name: **Changmao Li**

Problem 1

1.a.

$E[(Y-\hat{f}(x))^2][X=x] = E[(y-\hat{f}(x))^2)]$

$= E[y^2] + E[\hat{f}(x)^2] - E[2y\hat{f}(x)]$

$= E[y^2] - E[y]^2 + E[y]^2 + E[\hat{f}(x)^2] - E[\hat{f}(x)]^2 + E[\hat{f}(x)]^2 - 2f(x)E[\hat{f}(x)]$

$= E[(f(x)+\epsilon - f(x))^2] + E[(\hat{f}(x) - E[\hat{f}(x)])^2] + (E(\hat{f}(x) - f(x))^2$

$= \sigma_\epsilon^2 + (E(\hat{f}(x) - f(x))^2 + E[(\hat{f}(x) - E[\hat{f}(x)])^2]$

1.b.

σ_ϵ^2 is $Noise$, and here it is gaussian noise of the data

$(E(\hat{f}(x) - f(x))^2$ is $Bias^2$, it is a measure of whether the predictors we choose approximates $f(x)$ well

$E[(\hat{f}(x) - E[\hat{f}(x)])^2]$ is $Variance$, it is a measure of whether a predictor is susceptible to changes

Problem 2

2.a.

Assume Z is the standard normal distribution and the X is a $M \times 1$ random vector with covariance Σ, and we know its mean is μ, then we have,

$X = \mu + VZ$, where V is a $M \times M$ invertible matrix such that $\Sigma = VV^T = V^TV$Then we have,

$f_X(x) = \frac{1}{|det(\Sigma)|} f_Z(V^{-1}(x-\mu))$

$= \frac{1}{|det(V)|}(2\pi)^{-\frac{M}{2}} e^{-\frac{1}{2}(V^{-1}(x-\mu))^T(V^{-1}(x-\mu))}$

$= |det(V)|^{-\frac{1}{2}} |det(V)|^{-\frac{1}{2}} (2\pi)^{-\frac{M}{2}} e^{-\frac{1}{2}(x-\mu)^T(V^{-1})^T V^{-1}(x-\mu)}$

$= (2\pi)^{-\frac{M}{2}} |det(V)det(V)|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T(V^T)^{-1}V^{-1}(x-\mu)}$

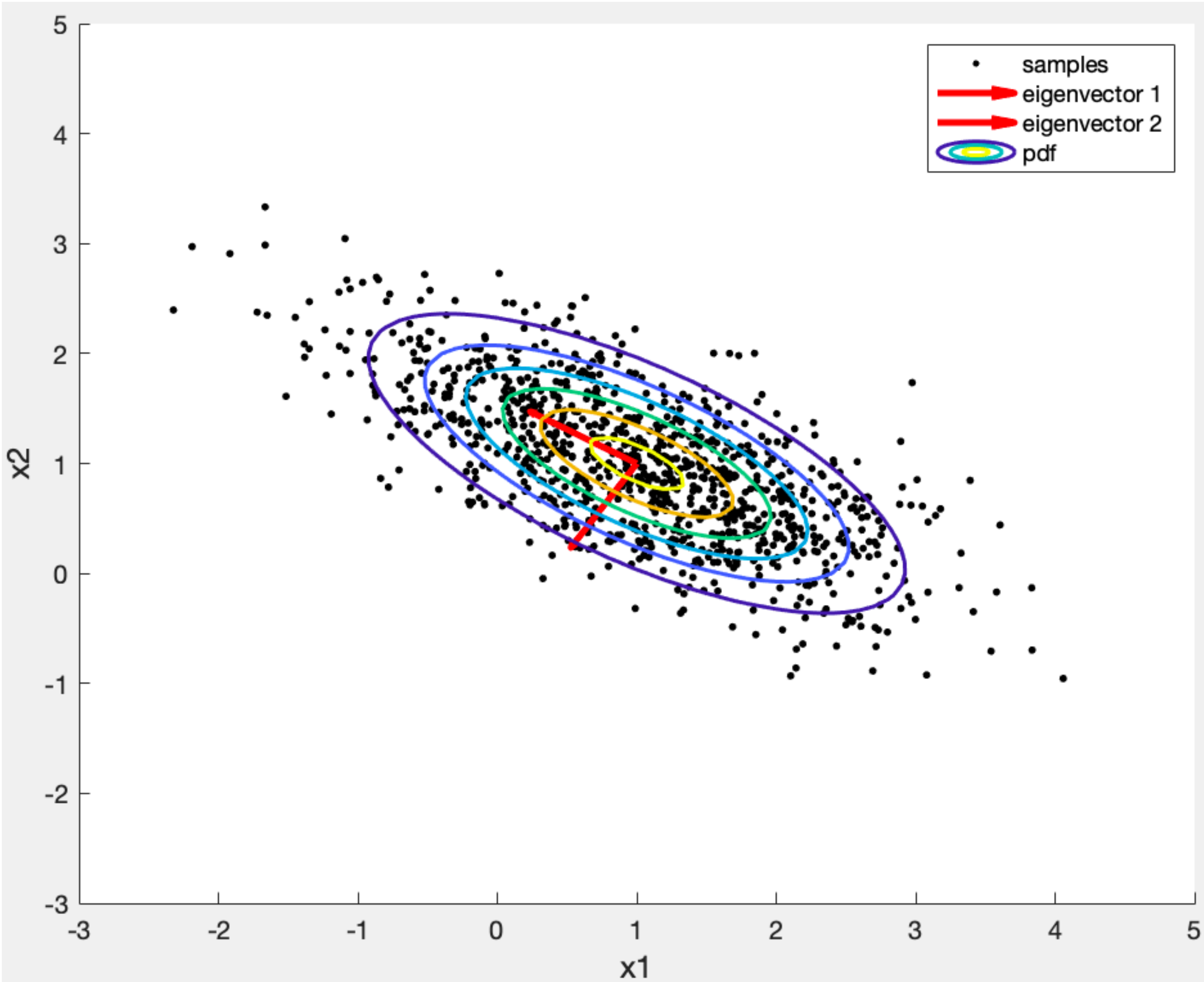
$= (2\pi)^{-\frac{M}{2}} |det(V)det(V^T)|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T(V^T)^{-1}V^{-1}(x-\mu)}$

$= (2\pi)^{-\frac{M}{2}} |det(VV^T)|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T(VV^T)^{-1}(x-\mu)}$

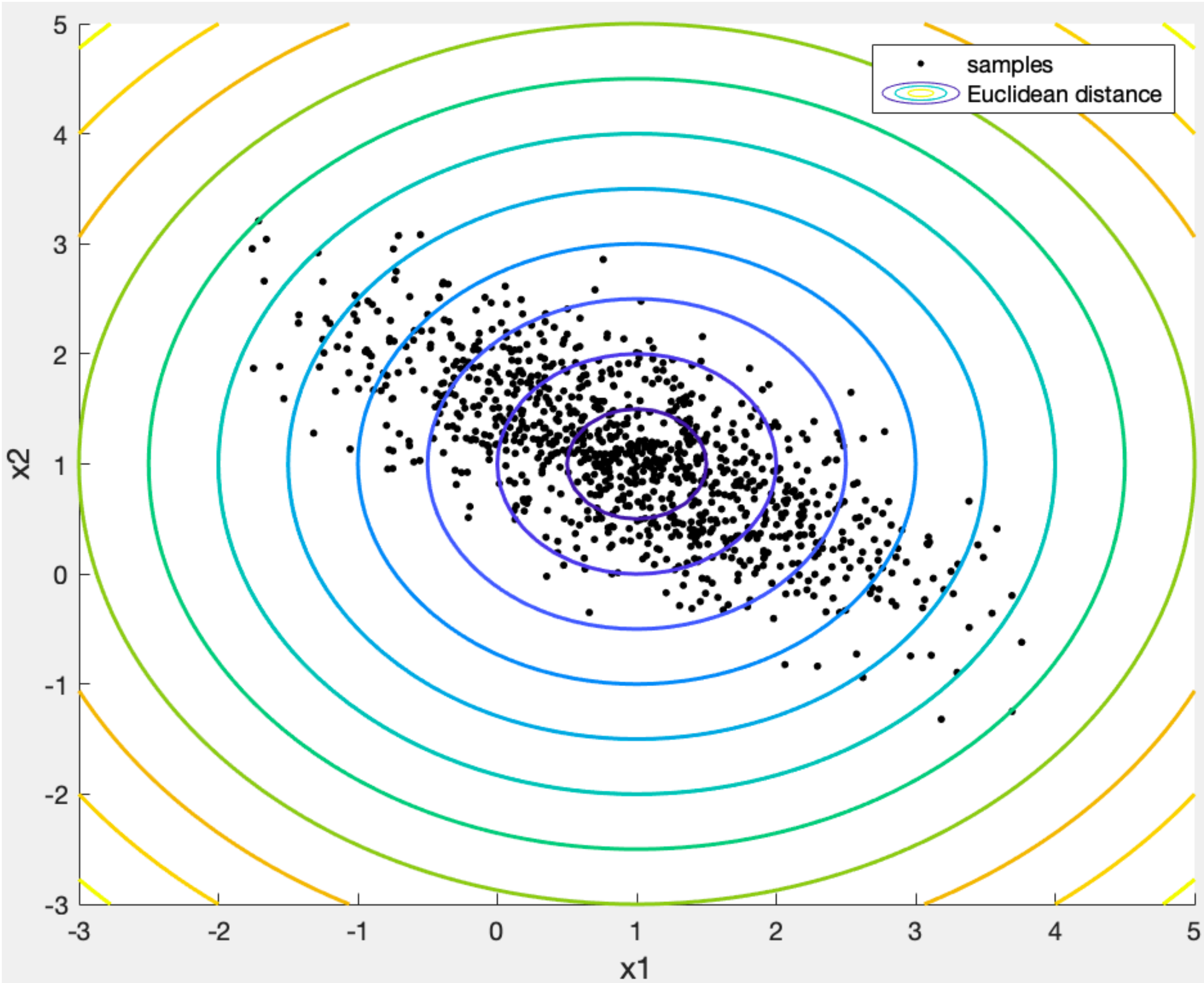
$= (2\pi)^{-\frac{M}{2}} |det(\Sigma)|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$

The existence of a matrix V satisfying $\Sigma = VV^T = V^TV$ is guaranteed by the fact that Σ is symmetric and positive definite.

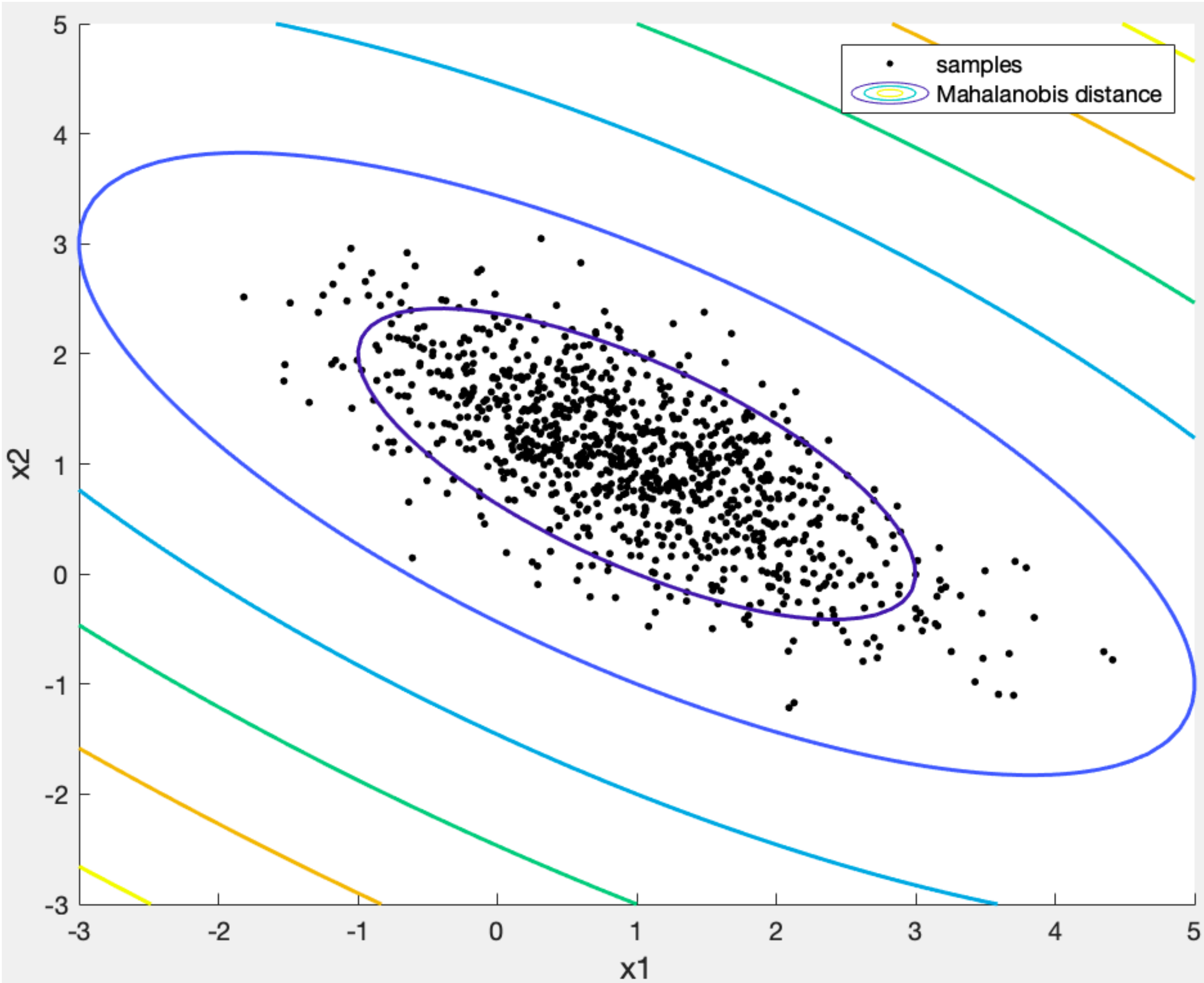
2.b.



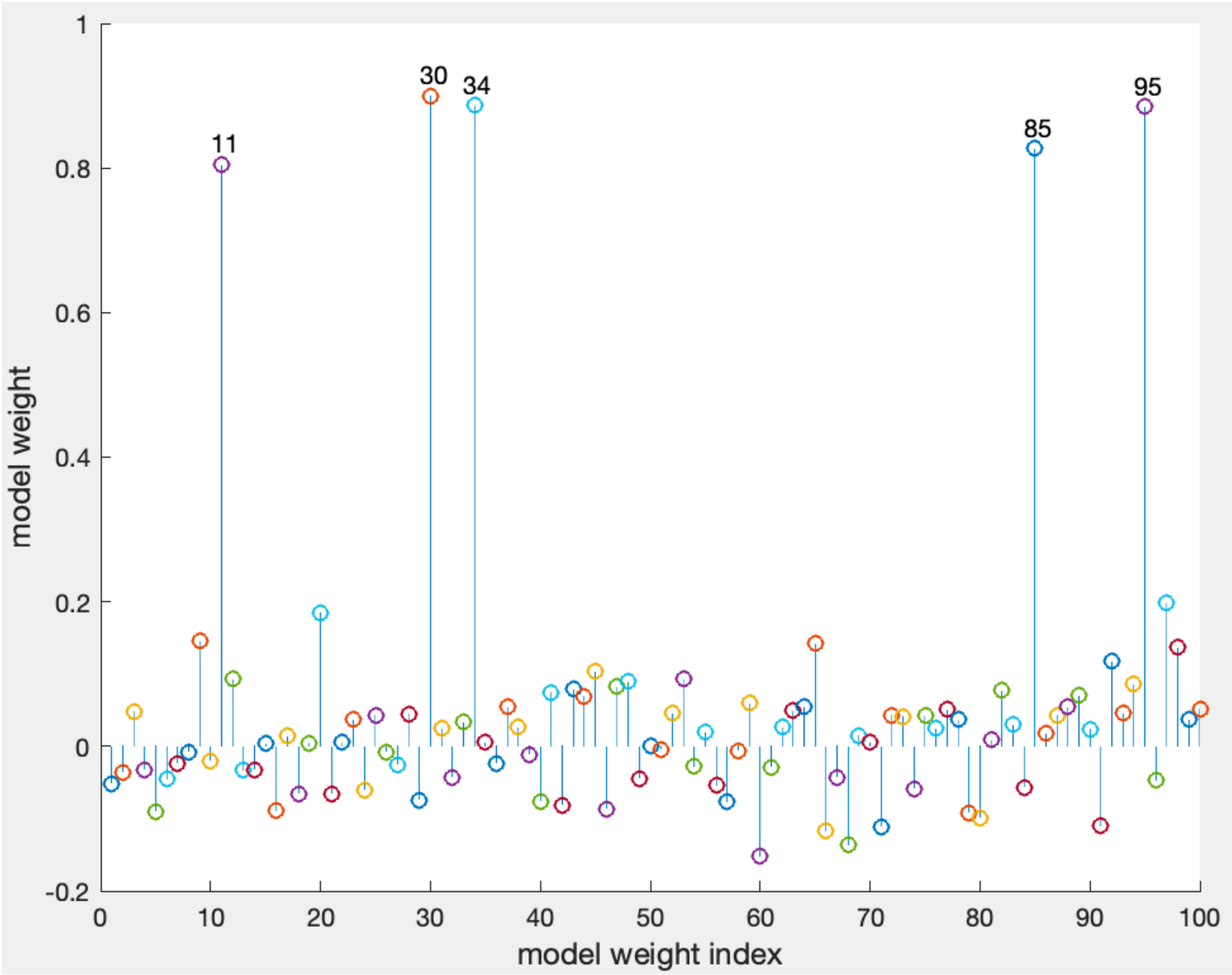
2.c.



2.d.

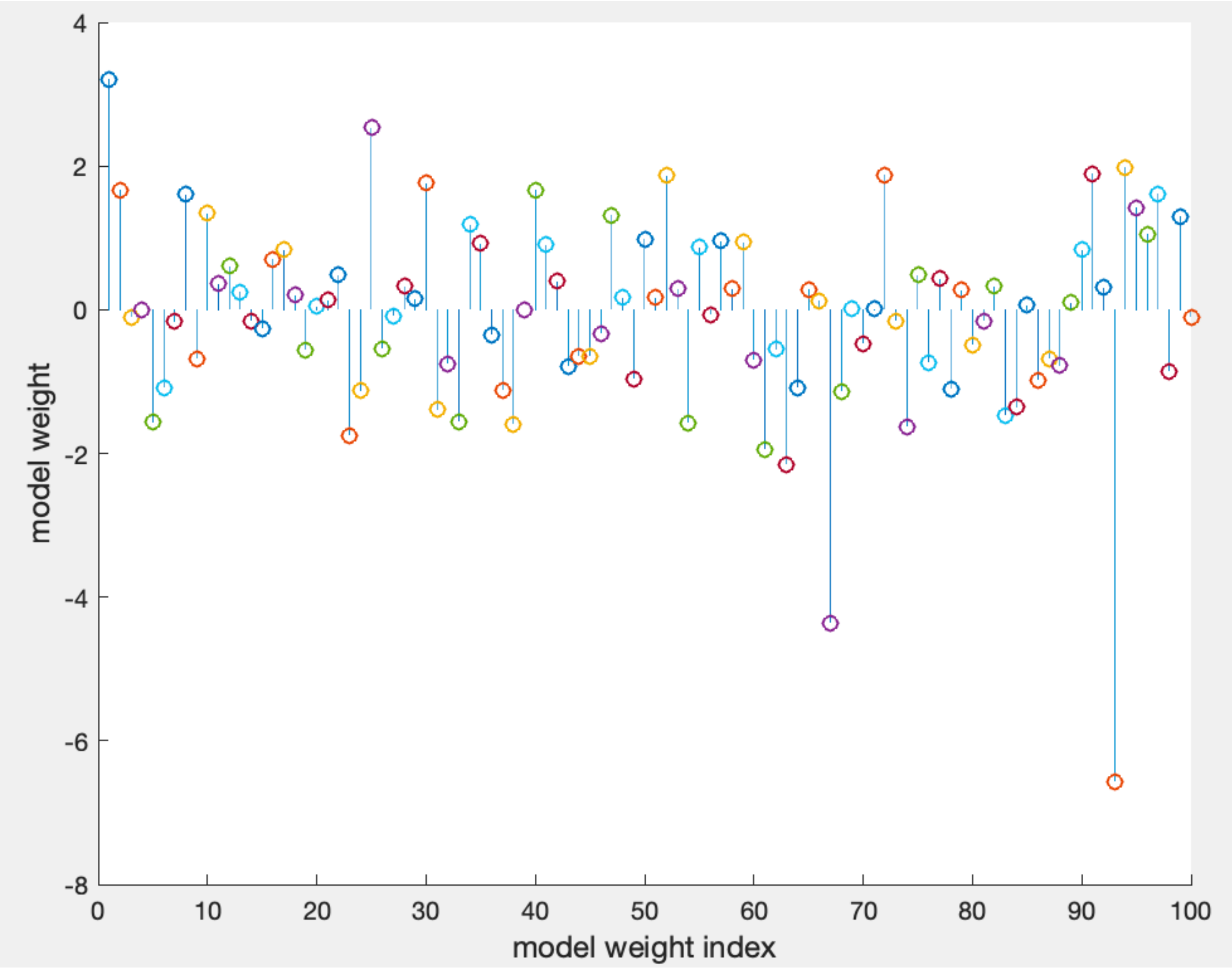


3.a.



From the graph we can see the indices of the nonzero model weights are 11,30,34,85,95

3.b.



3.c.

For ordinary least square model:

mean squre error:46.69630 |bias|:6.257846 variance:8.046464

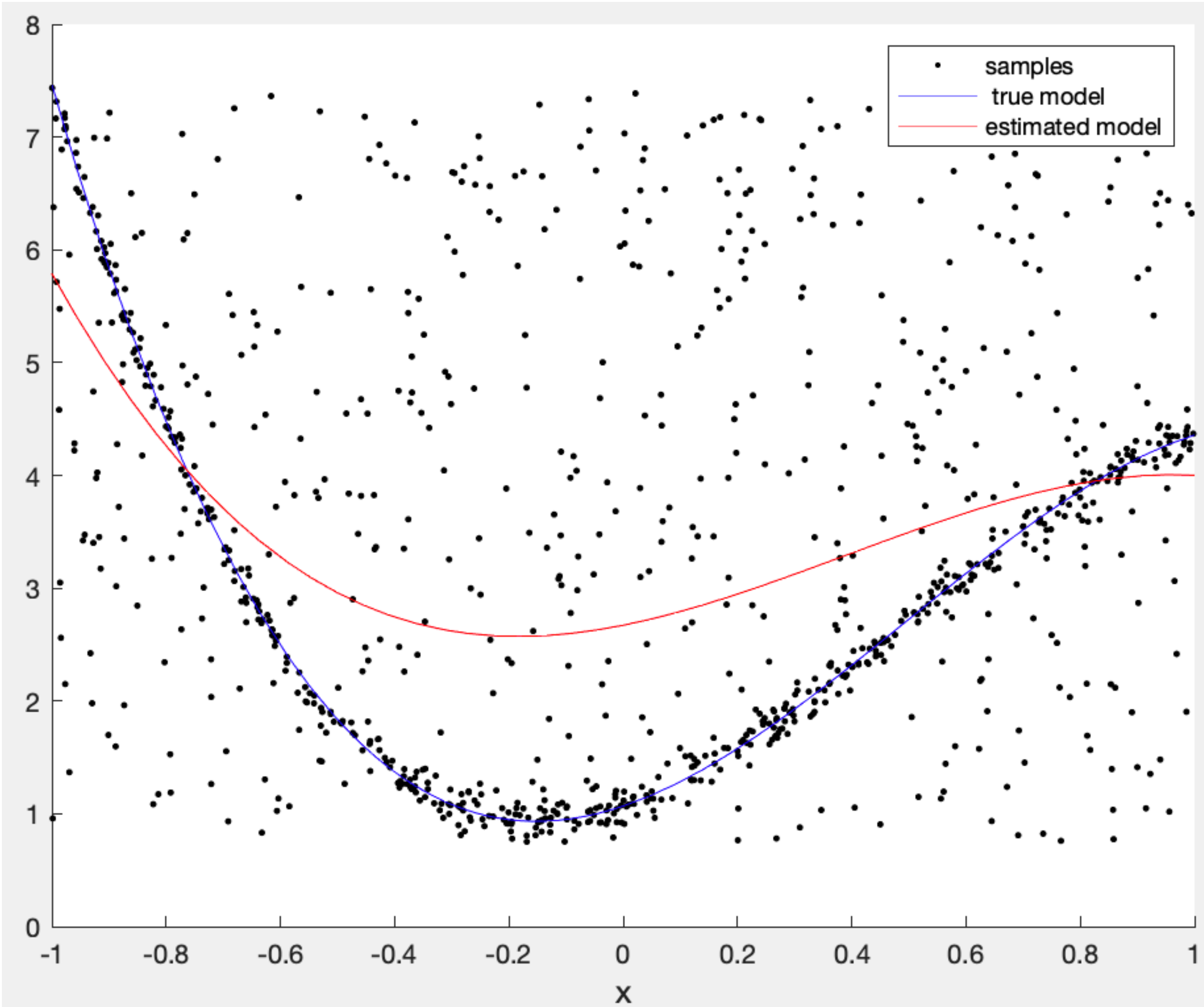
For LASSO:

mean squre error:1.864484e-04

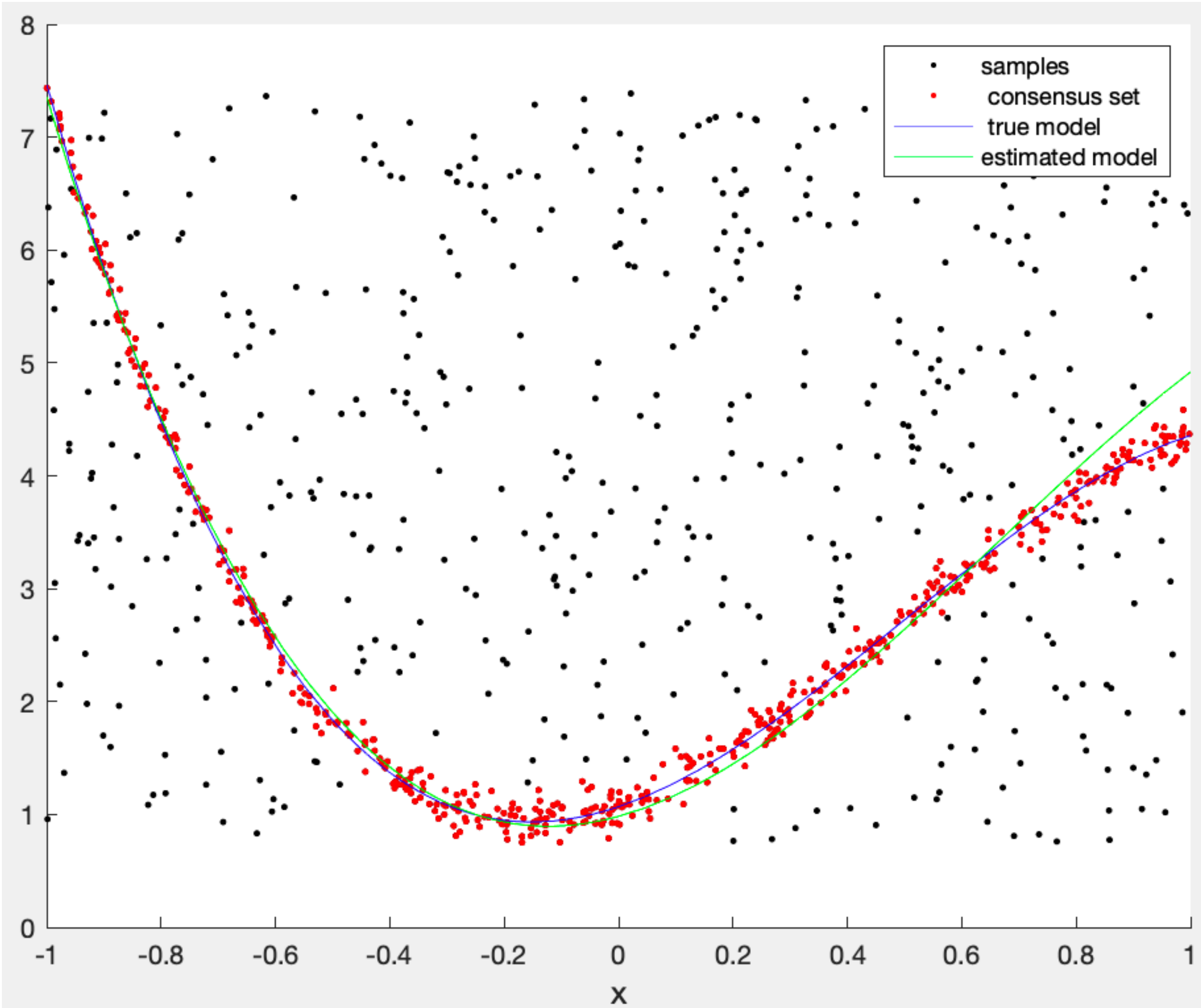
|bias|:8.174617e-05 variance:4.044406e-01

The above result is from the 3.a and 3.b program. Apparently the performance of LASSO is far better than oridinary least square model.

4.a.


samples
true model
estimated model
8
7
6
5
4
3
2
1
0
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
x

4.b.


samples
consensus set
true model
estimated model
8
7
6
5
4
3
2
1
0
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
x