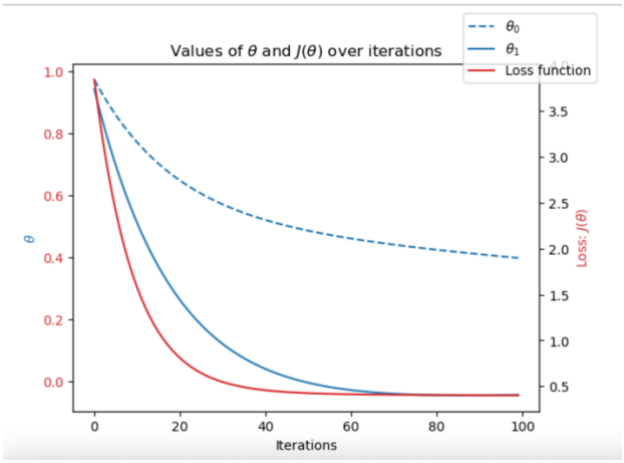
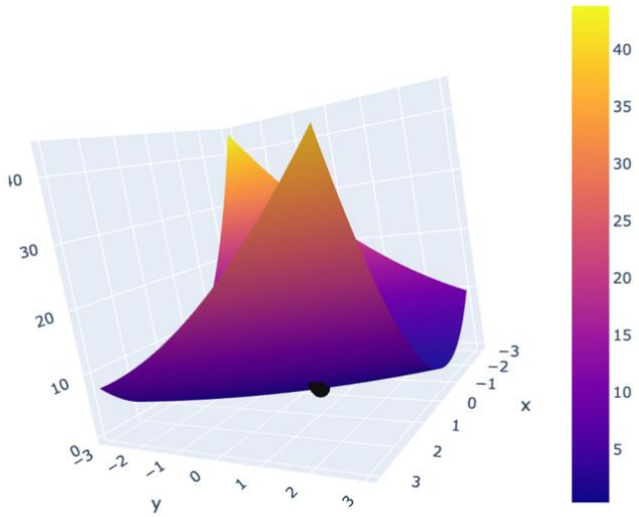
Machine Learning Analysis: Gradient Descent Results

Station #1: Basel, 2000

	Starting	Ending
Iterations	100	10
Theta0	1	-10
Theta1	1	-5
Step size	0.01	0.1

Profile and Loss Function:

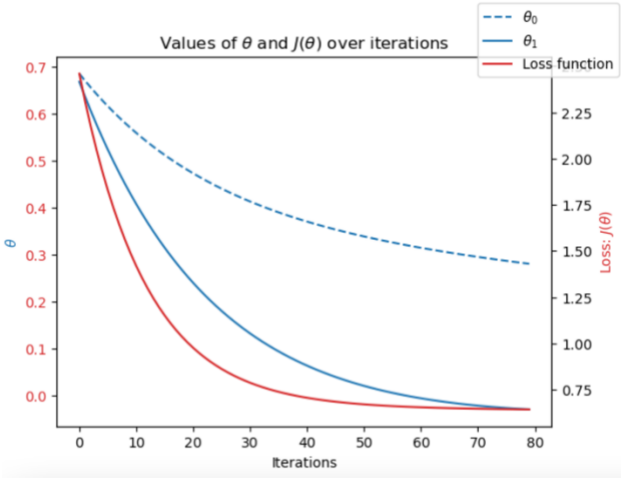


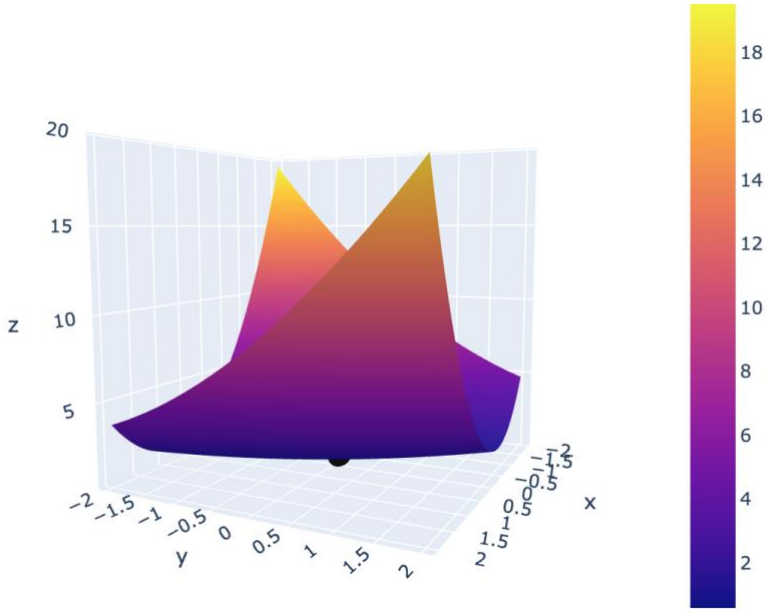


Station #2: Valentia, 1982

	Starting	Ending
Iterations	10	80
Theta0	1	0.06
Theta1	1	0.02
Step size	0.01	0.008

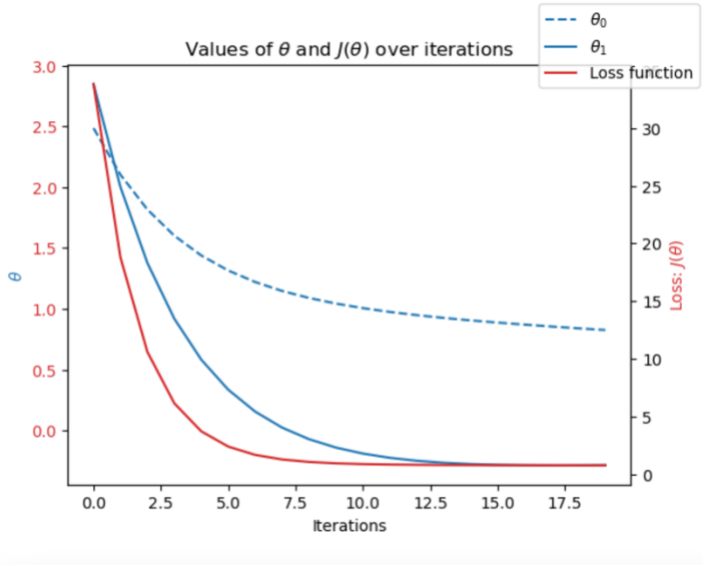
Profile and Loss Function:

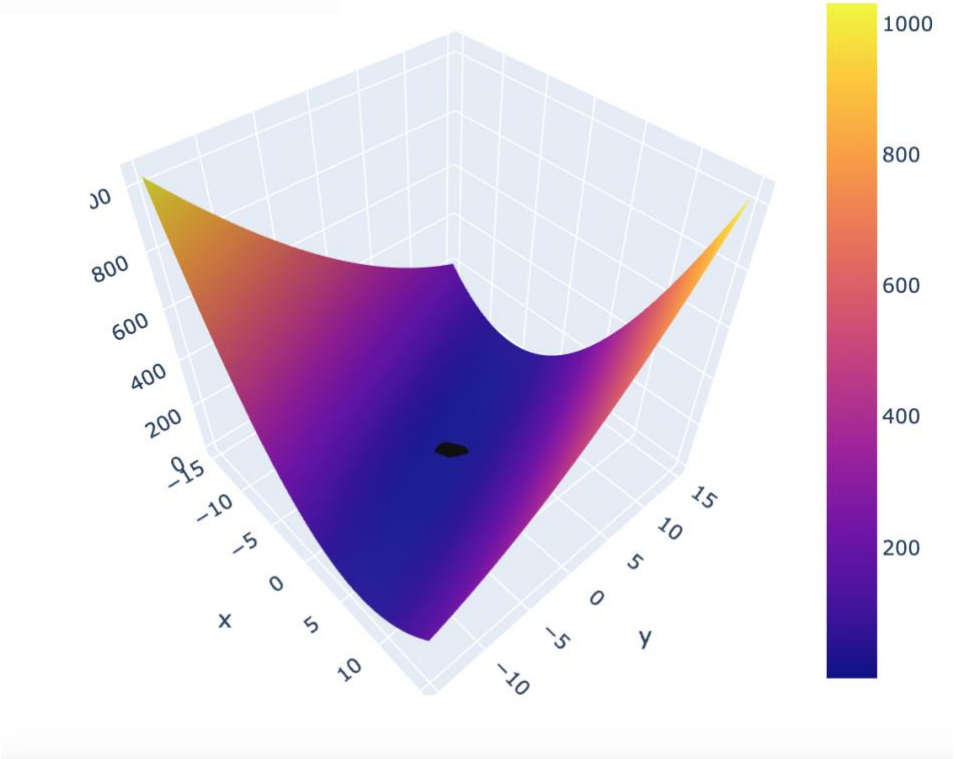




Station #3: Madrid, 1961

	Starting	Ending
Iterations	10	20
Theta0	1	1.16
Theta1	1	1.36
Step size	0.01	0.05





Observations

Sidebar: I don't understand how to make observations about the data itself from the gradient descent result. These answers come primarily from the initial data and scatterplots of temperatures per year.

1. Weather stations with more apparent fluctuations, based on the box and whiskers plot of each station over time, require more iterations to get to the local minimum.

2. It seems that the newer data (1982 and 2000) have more outliers, based on initial view of the scatterplots, but that could be because Madrid has a more stable climate overall.

3. There was a greater difference in the min and max temperatures I 1982 vs 2000, but again, the stations were different.