Basics of Machine Learning

1.3 Optimization in Relation to Problem Solving

Kendra Jackson

Original Loss Profiles

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Weather Station** | **Year** | **Theta0** | **Theta1** | **Iteration** | **Step Size** | **Result (Loss Path)** |
| Basel | 1960 | 3 | -10 | 500 | 0.01 | Large visual loss path (black line) |
| Basel | 1990 | -1 | 1 | 500 | 0.05 | Small loss path |
| Basel | 2020 | 0 | 0 | 500 | 0.05 | Very large loss path |
| Madrid | 1960 | 0 | 0 | 500 | 0.05 | Small loss path |
| Madrid | 1990 | 0 | 0 | 300 | 0.001 | Small loss path |
| Madrid | 2020 | 0 | 0 | 500 | 0.01 | Miniscule, a couple of dots |
| Stockholm | 1960 | 1 | 0 | 300 | 0.1 | Small loss path |
| Stockholm | 1990 | 0 | 0 | 100 | 0.01 | Very small loss path, basically a dot |
| Stockholm | 2020 | 0 | 0 | 150 | 0.1 | Small loss path |

Rerun optimizations (X,Y variables closer to objective based on loss path)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Weather Station** | **Year** | **Theta0** | **Theta1** | **Iteration** | **Step Size** | **Result**  **(Result Path)** |
| Basel | 1960 | -0.82 | 1.89 | 500 | 0.01 | Smaller, closer to goal loss path |
| Basel | 1990 | 0.07 | -0.07 | 500 | 0.05 | Incredibly small loss path |
| Basel | 2020 | 0.09 | 0.06 | 500 | 0.05 | Reduced loss path |
| Madrid | 1960 | 0.09 | -0.3 | 500 | 0.05 | Small loss path |
| Madrid | 1990 | 0.09 | -0.09 | 300 | 0.001 | Small loss path |
| Madrid | 2020 | 0.1 | -0.03 | 500 | 0.01 | Small loss path |
| Stockholm | 1960 | 0.32 | -0.68 | 300 | 0.1 | Smaller loss path closer to goal |
| Stockholm | 1990 | 0.052 | 0.029 | 100 | 0.01 | Loss path resembles a dot, guess is close to minimum and gradient descent has converged on answer quickly |
| Stockholm | 2020 | 0.18 | -0.05 | 150 | 0.1 | Small loss path |

General Statistics and Theta Values over Time

|  |  |  |  |
| --- | --- | --- | --- |
|  | Basel | Madrid | Stockholm |
| Min Temp | | | |
| 1960 | -3.19 | -2.32 | -2.35 |
| 1990 | -1.18 | -1.74 | -1.52 |
| 2020 | -1.68 | -1.63 | -1.07 |
| Change | +1.51 | +0.96 | +1.28 |
| Max temp | | | |
| 1960 | 1.82 | 1.68 | 1.68 |
| 1990 | 1.99 | 2.16 | 1.98 |
| 2020 | 2.30 | 2.32 | 2.15 |
| Change | + 0.48 | + 0.64 | + 0.47 |
| Difference between min and max | | | |
| 1960 | 5.01 | 4 | 4.03 |
| 1990 | 3.17 | 3.9 | 3.5 |
| 2020 | 3.98 | 3.95 | 3.22 |
| Theta 1 Rerun (X) | | | |
| 1960 | -0.82 | 0.09 | 0.31 |
| 1990 | 0.07 | 0.09 | 0.052 |
| 2020 | 0.09 | 0.1 | 0.18 |
| Theta 0 Rerun (Y) | | | |
| 1960 | 1.89 | -0.3 | -0.67 |
| 1990 | -0.07 | -0.09 | 0.029 |
| 2020 | 0.06 | -0.03 | -0.05 |

Observations about the data:

1. The mean temperature of each station increased over the time periods looked at (approximately 60 years)
2. Madrid had the largest increase in maximum temperature (+0.64) between the 3 stations looked at
3. Basel had the largest increase in minimum temperature (+1.51) between the 3 stations
4. The difference between the minimum and maximum temperature has decreased over the span of years looked at

With this data set, the starting points for gradient descent were already very close to the discovered minimum. The range of the data was very low with minimums in the -2s and maximums in the positive 2s. While looking at this data, I determined a good starting guess for my theta values would be around 0. This led to the loss paths being relatively small as they were close to their goal (the minimum). A couple of initial Theta 1 and theta 0 “guesses” were changed to demonstrate what larger loss paths may look like (see Basel 1960). With retrying the optimization with the respective X,Y coordinates found in the original graphs of each weather station and year looked at, the loss function did improve. Since the loss function decreases, we can say that the model, gradient descent is suitable to determine patterns in the temperature data over time.

Additionally, to further solidify the suitability of gradient descent for this data, we can look at parameter changes over integrations (Theta values). As demonstrated in the line graphs below, most of the parameters for the initial guesses converge and stabilize on a specific value as the algorithm approaches a solution (loss stabilizes/approaches 0)

**Graphs**

*The top graph demonstrates the changes in parameters (Theta0/Theta1) over iterations. Convergence and stabilization on a specific value as the loss stabilizes indicates the algorithm has/is determining a solution. (See first graph in each subsection)*

*The left-hand graph depicts the loss function with the initial theta guess. Some theta guesses were close to the minimums that we were looking for with gradient descent.*

*The right-hand graph depicts the loss function with the revised theta values based on the lower most point of the black line on the left hand graph. These loss functions are typically smaller as we are minimizing loss and more accurately predicting data.*

*Due to initial theta0 and theta1 “guesses” being close to the goal, loss paths are small however there are slight differences that can be noted between the left and right graphs below.*

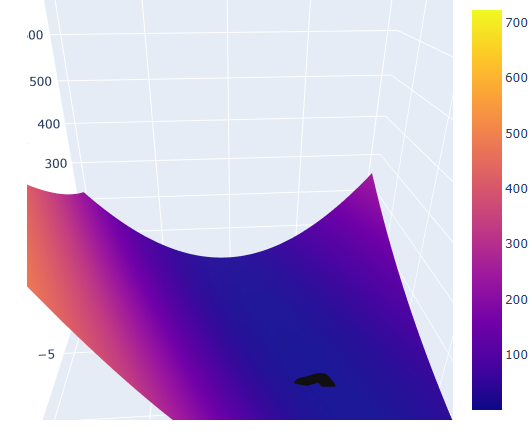
**Basel 1960**

A graph of values and loss

Description automatically generated

Original Loss profile with line towards minimum After optimization closer to objective

A graph with a curved curve

Description automatically generated with medium confidence ****

**Madrid 1960**

A graph of values and values

Description automatically generated

Original Loss profile with line towards minimum After optimization closer to objective

A graph of a graph

Description automatically generated with medium confidence A graph of a function

Description automatically generated with medium confidence

Stockholm 1960

A graph of a curve

Description automatically generated with medium confidence

Original Loss profile with line towards minimum After optimization closer to objective

A graph of a function

Description automatically generated with medium confidence A graph of a function

Description automatically generated with medium confidence

Basel 1990

A graph of values and loss

Description automatically generated

Original Loss profile with line towards minimum After optimization closer to objective

A graph of a function

Description automatically generated A graph of a function

Description automatically generated

Madrid 1990

A graph of values and values

Description automatically generated

Original Loss profile with line towards minimum After optimization closer to objective

A graph with a gradient color

Description automatically generated with medium confidence A graph of a function

Description automatically generated

Stockholm 1990

A graph of values and loss

Description automatically generated

Original Loss profile with line towards minimum After optimization closer to objective

A graph of a function

Description automatically generated A graph of a function

Description automatically generated with medium confidence

Basel 2020

A graph with a line and a red line

Description automatically generated

Original Loss profile with line towards minimum After optimization closer to objective

A graph of a function

Description automatically generated with medium confidence A graph of a function

Description automatically generated

Madrid 2020

A graph of values and loss

Description automatically generated

Original Loss profile with line towards minimum After optimization closer to objective

A graph of a function

Description automatically generated A graph of a function

Description automatically generated

Stockholm 2020

A graph of values and loss

Description automatically generated

Original Loss profile with line towards minimum After optimization closer to objective

A graph of a function

Description automatically generated with medium confidence A graph of a function

Description automatically generated with medium confidence