NEW YORK UNIVERSITY
TANDON SCHOOL OF ENGINEERING

Homework #1

Student name: *Zihan Zeng*

Course: *ECE 6143 Machine Learning* – Professor: *Yury Dvorkin*
Due date: *Sep 23th, 2021*

Problem 1

Answer.

(*a*) In my experiment, I set the value of d ranges from 1 to 50, and then output 50 figures. But as the value of d gets larger and larger, the pictures gradually become the same.

So I chose the first 18 pictures for display.

We can see from the Figure 1 and Figure 2 that the red line is moving closer to the point set, their shapes are becoming similar too. Which means that as d changes, the degree of fitting is getting higher.

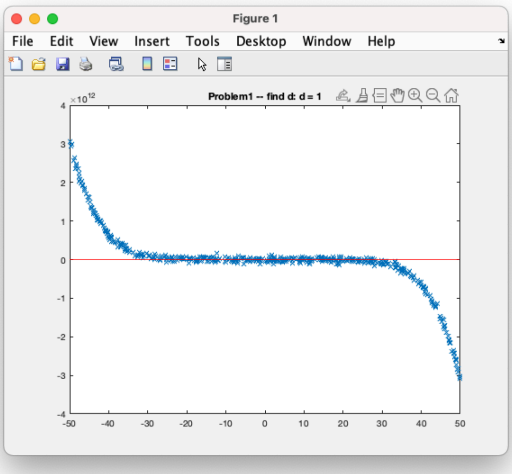


Figure 1: When d = 1

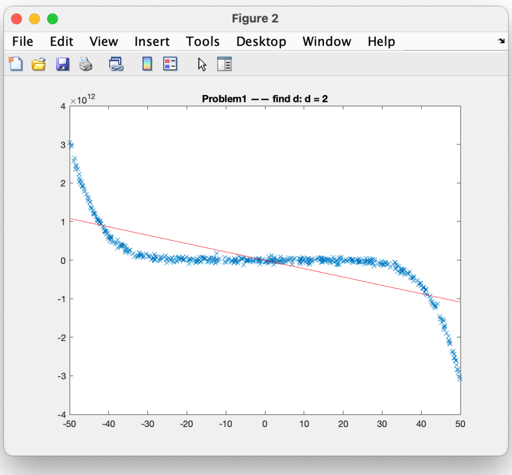
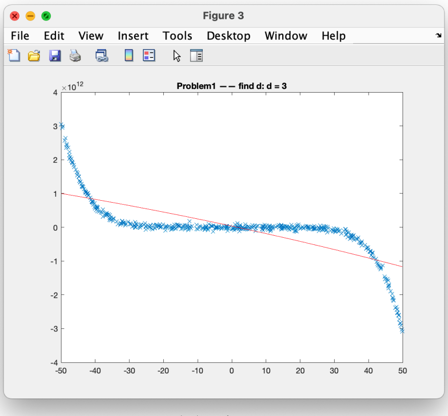


Figure 2: When d = 2

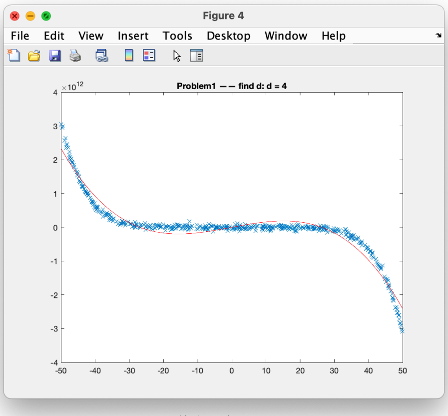
But as d continues to increase, the fitting seems to stop (according to my observation of those figures, it seems that the fitting stops when d=9 or 10). As the Figure3 shown.

Figure 4 shows that after d is greater than 10, there is basically no change in the figures.

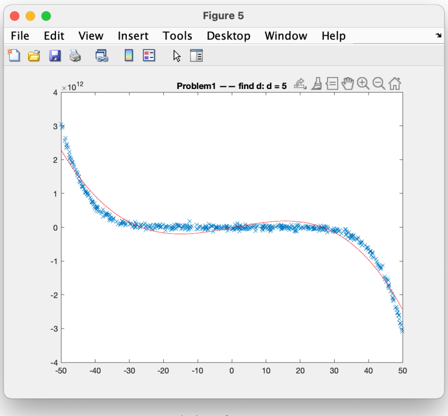
So I guess the best value of d is around 10.



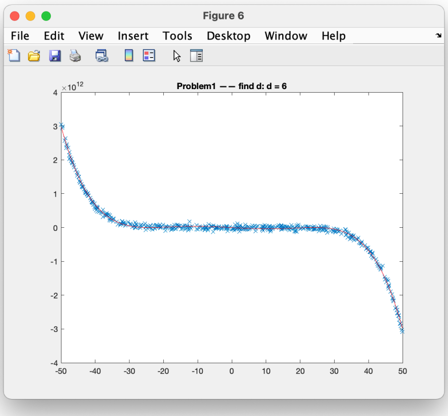
(a) d = 3



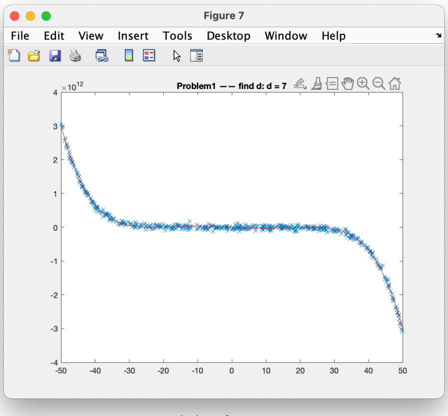
(b) d = 4



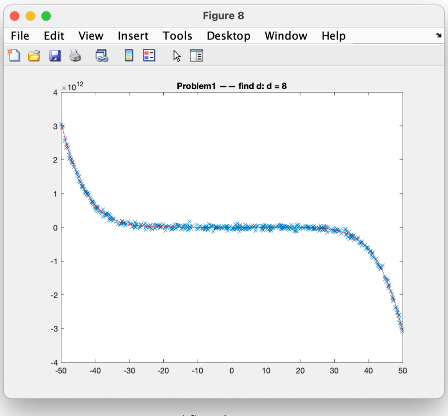
(c) d = 5



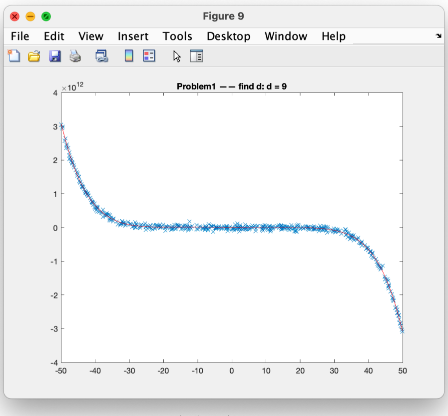
(d) d = 6



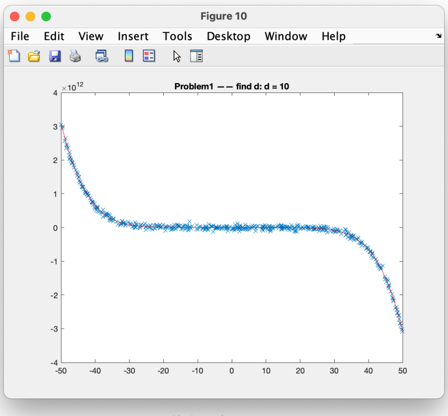
(e) d = 7



(f) d = 8

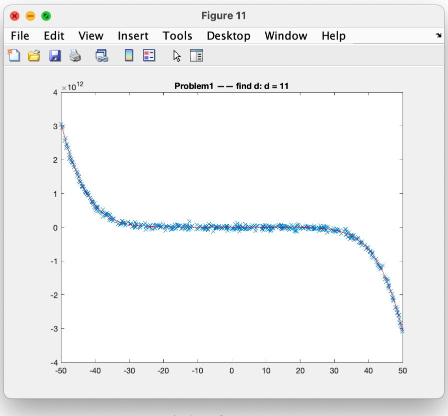


(g) d = 9

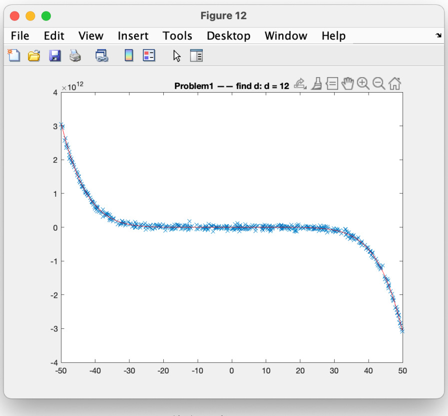


(h) d = 10

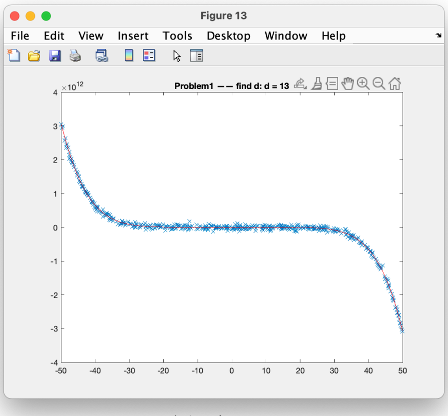
Figure 3: d from 3 to 10



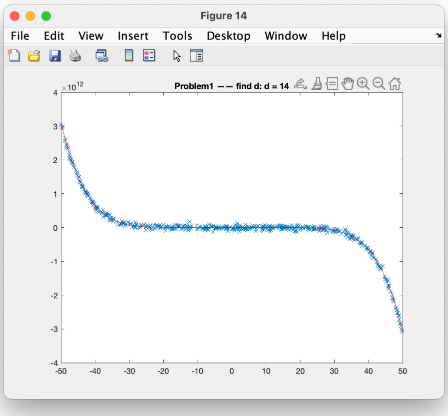
(a) d = 11



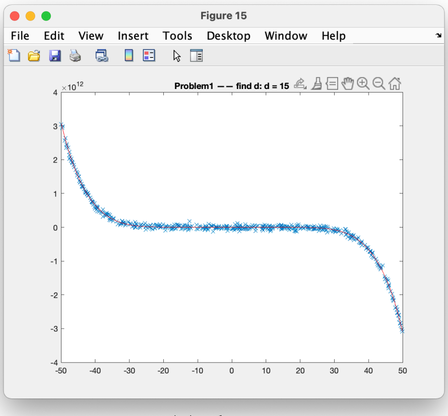
(b) d = 12



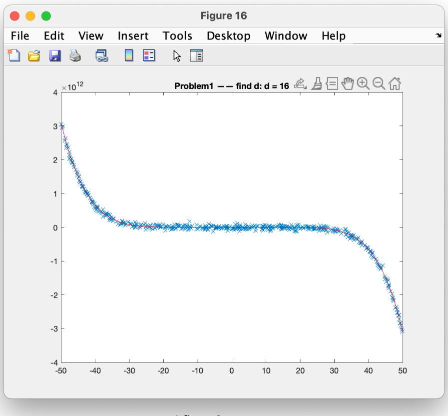
(c) d = 13



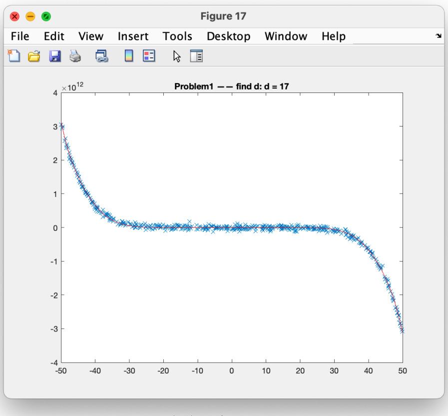
(d) d = 14



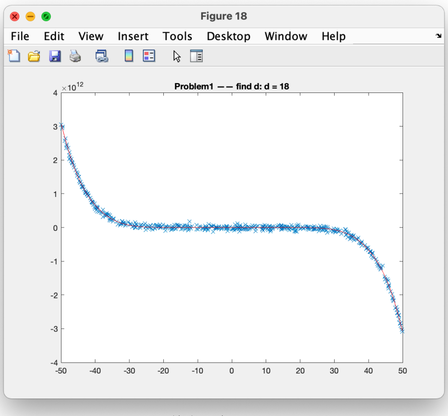
(e) d = 15



(f) d = 16



(g) d = 17



(h) d = 18

Figure 4: d from 11 to 18

(*b*) In a cross-validation experiment, the results obtained by running the code will change every time. After many experiments, the minimum value of error is when d is within the interval of 8 to 15. Figure 5 records some experimental results.

Figure 5: Problem1: Cross Validation Experiments Records

The following table records the percentage of different values of d in 100 experiments.

d =	8	10	**11**	13	14	15
Number of Times	21	19	**24**	15	12	9
Frequency	0.21	0.19	**0.24**	0.15	0.12	0.9

Table 1: 100 Experiments Results

So I guess the optimal value of d is 11.

Problem 2

Answer. In this experiment, I set the range of λ from 0 to 1000 with a step size of 1. Under these conditions, looking for the optimal value of λ and mark it in the generated figure.

The x-axis of the generated image(Figure 6) represents the value of λ, and the y-axis represents the errors. It can be seen from the figures that as the lambda increases, the error keeps decreasing.

However, in several runs, the optimal value of λ has changed a bit. I have listed these possible optimal values.

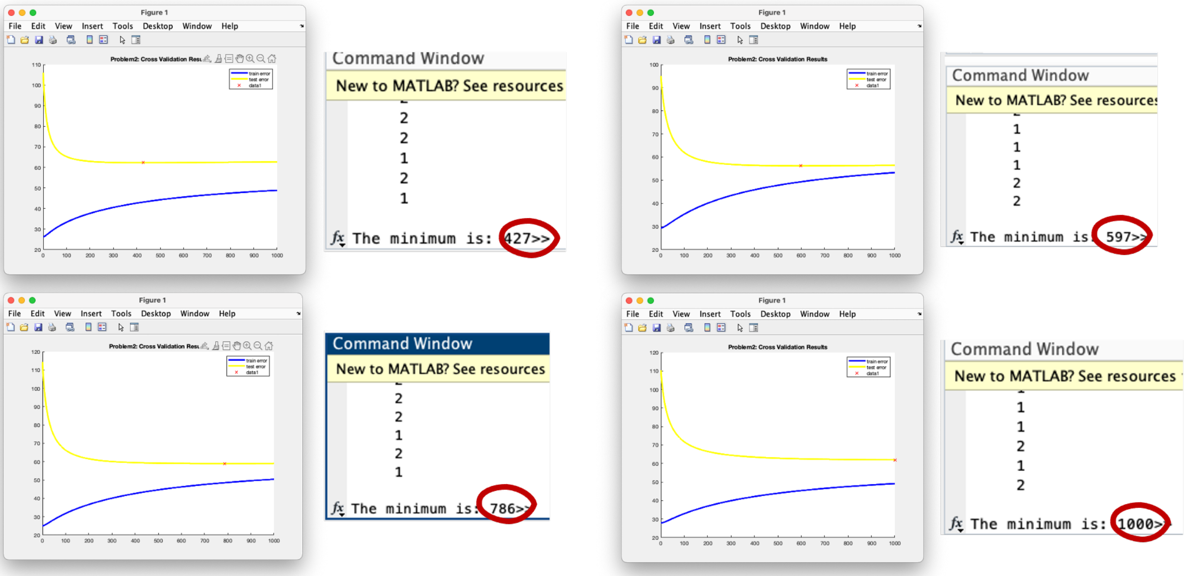


Figure 6: Problem3: Cross Validation Experiments Records

Problem 3

Answer.

(*a*) $g(-z) = 1 - g(z)$

$$g(-z) = \frac{1}{1+e^{z}} = \frac{1}{\frac{e^{-z}}{e^{-z}} + \frac{1}{e^{-z}}} = \frac{e^{-z}}{1+e^{-z}}$$

$$= \frac{1+e^{-z}-1}{1+e^{-z}} = 1 - \frac{1}{1+e^{-z}} = 1 - g(z)$$

(*b*) $g^{-1}(y) = \ln \frac{y}{1-y}$

$$\ln \frac{y}{1-y} = \ln y - \ln(1-y) = \ln\left(\frac{1}{1+e^{-z}}\right) - \ln\left(\frac{e^{-z}}{1+e^{-z}}\right)$$

$$= \ln 1 - \ln(1+e^{-z}) - \ln e^{-z} + \ln(1+e^{-z})$$

$$= -\ln e^{-z} = z$$

i.e. $g^{-1}(y) = z$

Problem 4

Answer.

(*a*) We need to calculate the gradient. Figure 7 shows the proof.

(*b*) Stepside and tolerance are very important in machine learning and need to be adjusted based on experience in some cases. Therefore, I only listed the running results in two different situations as examples. When the learning rate is constant, reducing the value of tolerance will make the result more accurate. As shown in Figure 8.

$$
\begin{aligned}
\nabla_\theta R &= \frac{1}{N}\sum_{i=1}^{N}(y_i-1)\frac{d}{d\theta}\log(1-f_i)-y_i\frac{d}{d\theta}\log f(i)\\
&= \frac{1}{N}\sum_{i=1}^{N}(y_i-1)\frac{d}{d\theta}\log\left(1-\frac{1}{1+e^{-\theta^T x_i}}\right)-y_i\frac{d}{d\theta}\log\left(\frac{1}{1+e^{-\theta^T x_i}}\right)\\
&= \frac{1}{N}\sum_{i=1}^{N}(y_i-1)\frac{d}{d\theta}\left(\log e^{-\theta^T x_i}-\log(1+e^{-\theta^T x_i})\right)-y_i\frac{d}{d\theta}\left(-\log(1+e^{-\theta^T x_i})\right)\\
&= \frac{1}{N}\sum_{i=1}^{N}(y_i-1)\left(-x_i+\frac{x_i e^{-\theta^T x_i}}{1+e^{-\theta^T x_i}}\right)-y_i\left(\frac{x_i e^{-\theta^T x_i}}{1+e^{-\theta^T x_i}}\right)
\end{aligned}
$$

Figure 7: Problem4: Proof of Gradient

```
tolerence = 0.1;
stepsize = 1;
```

```
tolerence = 0.0001;
stepsize = 1;
```

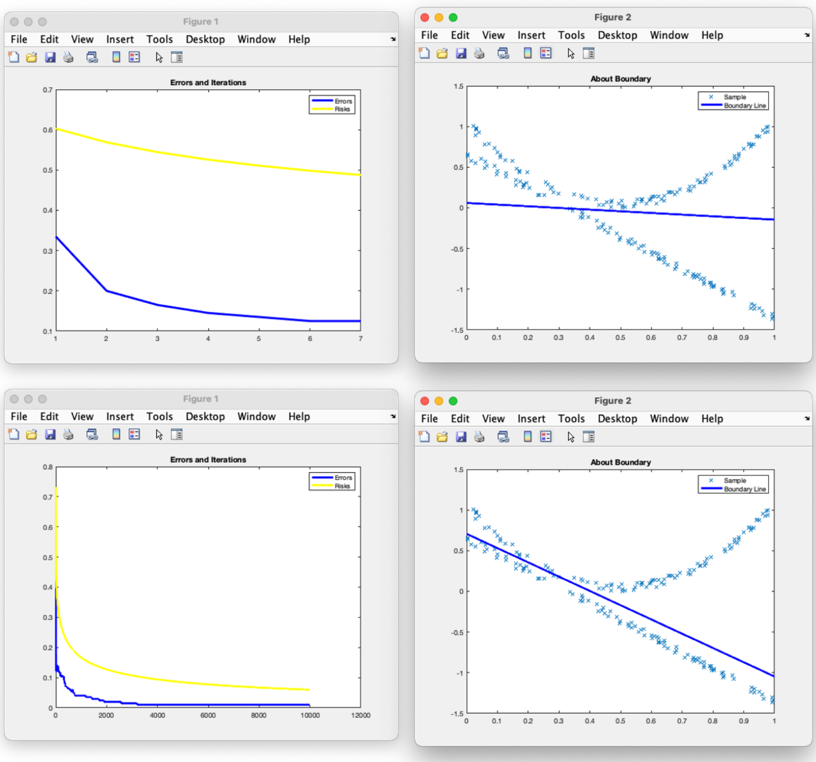


Figure 8: Problem4: Training Results