Wildfire Burn Area Predictions

Eliza Klyce, Cuthbert Sun, Rysen Otomo ({eklyce, csun3, rysotomo}@stanford.edu)

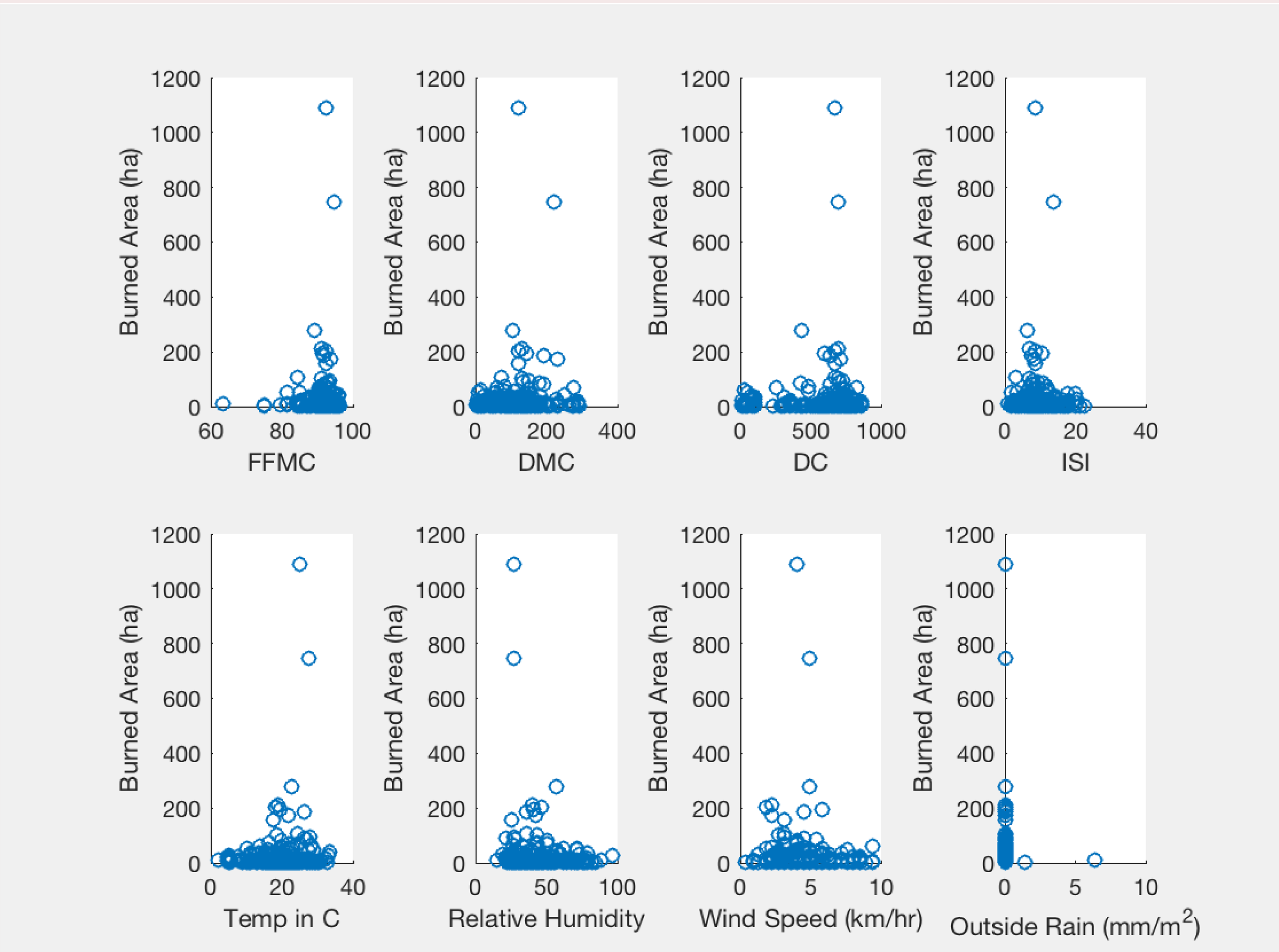
Stanford University CS 229

Motivation

- Project Goal: We wish to learn how to accurately classify the size of a wildfire using techniques such as, regressions, SVM, and neural networks.
- Intuitively, factors such as, temperature, precipitation, and wind would be the factors which impact the size of these fires.
- In addition to these, we considered other factors and allowed our algorithms to determine which features are the most relevant to predicting wildfire size (burn area).

Dataset and Methods

- Initial predictions on wildfire size were not good, and we realized that our features did not have a correlation to the output. PCA shows that most of our variance (more than 99%) comes from the first 4 features.



Principle Component Analysis:

Feature	Percent of Total Variance
FFMC (Fine Fuel Moisture Code)	85.0670
DMC (Duff Moisture Code)	11.3705
DC (Drought Code)	3.0889
ISI (Intraseasonal to Interannual)	0.3601
Temperature (°C)	0.0702
Relative humidity (percentage)	0.0227
Wind speed (km/h)	0.0160
Outside rain (mm/m^2)	0.0044

Methods and Results

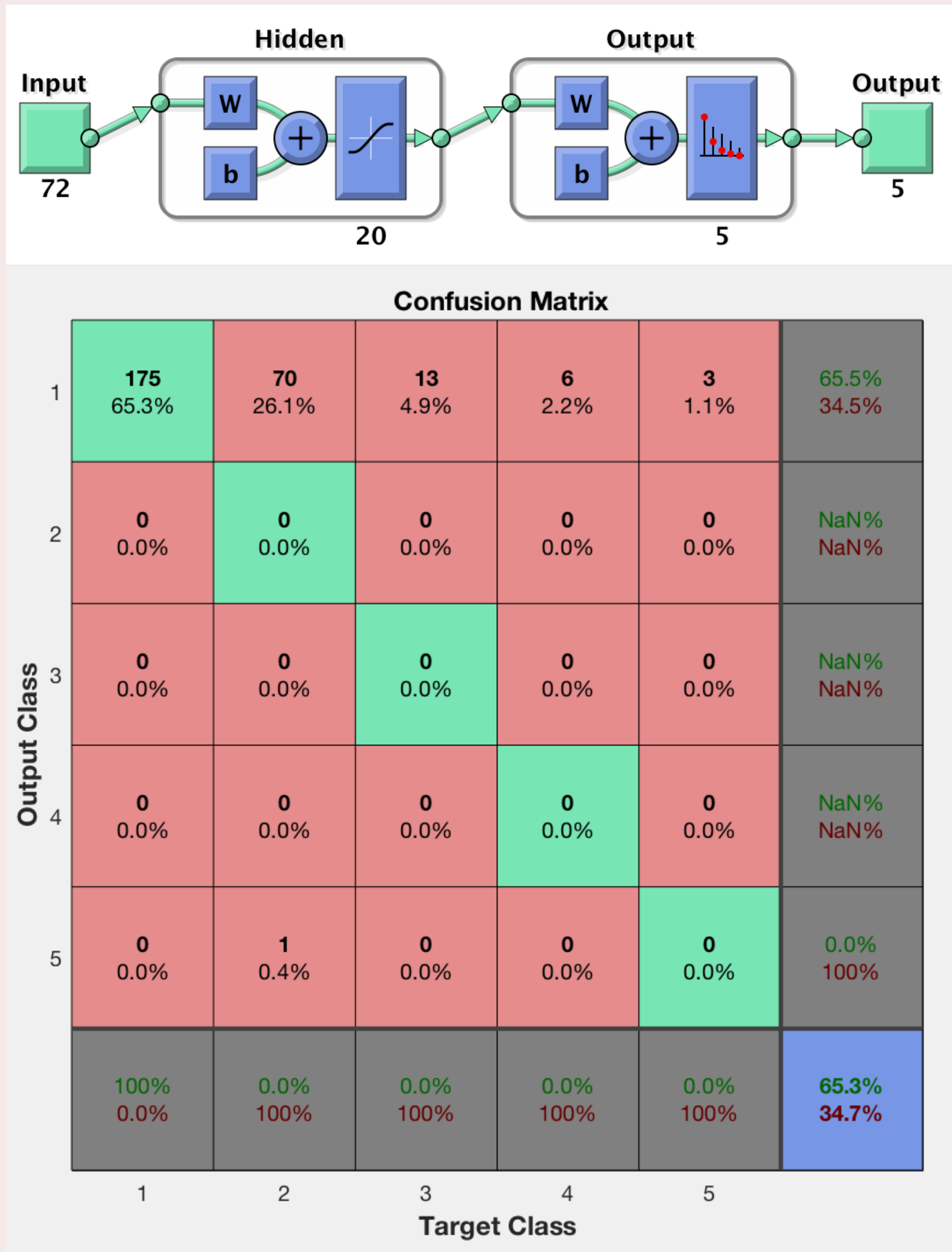
Discriminant Analysis:
One of our most successful approaches was to use Linear Discriminant Analysis. We used a fairly generic approach generalized by the algorithm below:
$f(x|\pi_i) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} exp[\frac{1}{2}(x-\pi_i)\Sigma^1(x-\pi_i)]$

Logistic Regression:
The logistic regression can be generalized with the following algorithm:
$$\pi(X) = \frac{e^{\beta_0+\beta_1 X_1+\ldots+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+\ldots+\beta_p X_p}}$$

SVM:
With SVM we tested both a linear and Gaussian kernel.

Naive Bayes:
Least successfully, we attempted a standard Naive Bayes algorithm.

Basic Neural Network Implementation:
All results are from an average of 20 runs of the neural network.



Classification of Burned Area:

Algorithm	Percent Misclassification (LOO)
Discriminant Analysis (Linear)	34.33
Linear classification (Logistic)	34.33
Naive Bayes	37.31
SVM (Linear Kernel)	34.32
SVM (Gaussian Kernel)	35.45
NN (8 features, 10 neuron hidden layer)	34.37

Discussion

- Using a Linear Kernel SVM provided the best classification out of Linear Discriminant Analysis, Logistic Regression, Naive Bayes, and Gaussian Kernel SVM when a leave-one-out training method was used.
- The neural network which provided the best classification was one with 8 features and 10 neurons in the hidden layer.
- Both the implemented neural network and the Linear Kernel SVM performed very similarly. The Linear Kernel SVM was implemented with the 4 dominant features as shown by PCA.
- The averaged runs of our differently parameterized neural networks all provided similar results to each other and the Linear Kernel SVM, which could mean that the neural network learns which features are most useful in our problem.
- The relatively small amount of data points may have limited training and importance of certain features within the dataset.

Conclusion and Future Work

- Regularization did not really make any significant improvement.
- Using variously structured neural nets all produced similar results.
- We would like to do a little more research to find a few more features that could be relevant to our classification problem.
- Given more time or resources, we really want to be able to get a larger dataset to work with. Currently, most of data we have is for fairly small fires. Finding a large amount of data for this project has been challenging. Ideally we would be able to find another dataset that had similar features to our own.

Acknowledgement

We would like to gratefully acknowledge the CS229 teaching staff for providing all the help, insight, and advice in making this project a great learning experience. The suggestions helped us determine a better approach to meet the goals of this project.

References

[1] Cortez, P., Morais, A., A Data Mining Approach to Predict Forest Fires using Meteorological Data