Wildfire Burn-Area Prediction

**Abstract**

- **Context:** Wildfires destroy billions of dollars worth of infrastructure. Difficult to predict
- **Purpose:** Design model to predict the size of a wildfire from basic starting conditions.

**Data**

- 1.88 million historic US fires (Kaggle)
  - Mostly tiny fires. No weather
- 600 fires from park in N. Portugal (UCI)
  - Weather features, but Geo-specific, few samples in dataset, more balanced

**Model/Results:**

- Predict final area burned with neural network, SVM, and others
- Best prediction accuracy from SVM: Mean-Absolute-Percent-Error = 78%

**Baseline**

- Linear regression

**Other Models**

- SVM
- K Nearest Neighbors
- Decision Tree
- Gaussian Process regressor

**Neural Network**

- Adam optimization, relu activation, 10 epochs, batch 64

**Models**

**Baseline**

- Normalized Prediction Plots

**Other Models**

- Ensembling
  - Stacked regressors
  - Random Forest
  - Adaboost

**Neural Network**

- 500 hectares

**Results/Discussion**

- Balancing dataset improves performance
- In general, models perform worse on Kaggle due to data imbalance and fewer highly correlated features
- SVM, neural net, and stacked regressors perform best

- Kaggle has weak correlations of features to fire size

**Future Work**

- Extend datasets to include additional salient features
- Merge additional weather features with Kaggle dataset to tackle high variance issue
- Predict fire size class instead

**Acknowledgements**

We would like to thank our TA, Leo Mehr, for his advice and support throughout the project as well as the teaching staff for helpful comments and ideas.