Feature Engineering for Deep Learning to Forecast Solar Events

Maxine Hartnett

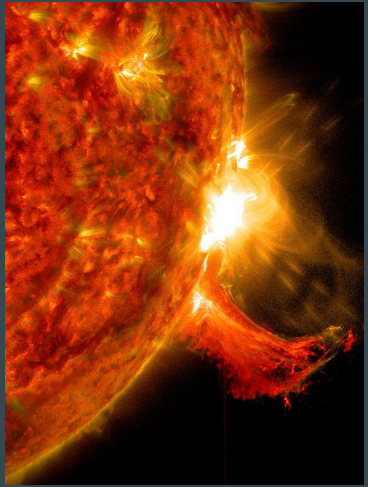
About Me

I'm going into my senior year as a Computer Science major at CU Boulder. I've worked at LASP for 2 years in the Data Systems software engineering group.

Interests include machine learning, sci-fi, and music.

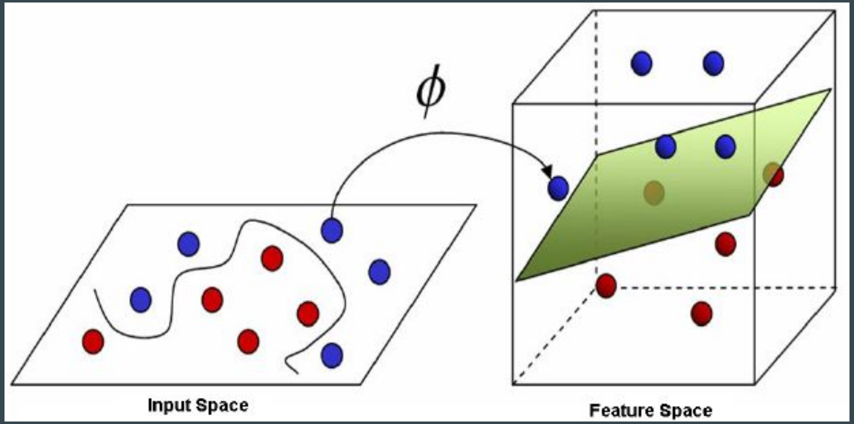
Motivation

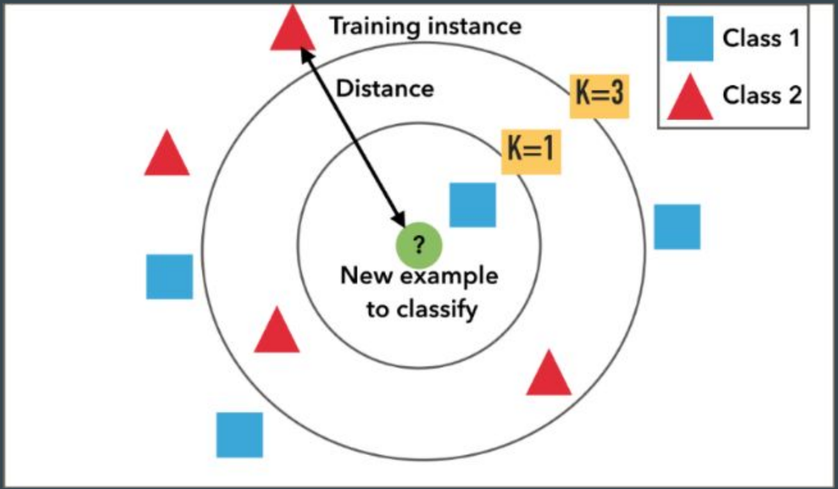
- The goal for this project is to create a system using deep learning to predict the likelihood of a flare happening in the next *n* hours.
- Currently, there are no systems in place for accurately forecasting large solar flares, and most research has been done on simple machine learning algorithms
    - "Simple" machine learning algorithms include SVM, KNN, and shallow neural nets with limited numbers of hidden layers.



Previous Exploration

- Commonly used systems: SVMs, Basic Neural Nets, KNN, or randomized forests
- True Skill Score (TSS) scores generally around the 0.6 or 0.7 range





Goals

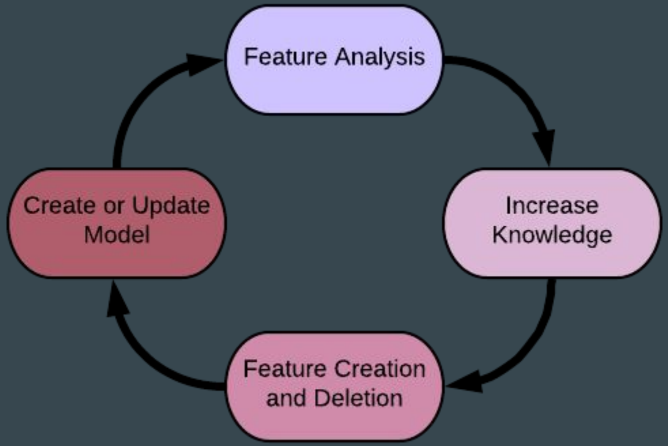
- Create an effective solar flare prediction system using deep learning to explore data in a deeper manner
- Analyze and improve upon the system using feature engineering and advanced machine learning techniques
- Advance scientific knowledge and guide future investigation with feature engineering

Project Methods

- **Multi-layer Perceptron Neural Net:** Deep learning method used to analyze numerical data
- **Convolutional Neural Net:** Deep learning method to analyze images, particularly vector and line-of-sight magnetograms
- **Feature Engineering:** Analyze and improve upon existing features, and generate new ones to capture elements the model might not be able to find and improve the accuracy of the models

What is feature engineering?

- Feature engineering explores existing features in the data, evaluates the usefulness of different features, and creates additional ones.
- Feature engineering is used to elevate a machine learning model using intelligent, human created features rather than raw data.

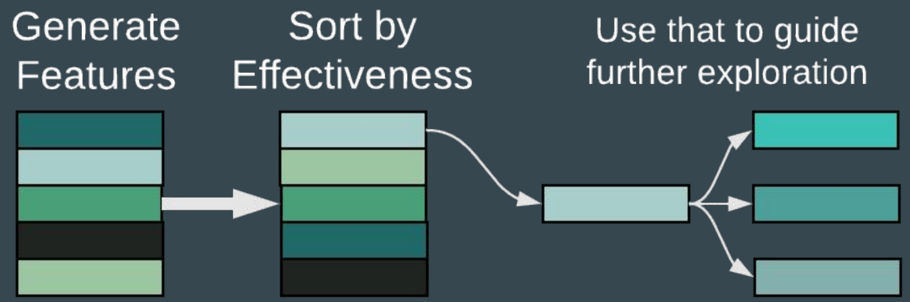


Feature Engineering and Science

- Machine Learning and especially deep learning is a black box - we don't know how a model gets its answers
- By analyzing features and working on creating new ones, we find out what data the model finds important
- Particularly with scientific models, this can guide future study and allow insight

Feature Engineering Goals

- Improve accuracy
- Reduce complexity
- Refine model based on scientific knowledge rather than just machine learning techniques
- Discover areas that could benefit from additional study



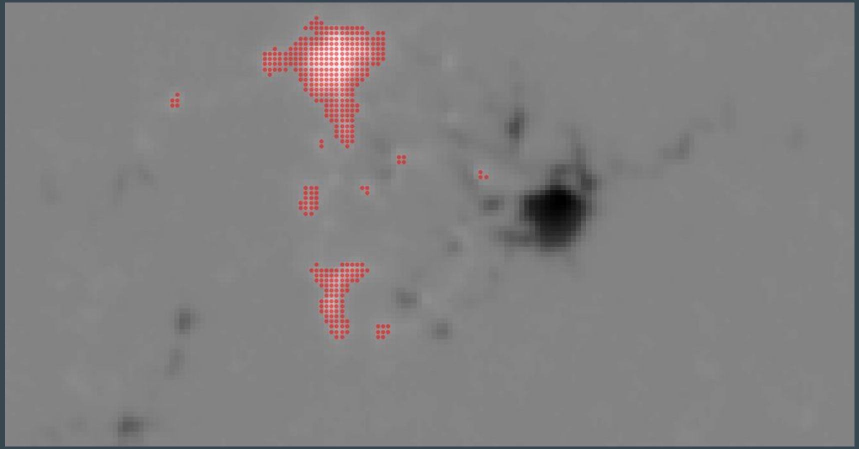
Maxine Hartnett, 2018

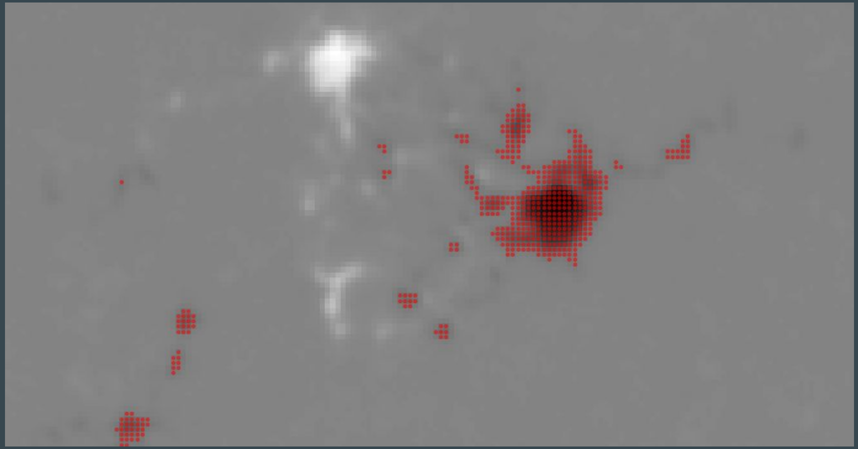
What does this mean for this project?

- Reading through papers to determine features that can be calculated from image data and existing features to improve accuracy
- Evaluating features within SHARPs for usefulness to see where we can eliminate complexity
- Creating features to improve the model based on existing knowledge
- Simplifying and refining complex data like images into clear, numerical features

Creation of Additional Features

- Strong field polarity lines
  - Found by many papers to be effective in machine learning models - extracted from magnetograms
  - Also referred to as Magnetic Neutral lines
- Flare history
- Shape, size, and area



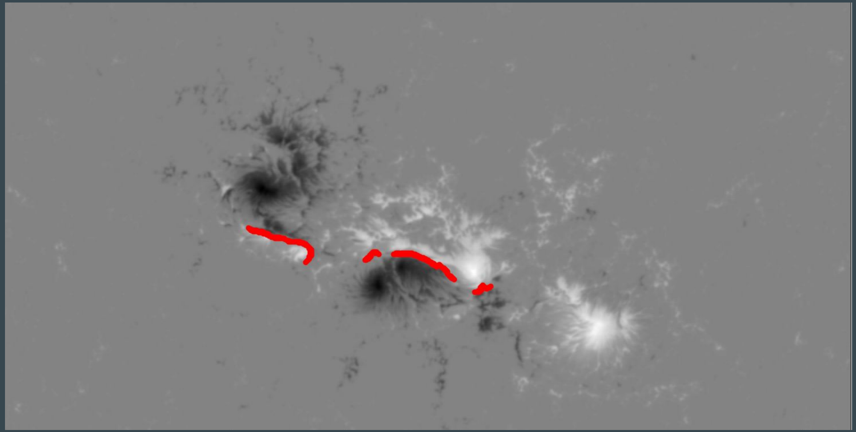


Polarity Inversion Lines (PIL)

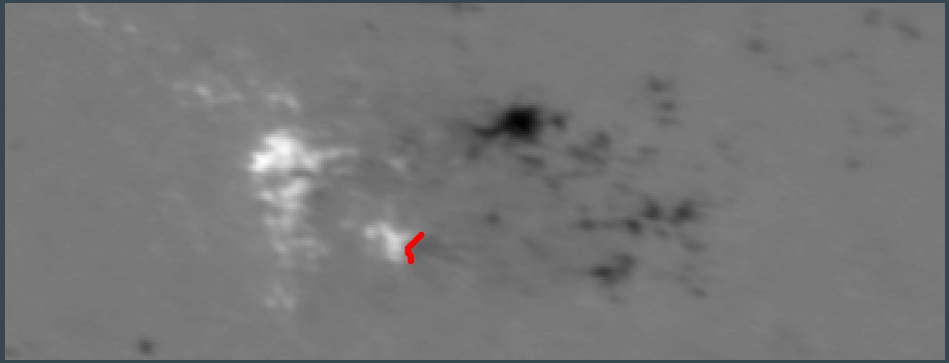
- Polarity inversion lines separate areas of opposite polarity on the sun
- These are often associated with *filaments*
  - A long ‘tongue’ of relatively cool material (10 000 K) suspended in the much hotter solar corona (2 million K). *(Ridpath, A Dictionary of Astronomy)*
- Filament length has also been found to be an indicator of solar eruption (Aggarwal et al, 2018)
- A strong PIL has several useful features to extract from the magnetogram images

Calculating Polarity Inversion Lines

- Code created by Sadykov and adapted for use by me
- Features that can be extracted include: Length, area, flux, and magnetogram gradient
- Only works on line-of-sight magnetograms, not vector magnetograms



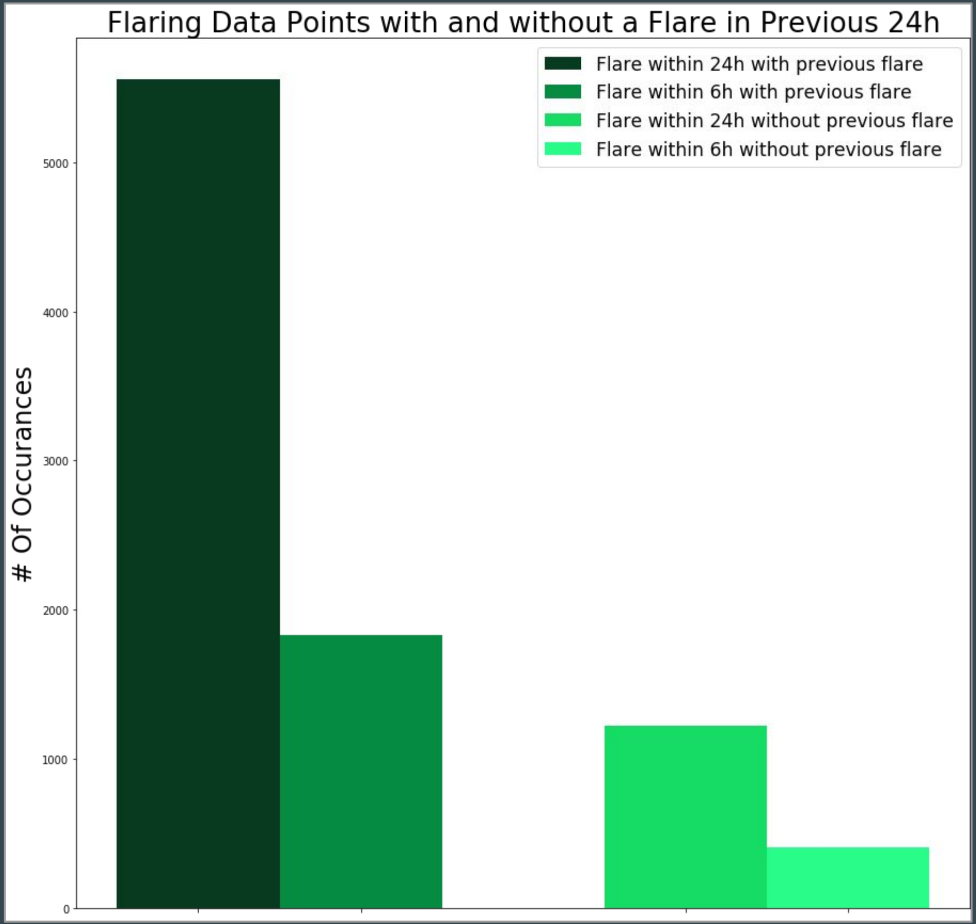
PIL of an AR with an X-class flare within 6h



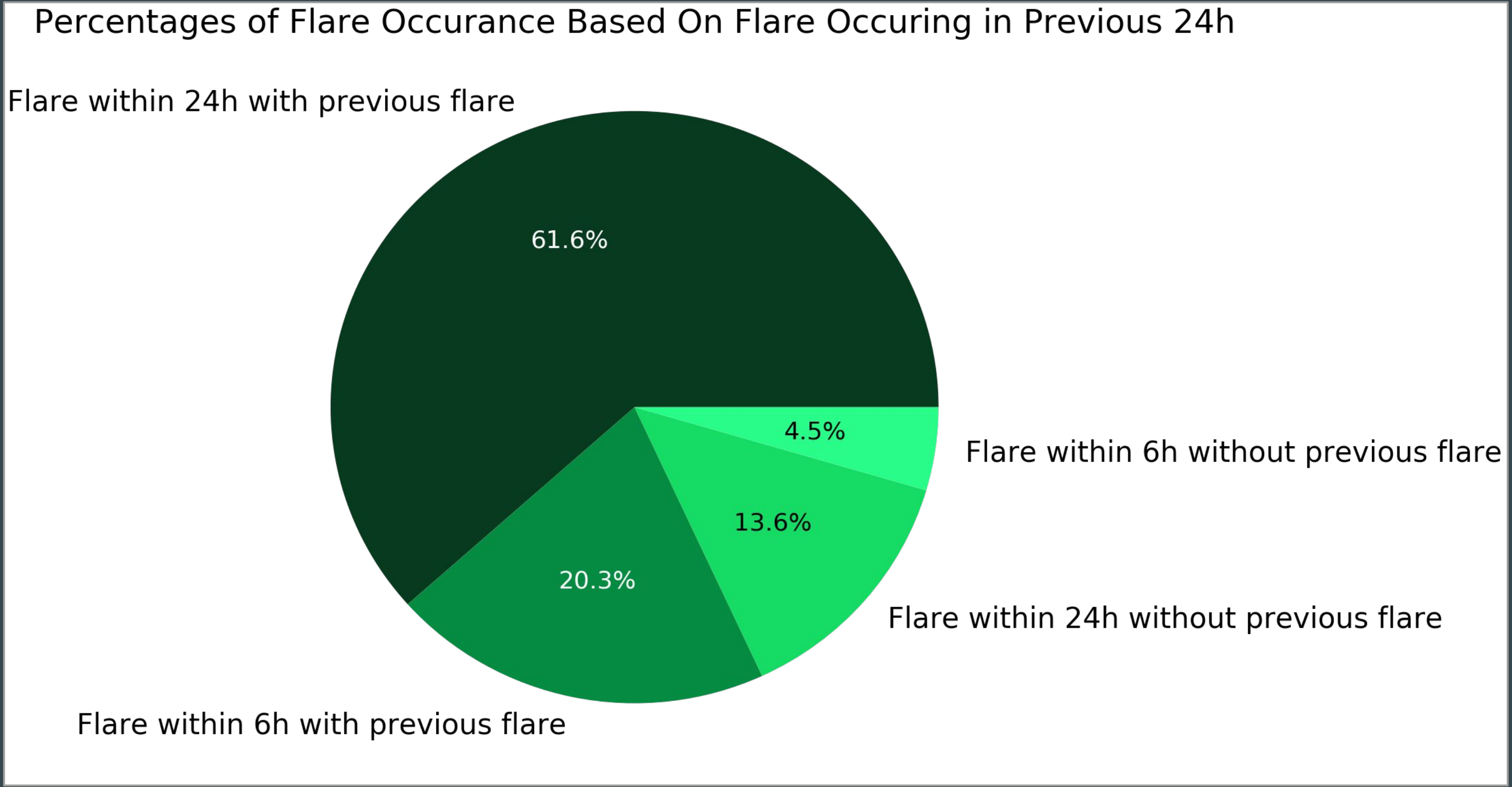
PIL of an AR that didn't flare within 48h

Previous Flare History

- Again, found by many researchers to be highly effective predictor for future flares
- Computed: # previous flares by class, average length of previous flares by class, and the time since the most recent flare by class

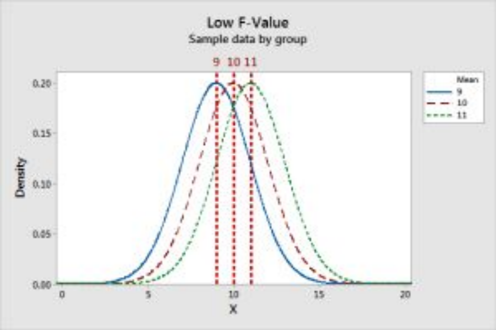


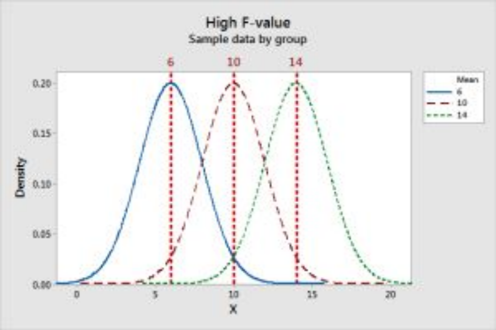
Previous Flare History By Percentage



Evaluating Feature Effectiveness

- ANOVA F-test - used to find if the means of two populations are statistically significant
  - Evaluate the flaring vs non-flaring means of two features to see if there's a significant difference that can be exploited in a model
- Correlation analysis - if two features are highly correlated, it might not be necessary to use both
  - If both features describe the same aspect of a sample, then they don't both need to be included in the model





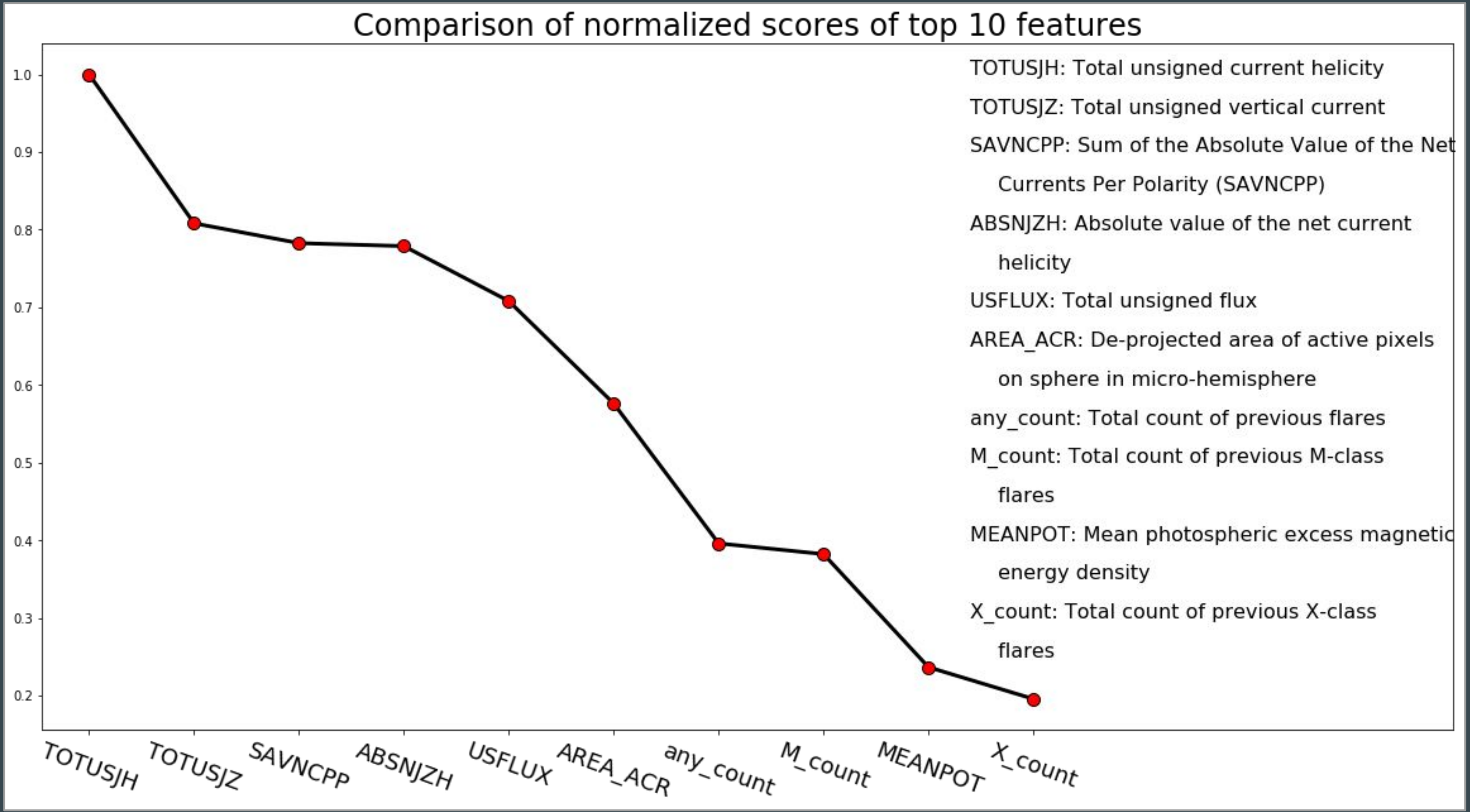
Standardization

- Standardizing the data reduces numerical error - large differences in the scale of the data can result in numbers too small for the computer to store.
- Standardizing methods: Take the log of the features with large values, then do *Z-standardization*
    - Measures the number of standard deviations from the mean a value is - where a value is compared to the population

$$Z_i = \frac{X_i - \bar{x}}{\sigma}$$

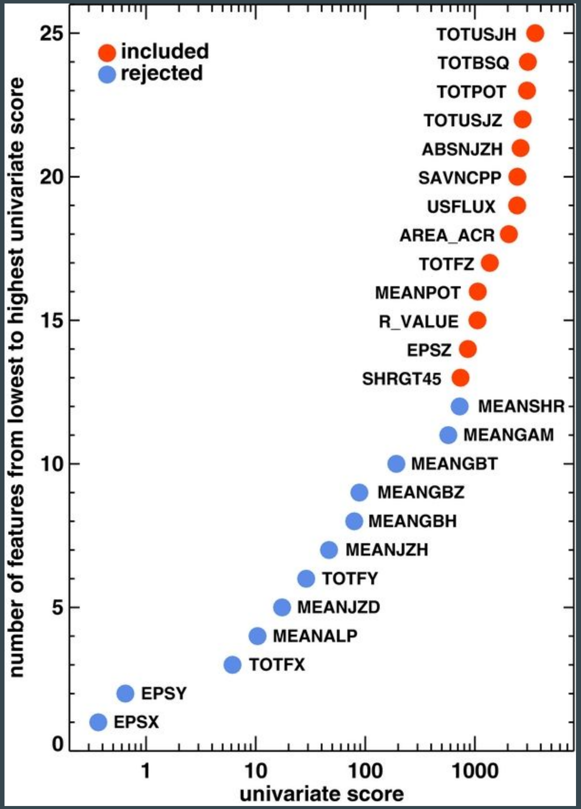
Z_i: The new value
X_i: value to standardize
X bar: Mean of full set
σ: standard deviation of full set

Feature Comparison Example



Comparisons to existing feature comparison results

TOTUSJH	Total unsigned current helicity
TOTUSJZ	Total unsigned vertical current
SAVNCPP	Sum of the Absolute Value of the Net Currents Per Polarity (SAVNCPP)
ABSNJZH	Absolute value of the net current helicity
USFLUX	Total unsigned flux
AREA_ACR	De-projected area of active pixels on sphere in micro-hemisphere
any_count	Total count of previous flares
M_count	Total count of previous M-class flares
MEANPOT	Mean photospheric excess magnetic energy density
X_count	Total count of previous X-class flares
X_average_time	Average duration of all X-class flares
MEANGAM	Mean inclination angle, gamma
M_most_recent	Time since most recent M-class flare
MEANGBT	Mean value of the total field gradient
any_most_recent	Most recent M or X-class flare
X_most_recent	Time since most recent X-class flare
MEANGBZ	Mean value of the vertical field gradient
MEANGBH	Mean value of the horizontal field gradient
LON_FWT	Stonyhurst longitude of flux-weighted center of active pixels
MEANJZD	Mean vertical current density
HARPNUM	HARP ID
LAT_FWT	Stonyhurst latitude of flux-weighted center of active pixels
MEANALP	Mean twist parameter, alpha
MEANJZH	Mean current helicity
M_average_time	Average duration of previous M-class flares
NOAA_AR	NOAA AR number that best matches this HARP
any_average_time	Average duration of previous flares



Bobra and Couvidat, 2015

Future Goals

- Apply magnetogram image analysis to all the magnetogram data
- Explore additional features to extract from magnetograms
  - E.g. symmetry, topology analysis, UV brightening...
- Explore the use of line-of-sight vs vector magnetograms
- Complete correlation analysis and eliminate unneeded features
- Analyze features using model, rather than more theoretical measures

Acknowledgments

I would like to thank Wendy Carande, Laura Sandoval, Tracey Morland, Kim Kokkonen, Tom Berger, Jim Craft and Andrew Jones for supporting this project and helping us improve.

Questions?

Email: maxine.hartnett@lasp.colorado.edu

Image Sources

https://www.space.com/27344-incredible-solar-flare-nasa-video.html

https://www.researchgate.net/figure/Figure-A15-The-non-linear-SVM-classifier-with-the-kernel-trick_fig13_260283043

https://gong.nso.edu/data/magmap/ifmodel.html

https://phys.org/news/2012-08-huge-solar-filament-sun.html

https://mapr.com/blog/real-time-credit-card-fraud-detection-apache-spark-and-event-streaming/

Citations

- N. Nishizuka et al. “Solar Flare Prediction Model with Three Machine-Learning Algorithms Using Ultraviolet Brightening and Vector Magnetogram.” *The Astrophysical Journal,* Nov 6 2016.
- M. G. Bobra and S. Couvidat. “Solar Flare Prediction Model with Three Machine-Learning Algorithms Using Ultraviolet Brightening and Vector Magnetogram.” *The Astrophysical Journal,* January 2015.
- B. T. Welsch and Y. Li. “On the Origin of Strong-Field Polarity Inversion Lines.” Oct 2007.