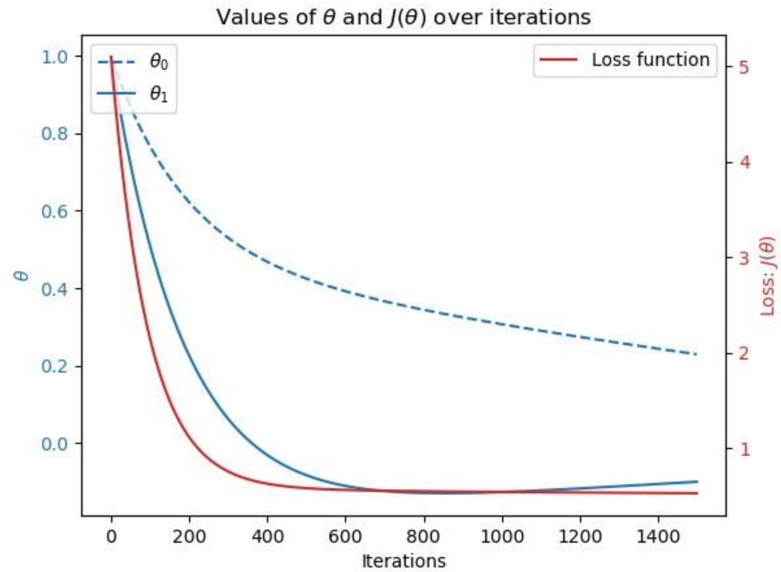
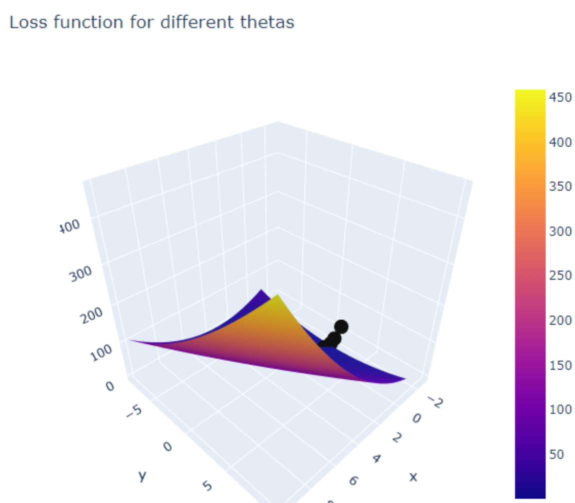
Gradient Descent Optimization

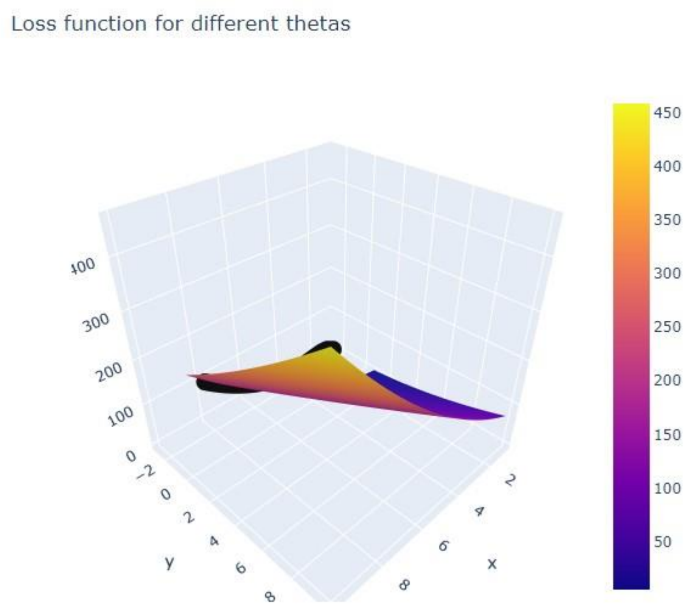
Round 1: Year- 1980; Station- Madrid



Loss over iterations: We see that theta1 declines steeply at first, then gradually flattens out along the loss function around 700 iterations, and then later has a small uptick. Overall, I think this small uptick is fine but keeping the model at 1000 iterations rather than 1500 would probably produce similar results. Theta0 acts similarly, declining steeply and then eventually getting closer to a flat line by the end.



Loss profile: the below loss profile is the final product for the gradient descent optimization. However, it was not the first profile I attained in the process. To keep the code clean and concise, I rewrote over original parameters when I needed to adjust the learning rate or iterations, rather than start a new code for the next iteration. The above profile shows the starting point I came up with, after using a learning rate of 0.1 and 100 iterations. As you can see, the gradient descent is not yet converging toward the global minimum. The learning rate was too high, and my theta values needed adjustment as well. A too-high learning rate (or step size) can lead to the model taking too big of steps to find the global optimum, and there's an increased chance it overshoots it. Smaller learning rates let us "creep down the hill" in a much slower, iterative fashion, taking more time but increasing the chances of the model finding the global optimum accurately. After a few more back-and-forths, I finally arrived at the parameters listed below.



Starting theta0: 1

Ending theta0: 0

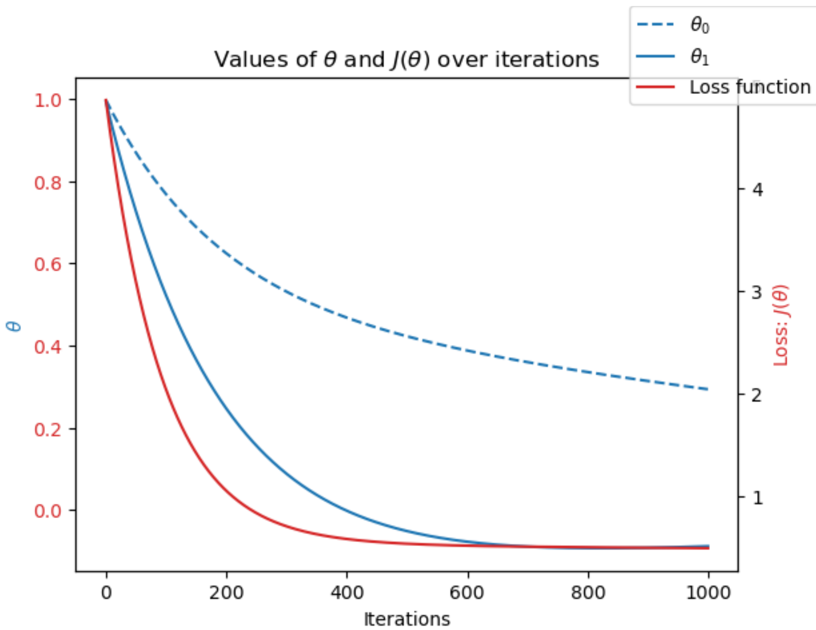
Starting theta1: 1

Ending theta1: -2

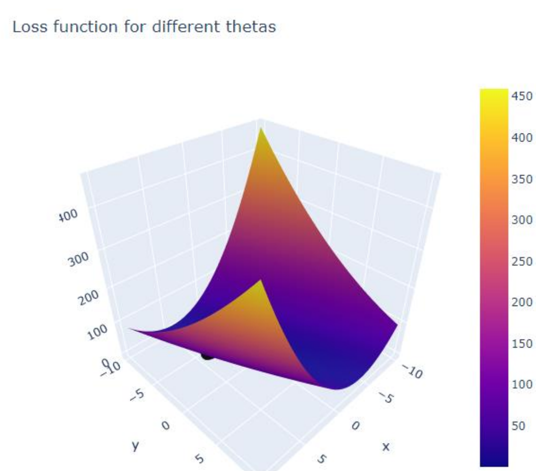
# of iterations: 1500

Step size: 0.001

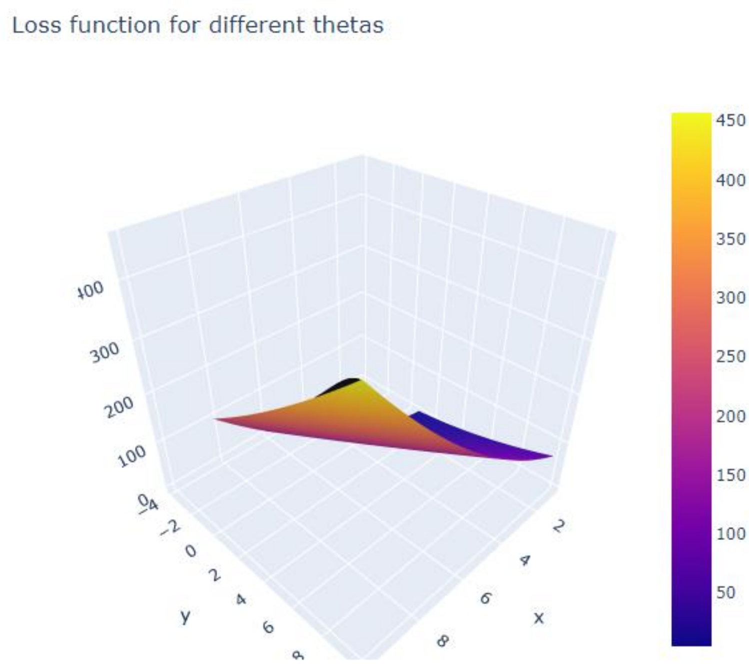
Round 2: Year- 1960; Station- Budapest



Loss over iterations: Besides having 1000 rather than 1500 iterations, this graph looks similar to the first one with Madrid. We can see convergence happens around 700 iterations, and theta1 remains along the loss function. Theta0, once again, does not completely flatten but gets flatter and flatter with more iterations. To my understanding, seeing a stabilization of the theta values is a good sign that our parameters are close to what we want for optimal model performance.



Loss profile: The above profile demonstrates my starting point for this GDO: a steep loss surface, indicating high gradient (step) values. After consultation with ChatGPT, we decided to reduce the learning rate and increase the number of iterations; after a few more attempts, I finally arrived at the below loss profile, with the marker much closer to the minimum, and flatter regions around the marker, indicating the loss values are stabilizing. Similar to the first example, the learning rate needed adjustment (by two orders of magnitude) so that the model could accurately find the global minimum for this dataset.



Starting theta0: 1

Ending theta0: 1

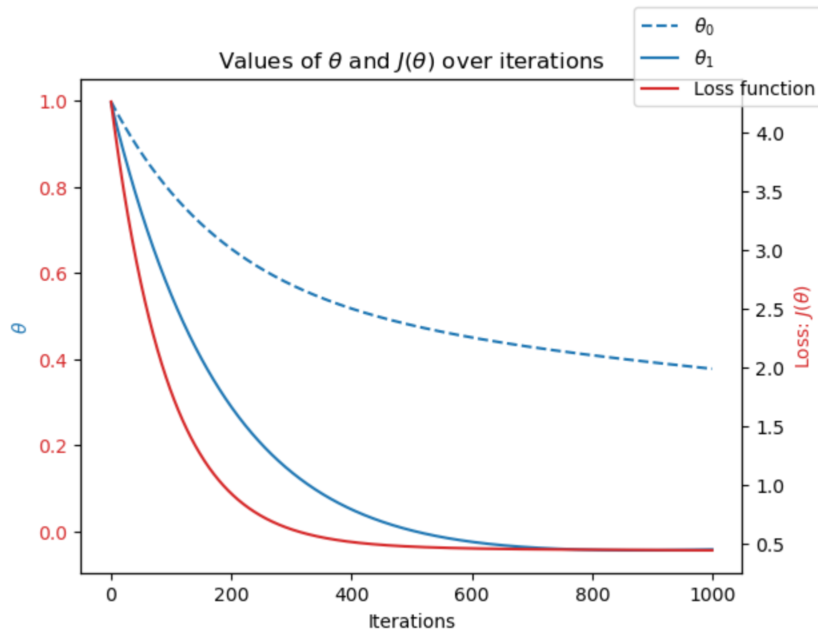
Starting theta1: 1

Ending theta1: -4

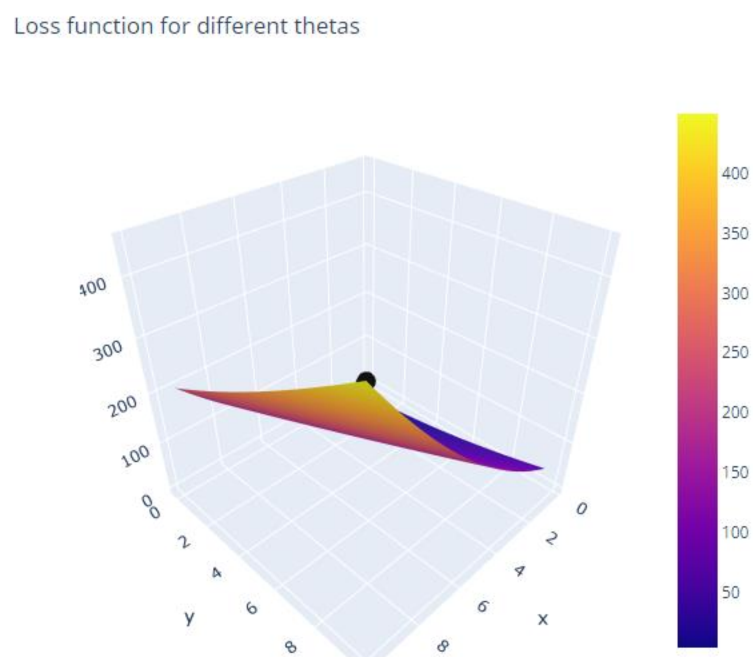
# of iterations: 1000

Step size: 0.001

Round 3: Year- 2019; Station- Stockholm



Loss over iterations: Once again, this graph looks similar to the others: both theta values decrease rapidly and then stabilize over time, with stability around 700 iterations. I will make further observations about the similarities in loss over iterations between the stations in the bottom section.



Loss profile: Here we see the black marker close to the global minimum, indicating the algorithm is approaching optimal values for theta. Also, the surface is flatter near the minimum, telling us we've reached the optimum number of iterations as well. A smaller learning rate allowed the model to take small steps and ensure it didn't overshoot the global minimum it was trying to find.

Starting theta0: 1

Ending theta0: 0

Starting theta1: 1

Ending theta1: 0.4

# of iterations: 1000

Step size: 0.001

Part Two: Observations

1. Across the 3 weather stations and the 3 different years assessed, in 2019, Stockholm demonstrated the highest temperatures, then 1980 Madrid, then 1960 Budapest. While this at first may be an easy connection to make: the later years see higher average temps, which is reflective of the increasing temperatures. However, I

analyzed data from *three different locations*. Looking at average yearly temperatures between Budapest, Madrid, and Stockholm, I found that Stockholm is the coldest city, followed by Budapest, then Madrid. Despite 1960 Budapest being colder than 2019 Stockholm, there were 59 years of climate change between the data collected and this must be taken into account.

2. The loss over iterations profile for each of the three years and locations was very similar. What does this tell us? Well, the complexity of the data used for the different GDOs was similar. That makes sense: we were using average temperatures in each case. Also, the three years picked (1960, 1980, and 2019) most likely were very predictable, stable years that didn't see a lot of volatility or extreme weather events in those locations. For that reason, we were able to use similar learning rates and iterations in each case. While the climates in each location were different, they all followed relatively smooth trends without large fluctuations. If I had picked other years or locations that saw more extreme events or volatile behavior, the loss over iterations profile most likely would have looked different.
3. Given that the end theta1 values for Madrid and Budapest were negative, this tells us that the temperatures in those areas were trending down over the period studied. As we know, theta1 is an indication of the slope of the line that the gradient descent model fits to the data. This can tell us a few things: firstly, these downward changes could be simply statistical noise; a too-close snapshot of Europe's climate that doesn't provide enough context. It's not uncommon to have occasional years with cooling trends- oceanic patterns like La Nina are a good example of this. Zooming out, would we see an upward trend in temperature in the years following, thus tracking along the same gradual increase in global temperatures? On the flip side, it could be a mistake in data collection, and the temperatures for those years were tracking up as well. I think this is far less likely than the first: given the Earth changes on such large timescales, I think it could be expected to see some small downward changes along the way. And these small changes are not reflective of the greater pattern at play- like the Earth as a whole, Europe's temperatures are gradually increasing over time.