Climate Classifier Using Weather Station Data

Brayden Smith
CS 472, Winter 2021
Department of Computer Science
Brigham Young University

Abstract

In hopes of better understanding how seemingly small features of an area contribute to the general nature of an entire region, this project aims to apply understanding of machine learning techniques to classify data received from weather reporting stations around the United States and inference the type of climate an input is coming from. Through determining which features are vital to differentiating the different climates I believe that attention can be brought to those small changes that may lead to more drastic changes in the health of a region. I hope that a model such as the one I have built could serve as an advisory tool to the monitoring stations around the world, serving as early-detection systems to catch drastic changes in key indicator features, but also as possibly useful forecasting systems given calculated data.

In this report, I found that two models, the decision tree classifier was extremely useful in trimming down the wide range of numerical features present in an environement, and also a backpropogating multi-layered perceptron classifier was then trained on data spanning months of reports, I achieved the highest consistent model accuracy beyond .9.

1 Introduction

Due to the large amounts of regional system data from many remote observation sites, machine learning has oftentimes been a tool repeatedly used to understand the vast amounts of data available. Notable research groups from universities such as those at MIT and Stanford [Faizollahzadeh ardabili et al., 2019] have built models that through the use of convolutional neural networks and deep learning predict the region and climate from image data. However, the interpretability of these models is oftentimes not very useful. Thus the models that I will discuss stretch the abilities of blind machine learning models, mainly processing on raw numerical input data as provided by weather stations.

ML has proved repeatedly to perform extremely well tackling similar problems, for example multi-layered perceptron classifiers have been shown to improve statistical forecasting [McGovern et al. 2017] and other purely data driven climate models have improved post-processing [Rasp and Thuerey, 2020]. Using these data driven models can be extremely useful, but many other factors into the current situation of a region are not always able to be numerically measured, for example things such as cloud condensation levels, thus models such as the CNNs are used to improve climate and weather forecasting. While all the models are useful, climate scientists still need to be able to interpret the results of the models, thus though the numerical driven models can be less versatile, they remain as the benchmark of data-driven climate analysis. [Schneider et al., 2017]

In this project I hope to build off of the knowledge of those models built before and evaluate more ways in which the numerical data continuously being collected can be interpreted. Our primary goal is to evaluate just how viable a machine learning model would be in classifying regions off of local weather and geological reporting data. Since data is widely available and continuously being reported, a model would have ample time and information to advise climate scientists as to those features that are changing and are indicative of a changing climate in a local region.

2 Methods

In this section, I cover our methods for building our most accurate model capable of classifying a region's climate given regional geological and meteorological data.

2.1 Data Source

The Natural Resources Conservation Service (NRCS) services a number of weather reporting stations around the United States, the publicly available databases contain measurements taken by these sites over the past decade of service. Through obtaining reports from these stations, I began to notice the wide range of features that oftentimes only a single station would report on, and on the other hand, missing data was frequent throughout the entries available to be trained on. Most of the data was that of monthly measurements from hundreds of stations from May 2011 to March 2021

2.2 Köppen Climate Classification Labels

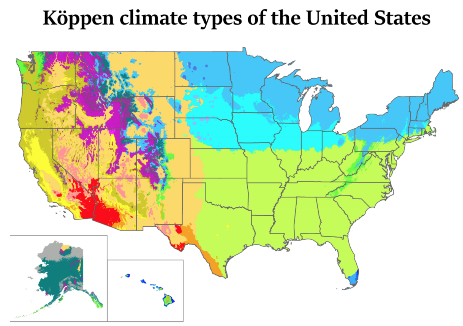
It appears that as each station reports data into the NRCS databases, it signs off with a signature that could be traced back to identify which weather station reported the data. Thus each report could be traced back to a specific location, and from there given the location of the station, the surrounding climate could be determined. In order to label the now known locations of the weather reporting stations I turned to the Köppen Climate classification system, a widely accepted system consisting of 30 labels for the various climates across the globe. Due to the fact that the climate groups are mainly separated by that of vegetation groups, I felt that the ecosystem conditions that the Köppen climate classification system focus on would be highly indicative of the climate and thus were useful labels for our purposes.

However, considering our goal was that of accurate climate classification, the 30 labels posed a difficult classification task for all of our future models. Nevertheless, models trained with the 30 labels were able to achieve fairly high levels of accuracy with respect to the large number of possible labels. However I did combine some of the climate groups into categories similar to that of how the Stanford classification study by Johnson et al. (2019) which simplified the large number of labels, the following table shows the 13 climate super-classes.

Table 1. Climate Superclasses

Superclass	Köppen Symbols
0. Arctic/alpine	EF, ET
1. Arid - cold	BWk, BSk
2. Arid - hot	BWh, BSh
3. Continental - hot	Dsa, Dwa, Dfa
4. Humid subtropical	Cwa, Cfa
5. Mediterranean	Csa, Csb, Csc
6. Ocean	Ocean
7. Oceanic	Cwb, Cwc, Cfb, Cfc
8. Subarctic (continental - cold)	Dfc, Dfd, Dsc, Dsd, Dwc, Dwd
9. Tropical monsoon	Am
10. Tropical rainforest	Af
11. Tropical savanna	Aw, As
12. Continental - warm	Dsb, Dwb, Dfb

The NRCS weather reporting stations scattered throughout the United States that report the data used for the following models were chosen off of the fact that they each resided in areas of different classification. The following graphic shows the Köppen Climate classifications for the United States. Notice the wide range of climate groups especially throughout the Western United States where most of the weather reporting sites were chosen.



2.3 Station data

Initially, when trying to build usable models with our data, I recognized the large number of features available to train on. Since early training results were abysmal, I were lead to the need to search for some subset of all of the features reported to those that were the most useful and consistent. Given the fact that many of the stations reported different subsets of data, and the fact that for the most part missing data was consistent throughout any feature, I initially picked out some of the features I felt were most consistent and likely to be important, and then proceeded to use forward selection to look for any features that were missing. The following table shows the list of features our final models were trained on.

Month of the year, initially year was included
Air Temperature Average, degrees Fahrenheit
Air Temperature Maximum
Air Temperature Minimum
Precipitation Accumulation, Inches
Average of Precipitation
Soil Moisture Percent, Percent
Relative Humidity, Percent
Solar Radiation Average, Watts per meter squared

Table 2: Final set of features used in training and evaluation

2.4 Missing Data

As has been mentioned previously, many of the rows on the weather stations reports were filled with missing values, while initially this was a frustrating problem to deal that that lead to inconsistent and varying results, the final list of features were some of the more frequently and consistently reported features as given by most of the weather stations around the United States. Thus, in the few cases where a missing feature was had in these more consistently reported columns, monthly averages were inserted. Understandably, by including monthly averages, possible anomalies were covered up, but by using monthly averages I found the highest levels of classification accuracy.

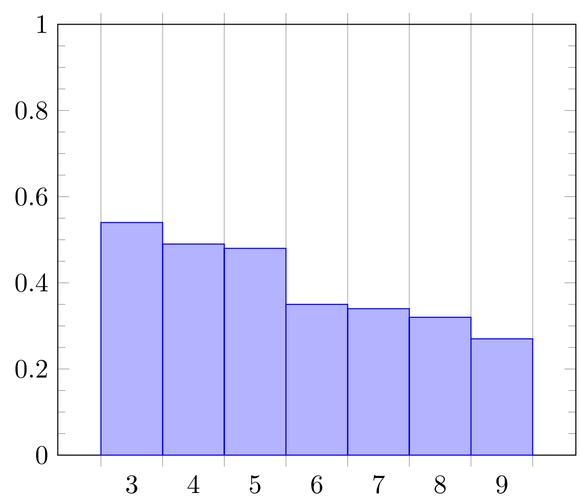
3 Results

The following section consists of our results, from initially trimming our features down through a Decision Tree Classifier, to eventually building a notable Multi-Layered Perceptron Classifier for tackling our problem of classifying weather data.

3.1 Initial Results

One of the first types of models that I built to begin understanding the relationships within the data was that of a Decision Tree Classifier, Decision Trees operate under the advantage that they are simple to understand and interpret, oftentimes perform well with larger datasets, and have in built feature selection. Whilst building different decision trees with different parameters, I repeatedly noted which features were used at the higher levels of the tree to split the data, then assuming those features carried large amounts of information, I ended up fine-tuning our list of usable features with the decision trees.

Figure 3: Accuracy given k CV folds for a DTC



From our decision trees, I found the highest level of accuracy to be consistently around .53 when classifying using the 13 super-classes as reported in Table 1. Highest levels of accuracy when using the 30 Köppen Climate Labels and a decision tree were consistently around .42. Since a decision tree that had been trained on the entire set of data will perform with 100% accuracy given any data from the original set, I repeatedly performed evaluations of decision trees through Sklearn's cross-validation package. Due to the fact that the results of cross-validation were fairly consistent until unnecessarily large numbers of folds, I feel confident that just our decision tree model was consistent enough to make fairly good predictions as to the type of climate given any sort of new data.

3.2 Multi-layered Perceptron

Despite our initial success with the decision tree, I felt that no matter the changes to the parameters of the model or the polishing of extra data from weather stations that I could ever break this fairly sub-par levels of accuracy by the tree. Thus, as I had originally intended to do, I turned to looking at the effectiveness of multi-layered perceptrons.

Multi-layered perceptrons are universal function approximators, and thus are often used in all areas of research. Given our large amount of easily available numerical training data I were fairly confident that I could easily build accurate models capable of performing our task of classifying climate given numerical data. Since I had already done our exploration of decision trees by the time that I began evaluating multi-layered models, I kept the scope of features for our entire work within the selected features as selected by the decision trees.

Initially, the results of perceptrons with only a fairly default settings and minimum number of hidden layers (4 hidden nodes) I achieved accuracies well below than what I was expecting, initial accuracy results came in around ranges of .10-.25 for our 30 Köppen climate classifications, and reached a maximum of around .32 even when using just the 13 super-classes. I then began an iterative search for looking for parameters that appeared to increase our generalization accuracy the most.

For better performing backpropogation, I eventually increased the learning rate a bit from the package's default settings and introduced momentum into the model's learning and found incremental improvements in accuracy. By initially starting with no hidden layers to 4 hidden layers with a maximum of 32 nodes per hidden layer. I frequently ran into the issue that with the larger models that the weights of the model would zero out entirely. But given enough training and reasonable model structure and parameters, our accuracy improved from the original average of .32 to .43.

After working with a dataset of 4500 reports from weather and geological stations from around the U.S. I report that the best classification model I found was that of a multi-layered perceptron of learning rate $1e{-2}$ and momentum of .9. The model should also contain 1 hidden layer of 8 nodes that learns using the stochastic gradient descent in order to update weights. Convergence appears to converge fairly quickly but early stopping of the model frequently produces fairly accurate models that don't appear to over-fit on the training data, as the model classifies our testing dataset of roughly 500 reports at an accuracy consistently around .43 averaged over 10 runs using the assumed optimal parameters and using the 13 super-class labels. Notably, small variations in the parameters usually greatly negatively impacted classification accuracies, so I felt extremely unsatisfied with our results, possibly I had found some unstable equilibrium that wasn't really good in reality.

3.3 Increasing Time Frame

Throughout our search efforts of building a better model, at each iteration the model is fed a single month's worth of data and then a prediction is made, but considering the source of our data, I had data that was collected at the same locations for multiple months in a row. It seemed extremely unlikely that a model would be able to accurately predict the climate of an area just based off of 1 data point from a month of data, so in search of a better classifying model I turned to exploring the abilities of a model that was fed multiple consecutive months of data and then made a prediction.

A large benefit of transforming multiple months of data is that the number of possibly useful features could be increased dramatically. I were fairly disappointed originally at the small number of usable features provided in our dataset. But with multiple months of valid data our model would now be separating even higher dimensional feature space.

In iteratively testing and evaluating these new models that used multiple months of data, I felt it important to not use too many months of data as that would mean a climate scientist using a model of this design would require more time and effort to be able to create the needed input into the model to make a prediction.

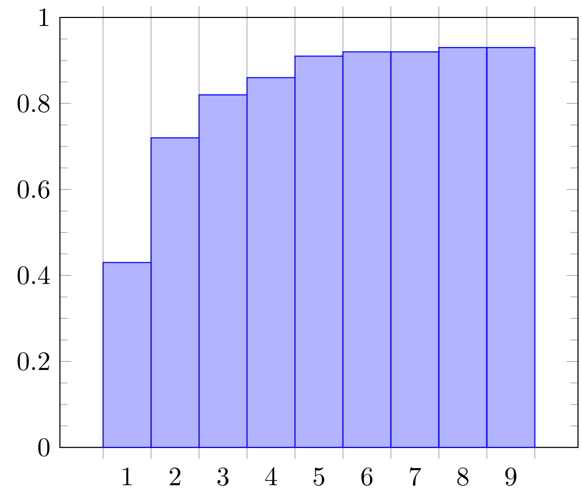
3.4 Final Model

Returning to the iterative search for the best model, I explored models that performed well on inputs consisting of input covering the span of 2 months to 10 months, while using both label systems as discussed. I verified the natural and intuitive trend that as the number of consecutive months increased that the general accuracy of the model improved drastically.

I began by returning to the decision tree and I ended up adding in some of the features that initially had been removed due to their apparent insufficient information gain. However, through testing on multiple sizes and shapes of models, I applied Occam's razor and returned to the simpler models of consecutive months using only the previously-mentioned features from Table 2 since the accuracies were generally better than those that had different features and many more missing attributes.

In similar shape to the model that was fed a single month of input, I found optimal models to consistently use a learning rate of 1e−2 and I found both Stochastic gradient descent and Adam to both be optimal learning strategies. As for the number of hidden layers, for roughly every 2 months added to the input, I would add another hidden layer where each layer had 128 nodes. Thus for example, our most optimal model for data of 3 consecutive months had 3 hidden layers, each with 128 layers, wheras our most optimal model for 10 consecutive months had 6 layers, consisting of 128 nodes for the first 3, 64 nodes for the last 3. All models were trained in a similar fashion, with 4500 inputs to be trained on and 500 new inputs to be evaluated on. All accuracies in Figure 4 are that of testing accuracy.

Figure 4: Classification Accuracy given k months



4 Conclusion

After exploring the limitations and abilities of these perceptron classifiers, I feel confident that the models I have built are pushing the edges of the abilities of these machine learning models. They have thus far proven to be adept classification models that given numerical weather data can quite accurately classify climate of local regions. Our most accurate models required that of multiple months of data, but given the continual collecting of data occurring at these weather stations, I feel these models could be promising advisory models to monitor the health and state of nearby areas. While models struggled to break .50 accuracy with only 1 month of data, a still fairly notable achievement considering the 13 possible labels, through testing and expansion of multiple months, I optimized our model to in some cases classify accurately with accuracy greater than .94.

5 Future Work

A large benefit of transforming multiple months of data into different pairs of inputs was that our possible usable data set's size increased. Given just 12 months of data from a single weather report one can create 66 unique pairs of data. For the purposes of our models, I kept only the pairs that were consecutive months since it seemed more realistic and useful to restrict the scope of data, thus 12 inputs become 11 consecutive pairs. I feel confident that through further training and data collection and sanitation that even the models that use an input of around 3 months would be able to match the performance of those models that use 6-10 months of input.

In restricting the scope of our work to that of just weather data from here in the United States I ignore the many climates and environments that cover other areas of the world. Interestingly, the span of the data I use in this work covered 9 of the 13 super-classes and 19 of the 30 Ködden climate labels. However, expanding the data to classify climate data from around the world would be invaluable.

References

[1] [Faizollahzadeh ardabili et al., 2019] Faizollahzadeh ardabili, S., Mosavi, A., Dehghani, M., & R. Várkonyi-Kóczy, A. (2019). Deep learning and machine learning in hydrological processes, climate change and earth systems: A systematic review. Inter-Academia. doi:10.20944/preprints201908.0166.v1

[2] [Johnson et al., 2019] Johnson, D., Ngo, T., and Fernández, A. (n.d.). Climate Classification Using Landscape Images (Tech.). Stanford University. doi:http://cs229.stanford.edu/proj2019spr/report/23.pdf

[3] [McGovern et al. 2017] McGovern, A., Elmore, K. L., Gagne, D. J., Haupt, S. E., Karstens, C. D., Lagerquist, R., . . . Williams, J. K. (2017). Using artificial intelligence to improve real-time decision-making for high-impact weather. Bulletin of the American Meteorological Society, 98(10), 2073–2090.

[4] [M. Kottek et al., 2006] M. Kottek et al., “World Map of the Köppen-Geiger climate classification updated,” Meterol. Z., vol. 15, pp. 259–263, 2006. DOI: 10.1127/0941–2948/2006/0130.

[5] [Rasp and Thuerey, 2020] Rasp, S., and Thuerey, N. (2020). Data-driven medium-range weather prediction with a resnet pretrained on climate simulations: A new model for weatherbench. arXiv preprint arXiv:2008.08626.

[6] [Schneider et al. 2017] Schneider, T., Teixeira, J., Bretherton, C. S., Brient, F., Pressel, K. G., Schar, C., and Siebesma, A. P. (2017). Climate goals and computing the future of clouds. Nature Climate Change, 7(1), 3–5.