WEATHER CONDITIONS AND CLIMATE CHANGE WITH CLIMATE WINS

JENNY MIELKE

10/17/24

PROJECT OVERVIEW

- **Background:**
  - Climate Wins is a nonprofit organization in Europe focused on addressing and reducing the effects of climate change. The goal is to utilize machine learning to forecast future weather patterns including potential extreme weather events which have raised concerns in the last 10-20 years across mainland Europe.
- **Objective:**
  - Conduct thorough research on supervised machine learning algorithms and determine which model would be most useful in providing insights on temperature trends, extreme weather occurrences, and general climate variability based on historical weather data.
- **Hypothesis**
  - Machine learning can be utilized to predict if weather conditions will be favorable on a given day.
  - If the data is scaled, the machine learning model accuracy will be higher.
  - Utilizing historical weather data, advanced machine learning models can effectively and accurately predict extreme weather events.

DATA SET

Data Source:

- The dataset comprises weather observations from 18 different weather stations across Europe covering the period from 1960-2022 and has been collected by the European Climate Assessment & Dataset Project.
- Some of the information collected includes temperature, wind speed, snow, global radiation, precipitation, etc.

Data Bias:

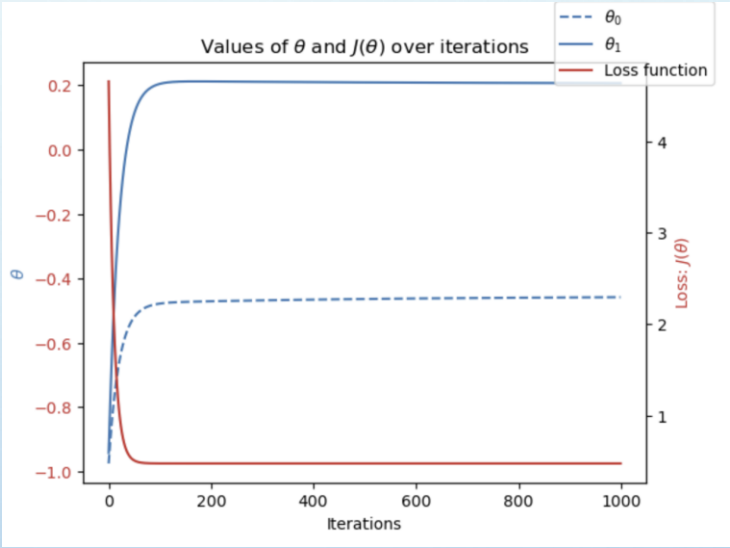
- **Sample bias**: This data is limited to only 18 weather stations and further excludes additional stations across Europe as well as global weather patterns.
- **Collection Bias**: There is a portion of data that contains dated observations which could impact the accuracy of a machine learning model.

Data Accuracy:

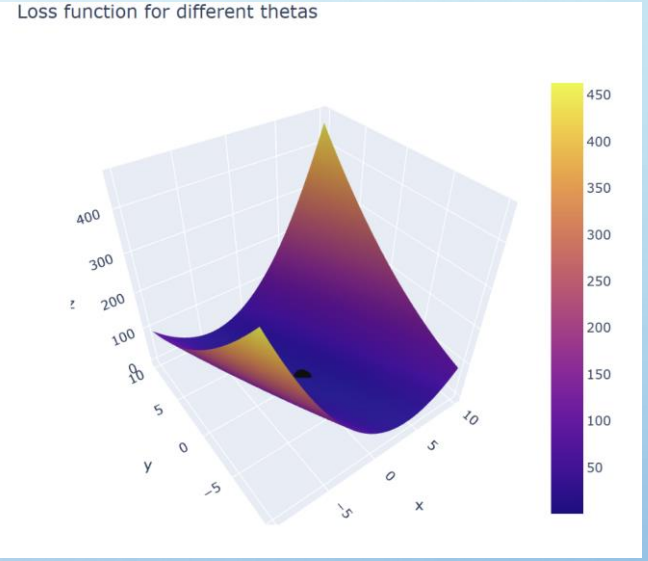
- The accuracy of the data depends on which machine learning model is used which will be discussed further in this presentation.

DATA OPTIMIZATION

- Purpose: Optimization techniques can be used to improve the performance of a machine learning model as well as ensuring accuracy of the training.
- Gradient Descent was the optimization algorithm used with the temperature data from the 18 weather stations to assess how well it performs and whether it can be incorporated into larger machine learning functions.
- Gradient descent is an iterative optimization method that minimizes a function by adjusting parameters in the direction of the gradient.



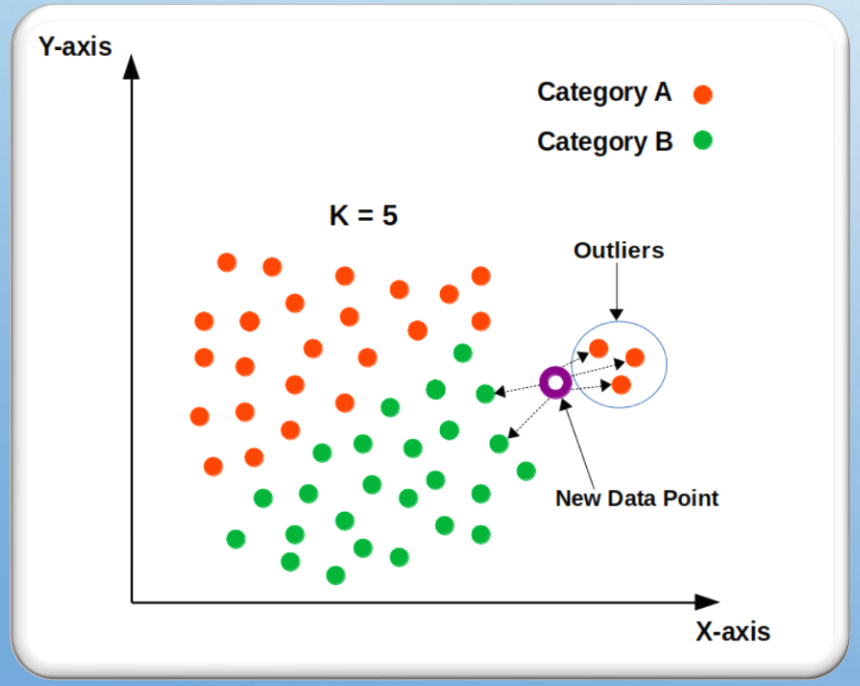
Fg. 2 illustrates where parameters theta0 and theta1 as well as where the loss function converge.



Fg. 3 For all weather stations, the loss function approached zero indicating a higher accuracy.

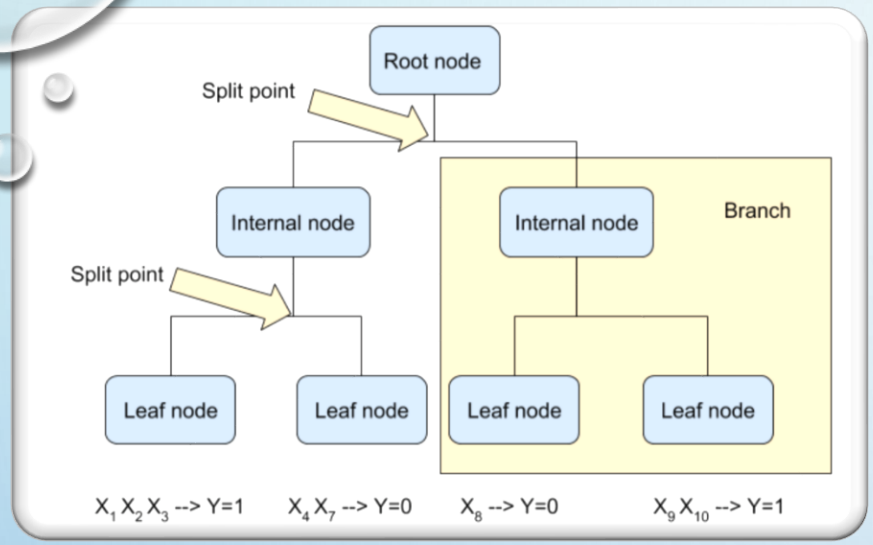
Weather Station	Year	Theta0	Theta1	Iterations	Step Size
Oslo	1960	-1	-1	1,000	0.009
	1992	0	0	600	0.01
	2019	0	0	500	0.1
Budapest	1960	-1	-1	600	0.01
	1992	0	0	500	0.1
	2019	0	0	500	0.1
Stockholm	1960	1	1	800	0.01
	1992	1	1	500	0.01
	2019	0	0	350	0.1

Fg. 1 contains the parameters used for optimal optimization function



SUPERVISED MACHINE LEARNING: MODEL 1

- **K-Nearest Neighbor (KNN)**
  - Instance-based algorithm used in machine learning that classifies data based on the similarity between points.
  - In terms of the Climate Wins analysis, the KNN model was utilized in the prediction of whether a day is pleasant or unpleasant.
  - Accuracy
    - Training: 86%
    - Testing: 88.12%

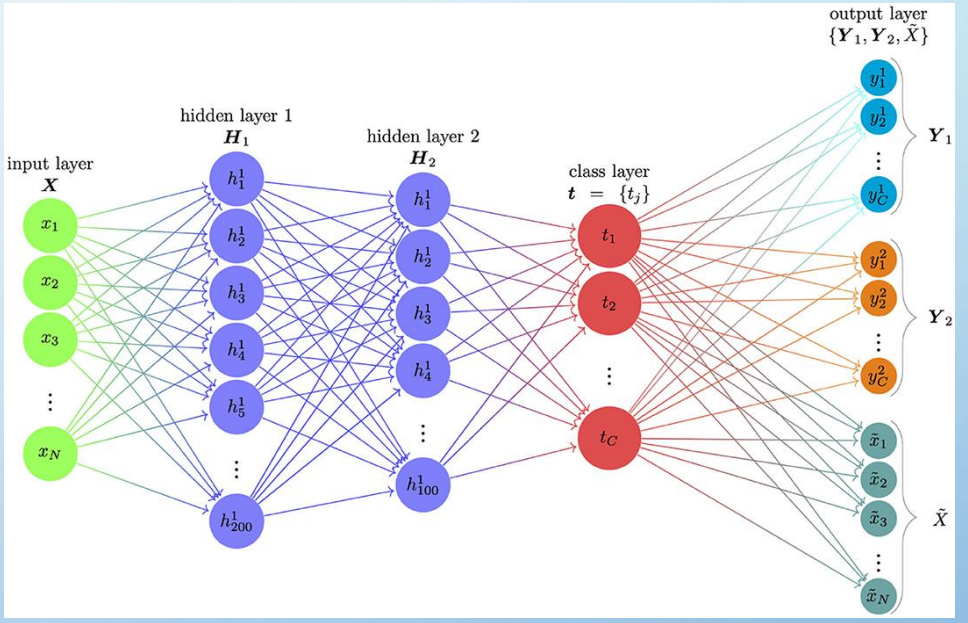


SUPERVISED MACHINE LEARNING: MODEL 2

- **Decision Tree**
  - Used to model decision-making by breaking down complex decisions into a series of simpler, conditional statements, helping to classify data or predict outcomes.
  - Specifically for Climate Wins, the decision tree can be used ask questions to understand and predict the complexity of weather on pleasant vs. unpleasant days.
  - Due to complexity of the model in this project, the decision tree requires pruning to make it easier to interpret.
  - Accuracy score:
    - Training: 100%
    - Test: 100%

SUPERVISED MACHINE LEARNING: MODEL 3

- **Artificial neural network (ANN)**
  - A model inspired by the human brain, consisting of interconnected nodes that process and learn from the data. It is efficient in classification and regression by adjusting parameters through training to minimize prediction errors.
  - This model dives even deeper and more complex into predicting the weather on a given day.
  - Accuracy score:
    - Training: 95%
    - Test: 99%



CONCLUSION / NEXT STEPS

- The findings thus far indicate that machine learning can be highly accurate in predicting extreme weather events on a given day which confirms the hypothesis.
- While the accuracy rates are slightly lower in the artificial neural network model, it is more efficient in the complexity of climate data in the identification of subtle patterns in comparison to the KNN model and decision tree.
- **Next steps**
  - Improve the accuracy of the ANN model by tuning hyperparameters
  - Analyze additional unsupervised machine learning techniques

THANK YOU!