Weather Conditions and Climate Change with ClimateWins

Colin Widner
8/21/24

[Video Presentation](#)

Agenda

- Objective and Hypotheses
- Data and Biases
- Data Optimization
- Supervised Machine Learning
- Summary and Recommendation

Objective and Hypothesis

- **Objective:** ClimateWins, a European nonprofit, wants to use machine learning to help predict the consequences of climate change around Europe, and potentially, the world.

- **Hypotheses:**
    - How is machine learning used? Is it applicable to weather data?
    - ClimateWins has heard of ethical concerns surrounding machine learning and AI. Are there concerns specific to this project?
    - Historically, What have been the minimums and maximums in temperature been?
    - Can machine learning be used to predict whether weather conditions will be favorable on a certain day? (If so, could it be possible to predict danger?)
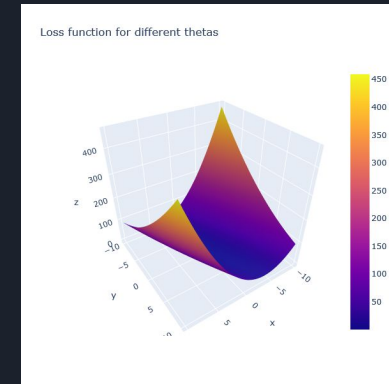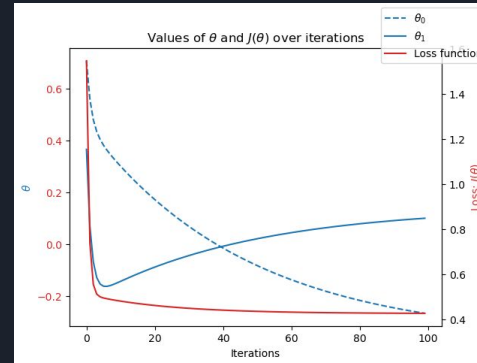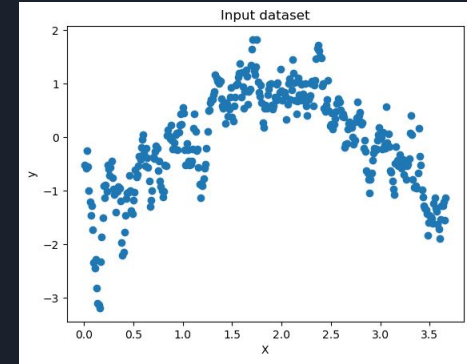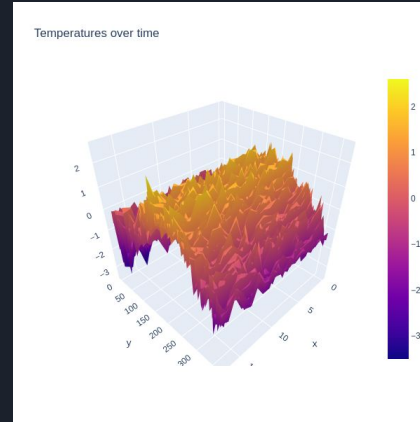
Data and Biases

- The data, collected by the European Climate Assessment and Data Set Project, was collected from 18 different weather stations throughout Europe.
- The data set contains records such as temperature, wind speed, and more every day from the late 1800s to 2022.

- **Biases**
  - **Collection Bias:** Changes in the instrumentation and measurement methods of the weather stations could lead to a collection bias.
  - **Sampling Bias:** With a sample size of only 18 stations, there could be a sampling bias when compared to the other numerous stations throughout Europe.
  - **Temporal Bias:** With a wide range of years in the dataset, some of the older data might not represent the current trends which might skew results.

Data Optimization

- The data was optimized through Gradient Descent to find the optimal parameters to minimize the error in our machine learning algorithm.
- First, the temperatures of all weather stations in a year were plotted .
- The temperature of one station for the year was then plotted with a scatterplot.
- A loss function was plotted to show the deviation between the predicted and actual temperature value.
- Finally, the loss function is plotted to find the local minimum, which optimizes the model.

Temperatures over time

Input dataset

Values of θ and $J(\theta)$ over iterations


Loss function for different thetas

K-Nearest Neighbors

- K-nearest neighbors is used to determine which group each data point belongs in based on what group it's nearest neighbors are in.

- Test Set Accuracy: **88%**
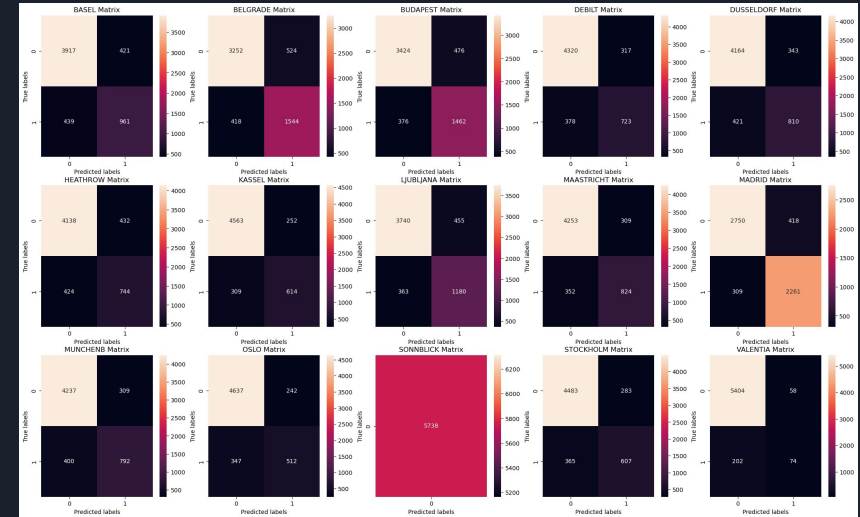
Decision Tree

- A decision tree is used to narrow down a solution by asking more and more specific questions

- Training Set Accuracy: **100%**
- Test Set Accuracy: **69%**

- There might be overfitting happening due to the high training set accuracy and low test set accuracy.
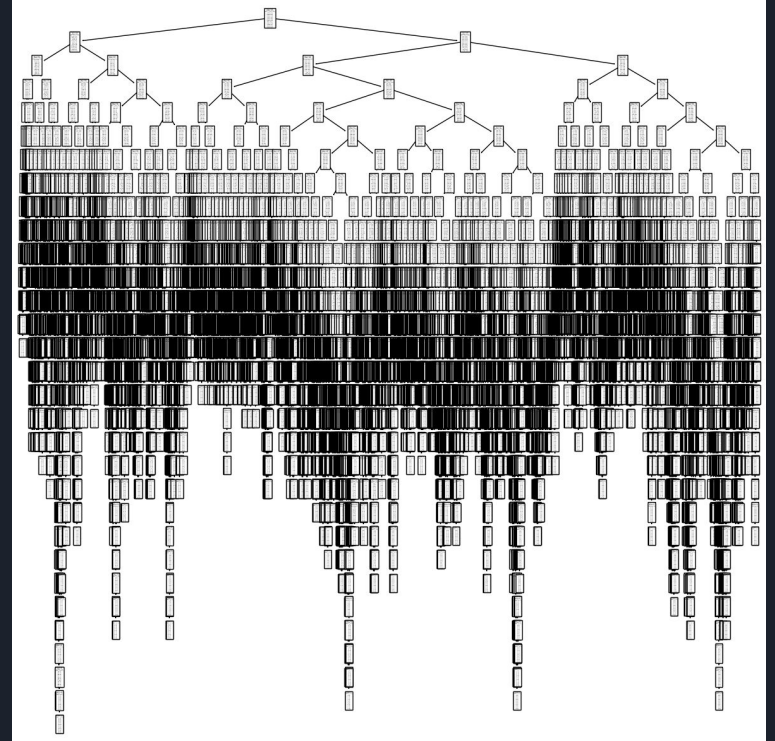
Artificial Neural Network

- An Artificial Neural Network (ANN) is used to compute and answer by using a linear combination of inputs that are then multiplied by weights.

- Training Set Accuracy: **87%**
- Test Set Accuracy: **83%**
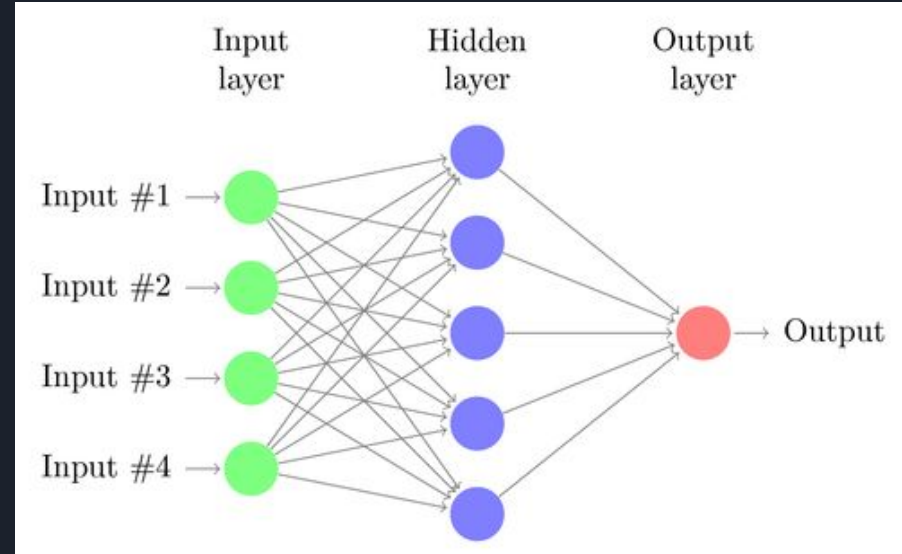
Conclusion and Next Steps

- Conclusions
  - Machine learning can be used to help predict weather patterns.
  - The KNN model had the highest accuracy rating when compared to the Decision Tree and ANN models.

- Next Steps
  - Continue testing supervised and unsupervised machine learning algorithms and optimize them.
  - Combine supervised and unsupervised models to create a climate model that helps predict future climate trends.
  - Prune decision tree to get better accuracy.

Thank You!

Colin Widner

GitHub