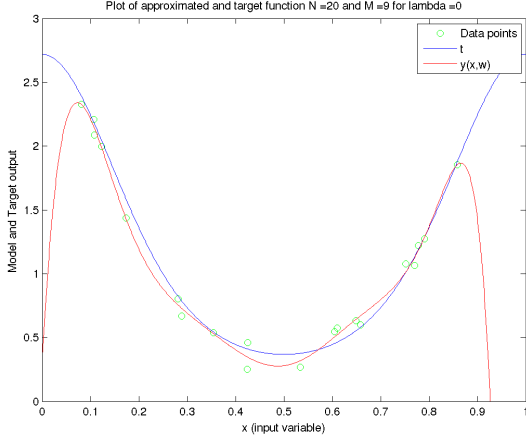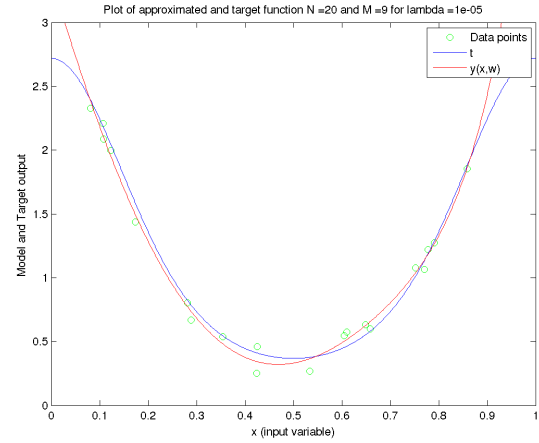


(a) $\lambda = 0$
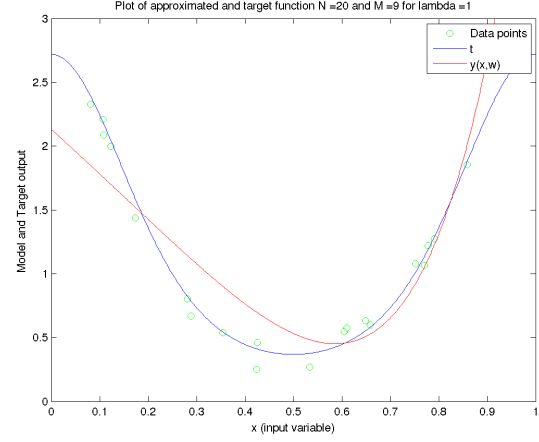
(b) $\lambda = 1$

(c) $\lambda = 0.1$

(d) $\lambda = 10^{-5}$

Observations:

- We can clearly notice the overfitting when regularization is not applied($\lambda = 0$) in the first plot.

- At the same time, $\lambda = 1$ in the second plot is a very high penalization of weights again leading to a relatively poor fit since less weightage is given to the error on the training points.

- The third plot corresponds to $\lambda = 0.1$. While this leads to a better fit than either of the above, it is still a high value penalization.

- The last plot corresponds to $\lambda = 10^{-5}$. This provides the closest approximation to the true function among these 4 plots. Thus typically small values of λ are investigated to minimize the overall error and produce the best fit, using the validation dataset.

2.3.2 Plots of RMS Error - Question 3

The following plots show the RMS Error obtained on training models of different complexities and different regularization parameters on training sets of different sizes. For each set of plots generated on varying 1 parameter (say M) , the optimal value of the other parameter (say λ) is obtained by cross-validation. The errors have been shown on all 3 datasets for each case (test, train and validation). The plots for 20 size training set undergo lot of variation for every new sampling due to the small size of the dataset, and have been added only for the sake of completion.
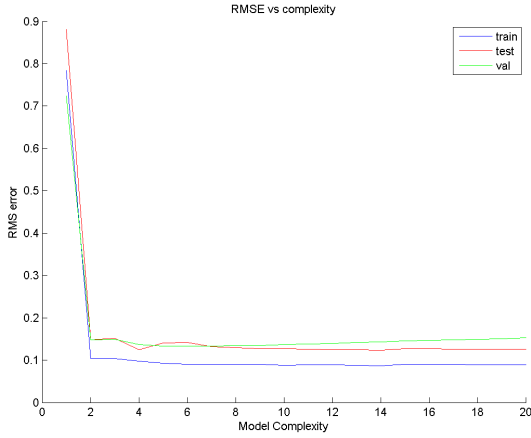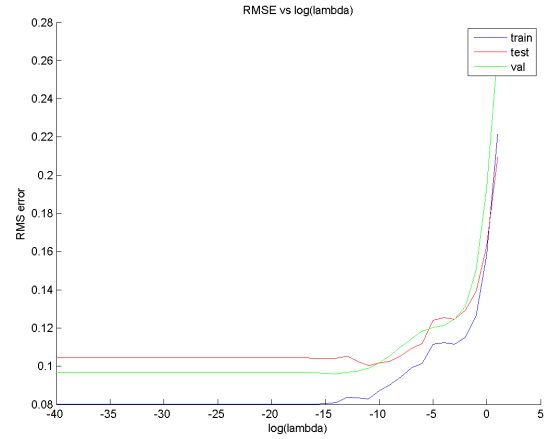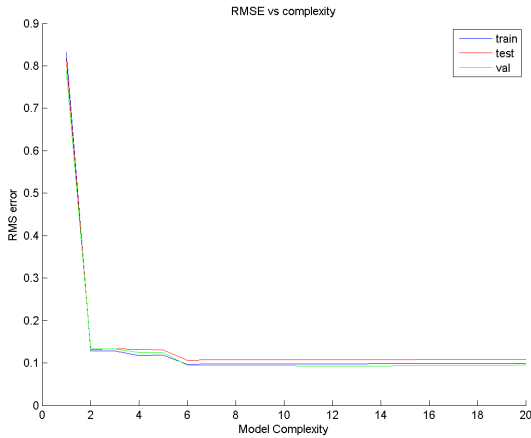
(a) Varying M, Optimal λ, 20 train size
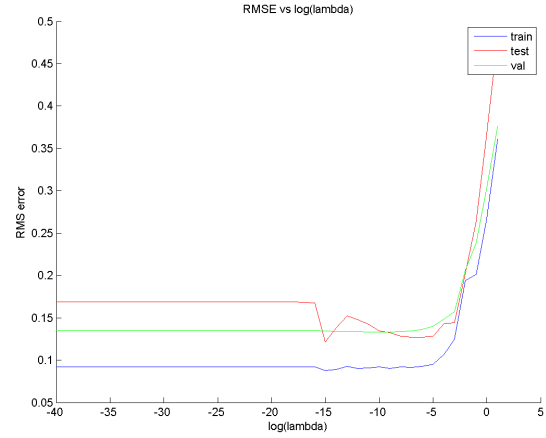
(b) Optimal M, Varying λ, 20 train size
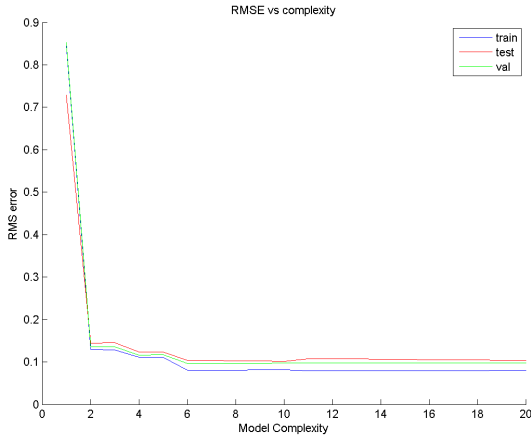
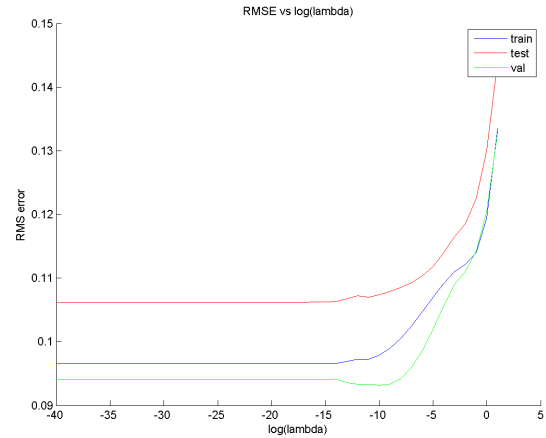(c) Varying M, Optimal λ, 100 train size


(d) Optimal M, Varying λ, 100 train size

(e) Varying M, Optimal λ, 1000 train size
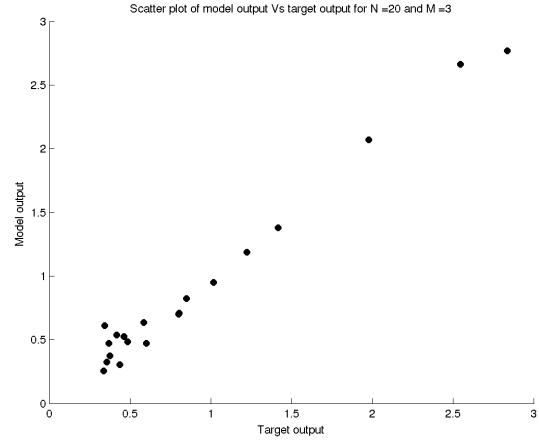

(f) Optimal M, Varying λ, 1000 train size

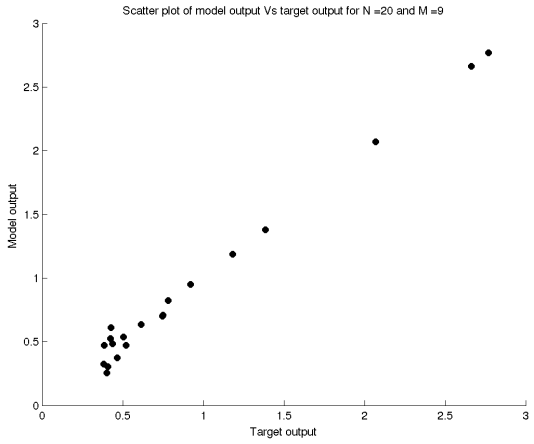### 2.3.3 Scatter plots of target and model output

The plots obtained here, have been generated in a process similar to that of question 1. The following plots show the scatter plots on the training data obtained on varying the degree of the polynomial M for training dataset size $N = 20$.
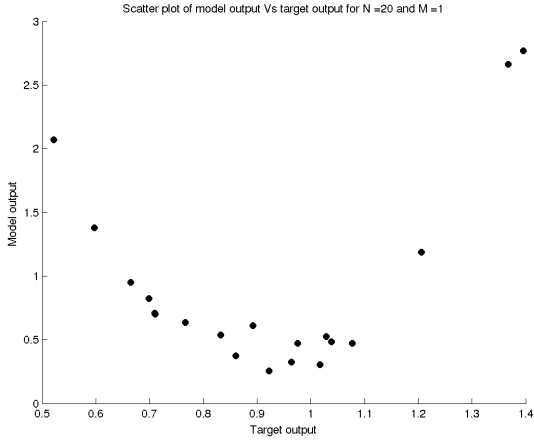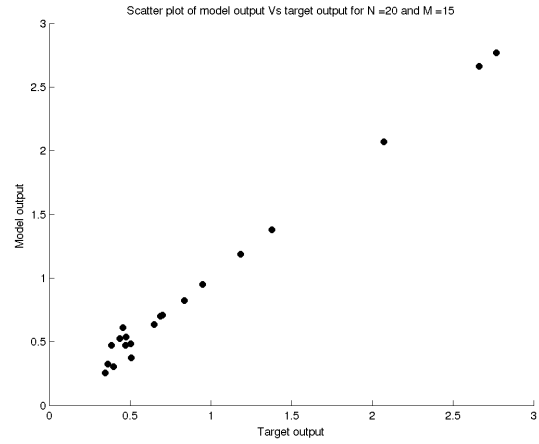
(a) N = 20 M = 1

(b) N = 20 M = 3

(c) N = 20 M = 9

(d) N = 20 M = 15

Observations :

- In case of a scatter plot, we would ideally like the points to lie along the $y = x$ line.

- We observe that for $M = 1$, we get a bad fit to the data, which leads to a bad plot.

- $M = 3$ does better, while $M = 9$ and 15 do much better as they fit the training data too well, leading to nearly perfect plots. This is because of overfitting to the training data, which leads to model output being very close to the target output, though the generalization ability is bad.

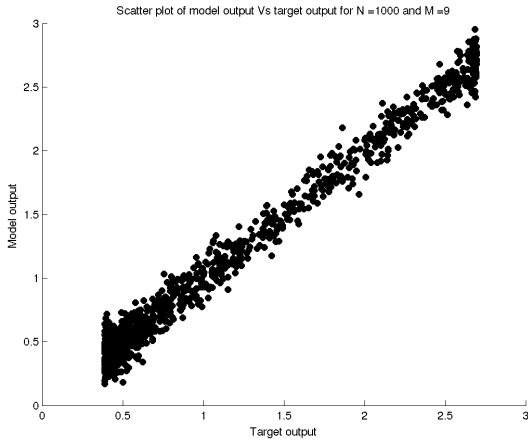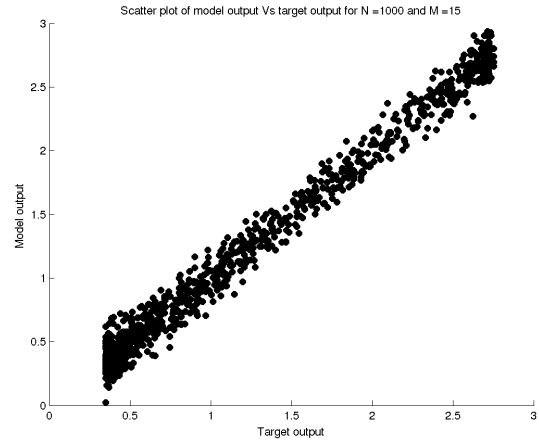The next set of scatter plots show the variation w.r.t the training set size N, namely $N = 100, 1000$.
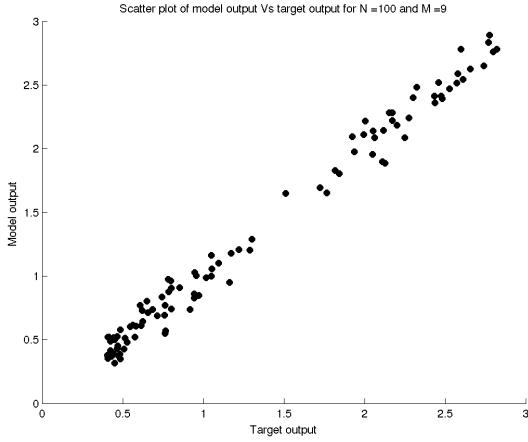
(a) N = 100 M = 9
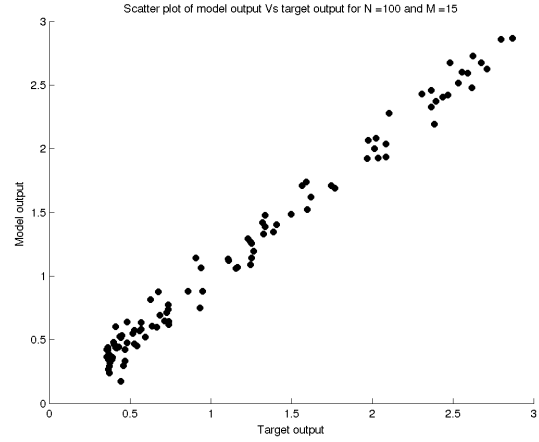
(b) N = 100 M = 15

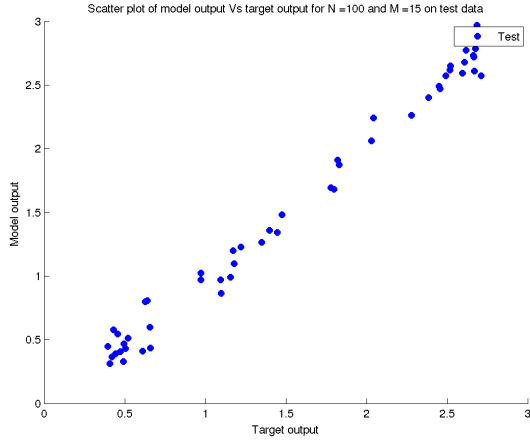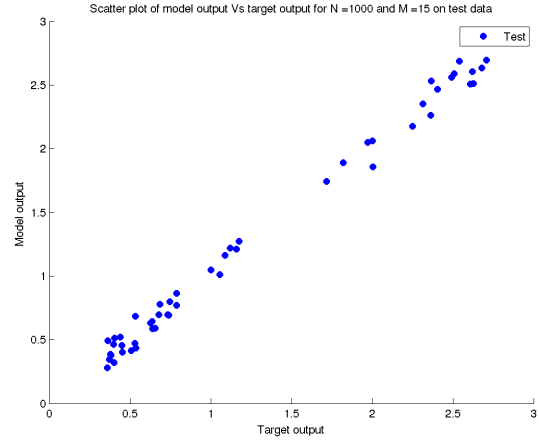(c) N = 1000 M = 9

(d) N = 1000 M = 15

Observations :

- For $M = 9$, we can see that both of the models ($N = 100, 1000$) perform well, as we have sufficient number of training examples to build a good model.

- Again for $M = 15$, we see that both models seem to give good results on the training data. Actually, overfitting happens in the case of $N = 100, M = 15$. To see this, we look at the scatter plots obtained on the test data below for the same model complexity $M = 15$.

- We can that the generalization ability is more, when we have a greater number of training examples, as exemplified by the better plot obtained when 1000 training examples are used. As we have overfitting when $N = 100, M = 15$, the model output doesn't follow the target output closely.
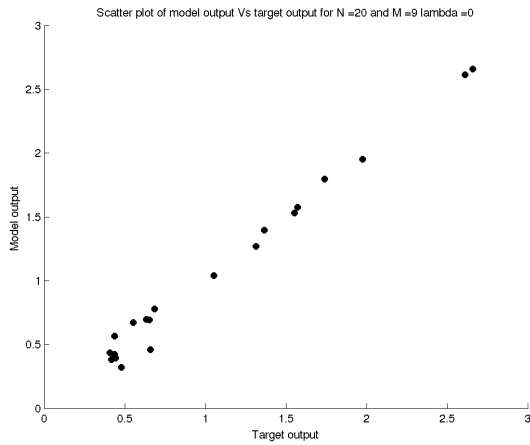
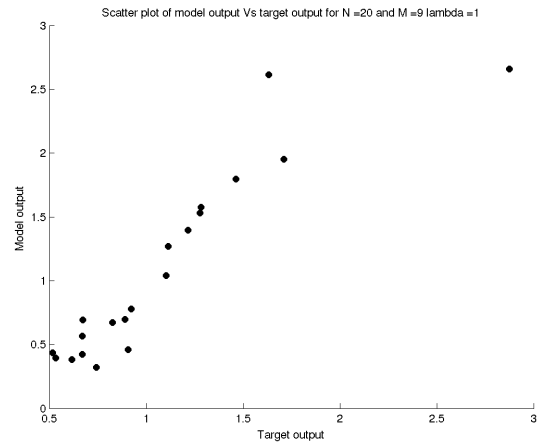- Next, the effect of λ is analyzed.
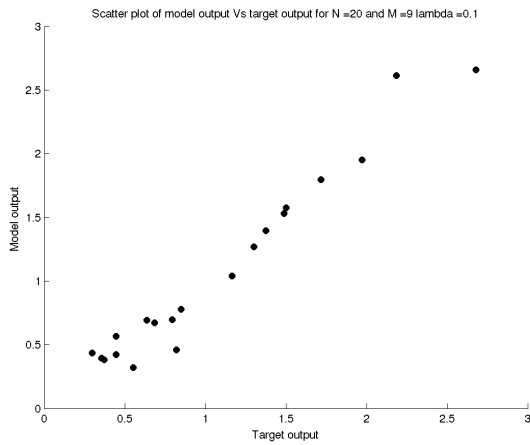
(a) N = 100 M = 15

(b) N = 1000 M = 15

The plots below show the scatter plots obtained on varying λ, for $N = 20$ and $M = 9$. The results obtained were very similar to that of question 1.
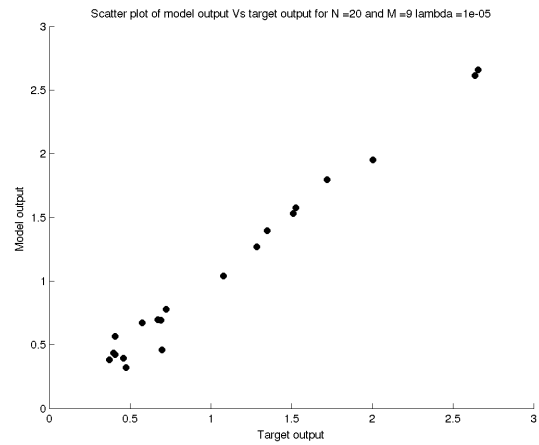

(a) $\lambda = 0$

(b) $\lambda = 1$

(c) $\lambda = 0.1$

(d) $\lambda = 10^{-5}$

10

Observations :

- Clearly, the first plot is the case where $\lambda = 0$, i.e. no regularization.

- Again for $\lambda = 1$, there is a very high penalization of weights leading to a bad scatter plot, while $\lambda = 0.1$ performs better.

- The last plot corresponds to $\lambda = 10^{-5}$, which gives a good scatter plot where the model output follows the target output closely and also gives a good generalization on test examples.

3 Dataset 2

3.1 Generation

3.2 Model creation

3.3 Results

3.4 Plots of approximated functions