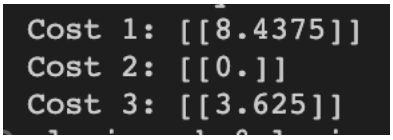
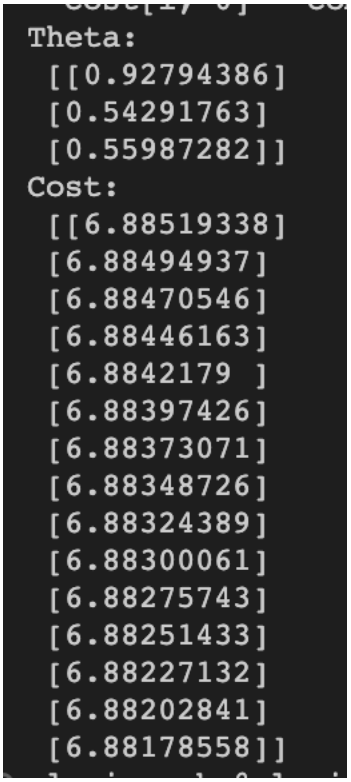
Alex Ivensky
ECE 1395
Homework 2

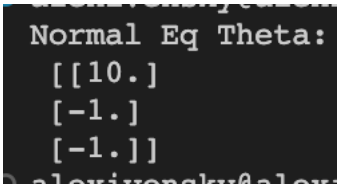
1.



2.

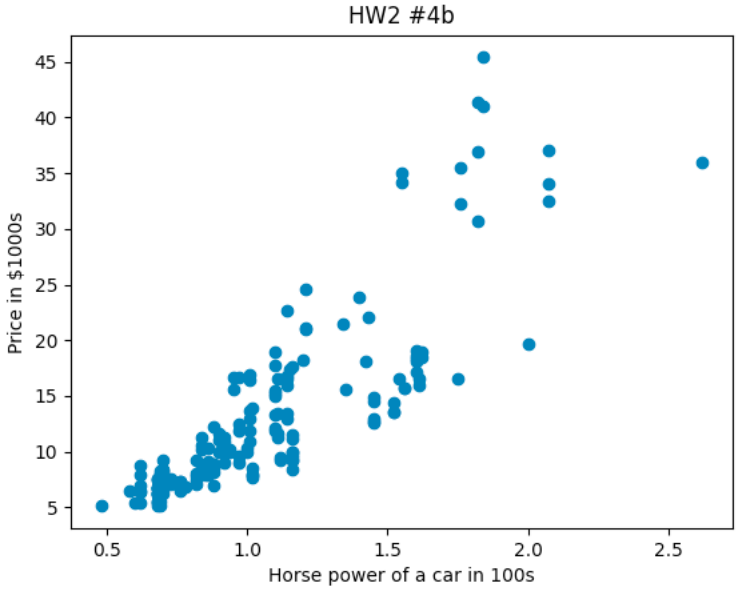


3.



There is a significant difference. The gradient descent conducted on this data set had too low of a learning rate due to the minimal decrease in the cost over the iterations, and too few iterations. Increasing both values makes the gradient descent response closer to the normal equation response.

4b.

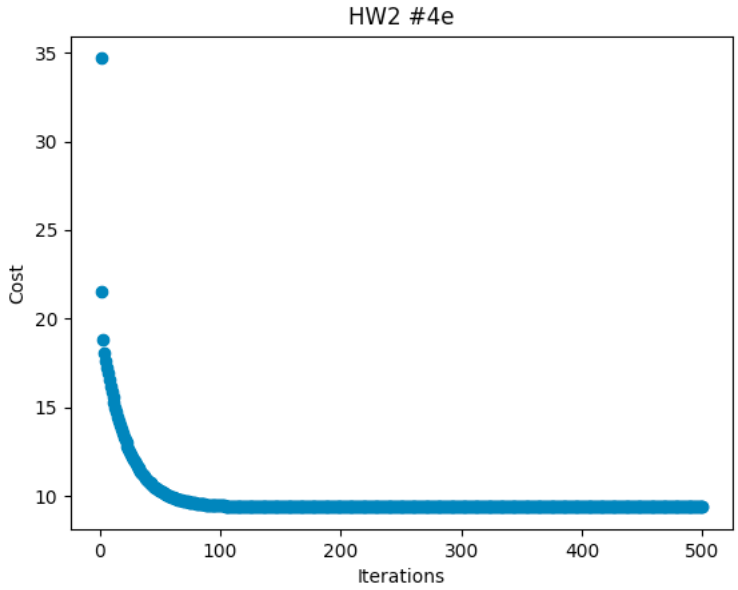


4c.

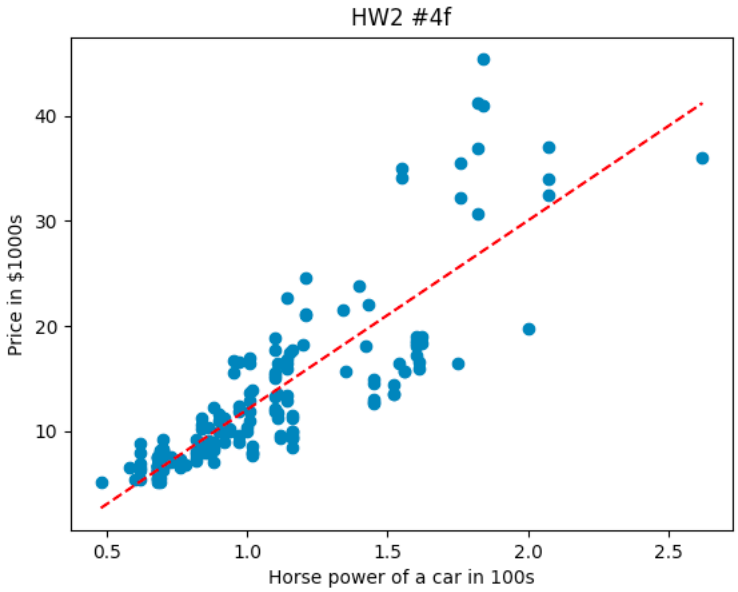
```
4c Dimensions:
X: (179, 2)
y: (179, 1)
```

4e.

```
4e theta:
 [[-5.98444595]
 [18.023411  ]]
```



4f.

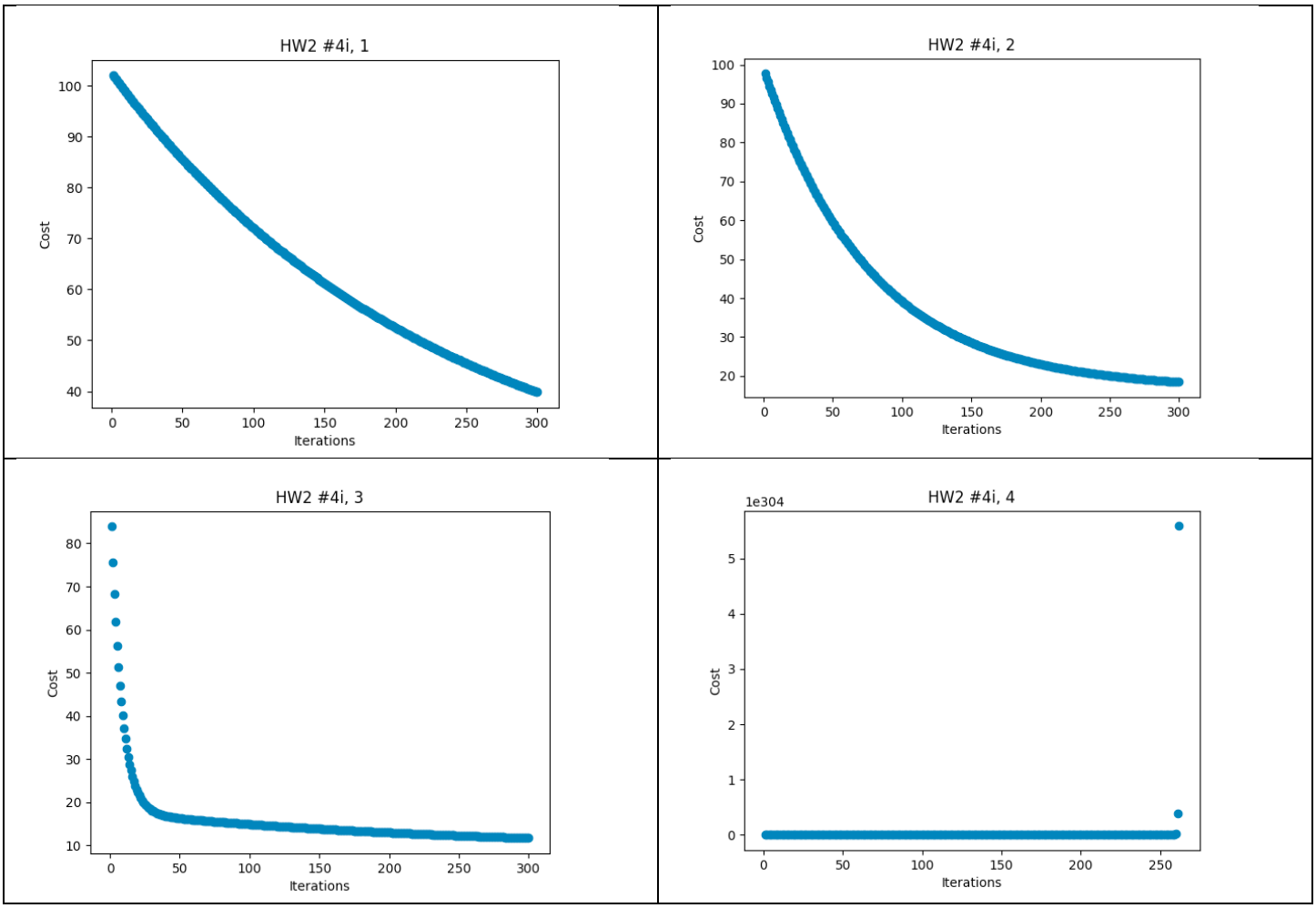


4g & 4h:

```
4g Error
 [[14780.76744006]]
4h Error
 [[37883.96352693]]
```

The errors in one run of the program will appear different, but running the program multiple times shows different results. This is due to the random initialization of theta. This shows that both algorithms on average produce the same error.

4i.



In figure 1, the learning rate is too low, so the gradient descent algorithm is slow to converge, and does not converge in the given 300 iterations. In figures 2 and 3, they both converge, but figure 3 converges much faster due to its higher learning rate. In figure 4, the algorithm diverges in the last two iterations because the learning rate is too high.

5a.

```
X1 mean: 1611.111111111111
X2 mean: 1292.2777777777778
y mean: 102.02777777777777
X1 stddev: 383.53456875762674
X2 stddev: 238.73737443185826
y stddev: 7.350306535977905
```

```
X dim (3, 1)
y dim (36, 1)
```

5b.

```
Theta:
 [[0.00052492]
 [0.42609413]
 [0.22645683]]
```

5c.

```
CO2 for 2300 engine size and 1300 weight:  106.44244365516667
```