Indian Institute of Technology, Madras
CS6464: Concepts In Statistical Learning Theory
Assignment-I Report

Rajan Kumar Soni - CS18S038

February 2019

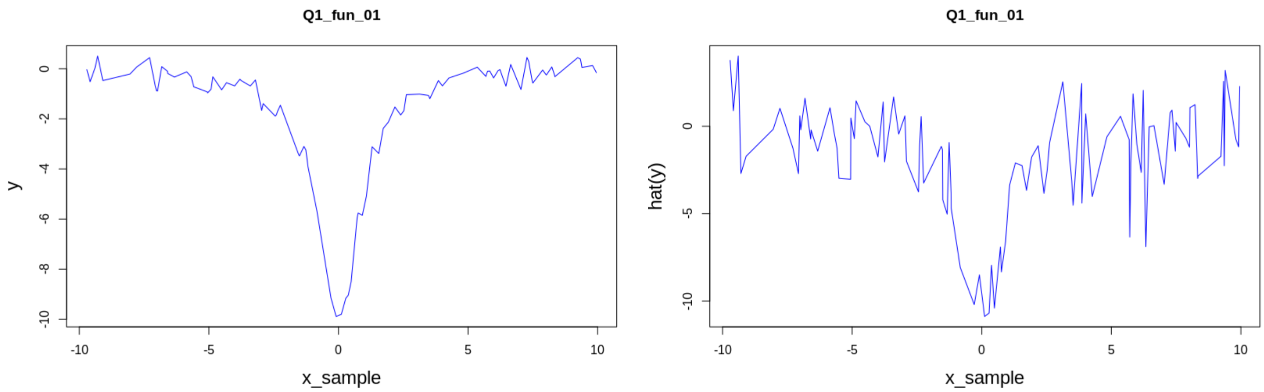
Contents

1 Question 1

Original and Noisy Data (sigma is taken as 2 unless specified)

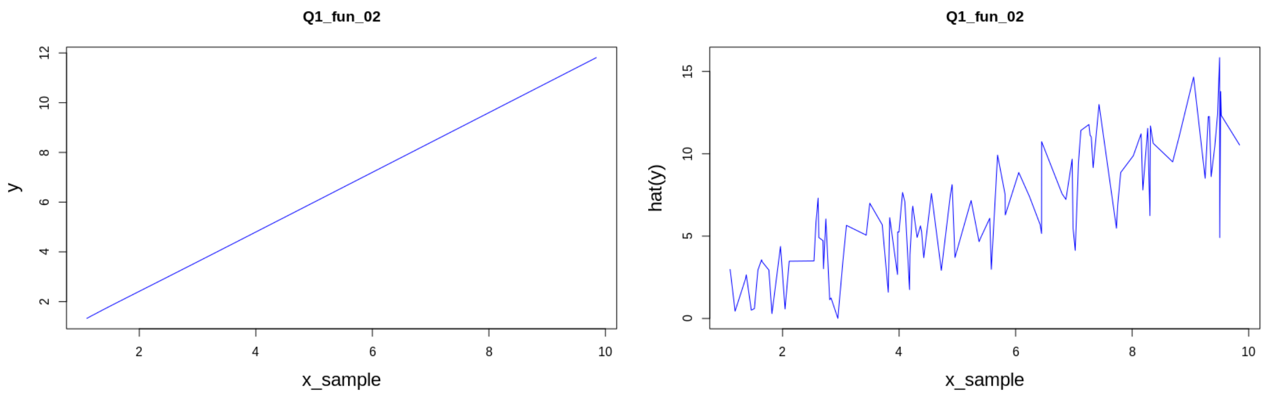
$$Y = f(X) + \epsilon$$

where $E[\epsilon] = 0$ and $V[\epsilon] = \sigma^2$.



(a) without noise (b) with noise $\sigma = 2$

Figure 1: plot of y = 0.3*cos(3*pi*x) - 0.4* cos(4*pi*x) - 10/(x**2 + 1)

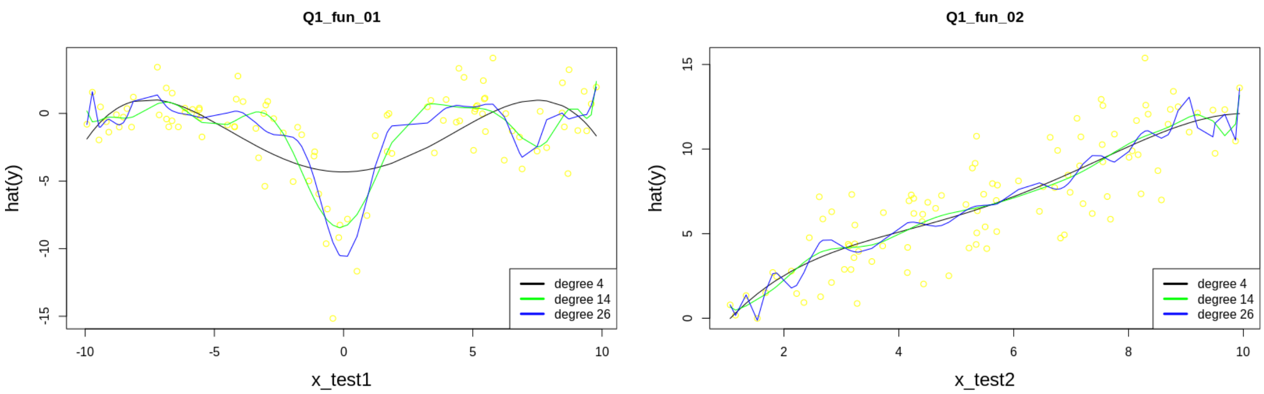


(a) without noise (b) with noise $\sigma = 2$

Figure 2: plot of y = 0.4*log(x**4+log(x-0.7)+exp(3*x))

Here first data is generated from actual function and the noise is added. The noise which is added is called irreducible error. This error occurs in nature because of assumptions. As we never have any idea about the true predictors on which true function, or it might be too large numbers of predictors.

1.1 Fitting the model on various degree on noisy data



(a) plot for Q1_fun_01

(b) plot for Q1_fun_02

Figure 3: Fitting the model on degree 4 , 14, 26

- As we can see from figure the figure as the degree keep on increasing models try to fit more and more.
- As the number of parameters increases or degree model learns complex functions.
- AS the degree increases model try to learn noise also which is not good for the model.
- If the model on increasing complexity learns the noise, i,e feature which is misleading also taken into consideration then it will not perform well on test.
- If the model has low degree then it will no bet able to learn data because of less parameters, i,e because less number of features would have been considered for fitting

1.2 Bias Variance for various degree

$$\widehat{\text{bias}}(\hat{f}(x)) = \frac{1}{n_{samples}} \sum_{j=1}^{n_{samples}} \frac{1}{n_{\text{sims}}} \sum_{i=1}^{n_{\text{stas}}} \left(\hat{f}_k(x) \right) - f(x)$$

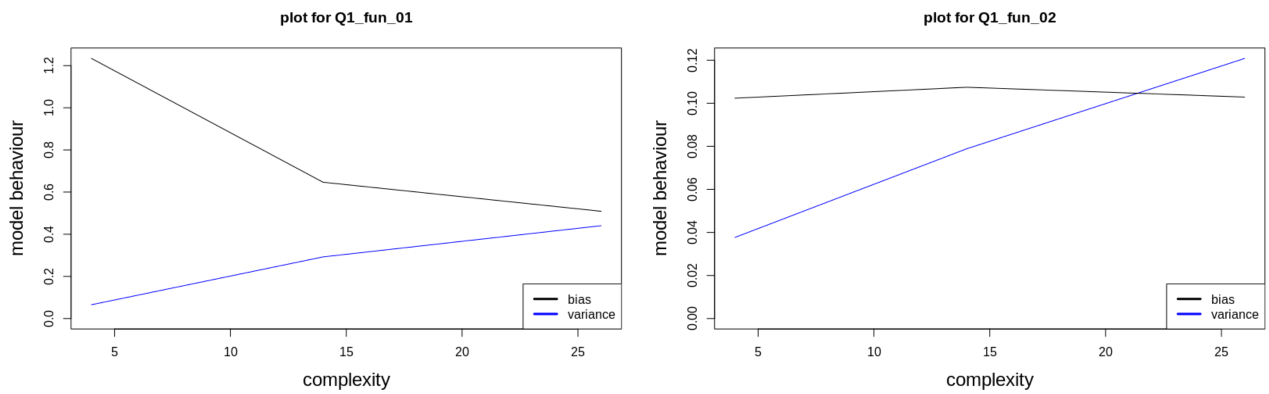
$$\widehat{\text{var}}(\hat{f}(x)) = \frac{1}{n_{samples}} \sum_{j=1}^{n_{samples}} \frac{1}{n_{\text{sins}}} \sum_{i=1}^{n_{\text{asam}}} \left(\hat{f}_k(x) - \frac{1}{n_{\text{sims}}} \sum_{i=1}^{n_{\text{num}}} \hat{f}_k(x) \right)^2$$

Bias variance of Q1_fun_01 for various degree		
Degree	Bias	Variance
4	1.276	0.0611
14	0.593	0.2737
26	0.506	0.401

Bias variance of Q1_fun_02 for various degree		
Degree	Bias	Variance
4	0.139	0.038
14	0.136	0.073
26	0.12	0.088

- From the above observation it is clear that as the degree increases bias decreases and variance increases.
- It is obvious from the above observation more the number of features more the complex model.

1.3 Bias- variance vs Complexity



(a) plot for Q1_fun_01

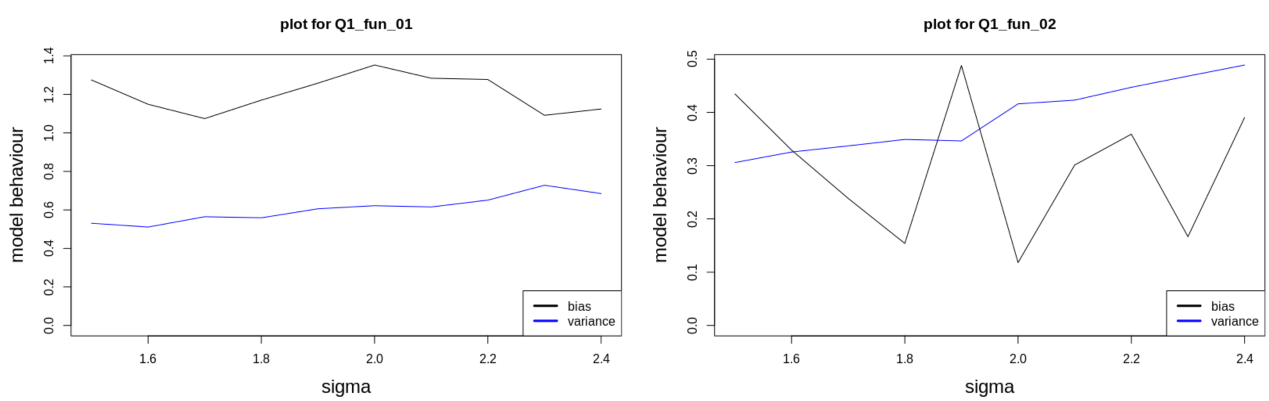
(b) plot for Q1_fun_02

Figure 4: Bias-Variance vs Model Complexity for $\sigma = 2$

- As discussed above it can be seen here bias is decreasing and variance is increasing as the degree of the polynomial increases or complexity increases

1.4 Plot of Bias-Variance vs Sigma

$$\text{MSE} = \text{E}_x \left\{ \text{Bias}_D[\hat{f}(x; D)]^2 + \text{Var}_D[\hat{f}(x; D)] \right\} + \sigma^2$$

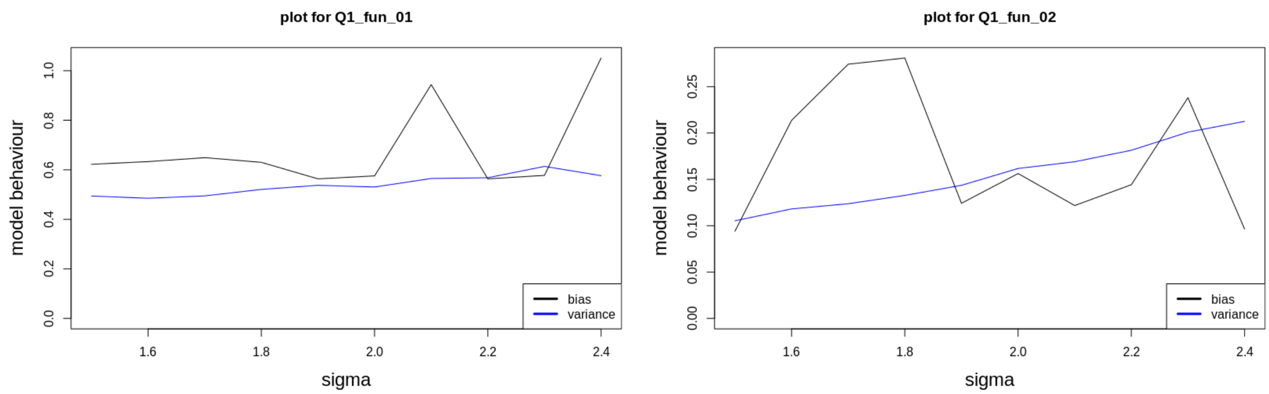


(a) plot for Q1_fun_01

(b) plot for Q1_fun_02

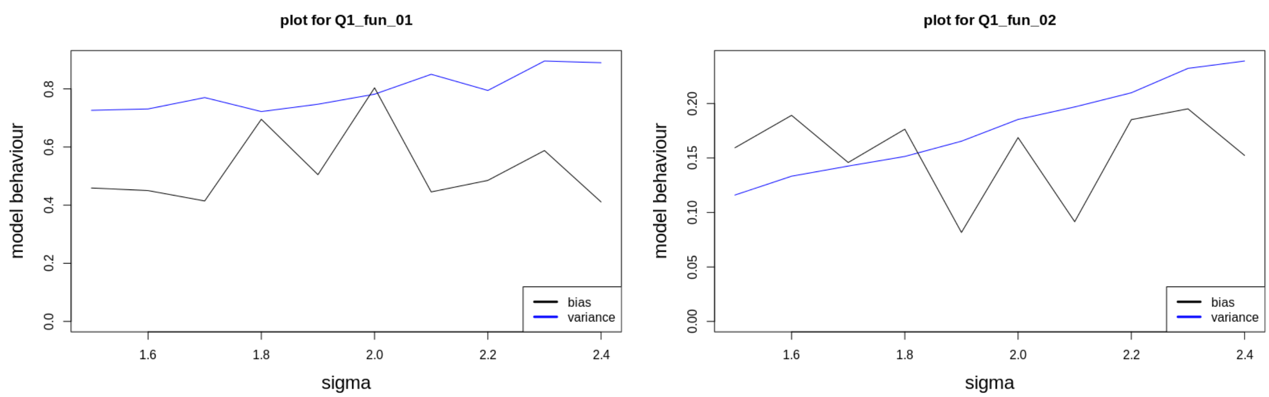
Figure 5: Bias-Variance vs sigma for degree = 4

- As the sigma increses
- AS we can see from the above equation bias variance and mean square error are dependent on each other.
- From here we can not see any pattern as the sigma variance increases but sigma does not show any pattern



(a) plot for Q1_fun_01 (b) plot for Q1_fun_02

Figure 6: Bias-Variance vs sigma for degree = 14



(a) plot for Q1_fun_01 (b) plot for Q1_fun_02

Figure 7: Bias-Variance vs sigma for degree = 26

2 Question2

2.1 Ploting average mean and camparing with true mean

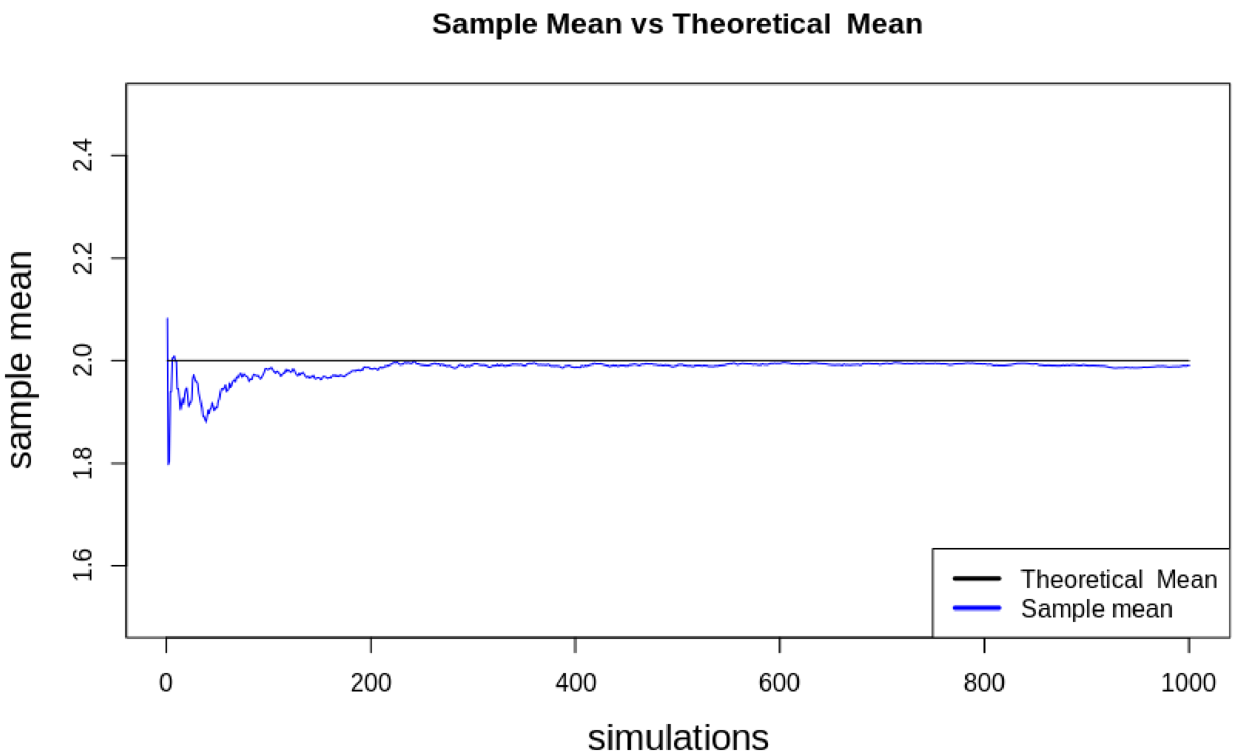


Figure 8: Theoretical vs Sample mean

Redline corressponds to sample expected sample mean Theoretical mean is $\frac{1}{\lambda} = 2$ and and expected sample mean we got is 2.001. As we can see from the above figure as the number of sample increse expected mean tend to converge the true mean. It satifies central limit theorem.

2.2 Comparing sample variance and theoretical variance

Theorectical variance is $\frac{1}{\lambda^2} = .1$ and sample variance $((1/0.5001)/\text{sqrt}(40))^2 = .10009. Aswecanseefromformherevarianceisapp$

2.3 Showing that the distribution is approximately normal using plot

As we can see from figure 9 most of the sample have mean around there true mean which should happen accoriding to central limit theorem. As the sample size and simulation will increase figure will more look like guassian.

2.4 Calculating paramters

Entropy is given by $1 - ln(\lambda) = 1.693147$

Fisher Information is given by $= \frac{number\ of\ sample}{\lambda^2} = 79.005$

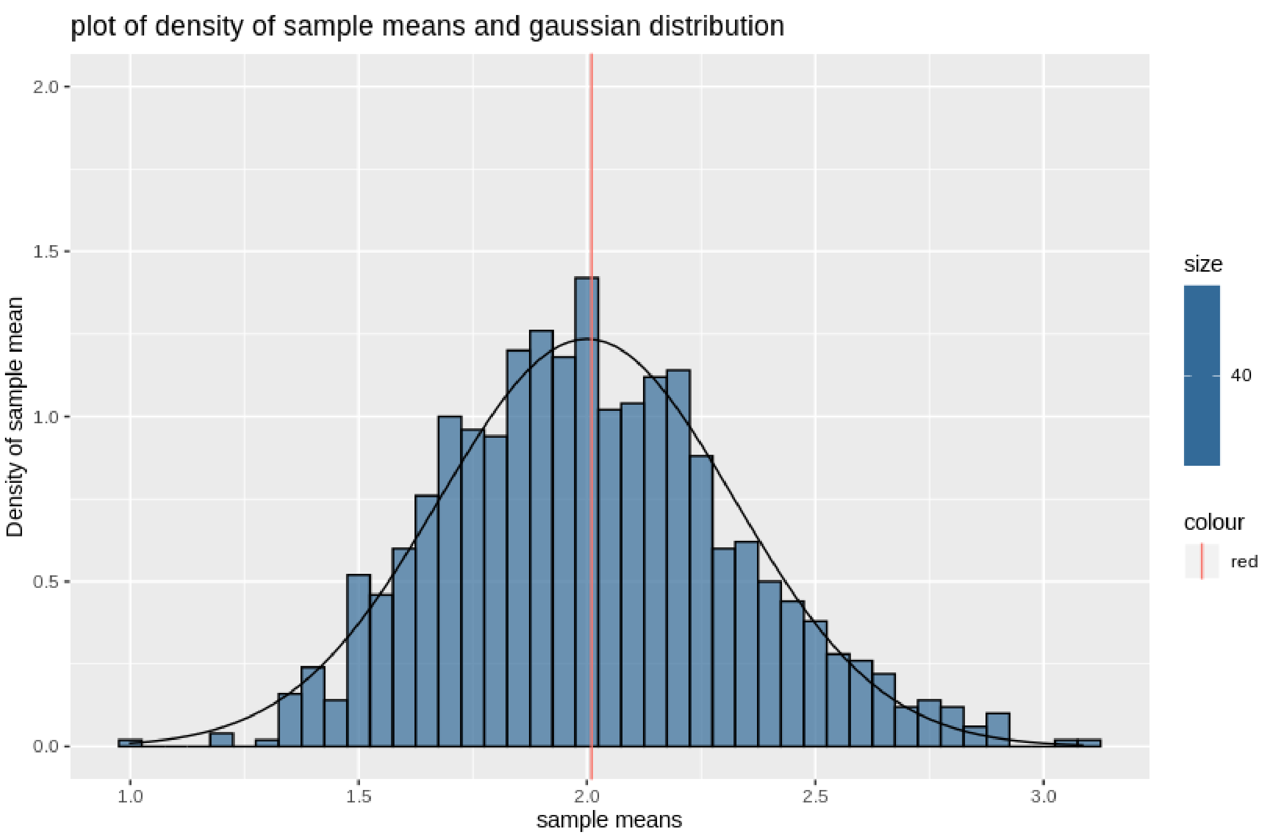


Figure 9: plot of averages of sample