Machine Learning for Computer Vision (EE5177)
Programming Assignment 3 : Regression Models
Problem #1

Akshit Kumar
EE14B127

29th March 2017

Contents

1 Introduction

1.1 Goal

The first problem of the programming assignment requires us to use non linear regression to fit a *sinc* function (non-linear curve) using an RBF kernel and arctan kernel.

1.2 Approach

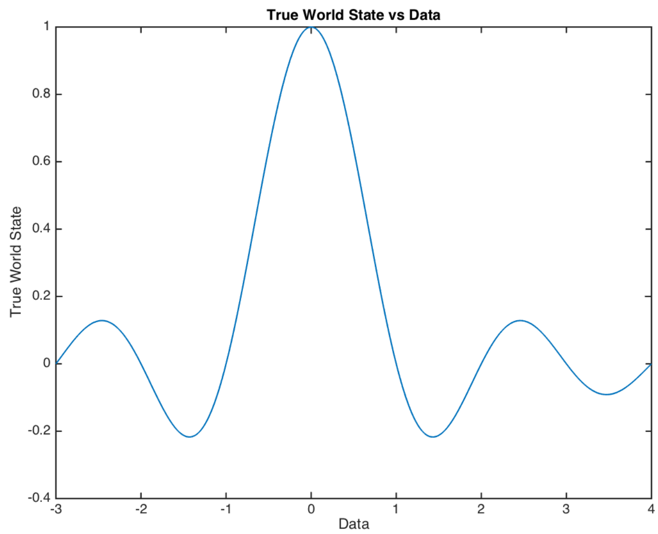
To fit a non-linear function, we first generate our ground truth data by setting the world states to $sinc(x)$, in order the generate synthetic data for curve fitting, we add noise to the ground truth data and use RBF kernel and arctan kernel to fit the data to it.

1.3 Results

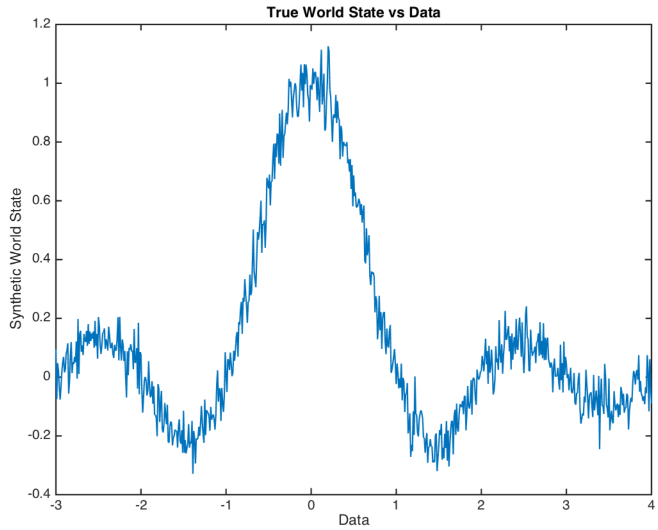
The number of kernels is initially taken to be 7. The number of kernels can be varied to minimise the error. It can be noticed that beyond a certain number of kernels, the error does not decrease.

2 Data Generation

2.1 Plot of x vs. y (Ground Truth Data)



2.2 Plot of x vs. w



2.3 Rational for using non linear regression

The relationship between the world states w and the input data x is not linear (sinc is not a linear function). Therefore we transform our input data through a nonlinear transformation as we would like to retain the mathematical convenience of the linear model while extending the class of functions that can be described.

3 Non Linear Regression using Radial Basis Function (Gaussian)

3.1 Transformation Equation

The input data x is transformed using the non linear function given by

$$z_i = f[x_i]$$

$$z_i = \begin{bmatrix} 1 \\ exp[-(x_i - \alpha_1)^2/\lambda] \\ exp[-(x_i - \alpha_2)^2/\lambda] \\ exp[-(x_i - \alpha_3)^2/\lambda] \\ exp[-(x_i - \alpha_4)^2/\lambda] \\ exp[-(x_i - \alpha_5)^2/\lambda] \\ exp[-(x_i - \alpha_6)^2/\lambda] \end{bmatrix}$$

3.2 Choice of α's

The αs are chosen as means of the intervals, partitioning the dataset into appropriate number of intervals depending upon the number of kernel functions to be used.

3.3 Choice of λ

The value of λ is initially arbitarily chosen to be 1 and then varied from 0.01 to 7 in steps on 0.001 and optimal λ value is found which minimizes the prediction error and fitting error.

3.4 Weight Estimation using MLE

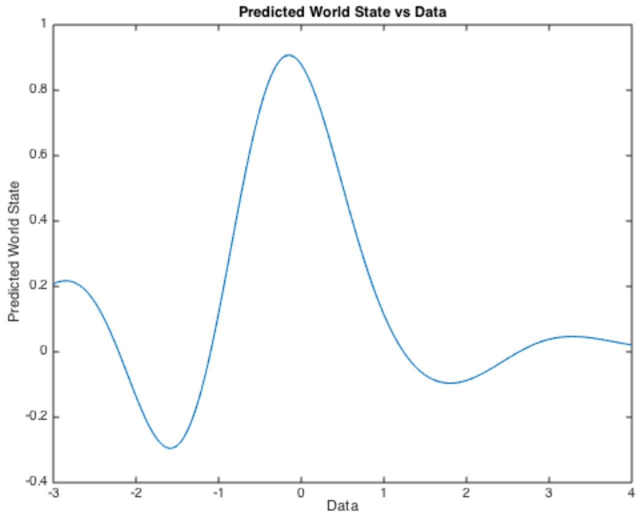
The optimal weights can be computed using

$$\phi = (ZZ^T)^{-1}Zw$$

where the matrix Z contains the transformed vectors $\{z_i\}_{i=1}^{I}$ in its columns.

3.5 World State Estimation (using $\lambda = 1$)

3.5.1 Plot of the Predicted World State vs Data



3.5.2 Prediction and Fitting Error

Prediction Error = ***0.1137***
Fitting Error = ***0.1255***

3.6 Error Analysis

3.6.1 Prediction Error

Predicition Error is the RMSE between the estimate of the weights and the ground truth value. It can be defined as

$$P.E = \left\|\phi^T z - y\right\|_2$$

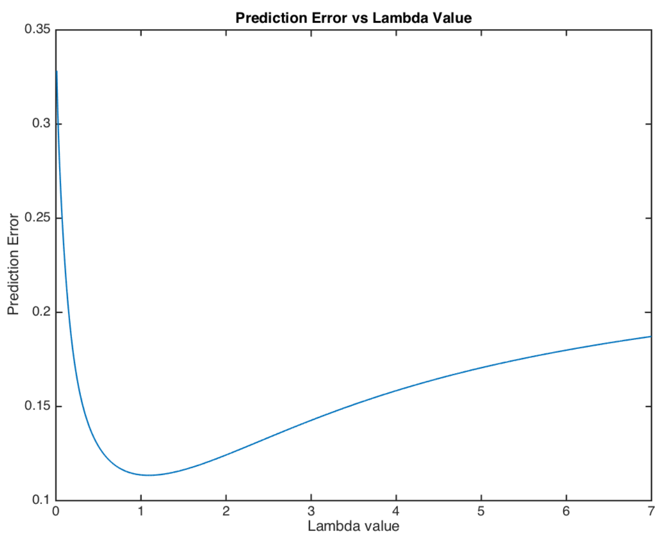
3.6.2 Fitting Error

Fitting Error is the RMSE between the estimate of the weights and the noisy data sample. It can be defined as

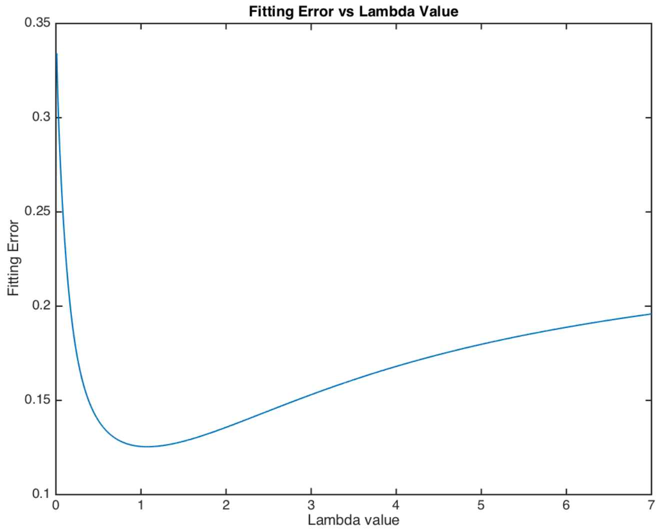
$$P.E = \left\|\phi^T z - w\right\|_2$$

3.7 Varying λ

3.7.1 Plot of Prediction Error



3.7.2 Plot of Fitting Error



3.7.3 Results

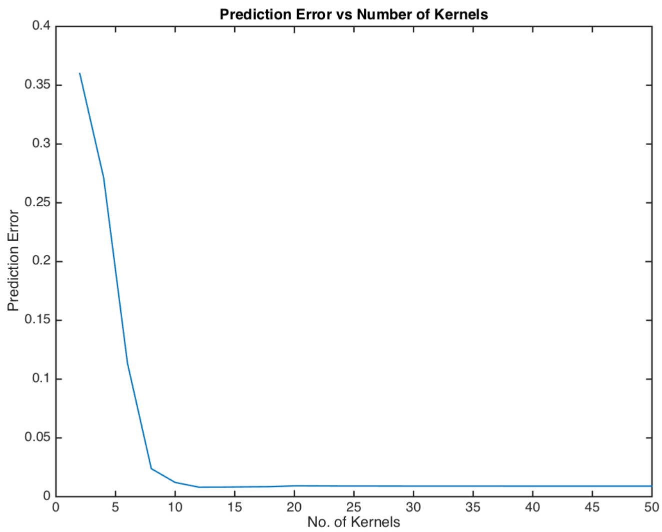
The value of λ corresponding to the minimum prediction error is ***1.0910***.
The minimum prediction error is ***0.1135***.
The value of λ corresponding to the minimum fitting error is ***1.0750***. Note, this is not a deterministic value as it depends on the noisy-ness of the data.
The minimum fitting error is ***0.1254***.

3.7.4 Explanation

For small values of λ the error decreases rapidly. For some value of λ the error hits the minimum and then continues to increase. λ represents an estimate of the variance of the data around the peaks given by α's. Once we get to the required variance by iterating through the different values of λ, we get the best prediction of the data and hence the least error.

3.8 Varying the number of kernels

3.8.1 Plot of Prediction Error



3.8.2 Optimal number of kernel functions

The optimal number of kernel functions obtained is ***14*** with $\lambda = 1.0910$

3.8.3 Explanation

Getting the number of kernels required for minimum prediction error implies that 14 components are sufficient for explaining the data and better compared to the use of a larger number of components. Thus adding more degrees of freedom to the vector ϕ is not going to improve the estimation and reduce the prediction error.

4 Non Linear Regression using Sigmoid function(Arc tangent)

4.1 Transformation Equation

The input data x is transformed using the non linear function given by

$$z_i = f[x_i]$$

$$z_i = \begin{bmatrix} exp[-(\lambda x_i - \alpha_1)] \\ exp[-(\lambda x_i - \alpha_2)] \\ exp[-(\lambda x_i - \alpha_3)] \\ exp[-(\lambda x_i - \alpha_4)] \\ exp[-(\lambda x_i - \alpha_5)] \\ exp[-(\lambda x_i - \alpha_6)] \\ exp[-(\lambda x_i - \alpha_7)] \end{bmatrix}$$

4.2 Choice of α's

The αs are chosen as means of the intervals, partitioning the dataset into appropriate number of intervals depending upon the number of kernel functions to be used.

4.3 Choice of λ

The value of λ is initially arbitarily chosen to be 1 and then varied from 0.01 to 7 in steps on 0.001 and optimal λ value is found which minimizes the prediction error and fitting error.

4.4 Weight Estimation using MLE

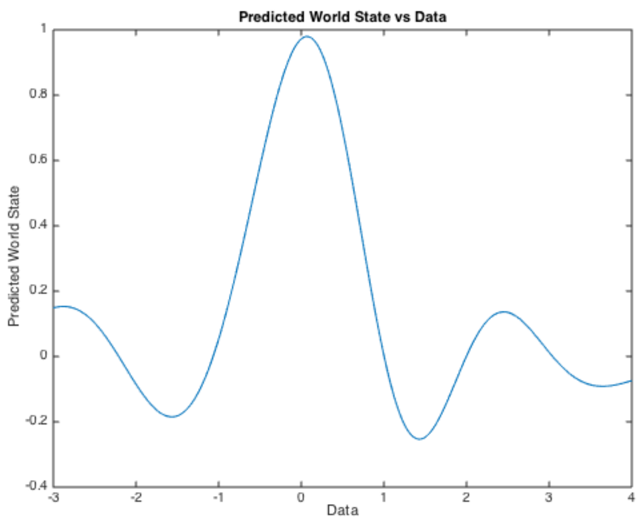
The optimal weights can be computed using

$$\phi = (ZZ^T)^{-1}Zw$$

where the matrix Z contains the transformed vectors $\{z_i\}_{i=1}^{I}$ in its columns.

4.5 World State Estimation (using $\lambda = 1$)

4.5.1 Plot of the Predicted World State vs Data



4.5.2 Prediction and Fitting Error

Prediction Error = ***0.0457***
Fitting Error = ***0.0718***

4.6 Error Analysis

4.6.1 Prediction Error

Predicition Error is the RMSE between the estimate of the weights and the ground truth value. It can be defined as

$$P.E = \left\|\phi^T z - y\right\|_2$$

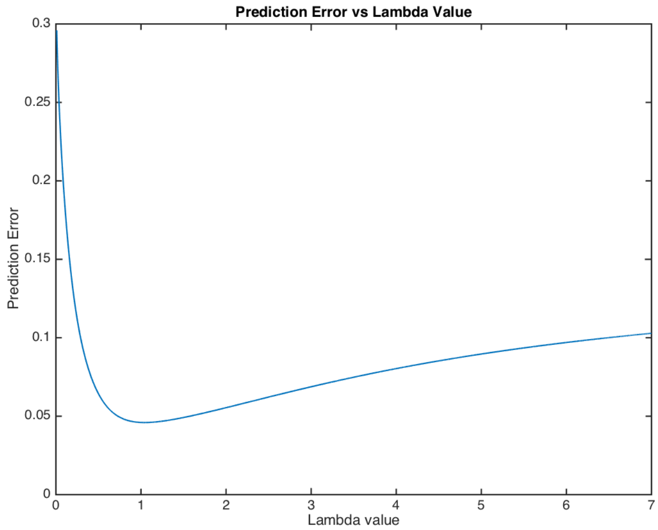
4.6.2 Fitting Error

Fitting Error is the RMSE between the estimate of the weights and the noisy data sample. It can be defined as

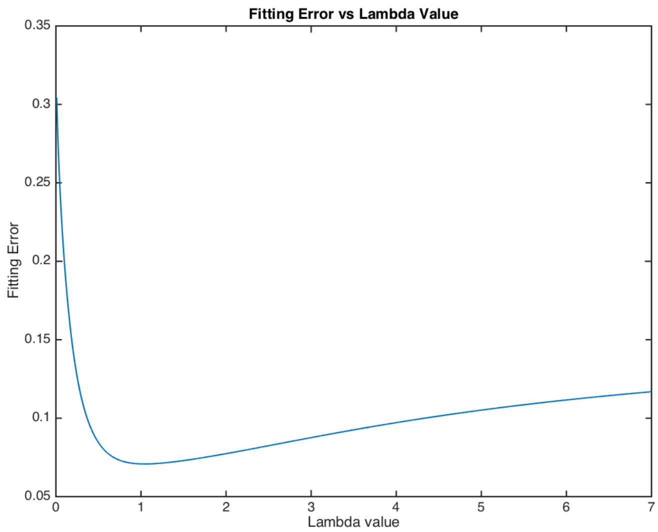
$$P.E = \left\|\phi^T z - w\right\|_2$$

4.7 Varying λ

4.7.1 Plot of Prediction Error



4.7.2 Plot of Fitting Error



4.7.3 Results

The value of λ corresponding to the minimum prediction error is ***1.0410***.
The minimum prediction error is ***0.0456***.
The value of λ corresponding to the minimum fitting error is ***1.0430***. Note, this is not a deterministic value as it depends on the noisy-ness of the data.
The minimum fitting error is ***0.0718***.

4.7.4 Explanation

For small values of λ the error decreases rapidly. For some value of λ the error hits the minimum and then continues to increase. λ represents an estimate of the variance of the data around the peaks given by α's. Once we get to the required variance by iterating through the different values of λ, we get the best prediction of the data and hence the least error.