HW4

Computer Exercise

**Part a**

In Solution PDF

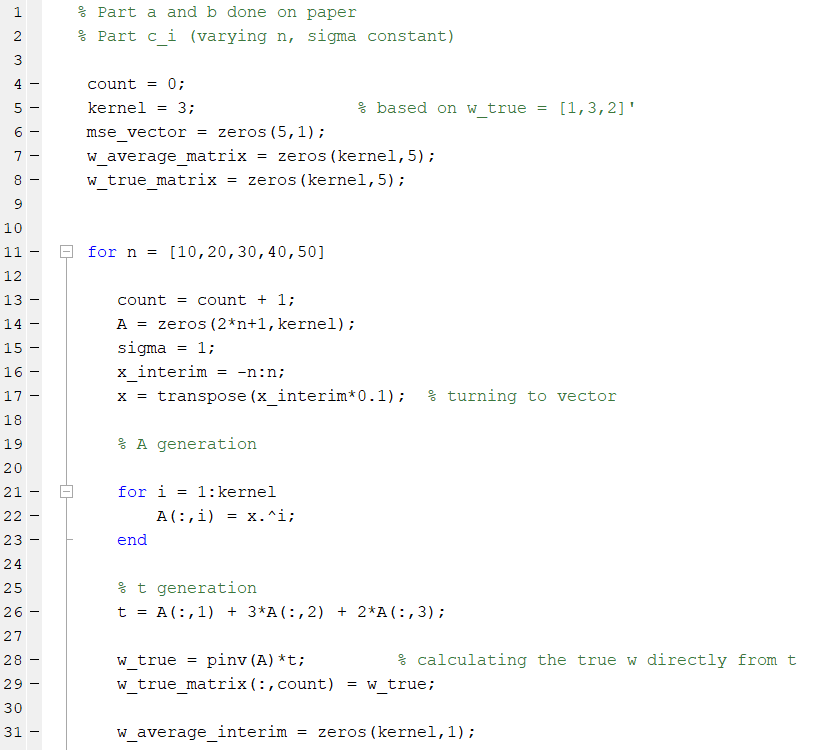
**Part b**

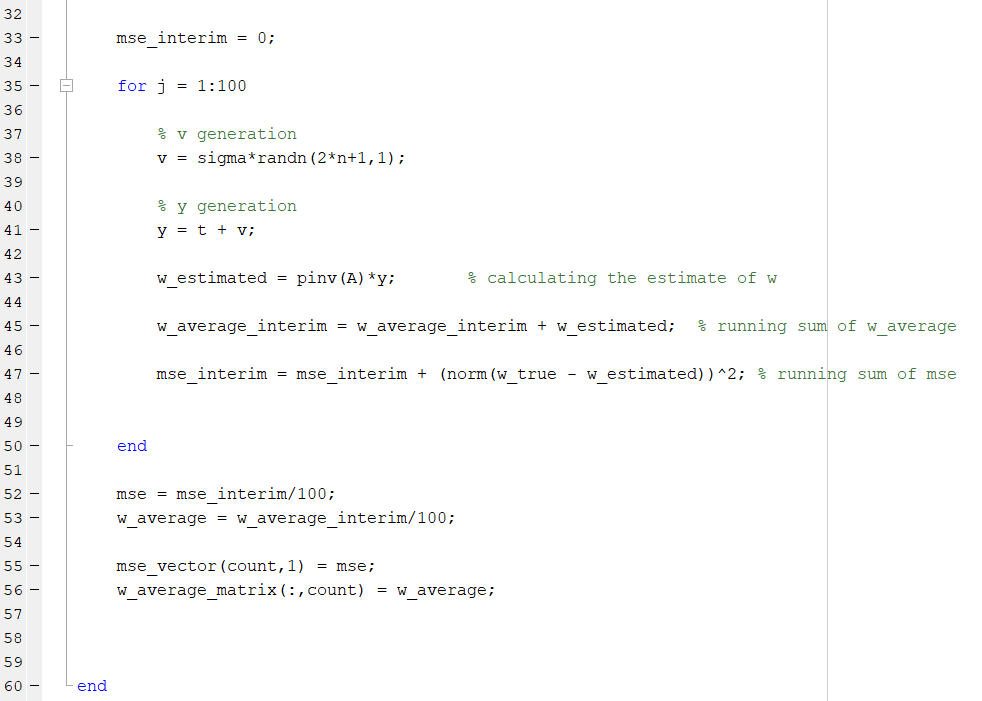
In Solution PDF

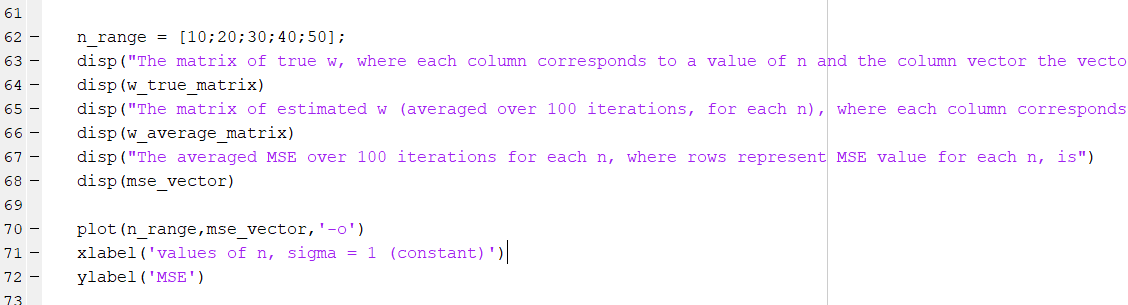
**Part c**

**i. Varying n, sigma constant**

Code

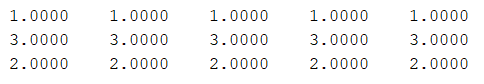






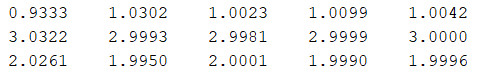
Results

The matrix of true w, where each column corresponds to a value of n and the column vector the vector of w, is,



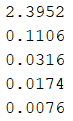
As expected, it is [1,3,2]

The matrix of estimated w (averaged over 100 iterations, for each n), where each column corresponds to a value of n and the column vector the vector of averaged w estimate, is,

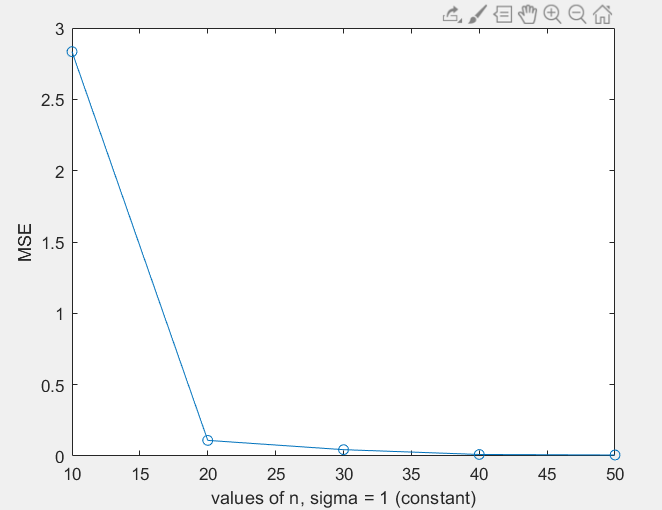


As n increases, the average (over 100 trials) of estimate of **w** becomes more accurate.

The averaged MSE over 100 iterations for each n, where rows represent MSE value for each n, is,



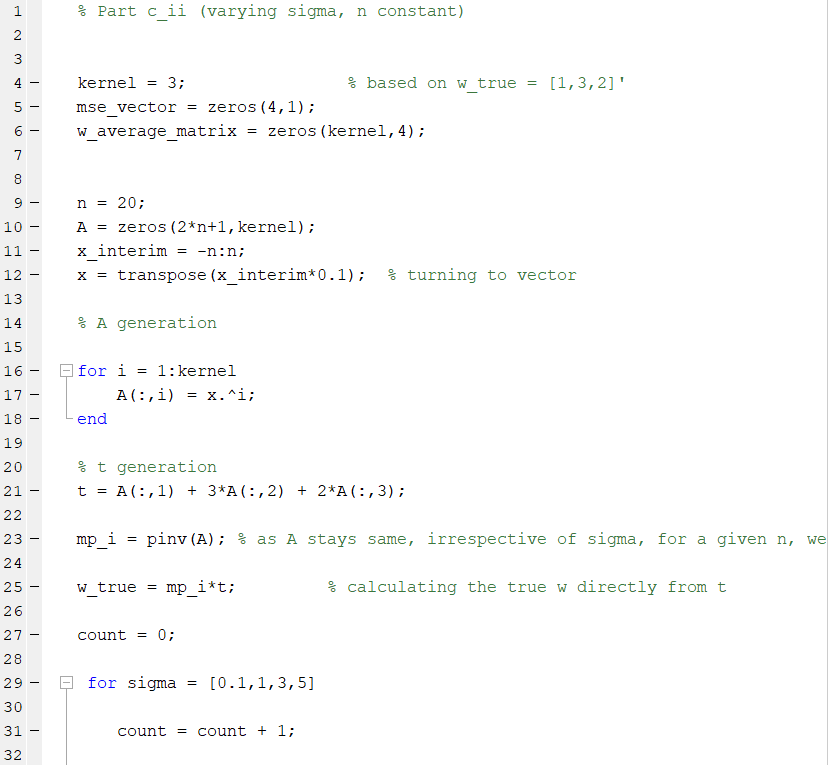
The MSE values also reduce as the n value increases.

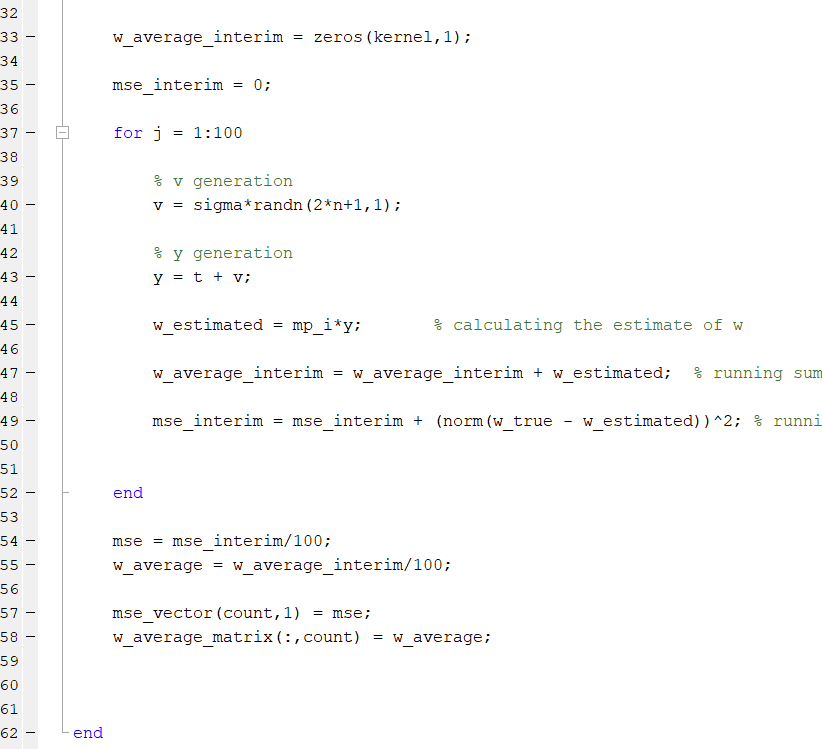


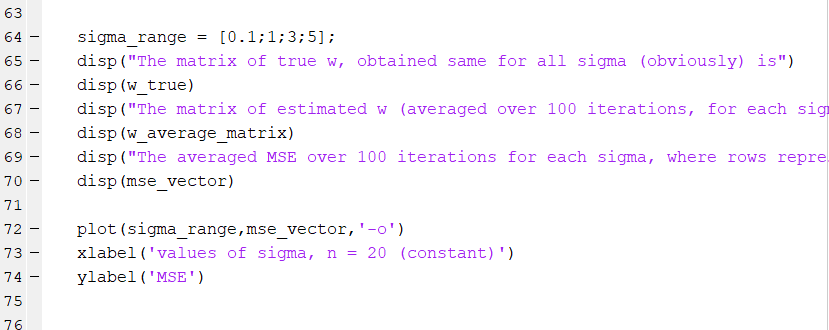
The above shows the MSE versus n plot with sigma constant

**ii. Varying sigma, n constant (n = 20)**

Code







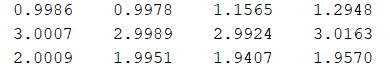
The code is pretty similar to c, Part i.

Results

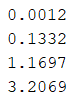
The matrix of true w, obtained same for all sigma (obviously) is,



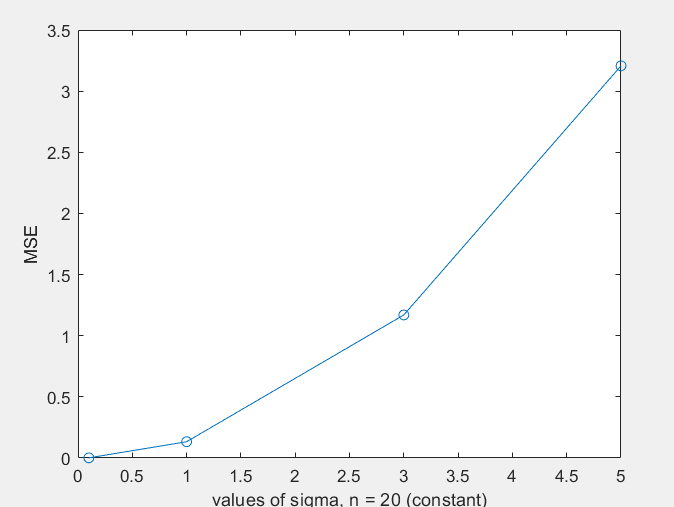
The matrix of estimated w (averaged over 100 iterations, for each sigma), where each column corresponds to a value of sigma and the column vector the vector of averaged w estimate, is,



The averaged MSE over 100 iterations for each sigma, where rows represent MSE value for each sigma, is,



The curve is,



As expected, as the variance of the added noise increases, the MSE increases and the estimate of **w** becomes more and more inaccurate.

Variance of z\*sigma = (sigma)2 Variance of z = (sigma)2

Since z is standard normal, with zero mean and unit variance.

Thus, as the value of added distortion to **t** increases, error in **w**, increases.

y = t + v = t + z\*sigma

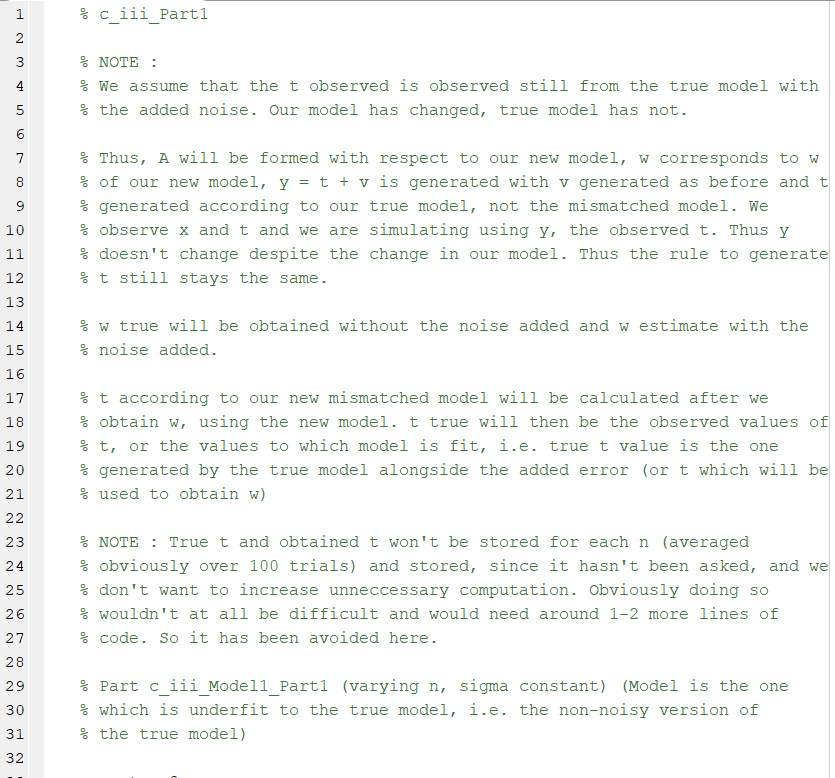
Thus, by increasing sigma, we increase both the variance (how much the noise varies) as well as the overall value of the noise (the probability of larger values appearing in noise term increases as the value of sigma increases).

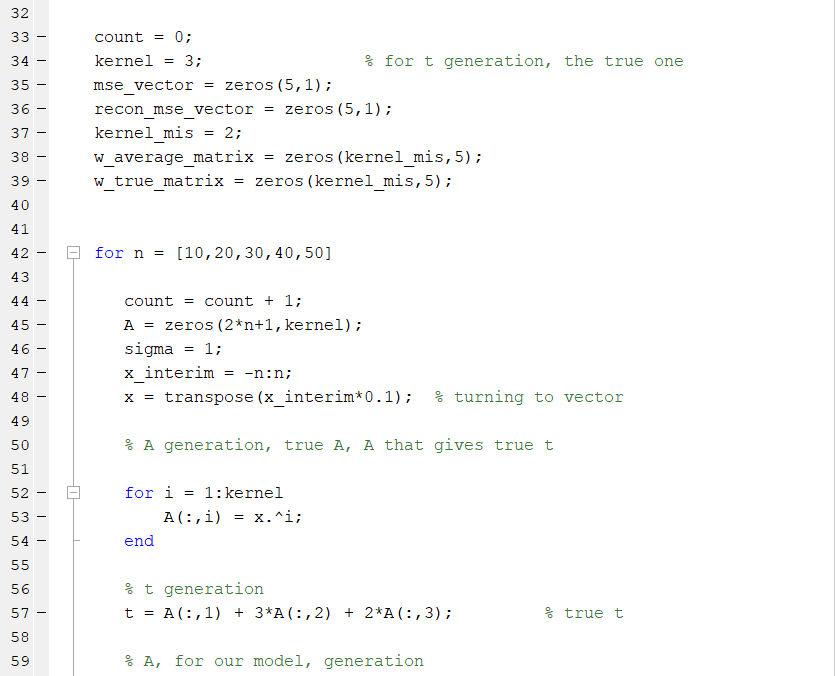
Thus, with increase in sigma, due to these reasons, MSE increases.

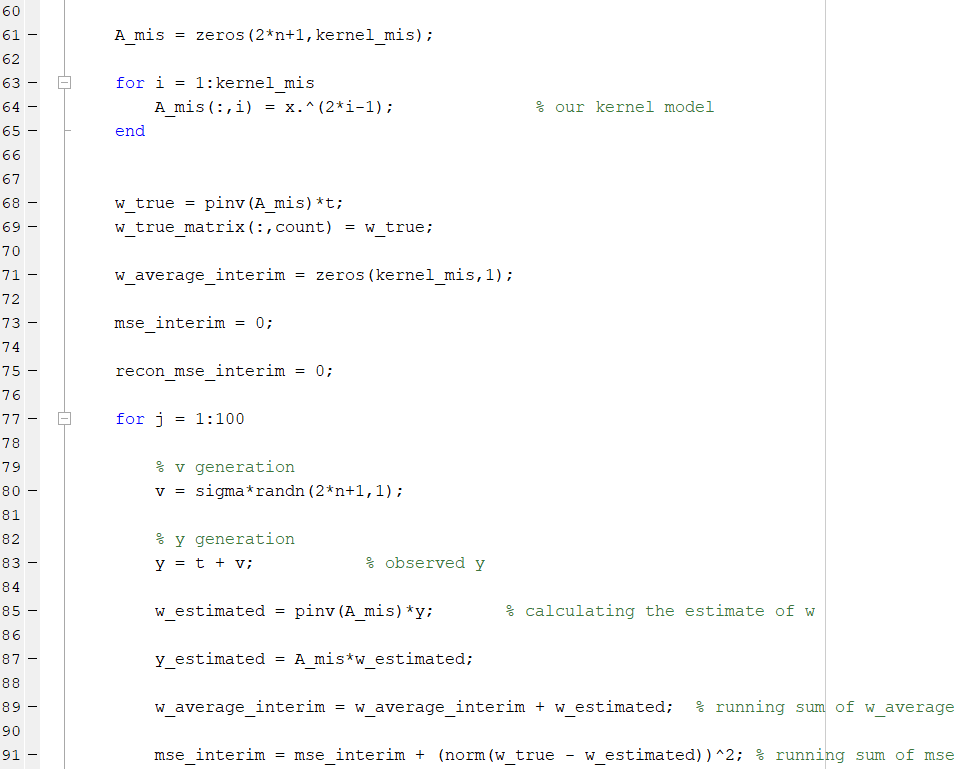
**iii. For Model with x and x3 (underfit)**

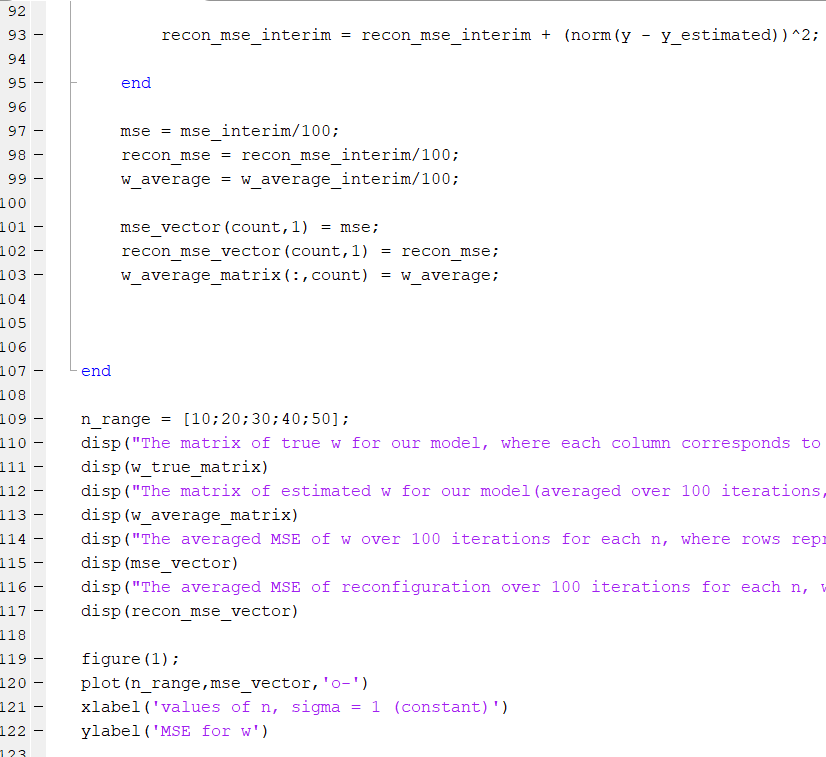
**PART 1 – Varying n with sigma constant**

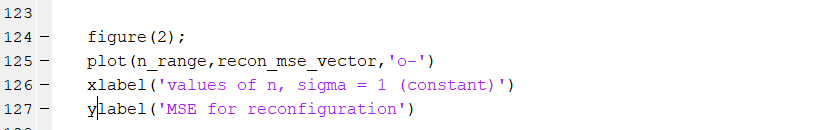
Code











Results

>> c\_iii\_Model1\_Part1

The matrix of true w for our model, where each column corresponds to a value of n and the column vector the vector of w, is,

1.0000 1.0000 1.0000 1.0000 1.0000

2.0000 2.0000 2.0000 2.0000 2.0000

The matrix of estimated w for our model (averaged over 100 iterations, for each n), where each column corresponds to a value of n and the column vector the vector of averaged w estimate, is,

0.9973 0.9585 1.0251 1.0157 0.9945

1.9959 2.0160 1.9950 1.9988 2.0002

The averaged MSE of w over 100 iterations for each n, where rows represent MSE of w value for each n, is,

2.3247

0.1036

0.0395

0.0149

0.0052

The averaged MSE of reconfiguration over 100 iterations for each n, where rows represent MSE of reconfiguration value for each n, is,

1.0e+05 \*

0.0006

0.0134

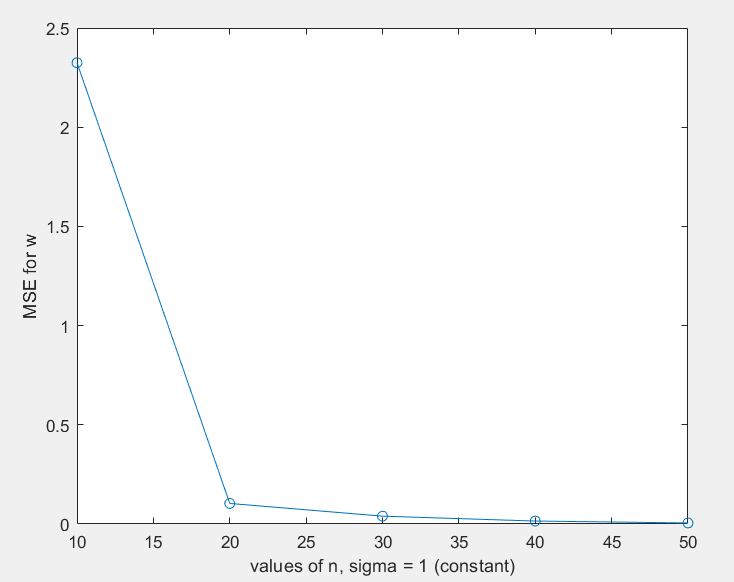
0.0955

0.3931

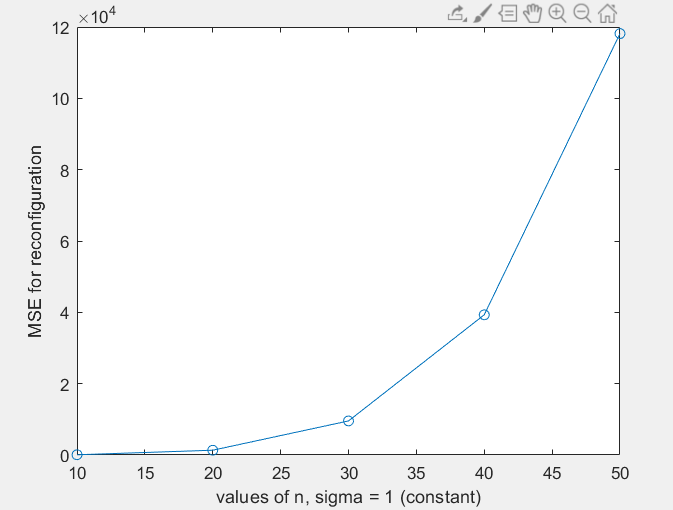
1.1815

Plots

MSE for w



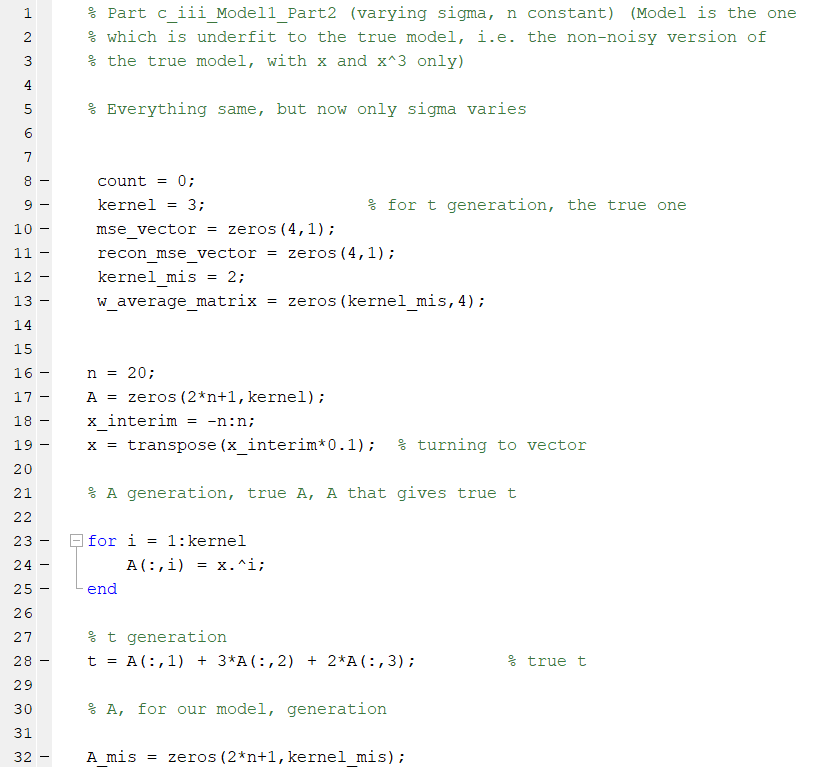
MSE for Reconfiguration

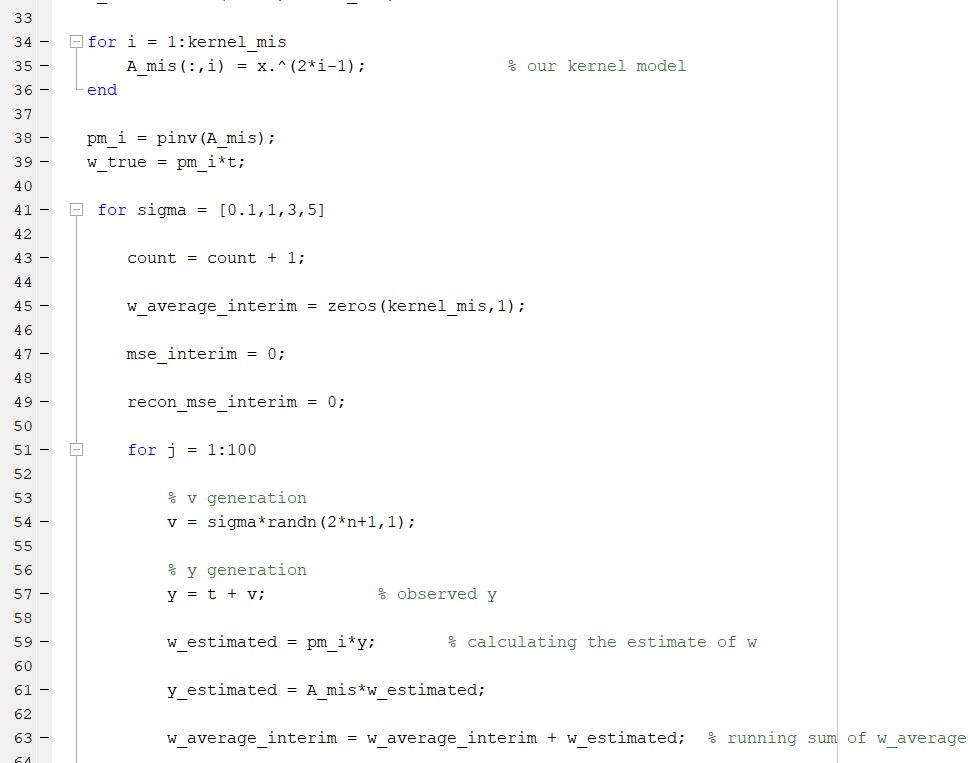


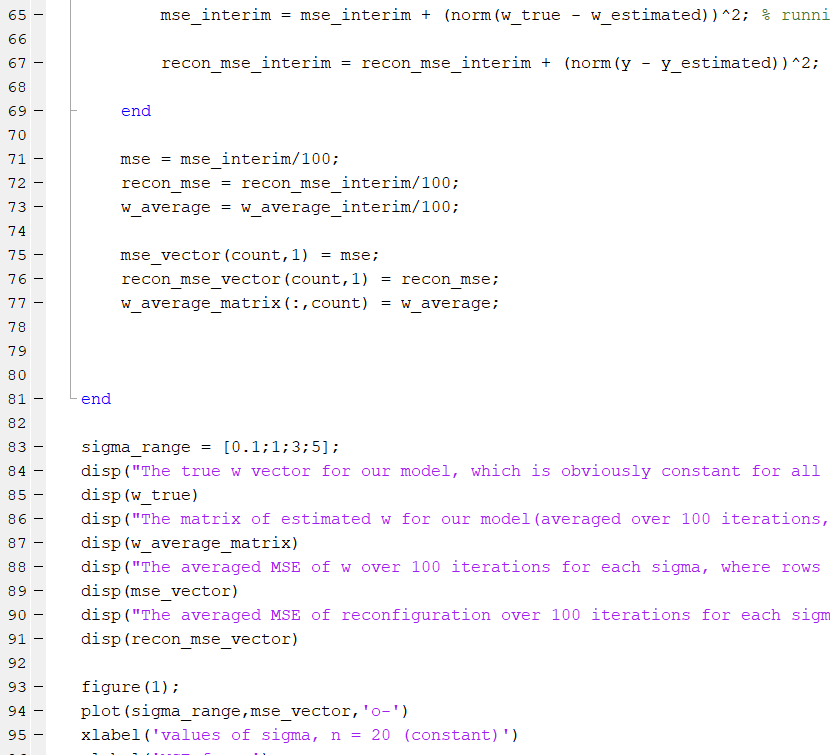
The MSE for reconfiguration is extremely large, since, the coefficients of x and x3 are still close to 1 and 2 as before, but now, we have the x2 factor missing due to which our estimated y deviates largely from our actual y in general.

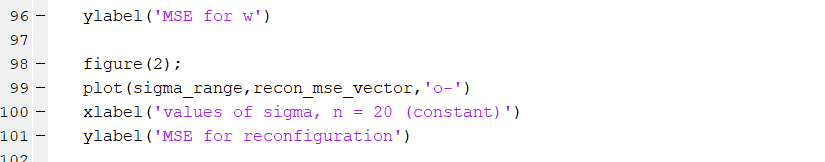
**PART 2 – Varying sigma with n constant, same model**

Code









Results

>> c\_iii\_Model1\_Part2

The true w vector for our model, which is obviously constant for all sigma, since n is constant is

1.0000

2.0000

The matrix of estimated w for our model(averaged over 100 iterations, for each sigma), where each column corresponds to a value of sigma and the column vector the vector of averaged w estimate, is

0.9991 0.9937 0.8778 0.7348

2.0004 2.0037 2.0665 2.1051

The averaged MSE of w over 100 iterations for each sigma, where rows represent MSE of w value for each sigma, is

0.0010

0.1453

1.0021

2.7495

The averaged MSE of reconfiguration over 100 iterations for each sigma, where rows represent MSE of reconfiguration value for each sigma, is

1.0e+03 \*

1.3007

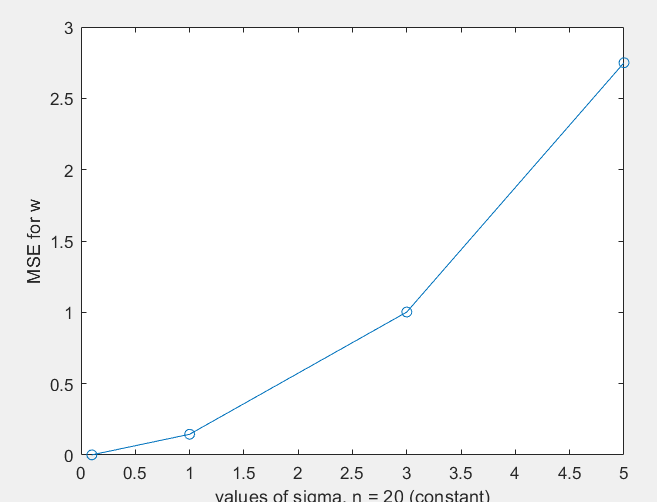
1.3314

1.6466

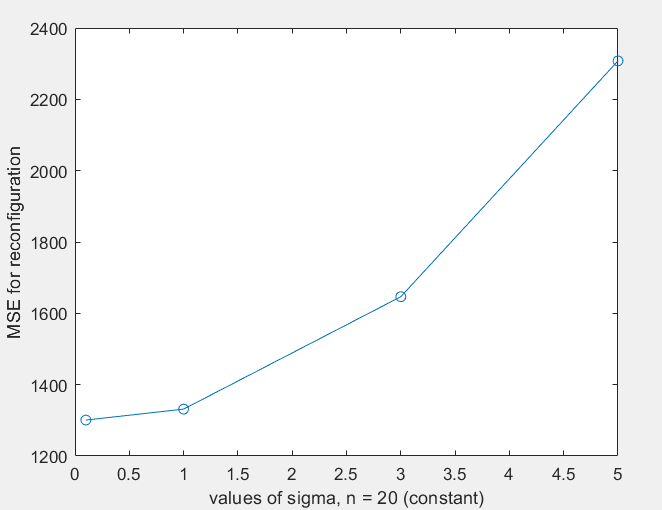
2.3078

Plots

MSE for w



MSE for Reconfiguration

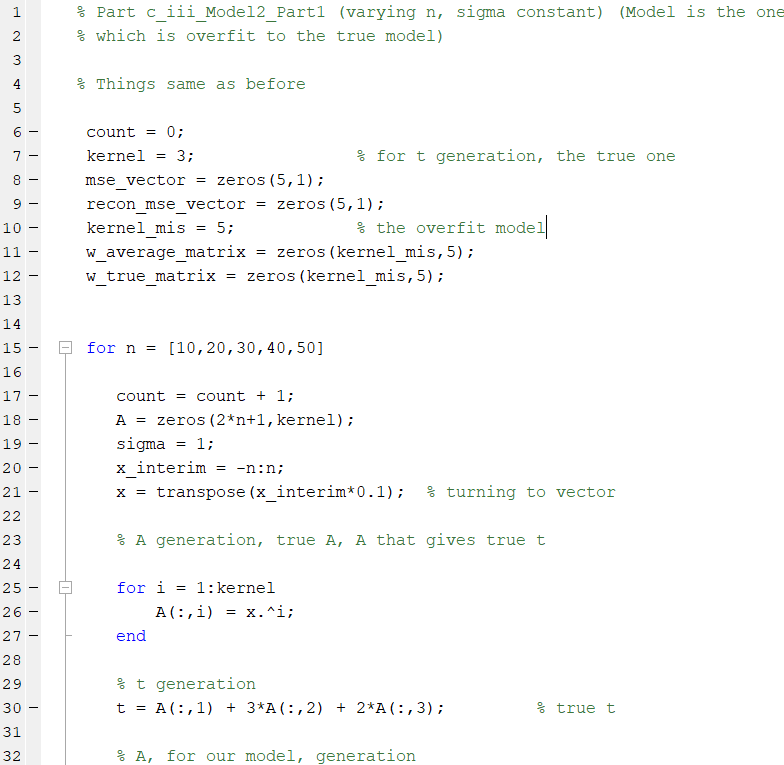


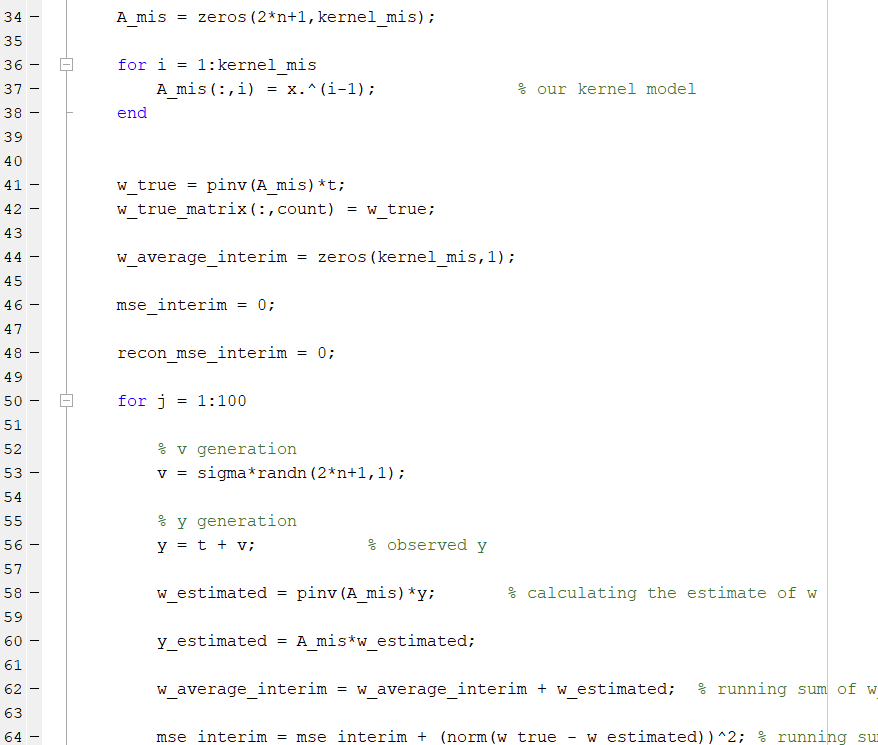
As seen before, as sigma increases, both the MSE has a tendency to increase. The same previous explanation goes here as well.

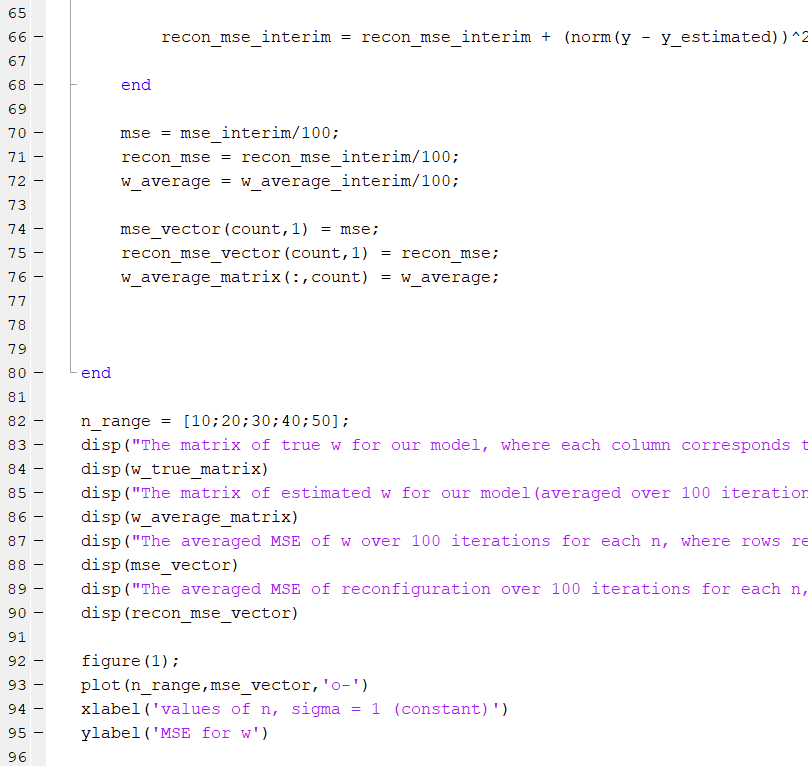
**iii. For Model with terms till and including x4 (overfit)**

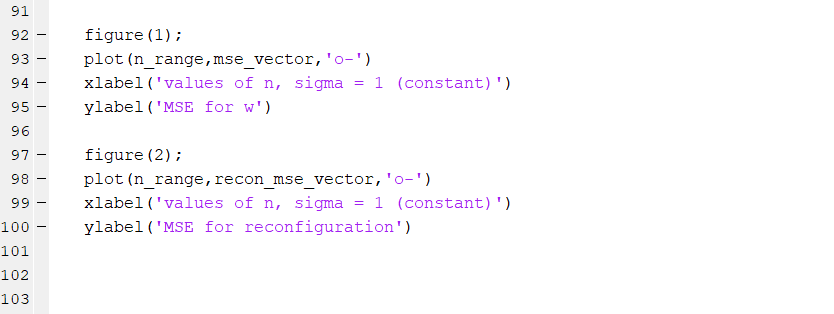
**PART 1 – Varying n with sigma constant**

Code









Results

>> c\_iii\_Model2\_Part1

The matrix of true w for our model, where each column corresponds to a value of n and the column vector the vector of w, is

0.0000 0.0000 -0.0000 0.0000 -0.0000

1.0000 1.0000 1.0000 1.0000 1.0000

3.0000 3.0000 3.0000 3.0000 3.0000

2.0000 2.0000 2.0000 2.0000 2.0000

0.0000 0.0000 -0.0000 -0.0000 -0.0000

The matrix of estimated w for our model(averaged over 100 iterations, for each n), where each column corresponds to a value of n and the column vector the vector of averaged w estimate, is

-0.0809 -0.0246 -0.0109 -0.0084 0.0028

0.9675 1.0193 0.9783 1.0082 0.9960

3.3731 3.0623 3.0005 2.9953 2.9994

2.0593 1.9947 2.0051 1.9995 2.0006

-0.2965 -0.0184 -0.0003 0.0005 -0.0000

The averaged MSE of w over 100 iterations for each n, where rows represent MSE of w value for each n, is

14.9549

0.4332

0.1175

0.0672

0.0372

The averaged MSE of reconfiguration over 100 iterations for each n, where rows represent MSE of reconfiguration value for each n, is

14.9236

35.6173

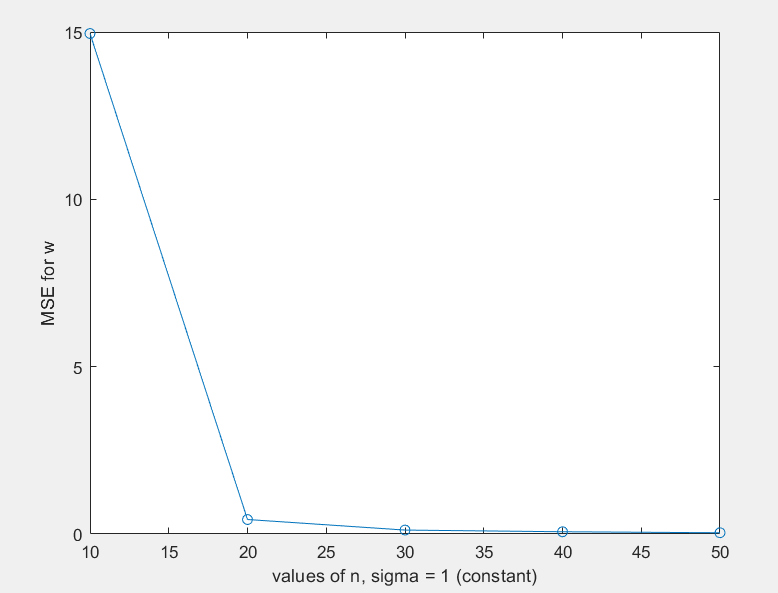
56.9790

76.7711

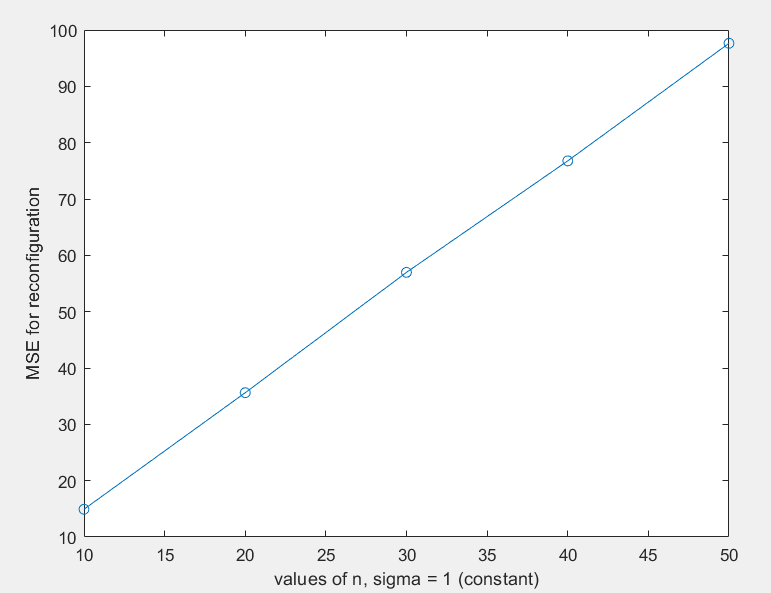
97.6454

Plots

MSE for w



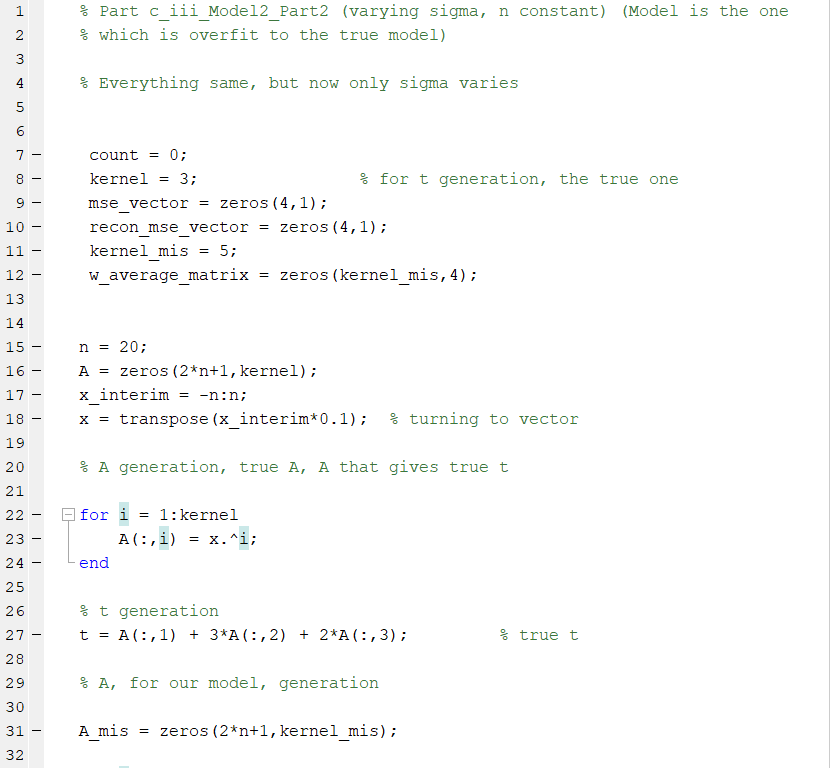
MSE for Reconstruction

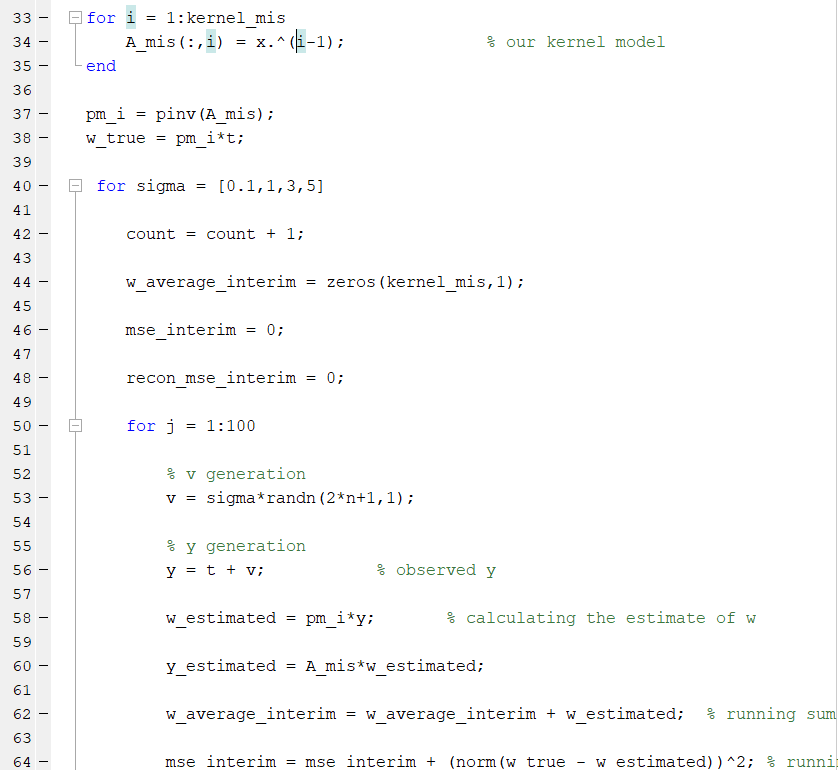


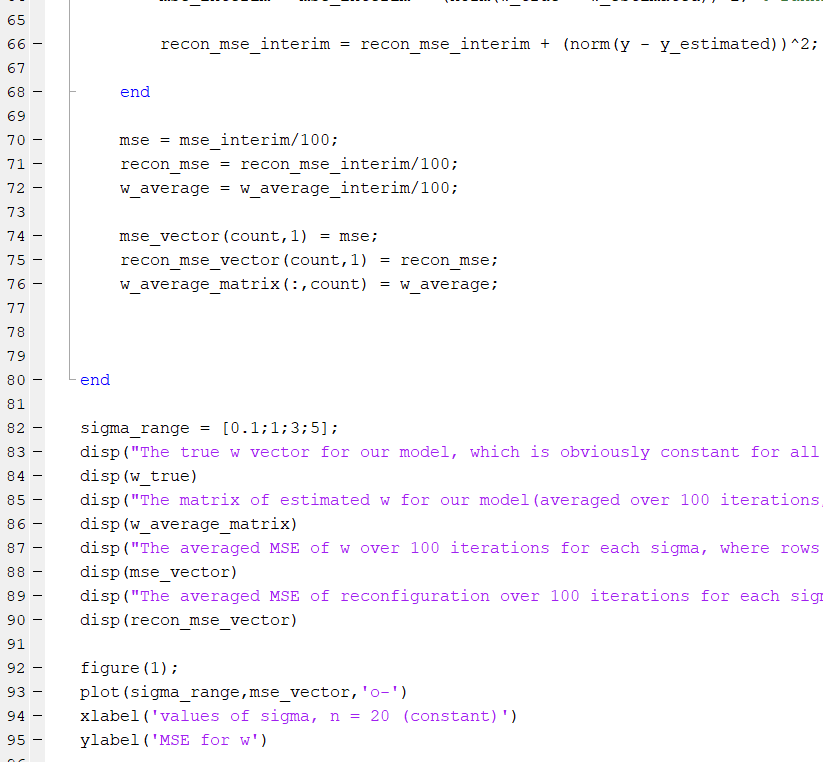
As can be seen the MSE for the overfit case is much less than that for the underfit case. Hence, it is obvious that if MSE of reconstruction is our main criteria than overfit model should be selected since it gives us an MSE which is extremely less in comparison to the underfit model.

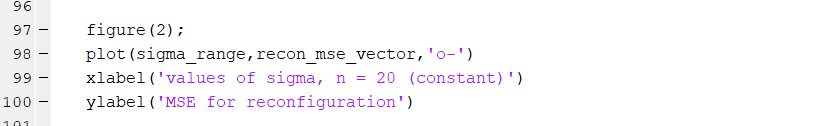
**PART 1 – Varying sigma with n constant**

Code









Results

>> c\_iii\_Model2\_Part2

The true w vector for our model, which is obviously constant for all sigma, since n is constant is

0.0000

1.0000

3.0000

2.0000

0.0000

The matrix of estimated w for our model(averaged over 100 iterations, for each sigma), where each column corresponds to a value of sigma and the column vector the vector of averaged w estimate, is

-0.0032 -0.0155 0.1085 -0.0525

1.0011 1.0133 0.8659 1.0348

3.0037 3.0665 2.9011 3.0538

2.0001 2.0015 2.0399 1.9652

-0.0007 -0.0222 0.0186 -0.0108

The averaged MSE of w over 100 iterations for each sigma, where rows represent MSE of w value for each sigma, is

0.0043

0.3726

3.6304

10.5548

The averaged MSE of reconfiguration over 100 iterations for each sigma, where rows represent MSE of reconfiguration value for each sigma, is

0.3584

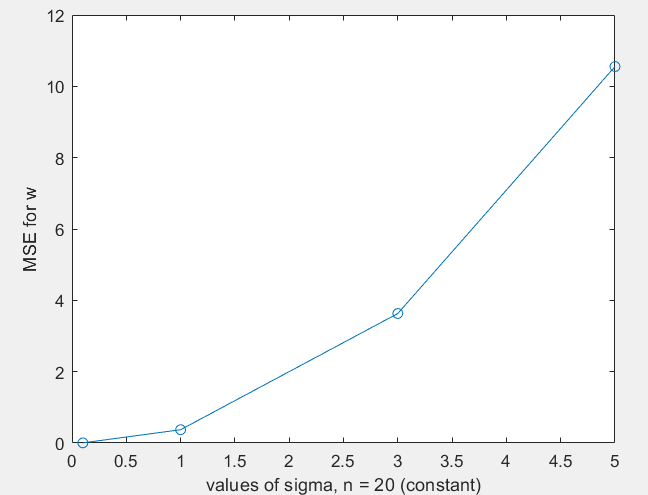
37.0618

318.1502

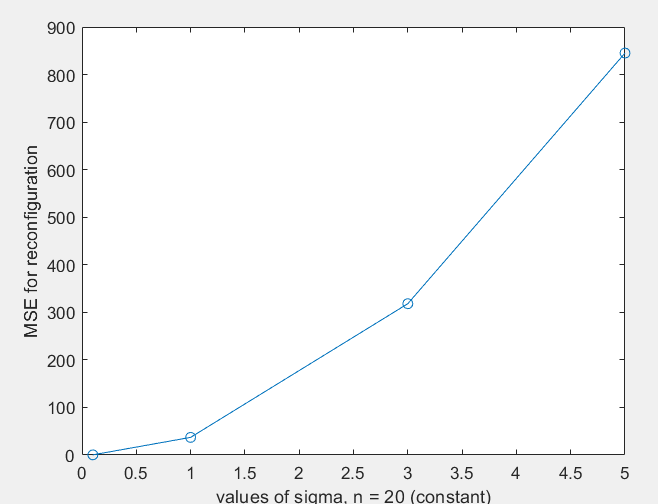
845.2264

Plots

MSE for w



MSE for Reconstruction



Similar results with similar explanation

**iv. Comment on Results**

It is definitely worse, MSE wise, to underestimate/underfit the model. Thus, the underestimated/underfit model had an excruciatingly worse MSE compared to the overfit/overestimated version which had a comparatively much better MSE.

However, with overfitting we do have an issue where the model is so overfit, that we can’t have good model generalization to the test data, increasing the test MSE. This fact is also known as the bias-variance trade-off in machine learning algorithms.

Underfit model, have, higher training MSE but generalize better and may have low test MSE, sometimes even lower than the Overfit model.

-----------------------------END-------------------------------------------------------------