Weather Predictions Using Machine Learning

Date: June 16, 2024

Nirav Bariya

Project Objective

Utilize machine learning techniques to predict the impacts and consequences of climate change.

Hypotheses

Machine learning can accurately predict historical extreme temperatures using temperature data from stations across Europe.

Supervised machine learning models can predict the likelihood of extreme weather events based on historical data.

Machine learning models can be used to predict whether weather conditions on a given day will be favorable or unfavorable for outdoor activities.

About the Dataset

The dataset used for this project comes from **European Climate Assessment & Dataset (ECA&D).**

High quality data provided by participation institutions.

Quality control and homogeneity procedures used to ensure reliability and consistency of the data.

Challenges with the Dataset:

- There may be errors with the most recent data which may not have been through validation processes.
- Instrument malfunction or calibration errors can lead to systematic errors.
- Extreme weather conditions might be considered as errors if they fall outside a certain range.

Observation:

Despite the challenges, the dataset we are using is from 1960-2022. The datasets are regularly updated to resolve any errors that might have been in the historical data.

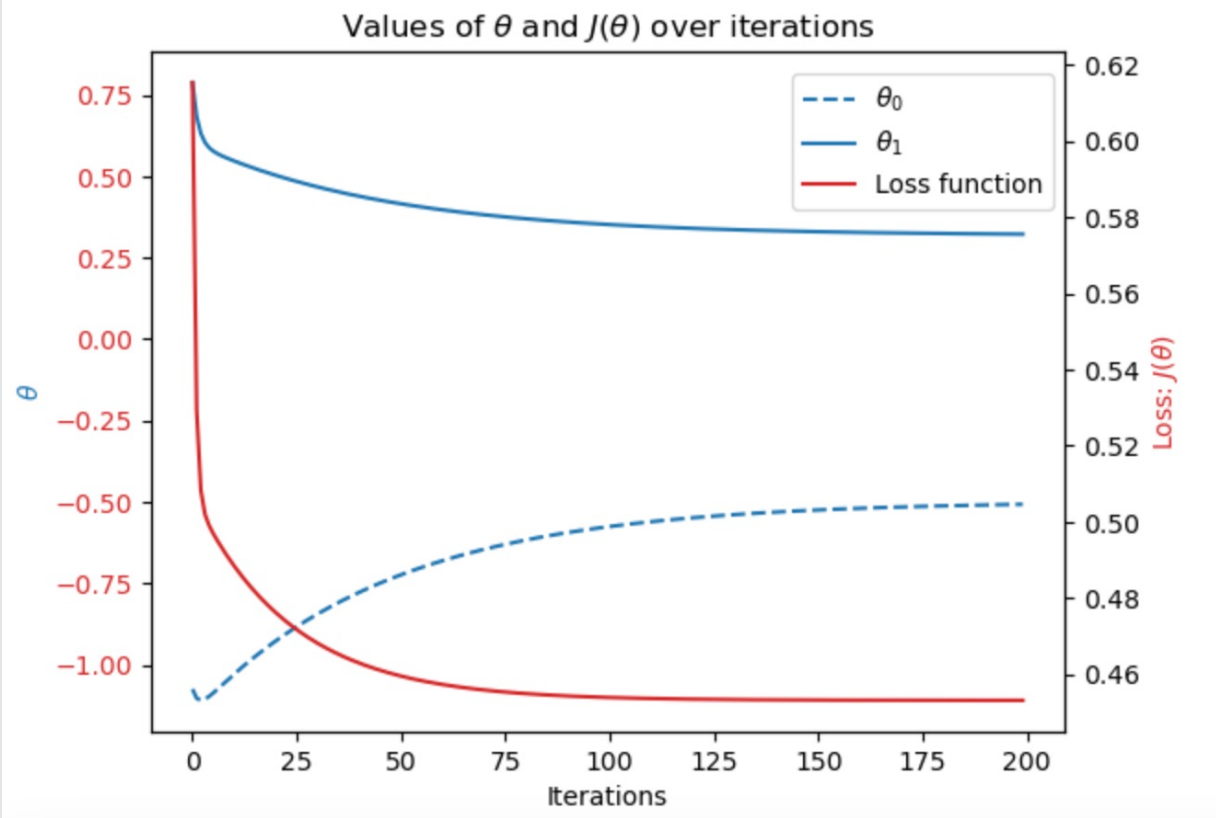
Predicting Temperature Using Optimization

Machine learning is a type of optimization problem.

There is an error while modeling a real-life situation using a mathematical equation.

One of the tasks after choosing a model is to minimize this error.

Using gradient descent to find parameters that best explains the relationship between temperature and day of the year.



Weather Station: Heathrow
Year: 2021

Supervised Machine Learning Algorithms

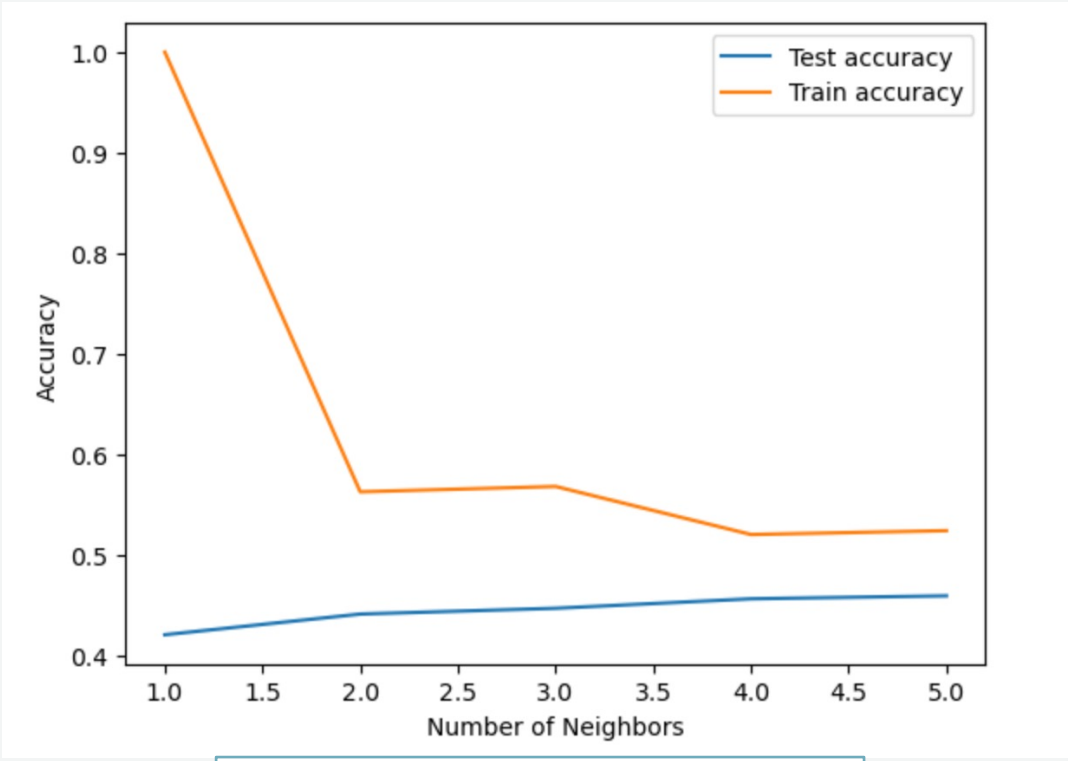
How do we predict whether a day will be pleasant or unpleasant?

- K-Nearest Neighbors (KNN)
- Decision Trees
- Artificial Neural Network (ANN)

How do we predict whether a day will be pleasant or unpleasant?

K-Nearest Neighbors (KNN)

- The algorithm identifies the k-nearest data points using distance formula for a given data point.
- The model's accuracy varies as we change number of neighbors.
- After a certain point adding more neighbors won't change the accuracy.



Training Accuracy: 52%
Testing Accuracy: 45%

How do we predict whether a day will be pleasant or unpleasant?

Decision Trees

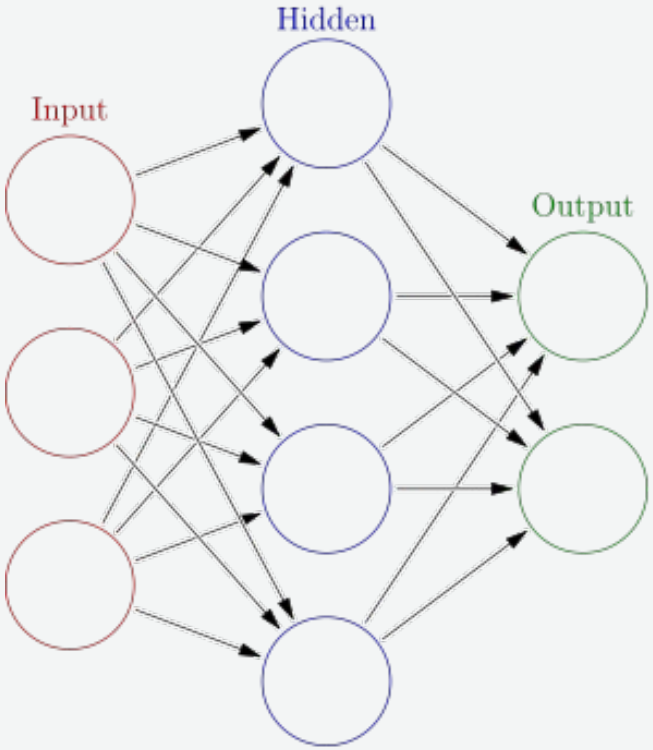
- Decision Trees work by recursively splitting the data into subsets based on input features.
- They try create homogeneous subsets by calculating impurity of a node and reducing the impurity to 0.
- The height of the decision tree affects the accuracy of the algorithm.
- Simplifying the decision tree by pruning the tree does not always help.
- Extreme weather events are rare and pruning the decision tree may consider these events as outliers. This will defeat the purpose of the model which is to predict extreme weather conditions.

Training Accuracy: 60%
Testing Accuracy: 63%

How do we predict whether a day will be pleasant or unpleasant?

Artificial Neural Network (ANN)

- Multilayer Perceptron Model has hidden layer(s) consisting nodes which takes input variables and computes an output by assigning weights to each input variable.
- Parameters such as number of layers, nodes per layer, number of iterations, and tolerance affects the accuracy of the model.
- Different values for the parameters were assumed to find out which set of values works the best.



Source: Wikimedia

How do we predict whether a day will be pleasant or unpleasant?

Artificial Neural Network (ANN)

Layers	Iterations	Tolerance	Training Accuracy	Test Accuracy
(25, 25, 25)	500	0.0001	68%	63%
(50,50,50)	1000	0.0001	78%	62%
(60,60,60)	1000	0.0001	80%	61%
(100,100)	500	0.0001	90%	59%
(100,100,10)	1000	0.0001	78%	58%
(200, 200, 50)	500	0.0001	100%	57%
(35,35,35)	**1000**	**0.0001**	**79%**	**70%**
(35,35,35)	1000	0.001	71%	60%
(35,35)	1000	0.001	74%	66%

Result:

- As demonstrated, increasing the number of nodes allows ANN to learn highly intricate patterns. But it can also overfit the data and catch the noise in the data. It doesn't necessarily increase the testing accuracy.
- In my view, ANN with 35 nodes in three layers, 1000 iterations, 0.0001 tolerance level gives the best training and testing accuracy.

Summary

Hypotheses

- Machine learning can accurately predict historical extreme temperatures
- Machine learning models can be used to predict extreme weather conditions
- Machine learning models can be used to predict whether weather conditions on a given day will be favorable or unfavorable

- Models used: KNN, Decision Trees, ANN
- Out of the three models we have chosen, ANN works the best so far with 35 neurons in three layers.

Best Models

Next Steps

- Incorporate regularization techniques to reduce overfitting in the neural networks.
- Use cost-sensitive learning to assign more cost for misclassification of smaller class (extreme weather events).
- Perform grid search to optimize hyperparameters

Thank you

Do you have any questions?

Contact Info:

Email: niravkb64728@gmail.com

LinkedIn: https://www.linkedin.com/in/nirav-bariya-a87984122/