# Assignment 01

Solution by Khaled Hamad

September 28, 2021

## Question 1

$$f(x) = x_2^3 + 2x_1x_3^2 + 3x_1^3$$

first we need to find the derivative with respect to the vector $\mathbf{x}$. In this case, $\mathbf{x} = [x_1, x_2, x_3]$

$$\frac{\partial f}{\partial \mathbf{x}} = \frac{\partial f}{\partial [x_1, x_2, x_3]} = [\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \frac{\partial f(x)}{\partial x_3}]$$

$$\frac{\partial f(x)}{\partial x_1} = 4x_3^2 + 9x_1^2$$

$$\frac{\partial f(x)}{\partial x_2} = 3x_2^2$$

$$\frac{\partial f(x)}{\partial x_3} = 4x_1x_3$$

$$\frac{\partial f}{\partial \mathbf{x}} = \frac{\partial f}{\partial [x_1, x_2, x_3]} = [4x_3^2 + 9x_1^2, 3x_2^2, 4x_1x_3]$$

by substituting the provided points to the derivative $\frac{\partial f}{\partial \mathbf{x}}$, all the derivative values should equal to zero for the point to be stationary.

$$\frac{\partial f}{\partial [1, 0, 0]} = [4(0)^2 + 9(1)^2, 3(0)^2, 4(1)(0)] = [9, 0, 0]$$

$$\frac{\partial f}{\partial [0, 0, 0]} = [4(0)^2 + 9(0)^2, 3(0)^2, 4(0)(0)] = [0, 0, 0]$$

$$\frac{\partial f}{\partial [3, -1, 5]} = [4(5)^2 + 9(3)^2, 3(-1)^2, 4(3)(5)] = [181, 3, 60]$$

$$\frac{\partial f}{\partial [1, 2, 3]} = [4(3)^2 + 9(1)^2, 3(2)^2, 4(1)(3)] = [45, 12, 12]$$

The only stationary point is when $\mathbf{x} = [0, 0, 0]$ because at that point all the derivative values are equal to zero.

## Question 2

$$g(\mathbf{v}) = \mathbf{v}^T\mathbf{S}\mathbf{v} + 3\mathbf{v}^T\mathbf{v} + \mathbf{x}^T\mathbf{v} + 10$$

$$\frac{\partial g}{\partial \mathbf{v}} = (\mathbf{S}\mathbf{v})^T + \mathbf{v}^T\mathbf{S} + 3(\mathbf{v}^T + \mathbf{v}^T) + \mathbf{x}^T$$

$$\frac{\partial g}{\partial \mathbf{v}} = \mathbf{v}^T\mathbf{S} + \mathbf{v}^T\mathbf{S} + 3(2\mathbf{v}^T) + \mathbf{x}^T$$

$$\frac{\partial g}{\partial \mathbf{v}} = 2\mathbf{v}^T\mathbf{S} + 6\mathbf{v}^T + \mathbf{x}^T$$

## Question 3

$$f_1(\mathbf{x}, \mathbf{w}) = x_1 sin(2\pi w_1) + w_2 cos(2\pi x_2) + w_1^3 w_2 x_1 + w_1 w_2^2 x_2$$

$$\frac{\partial f_1}{\partial \mathbf{w}} = \frac{\partial f_1}{\partial [w1, w2]} = [\frac{\partial f_1}{\partial w1}, \frac{\partial f_1}{\partial w2}]$$

$$\frac{\partial f_1}{\partial w1} = 2\pi x_1 cos(2\pi w_1) + 3w_1^2 w_2 x_1 + w_2^2 x_2$$

$$\frac{\partial f_1}{\partial w2} = cos(2\pi x_2) + w_1^3 x_1 + 2w_1 w_2 x_2$$

$$\frac{\partial f_1}{\partial \mathbf{w}} = [2\pi x_1 cos(2\pi w_1) + 3w_1^2 w_2 x_1 + w_2^2 x_2, cos(2\pi x_2) + w_1^3 x_1 + 2w_1 w_2 x_2]$$

# Question 4

# Plots when s=0.5 and M is varying from 1 to 20

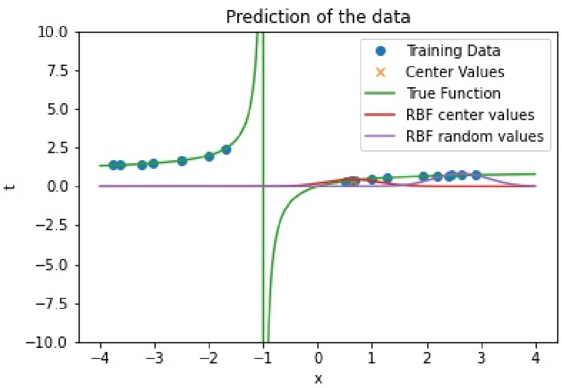


Figure 1: prediction M=1

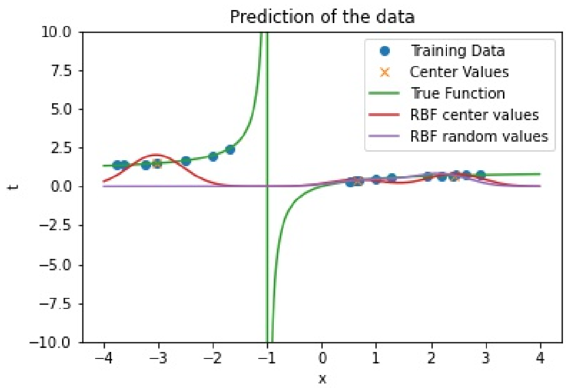


Figure 2: prediction M=3

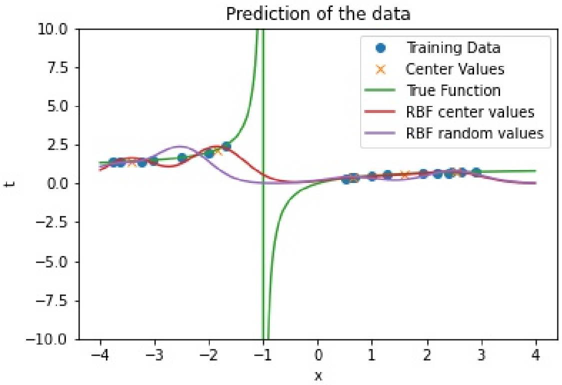


Figure 3: prediction M=5

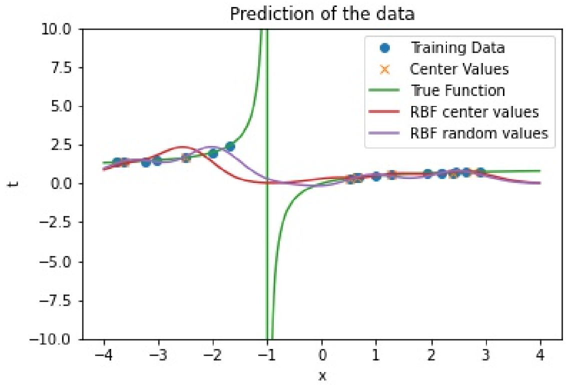


Figure 4: prediction M=7

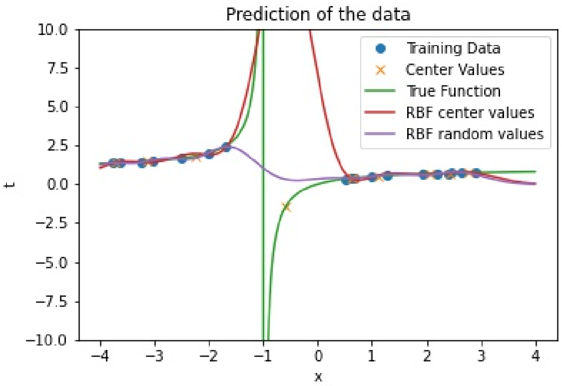


Figure 5: prediction M=10

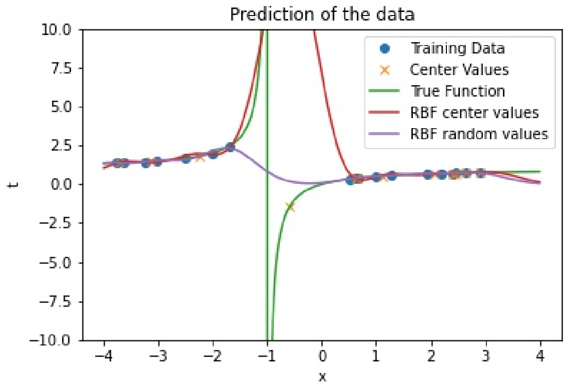


Figure 6: prediction M=12

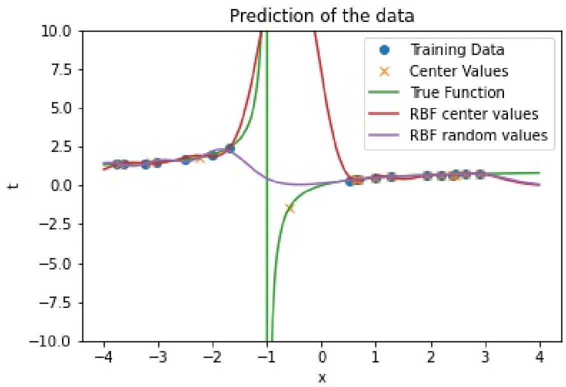


Figure 7: prediction M=15

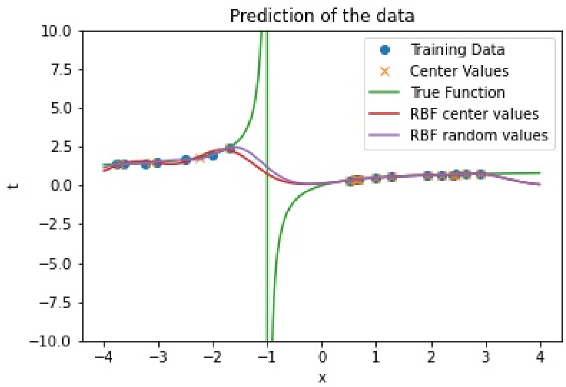


Figure 8: prediction M=17

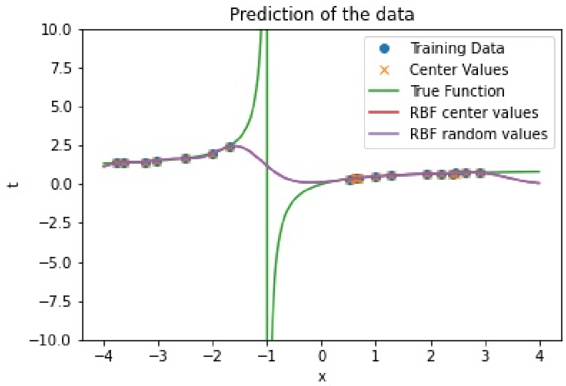


Figure 9: prediction M=20

**The error plot for this case would be**

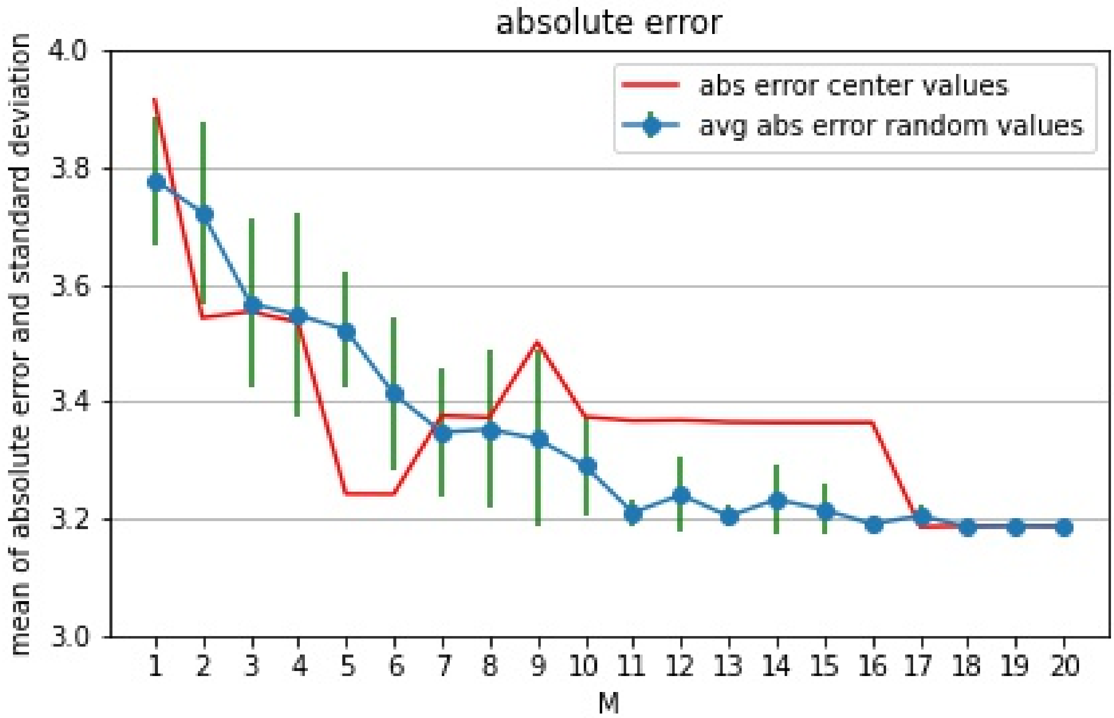


Figure 10: The error behavior as "M" increases

## Plots when M=5 and s is varying from 0.001 to 10

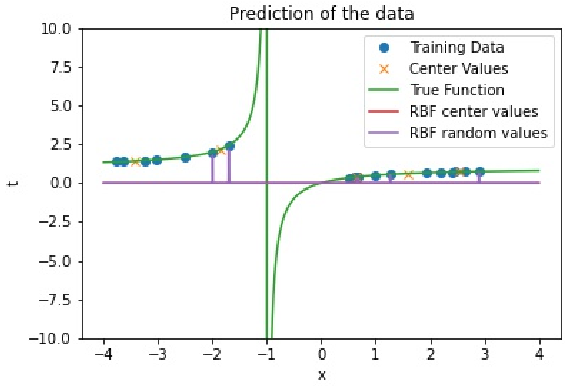


Figure 11: prediction s=0.001

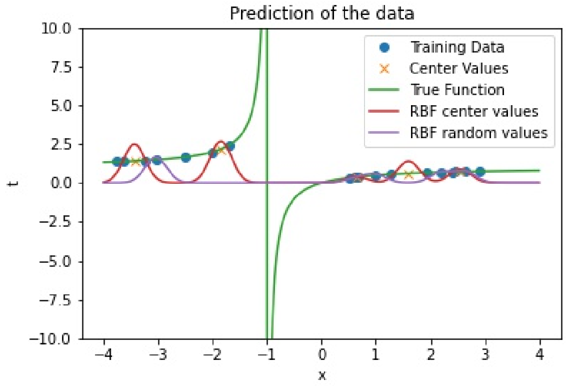


Figure 12: prediction s=0.2

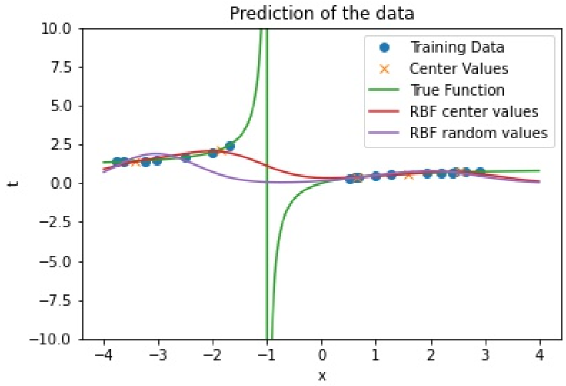


Figure 13: prediction s=0.8

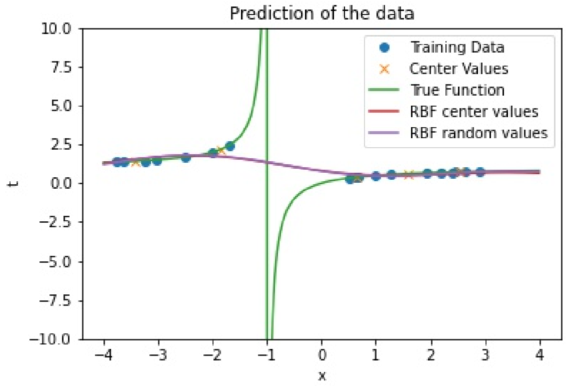


Figure 14: prediction s=2

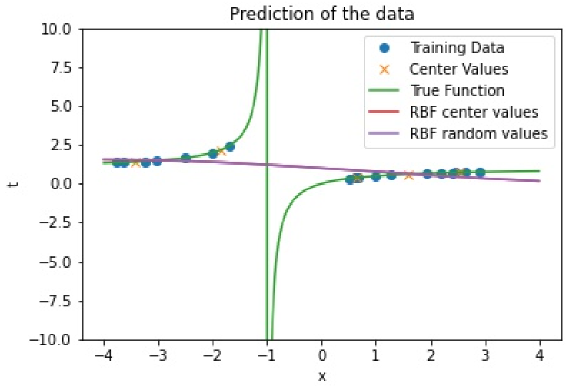


Figure 15: prediction s=5

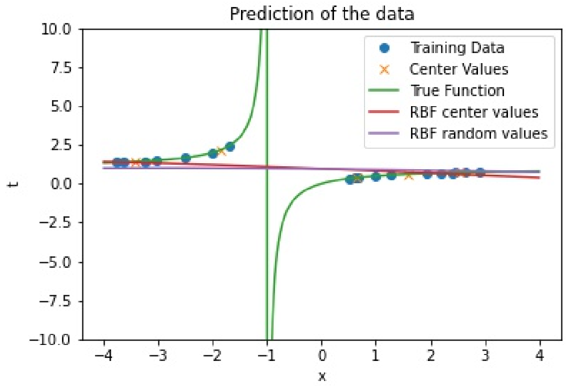


Figure 16: prediction s=10

**For the error plots:**

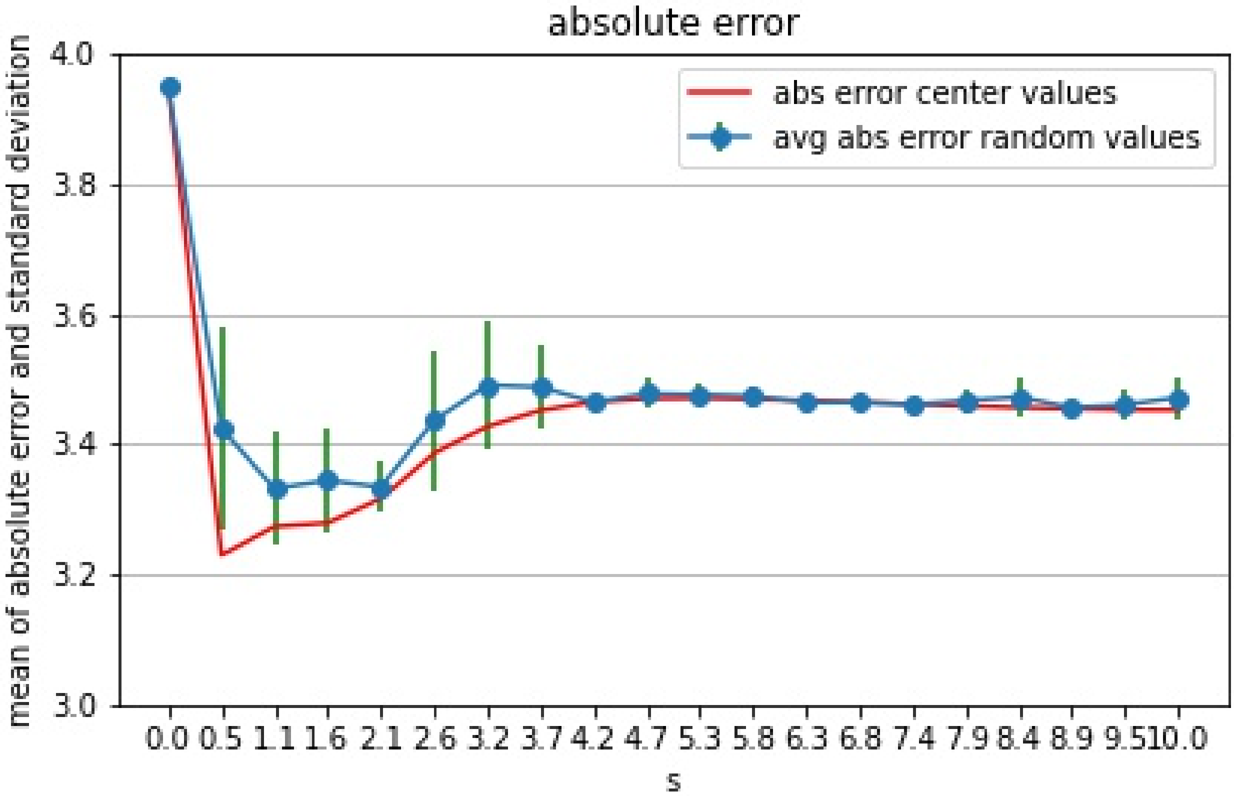


Figure 17: The error behavior as "s" increases

## Question Responses

### 1) Do the RBFs outperform (in terms of error between predicted and desired) the polynomial basis function with the same M value on the provided training and testing data? Why or why not? Are there cases where the RBF outperforms and cases where the polynomial basis outperforms? What are those cases and why does (or does not) this occur?

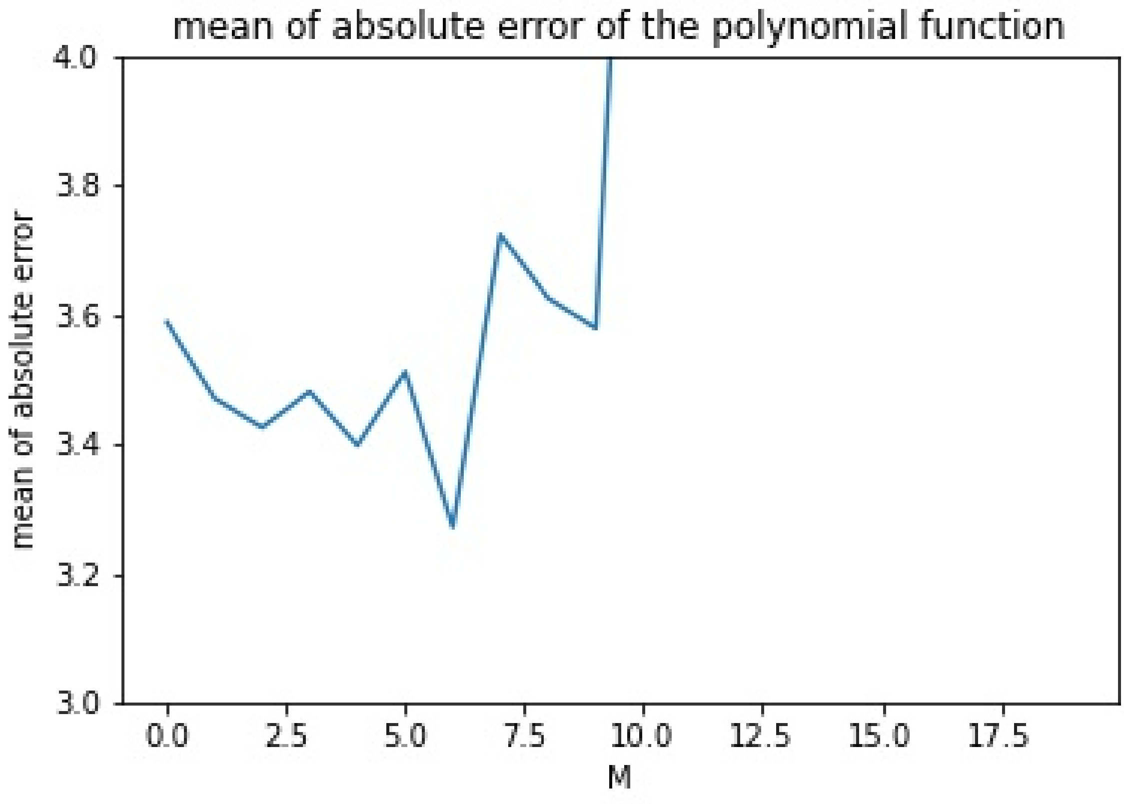


Figure 18: The error behavior as "M" increases for the polynomial curve fitting

avg_ploynomial_error=array([ 3.58845912, 3.46939484, 3.42820241, 3.48227302, 3.4161836 , 3.52117047, 3.31891796, 3.86734001, 3.26913179, 4.74561218, 4.061853 , 7.07365995, 4.98982492, 15.1983917 , 28.87032076, 4.60047535, 57.75620834, 66.06608148, 89.89176215, 13.64144499])

As the plot of the polynomial curve fitting error illustrates, the polynomial basis functions have a similar error to those from the RBF with M ranging from M=1 to M=9. After M=9, overfitting occurs with the polynomial basis function. For the RBF, as M increases, the error decreases for both the spread center values and the randomly selected center values. For the RBF, if s is too small or too large, the error increase. The best performance is obtained with an s value around 0.5, M at 20, using spread center values. This occurs because the spread centers ensure that all regions of the range of inputs is covered (with randomly selecting centers, sometimes all ranges are covered and sometimes they are not - so performance varies and, when centers are bunched, reduces performance). The s value determines the size of a neighborhood around each center that is impacted by a particular RBF, if s is too small, much of the range is not near to any RBF center. If s is too large, every value is in the neighborhood of every RBF center. An effective s value around ensures that each point in the range is only in the neighborhood of the most relevent RBF centers. For larger values of M with the RBF, this ensures that the full range is covered adequately. Neither the polynomial nor the RBF basis functions can adequately characterize the region around -1 since there is no training data in that area.

### 2) Do any of the generated plots show indication of overfitting in any of your results? Why or why not?

Yes, some of the plots did show indication of overfitting. For example, Figure 19 shows that the fit is going directly through the points. However it is best to verify overfitting by examining error on test data.

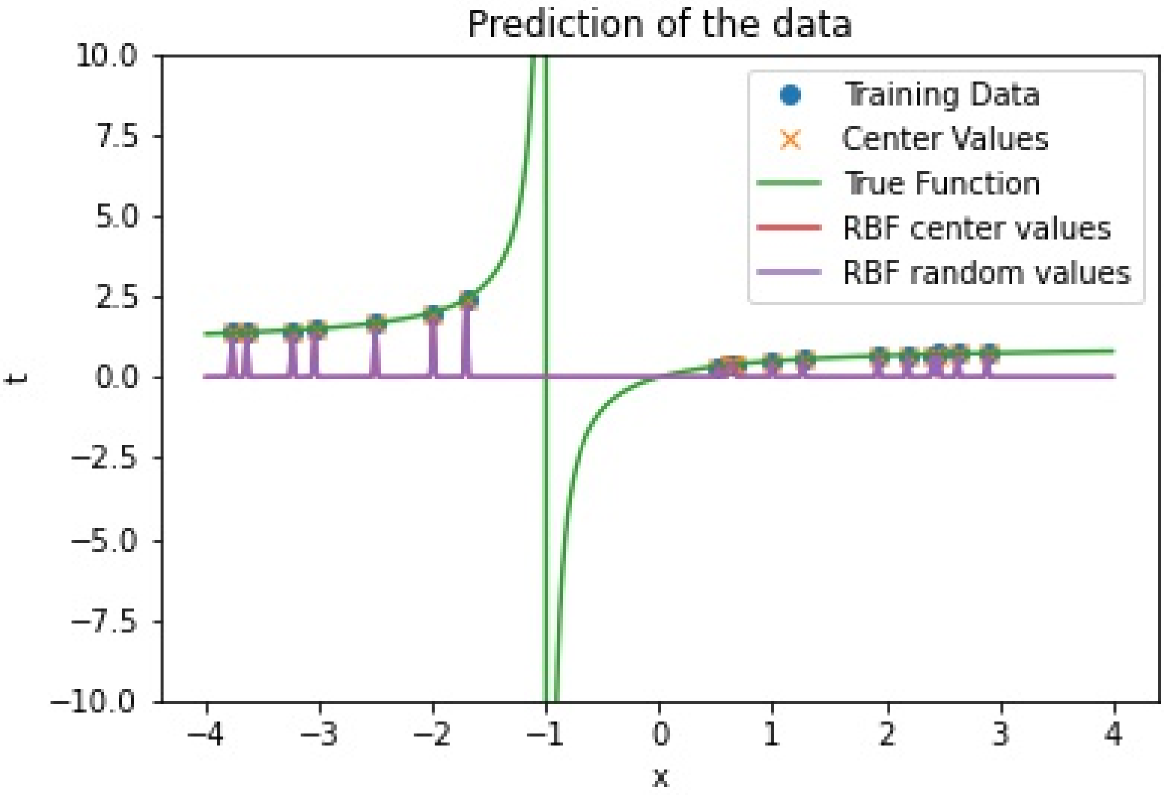


Figure 19: a case where the RBF is over fitting with s=0.01 and M=20

### 3) What is the role of the parameter s and how does the choice of s effect results?

The parameter "s" determines the neighborhood size around an RBF center. If s is too small, then only inputs very close to an RBF center will have a non-zero associated feature value. If s is too large, then many input points will have very similar RBF feature values.

### 4) How do the evenly space vs. randomly selected center values impact performance? When is one better than the other? Why?

There are cases cases in which the evenly spaced selected center values outperformed the randomly selected ones. If centers are bunched, there are some regions of the input range does adequately represented. Centered/spread values avoid this issue.