Name: Aragya Goyal
Student ID: 4541458
D Value = 58

Question 1:

1. My D value is 58. My cost calculation is below

HW #2

$$\theta = \begin{bmatrix} 0.5 \\ 2 \\ 1 \end{bmatrix}$$

x_0	x_1	x_2	y
1	0	1	$1.5 + (58/100) = 2.08$
1	1	1.5	4
1	2	4	8.5
1	3	2	$8.5 + \left(\frac{58}{50}\right) = 9.5 + \frac{8}{50} = 9.66$

Predictions

$$\underset{4\times 3}{\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1.5 \\ 1 & 2 & 4 \\ 1 & 3 & 2 \end{bmatrix}} \underset{3\times 1}{\begin{bmatrix} 0.5 \\ 2 \\ 1 \end{bmatrix}} = \begin{bmatrix} 0.5 + 0 + 1 \\ 0.5 + 2 + 1.5 \\ 0.5 + 4 + 4 \\ 0.5 + 6 + 2 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 4 \\ 8.5 \\ 8.5 \end{bmatrix}$$

err

$$\begin{bmatrix} 1.5 \\ 4 \\ 8.5 \\ 8.5 \end{bmatrix} - \begin{bmatrix} 2.08 \\ 4 \\ 8.5 \\ 9.66 \end{bmatrix} = \begin{bmatrix} -0.58 \\ 0 \\ 0 \\ -1.16 \end{bmatrix}^2 = \begin{bmatrix} 0.3364 \\ 0 \\ 0 \\ 1.3456 \end{bmatrix}$$

$$\sum \text{err} = 1.682$$

$$\text{Cost} = \frac{1}{2(4)}(1.682) = 0.21025 \checkmark$$

   a.

2. The cost from my function was 0.21025. This did match my manual calculation.
3. Theta from Gradient Descent: [[0.94396034] [2.358288] [0.72686973]]
4. Theta from Normal Equation: [[1.12537415] [2.36496599] [0.65278912]]
5. The two theta estimates are semi-similar. In order to make them converge to the same solution, we could potentially increase the alpha value or also increase the numbe rof iterations we do.

Name: Aragya Goyal
Student ID: 4541458
D Value = 58

Question 2:

A.

a. The dataset name was “Electrical Consumption Based on Weather Data” at: https://www.kaggle.com/datasets/sudhirsingh27/electricity-consumption-based-on-weather-data/data
b. There are four distinct features in this dataset including average daily wind speed (m/s), the daily precipitation (mm), the daily maximum temperature (℃), and the daily minimum temperature (℃). We are also given dates but these will be excluded from evaluation as the seasonal information is implicitly captured through the daily highs/lows of each day.
c. The output variable is the daily electricity consumption. While units are not explicitly given in the dataset, typically power consumption is measured in kWh which is what I’ll be using as well.
d. I chose this dataset because it has real world relevance and I was also curious to see how much different variables within weather end up affecting how much power is used. The data also fit the requirements necessary for this homework nicely which made preprocessing of the data relatively simple.
e. The total number of training samples in this dataset is 1433 examples. There are four numerical features.

B.

a. For the average daily wind speed, there is a slight positive correlation as the speed increases so does the energy consumption. With daily precipitation measurements, the relationship is unclear and looks like there is almost no correlation. With daily max temperature, there seems to be negative correlation where as the daily max temperatures increase, the energy consumption decreases. Finally, with the daily min temperature, there is also a negative correlation where as the daily min temperatures increase, so does the average energy consumption.
b. I believe that both of the temperature metrics seem to have a semi strong correlation with energy consumption. I believe this because even through my own personal experience, temperature is typically the main determining factor to whether or not I have my air conditioning or heating turned on. These also typically tend to make up the majority of the electricity bill. Based on this prior experience, I would imagine that these metrics will have a decent correlation to power consumption.
c. Based on my data, I do believe that linear regression will work well. This is because there seems to be either a linear correlation between my inputs and outputs or no correlation at all.

C.

a. The data means of the average wind speed, the daily precipitation, the daily max temperature, and the daily min temperature are respectively 2.64231312 m/s, 3.73427362 mm, 17.13272214℃, and 9.08582511℃. The standard deviations are respectively 1.139619 m/s, 10.86036564 mm, 10.15147761℃, and 9.04134466℃.

b. The size of X_train is [1134 x 5], the size of y_train is [1134 x 1], the size of X_test is [284 x 5], and finally the size of y_test is [284 x 1]. The X matrices have the bias columns included.

D.

a. I chose alpha as 0.5 for gradient descent here. I was able to select this through trial and error. I started off at 0.01 as the initial alpha and when I saw that the error was still relatively high and the graph had not yet converged, I kept increasing theta until I found the error to be a lot lower and the graph to look more converged.
b. My learned theta values were [[1555.96082712] [-318.86455622]].
c. The cost curve does indicate convergence because the error has gotten relatively low and the change between cost across iterations is not decreasing as much at the end.

E.

a. Theta from Gradient Descent (Multivariate): [[1556.31568412] [4.995762] [-27.15447935] [-185.75592883] [-137.62010266]]
b. Theta from Normal Equation (Multivariate): [[1556.31568412] [4.995762] [-27.15447935] [-185.75592884] [-137.62010266]]
c. The thetas are highly similar. This means that there is a good alpha rate and the thetas are converging.
d. At a glance, the graph does seem to have converged due to the flat nature of the line at the end of the graph. However, since the scale of the y axis is very large, even small slopes could indicate that the cost has not yet converged and would require some more detailed comparison between iterations to determine whether the cost has converged.

F.

a. The MSE for the univariate model was 304182.24, the MSE for the multivariate gradient descent model was 305399.16, and finally the MSE for the multivariate normal equation was 305399.16.
b. The univariate model performed the best on the test set.
c. Adding more features did not improve the performance of the models. This may be because some of them are not very representative of how much electricity gets used. Especially between average wind speed and precipitation. I believe that if I had used just the min/max daily temperatures, I'd be able to create a better model than if I used all four features.

G.

a. The four learning rates I used were [0.001, 0.5, 0.865, 1]
b. For the first learning rate, the cost decreased over iterations, however, they were very slow and the rate at which the cost decreased was almost linear. For the second learning rate, the cost decreased very quickly in the initial iterations and then started to reach a point of minimal gain where the cost would not decrease as much. The third learning rate dropped very quickly the first couple of iterations

Name: Aragya Goyal
Student ID: 4541458
D Value = 58

but then started decreasing linearly. This cost decreased very slowly but was still in a converging manner. The fourth learning rate increases exponentially as the number of iterations increases.

c. I would say that if the cost is decreasing very slowly from the get-go, this means that the learning rate is too small. If the cost explodes as iterations go up, the learning rate is too high. If the cost decreases like expected at first but then the rate slows down a lot, the learning rate might still be a little too high and can be decreased for improved performance of the algorithm.
d. Feature standardization is very useful as it allows for a more standardized learning rate range that will work for all of the different features. This is because when we have some features that are very high in value and some features that are very low, no learning rate will work well for both features without standardization. By standardizing, it allows for the user to be able to find a learning rate that will satisfy both features.