Accessible Drought Prediction

Dhamma Kimpara, Rick Nueve (Deep Learning class)

Problem Space

Task:

Predict future drought levels for 1 week to 6 weeks ahead.

Use widely available soil and meteorological data

Our Contribution:

Implementation and utilization of an **Ordinal Classification Estimator [2]**

- Uses simple models for complex predictions

Category	Description
D0	Abnormally Dry
D1	Moderate Drought
D2	Severe Drought
D3	Extreme Drought
D4	Exceptional Drought

Data

- 3100 prediction sites (US counties)

Each datapoint:
- 180 days of past meteorological indicators
- Weekly measured drought value (none to D4)
- Soil data for the prediction site

18 meteorological Indicators per day
32 soil features for each prediction site

Raw Data:
Train: 19 million rows
Validation: 2.2 million rows
Test: 2.2 million rows

After processing:
180 past days: 100k rows by 3810 features
14 past days: 1.4 million rows by 324 features

Example features of Raw data:

fips	date	PRECTOT	PS	QV2M	score
1001	2019-01-01	2.25	100.51	9.69	0.0

Approach

Problem:
- k ordered classes, WLOG call them 1, … , k

The classifier **(ORD)**:
Turn the problem into k-1 different binary classification problems. For each sample x, y train a model, for each class i, that answers:

$$\text{is } \; y > i?$$

for i = 1, … , k-1. Then the predicted class probabilities for the original problem are:

$$\Pr[\, y = 1 \,] = 1 - \Pr[\, y > 1 \,],$$
$$\Pr[\, y = 2 \,] = \Pr[\, y > 1 \,] - \Pr[\, y > 2 \,],$$
$$\vdots$$
$$\Pr[\, y = j \,] = \Pr[\, y > j\text{-}1 \,] - \Pr[\, y > j \,],$$
$$\vdots$$
$$\Pr[\, y = k \,] = \Pr[\, y > k\text{-}1 \,].$$

- Can use any binary classifier that can output probability estimates
- Can use a different classifier for each ordinal value

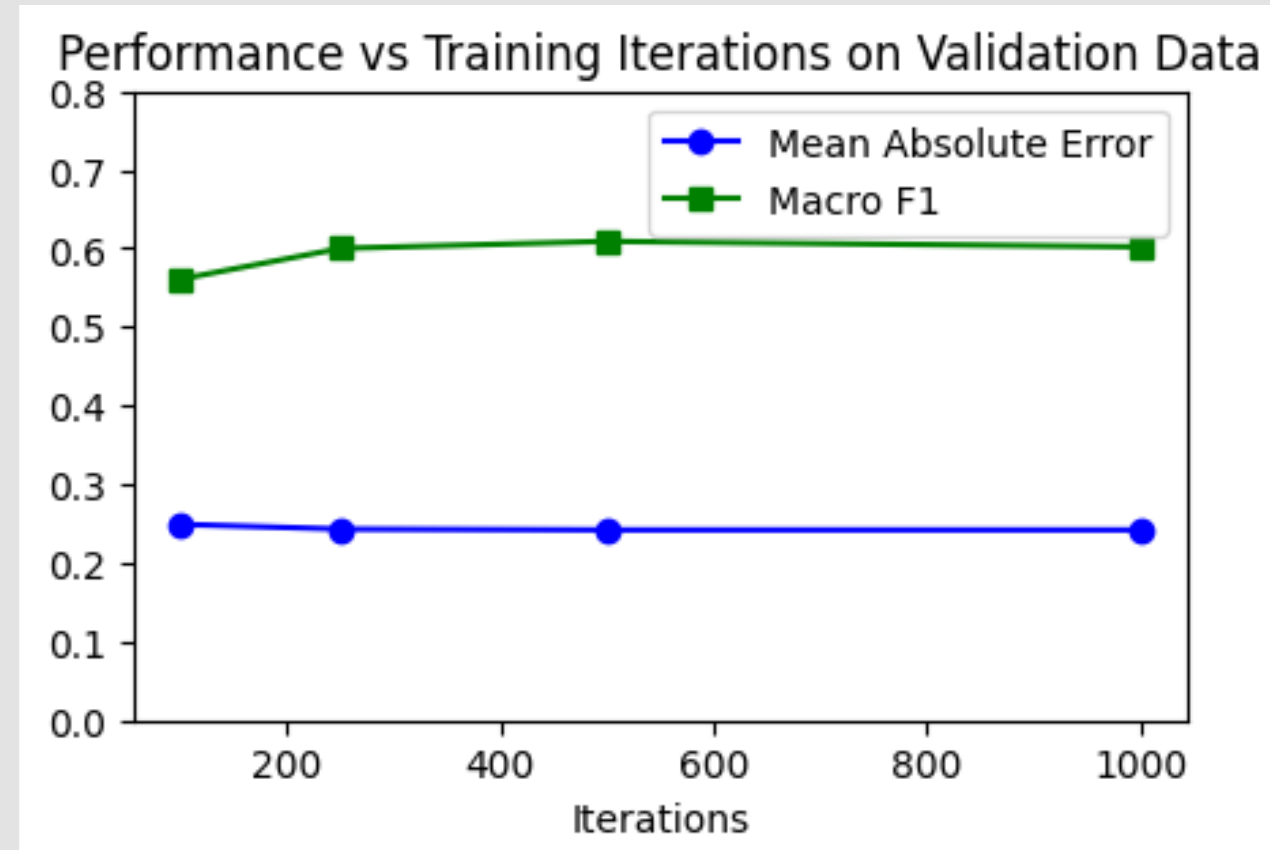
Our base classifier: Logistic Regression

Attempted Approaches
- Multi-layer Perceptron (MLP)
- Boosted Trees

* These approaches did not perform as well due to our computational limits

Results

Model	Macro F1	MAE
LSTM [3]	0.64	0.28
Ridge Regression [1]	0.58	0.26
Ordinal w/ Logistic Regression	**0.58**	**0.20**



Convergence Issues

None of the Logistic classifiers ever converged (sci-kit learn: Convergence Warning)

None of the remedies worked (Changing the solver, increasing training iterations, feature engineering

Discussion

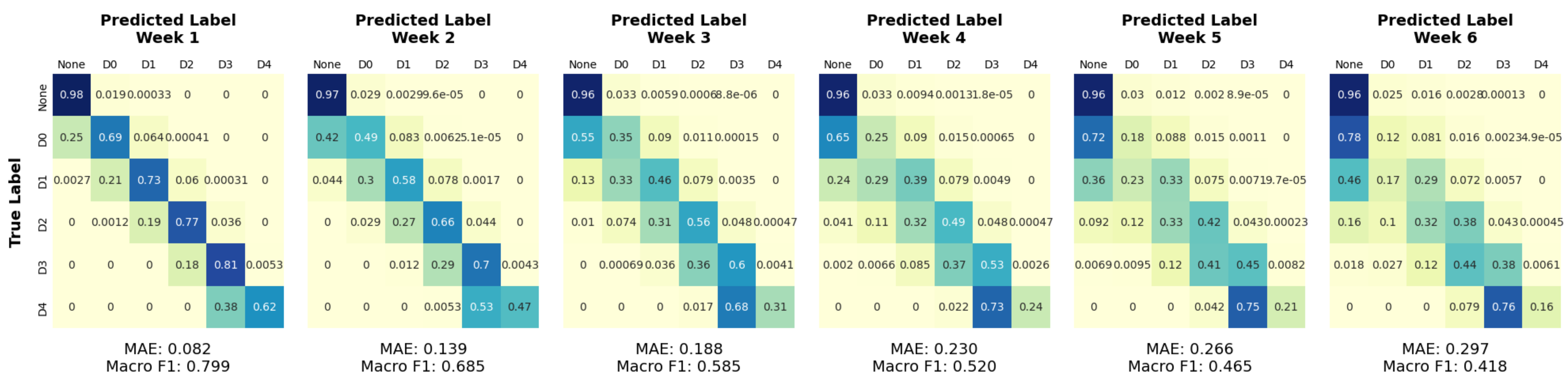
Results
- Non-convergence -> logistic regression model does not have enough modeling power to fit the data
- The further out the prediction, the less accurate

Impact
- We showed that a model made of simple models can have similar performance Deep Learning (DL) models
- The Ordinal estimator can also be used with DL models.

Extensions
- With more computational power, train more sophisticated models (MLP, Gaussian Process, etc.) with more data
- Take advantage of the time series nature of the data. e.g. use a RNN model
- Use auxiliary data, such as weather forecasts, to enhance predictions

References

[1] epistoteles (2021). Predicting Drought. https://github.com/Epistoteles/predicting-drought
[2] Frank, E., & Hall, M. (2001). A simple approach to ordinal classification. In *European conference on machine learning* (pp. 145-156). Springer, Berlin, Heidelberg.
[3] Minixhofer, Christoph (2021). Predict Droughts using Weather & Soil Data. From https://www.kaggle.com/datasets/cdminix/us-drought-meteorological-data



avg. MAE: 0.200
avg. Macro F1: 0.579