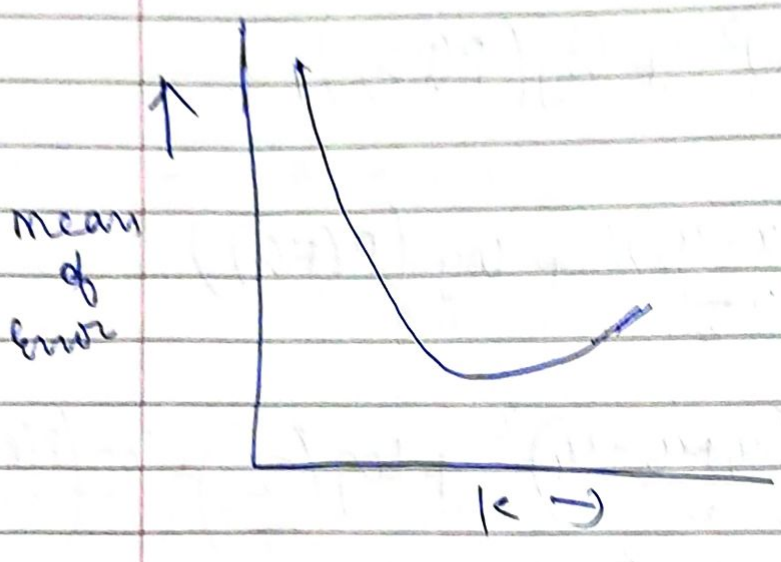
Q-1.

$$y = \sin(x) + N(0,1)$$

a)

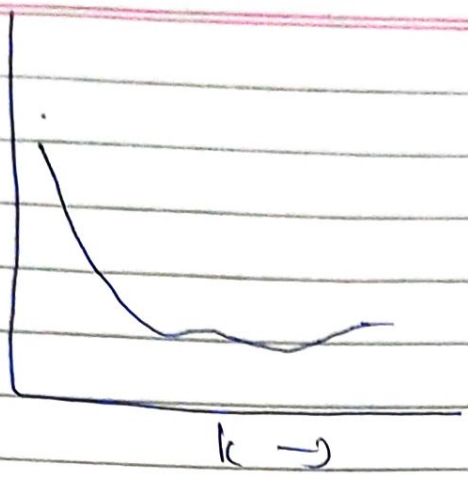
sample = 100



for less k, the ~~train~~ test set will be less.
training set high
↓
overfitting. so increase in error.

then once k is about 5-20% of data.
That is optimal.

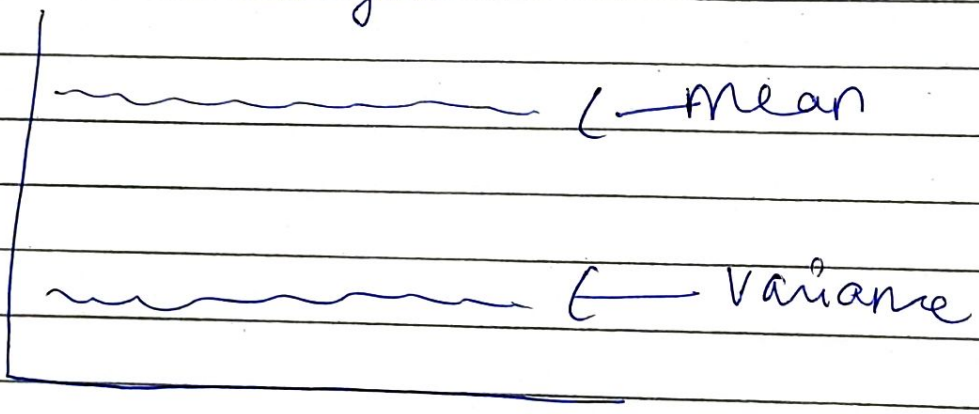
then as k increase,
training set decrease,
and test set inc.
↳ less training data so again error inc.



similarly,
Variance has
k will increase
upto certain
point, it will
decrease.

for huge dataset 10,0,00

mean & variance are quite constant throughout



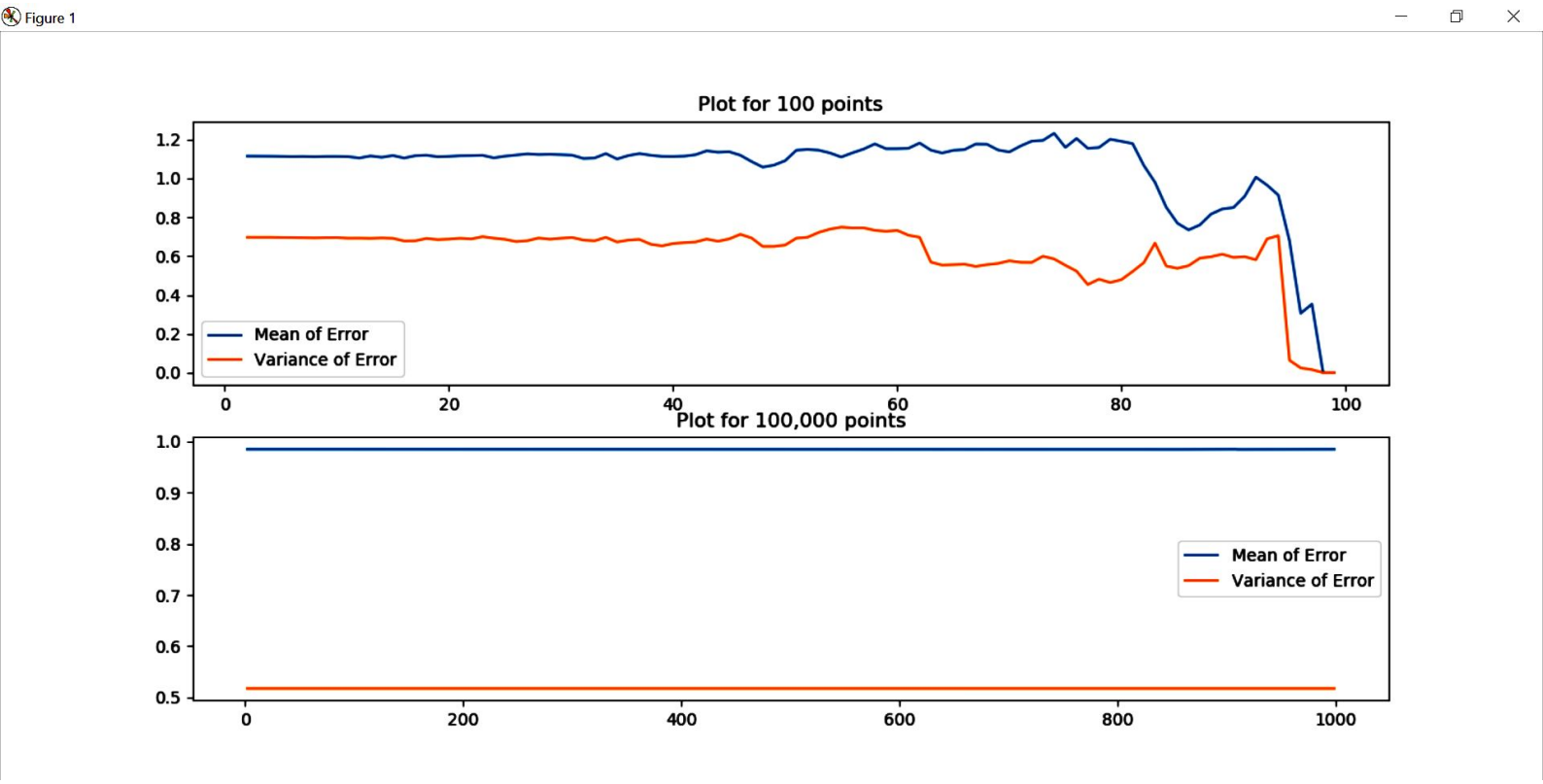


Figure 1
Plot for 100 points
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
Mean of Error
Variance of Error
Plot for 100,000 points
1.0
0.9
0.8
0.7
0.6
0.5
0
200
400
600
800
1000
Mean of Error
Variance of Error

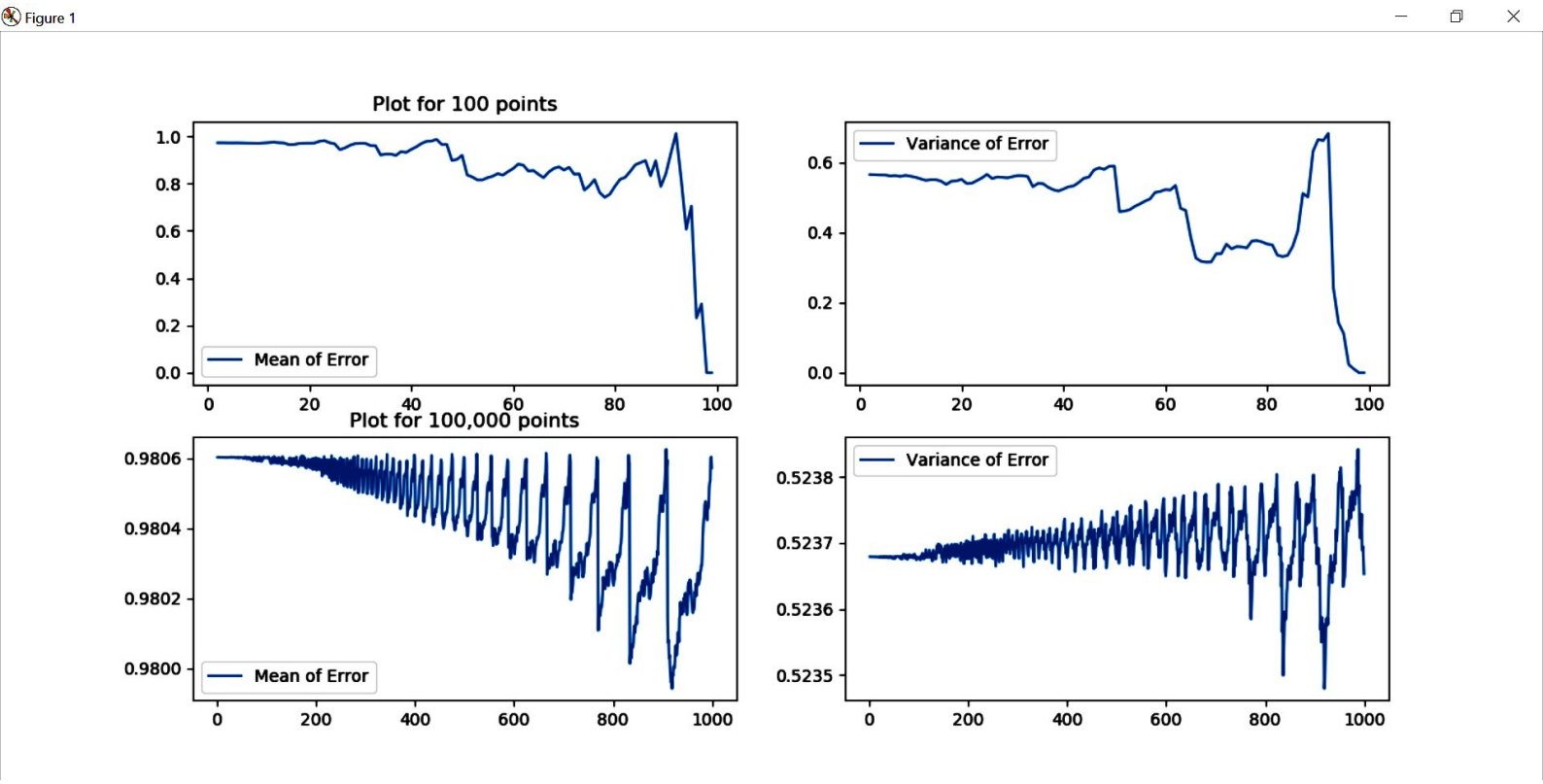


Figure 1
Plot for 100 points
Mean of Error
Variance of Error
Plot for 100,000 points
Mean of Error
Variance of Error

b) for huge ∞ dataset ⇒ results are perfect.

But for sample size = 100,
even for K close to 100,
the error is decreasing. which suggest that larger K is better. But that's not true