Nguyen, Theodore

704-156-701

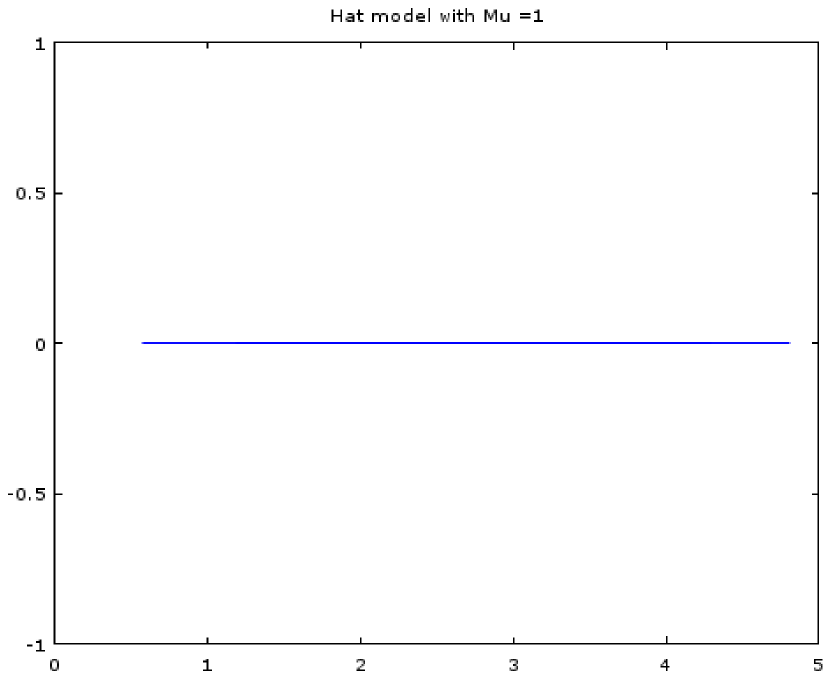
Math 156 Spring 2016

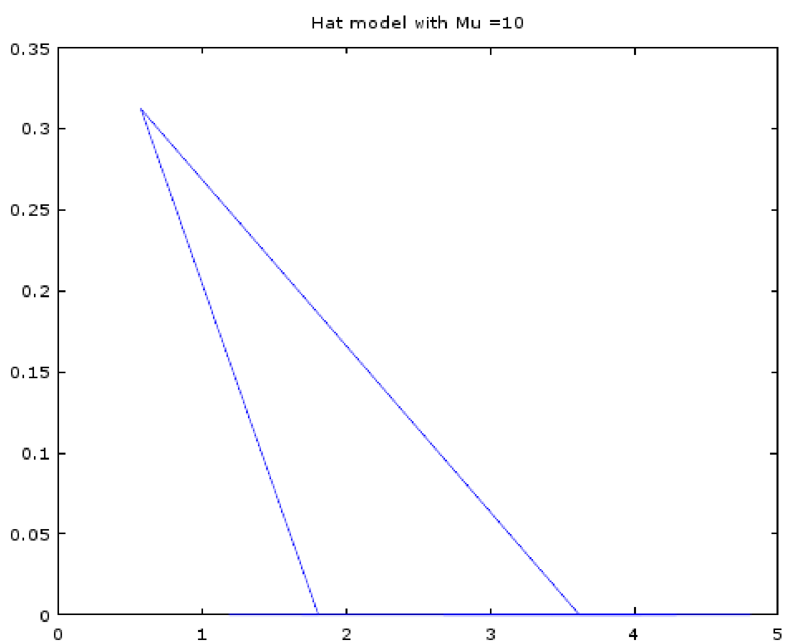
Lab 4: linear regression

2. (4) Train the model on the data in simple.mat using M = 10 hat functions and Mu = 10^5. Plot and turn in the learned model (the function fitted to the data) on the interval [0; 2pi].

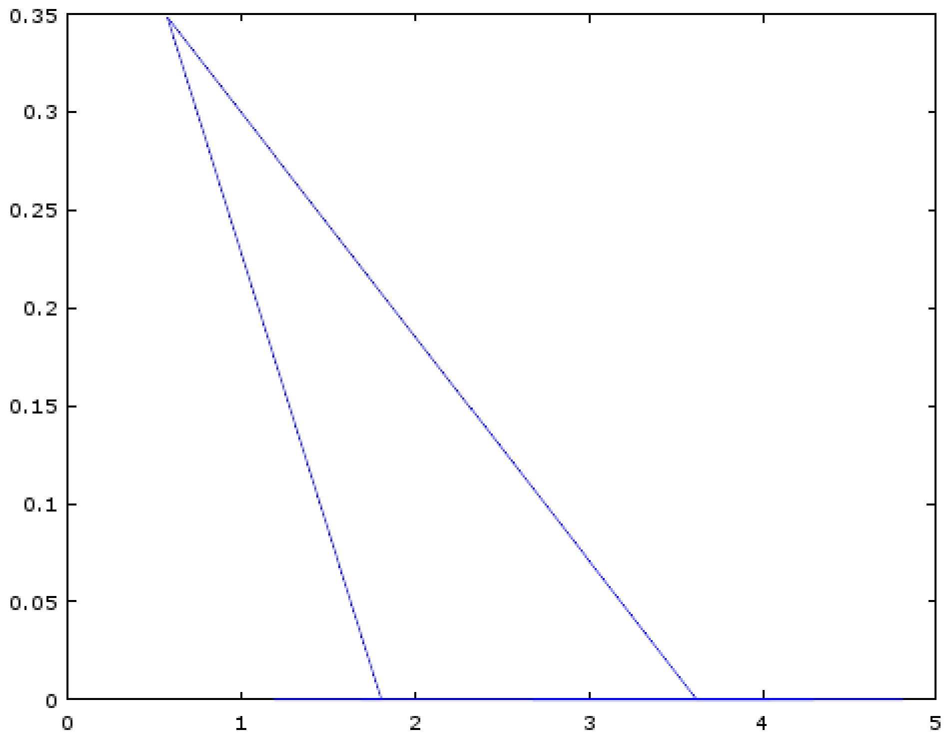
-- shown below (3) with the rest of the graphs

3. Do the same for other values of the hyperparameter such as Mu = 10 and Mu = 1.

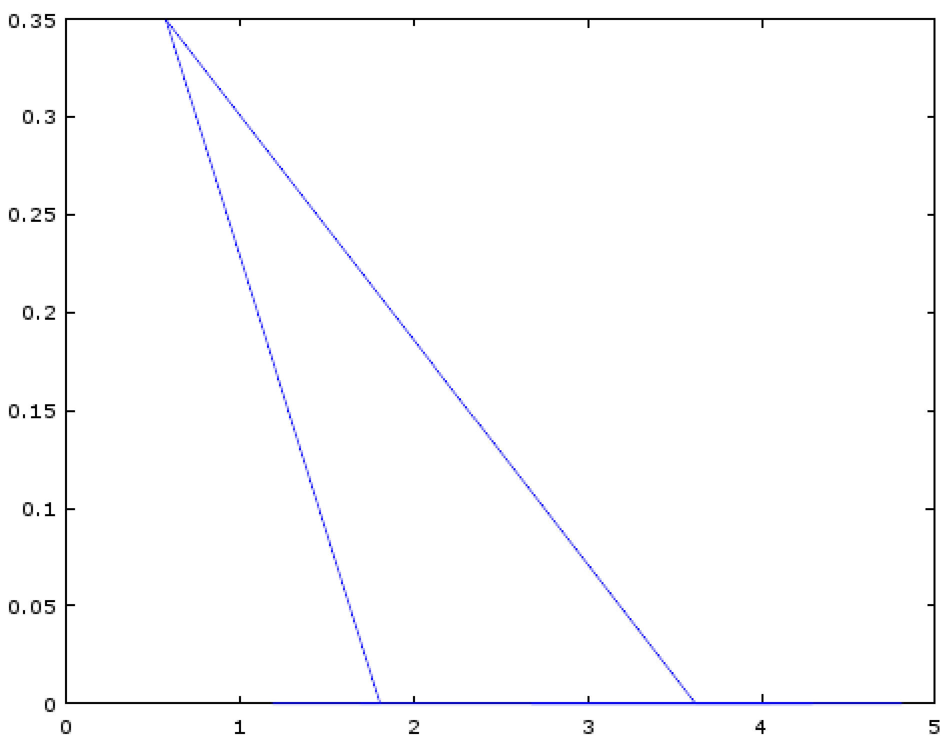


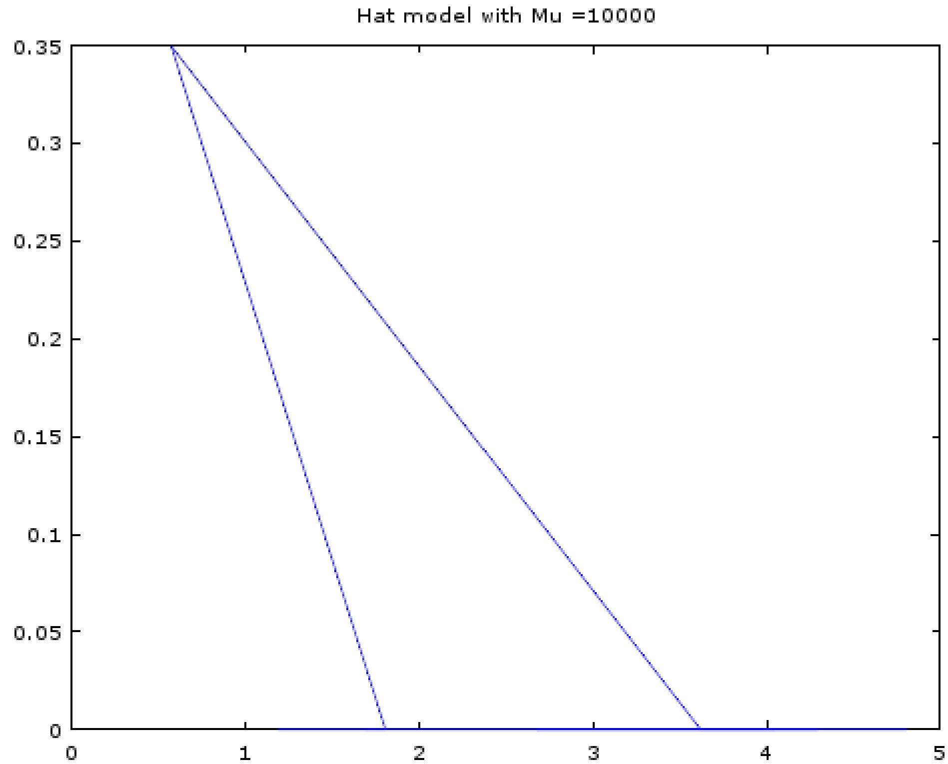


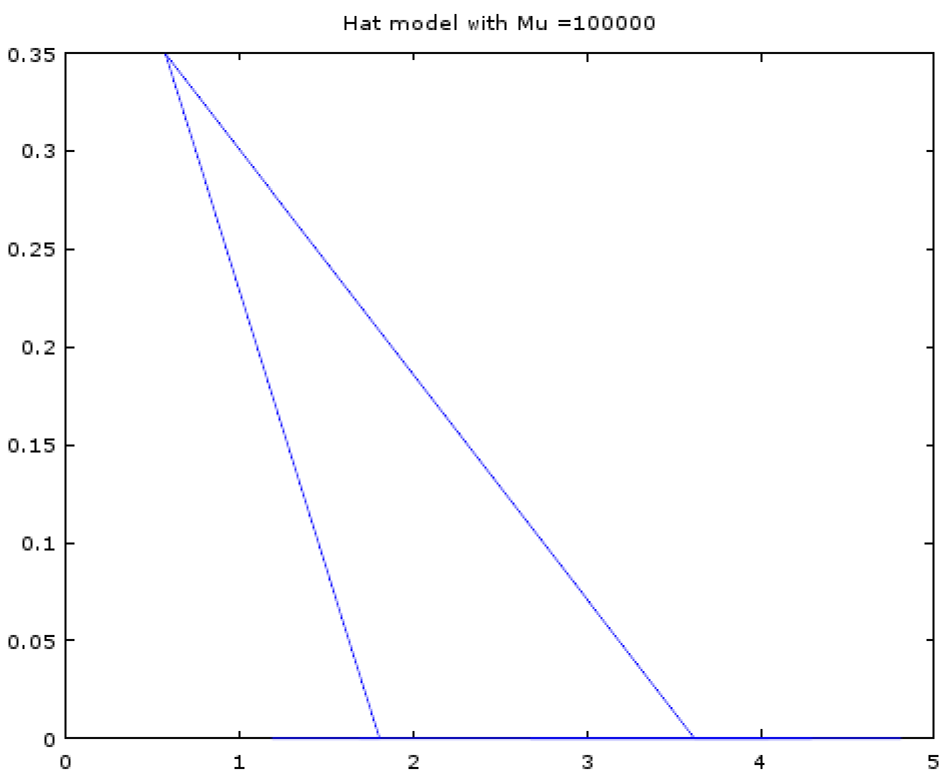
Hat model with Mu =100



Hat model with Mu =1000




Hat model with Mu =10000
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
1
2
3
4
5


Hat model with Mu =100000
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
1
2
3
4
5

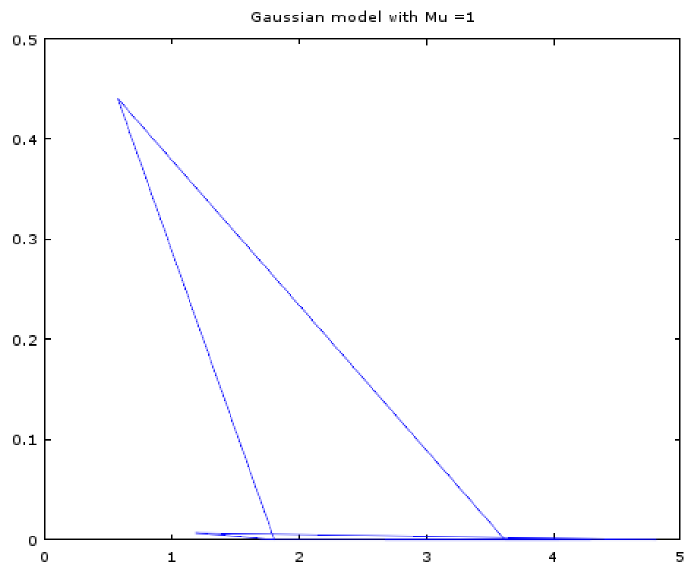
4. (2) What, if anything, is the hyperparameter controlling? How does it impact the solution to the problem?

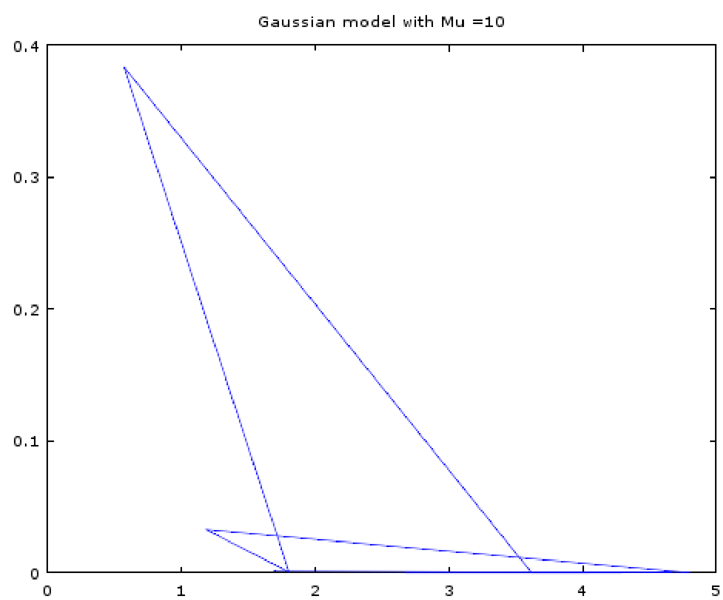
----The hyperparameter doesn't seem to influence the graph at all for large Mu, but has somewhat of an influence at lower Mu and great influence at Mu = 1. It appears to impact the solution to the problem by determining how much spread the solution has – of course, this is only for the cases where small Mu is used as mentioned just now. For large Mu, it seems to not affect the solution at all.

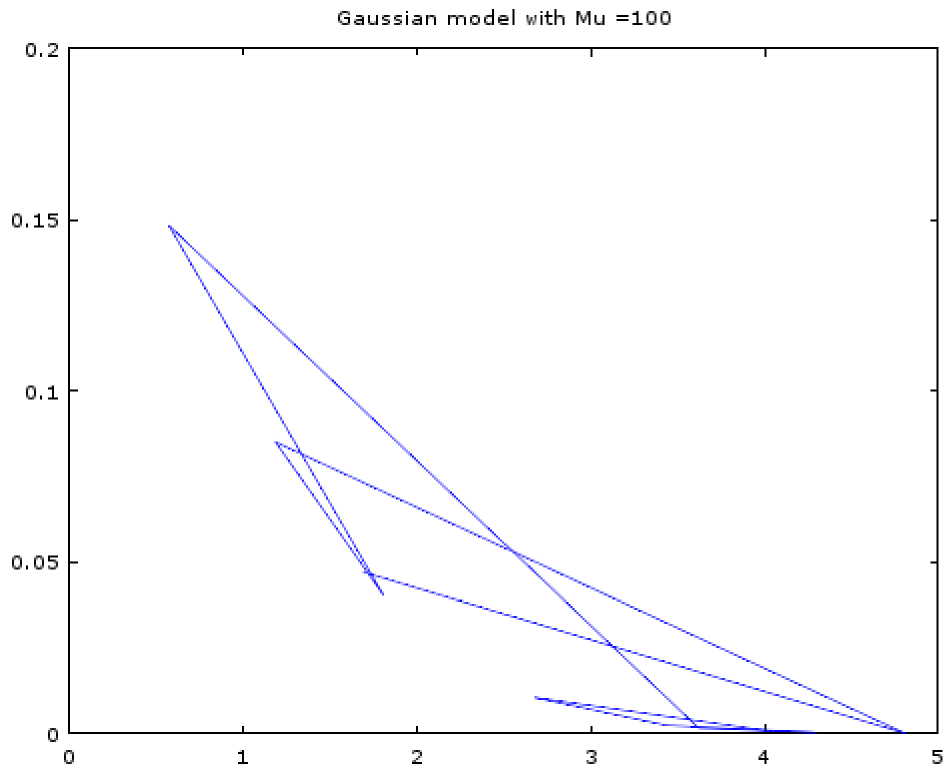
5. (2) Train the model with M = 10 Gaussian basis functions and Mu = 10^5, then plot and turn in the learned model evaluated (the function fitted to the data) on the interval [0; 2_].

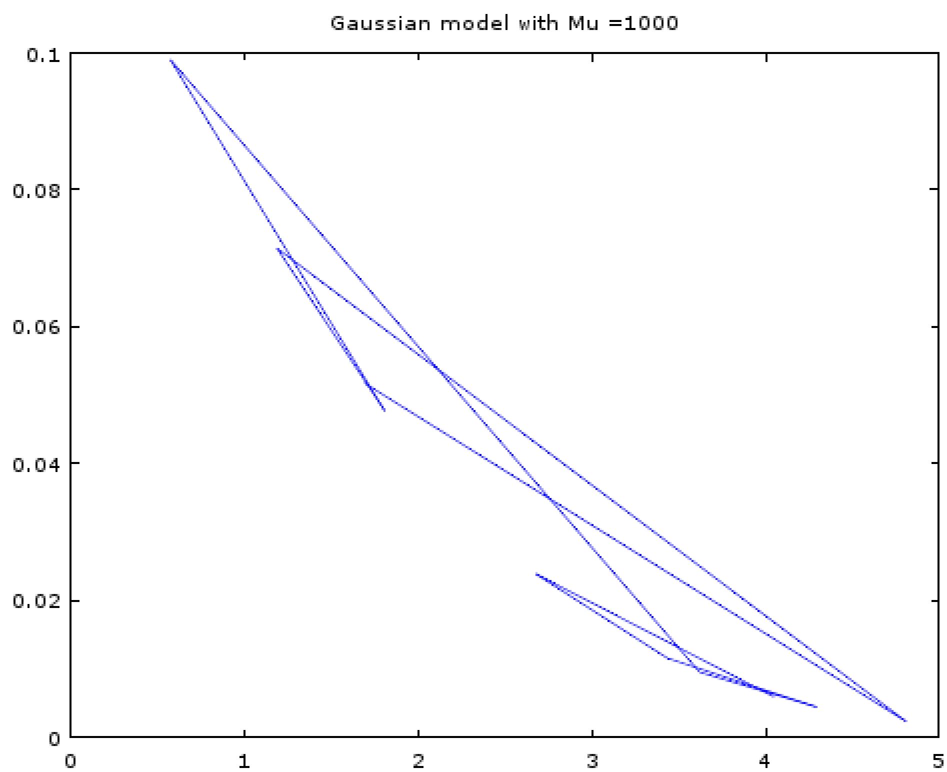
--- shown below (6) with the rest of the graphs

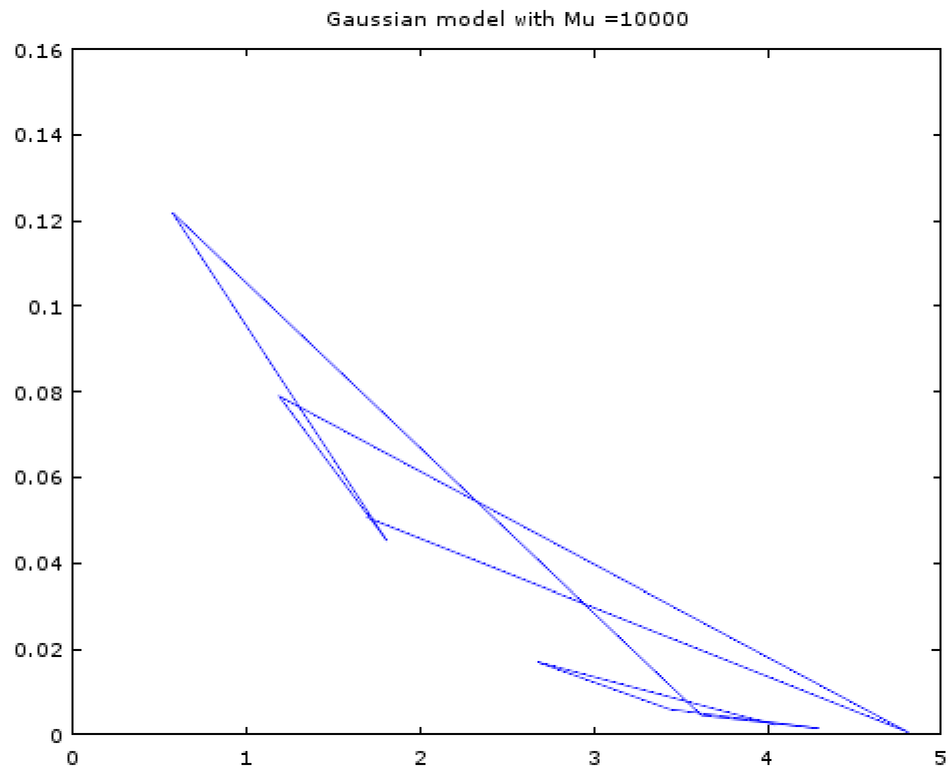
6. As before, explore other values of the hyperparameter such as Mu = 10 and Mu = 1.

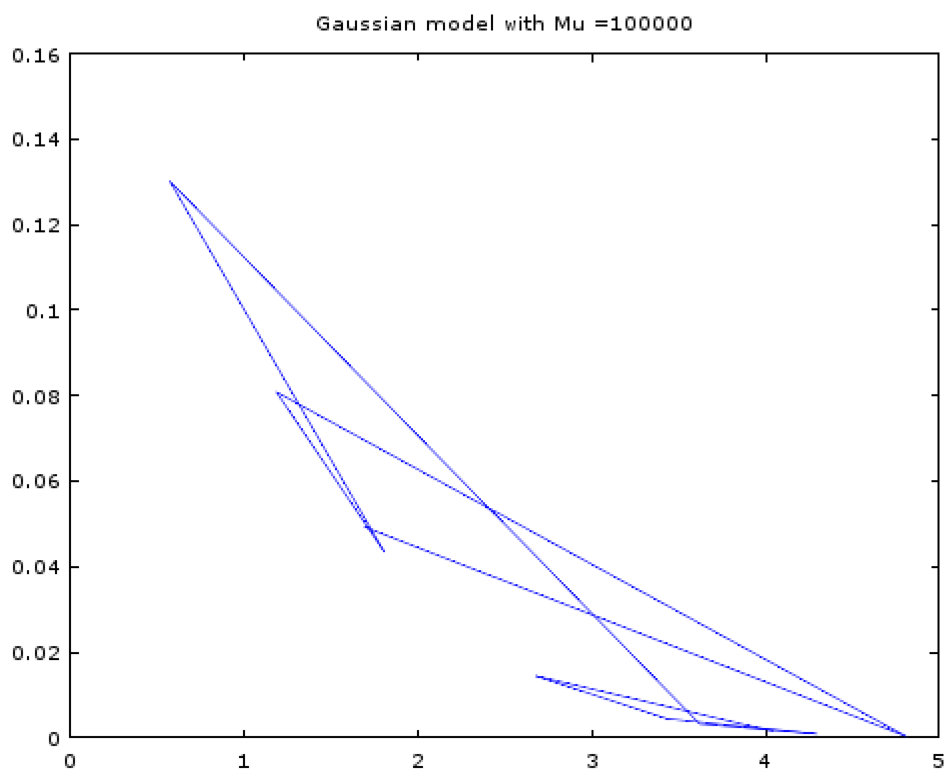





Gaussian model with Mu =100
0.2
0.15
0.1
0.05
0
0
1
2
3
4
5


Gaussian model with Mu =1000
0.1
0.08
0.06
0.04
0.02
0
0
1
2
3
4
5

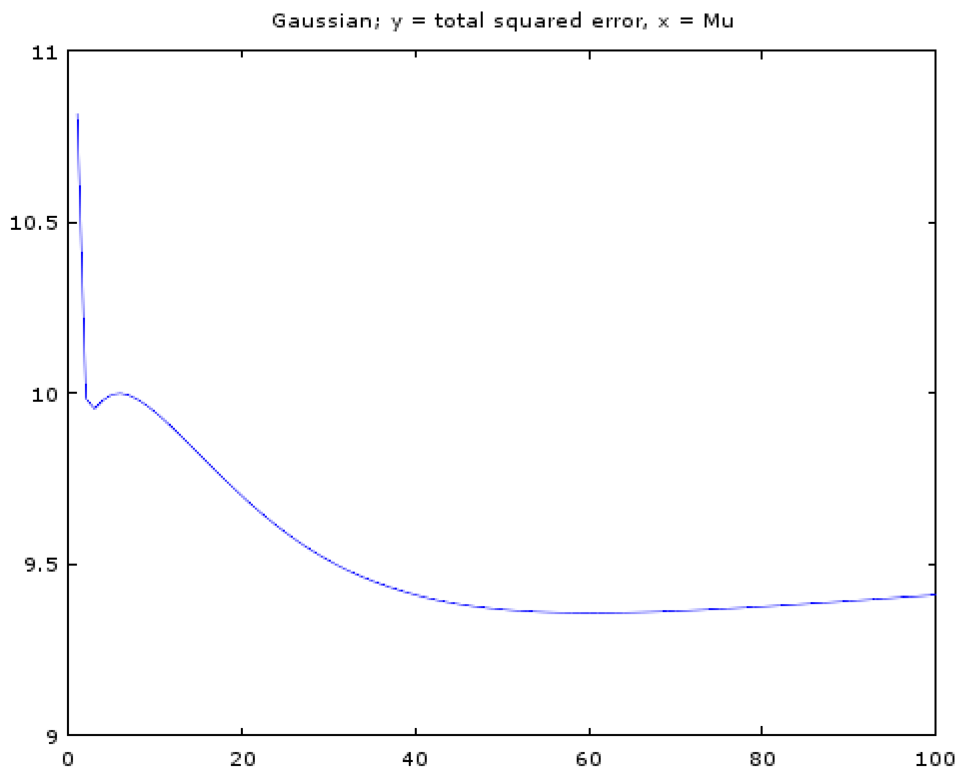




7. (2) What, if anything, is the hyperparameter controlling in this case? How does the dependence of the solution on the hyperparameter differ for this basis when compared to the hat basis?

---- The hyperparameter also seems to be influencing the spread in this case – for larger Mu, the spread of the solution has more verticality in general, while at smaller Mu, the solution is constained to less values on the bottom side of the vertical y spectrum. The solution here seems to be considerably more dependent on the hyperparameter, as it affects how far the solution can go vertically at all levels of Mu, while for the hat basis it only affected it for the small Mu.

8. (4) Fit the model with values of Mu in the range [1; 100] and using a Gaussian basis of ten elements on the data in simple.mat. For each model, calculate the squared error for the observations in test.mat (therefore testing the model). Generate and turn in a plot with Mu on the x-axis and the total squared model error on the test data along the y-axis.



9. (2) What value of Mu, when trained on the data in simple.mat, performs best on the data in test.mat? How do you know? Explain the shape of the plot you generated in the previous step.

---- I found that Mu = 60 performs best on the data in test.mat. I know this because as we plot Mu against the total squared error, the minimum squared error occurs when Mu = 60. This error is approximately around 9.385. The shape of the graph takes on massive error for small Mu; this might be because the solution has less positive values to center about on each side where Mu is. As Mu increases, the solution has more positive values to spread about on both sides of Mu, allowing the error to decrease suddenly and then level out.

10. (4) Repeat the process now fixing Mu to be the optimal value you found and varying the number of basis
elements from 1 to 100. Generate and turn in a plot with the number of basis elements on the x-axis and the error for the test data on the y-axis.