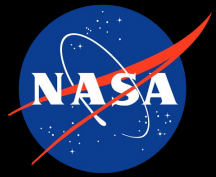
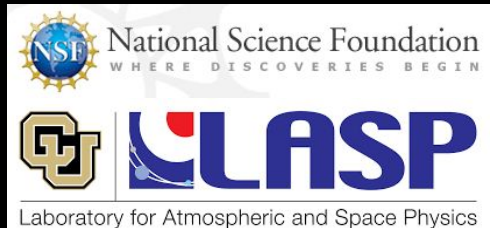
Classifying Active Regions Using Machine Learning

Megan Smith(Laboratory for Atmospheric and Space Physics)
Laura Sandoval(Laboratory for Atmospheric and Space Physics)
Stephane Beland(Laboratory for Atmospheric and Space Physics)
Andrew Jones(Laboratory for Atmospheric and Space Physics

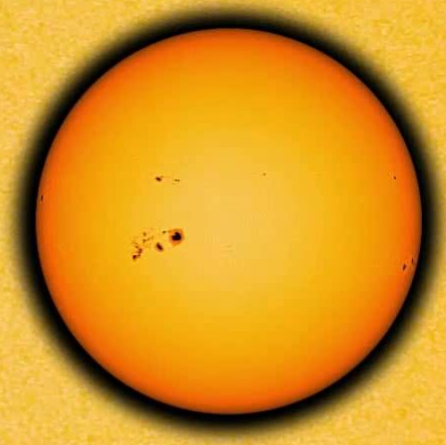


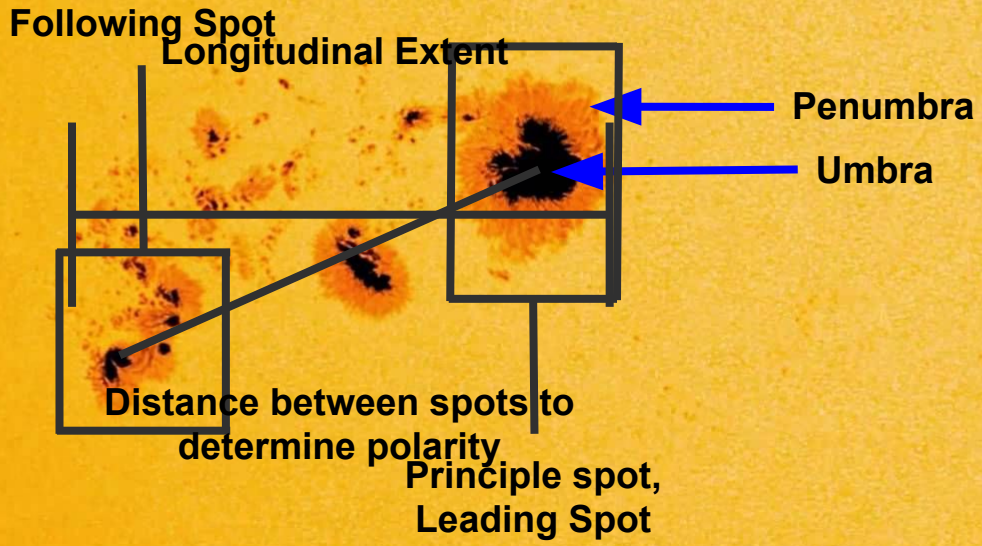


Motivation

- The McIntosh Classification system was created to use sunspots to predict solar flares
- Classifications are only published once per day
- Classifications are not precise
    - The definitions of classes are not all objective, some use words such as ‘large’, ‘small’, and ‘few’, which leaves room for disagreement on the classification of some spots

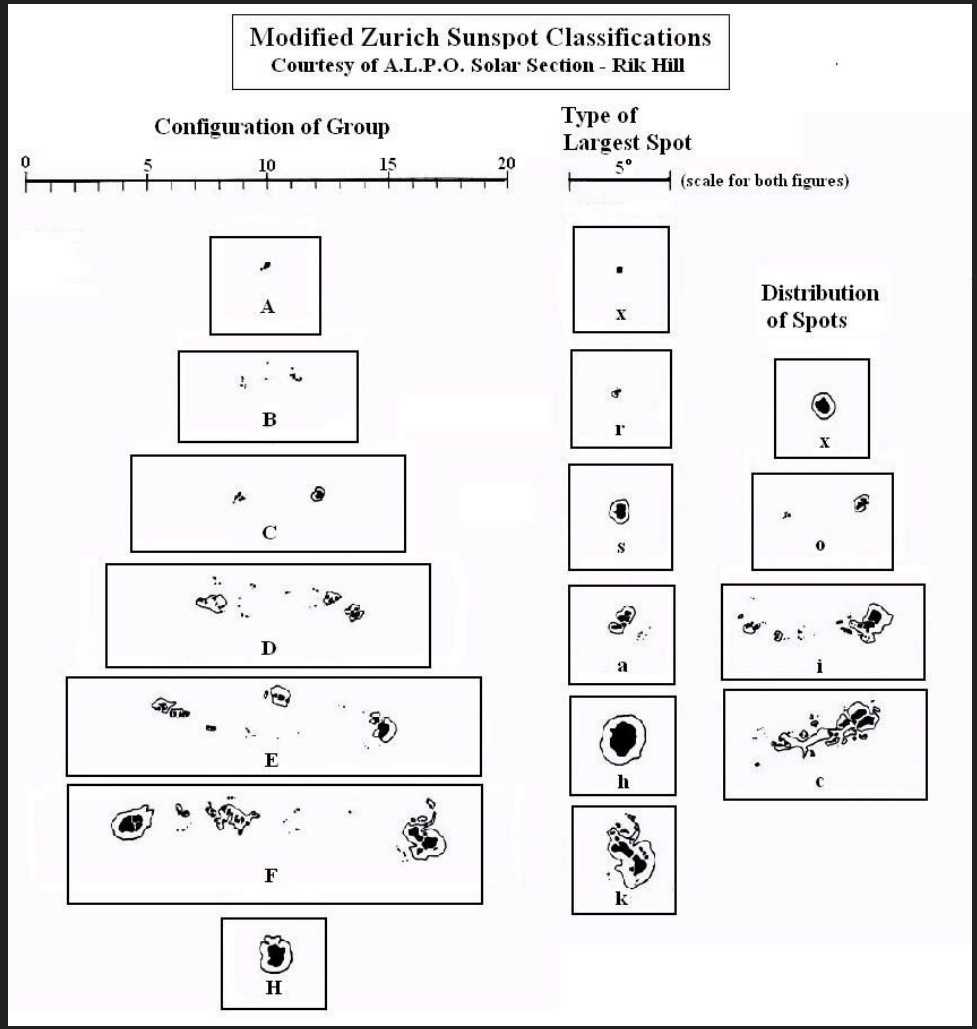
SDO/HMI Quick-Look Continuum: 20140106_073600





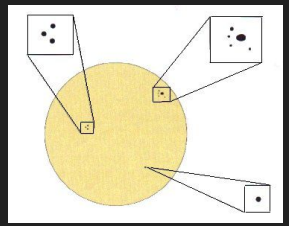
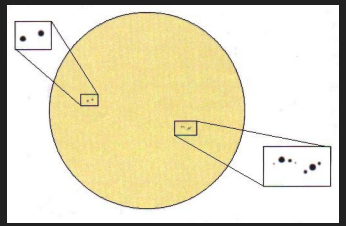
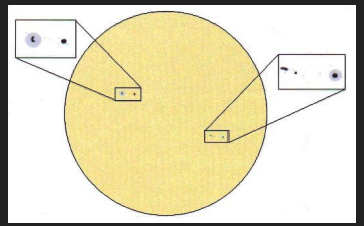
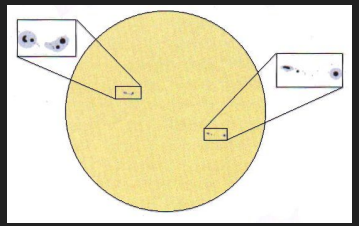
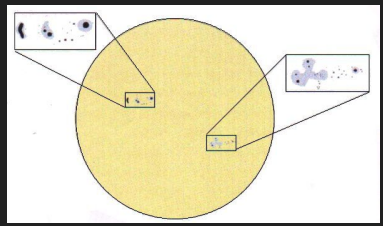
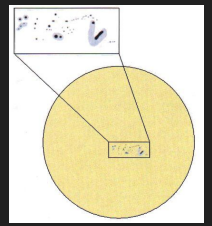
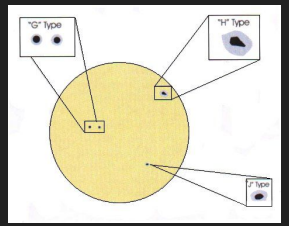
Classification

- Z value describes the group as a whole
  - Polarity
  - Longitudinal extent
  - Presence and distribution of penumbra
- P value describes principle spot
  - Size
  - Symmetry
  - Presence and maturity of penumbra
- C value describes compactness
  - Spots in between leading and following spot
  - Penumbras between leading and following spot



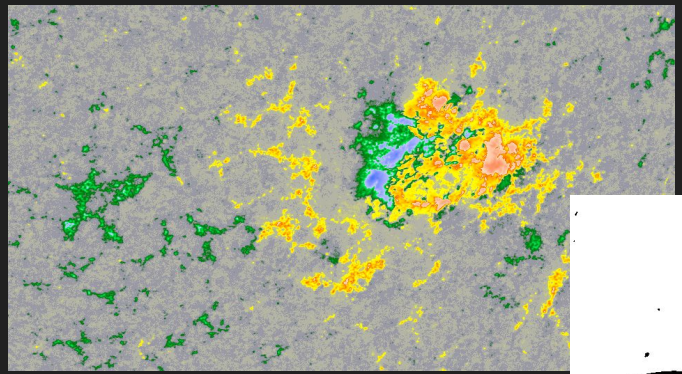
Z value

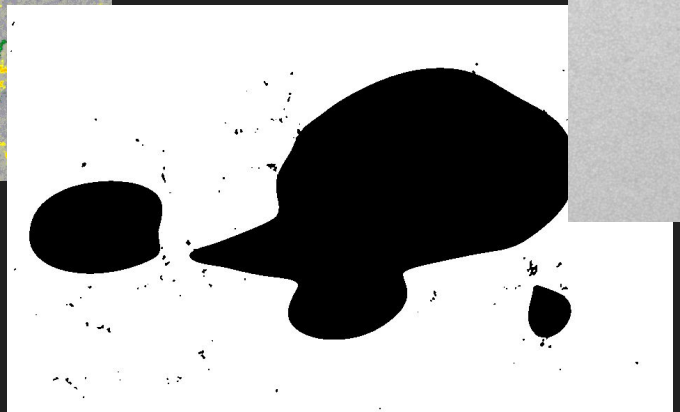
- The Z value is also called the Modified Zurich value
- The Zurich classification served a similar purpose to the McIntosh classification, but also described the history of an active region

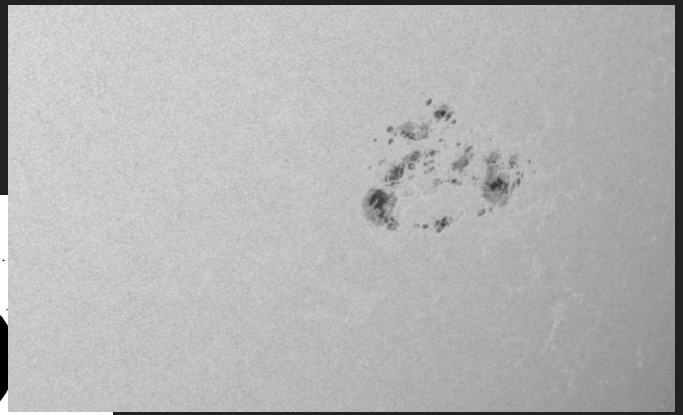








SHARP dataset

- Space Weather Hmi Active Region Patch
- Constructed using an automated process from full disk images (SDO)
- We used magnetograms, bitmaps, and intensitygrams







Parameters needed to determine Z value

- Polarity
  - Distance between the umbras of spots
- Longitudinal Extent
- Presence of penumbra
  - Which sides of region have penumbra for bipolar groups

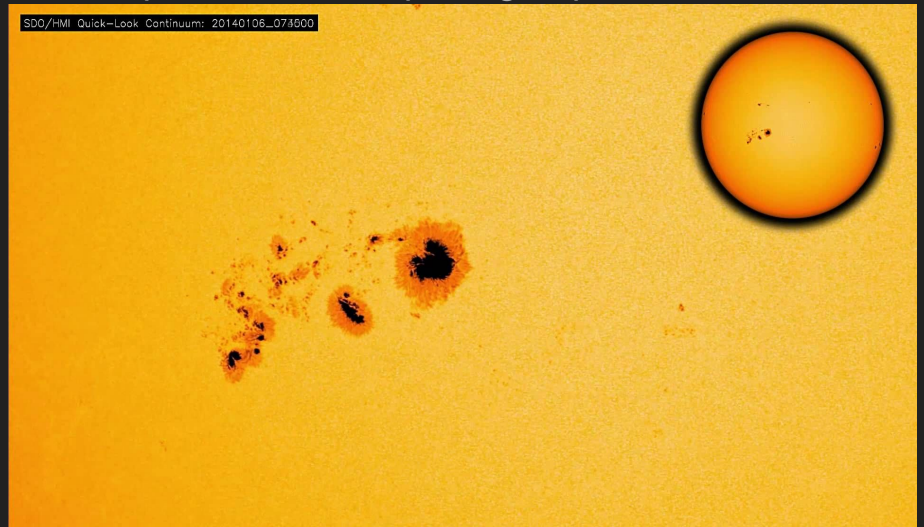
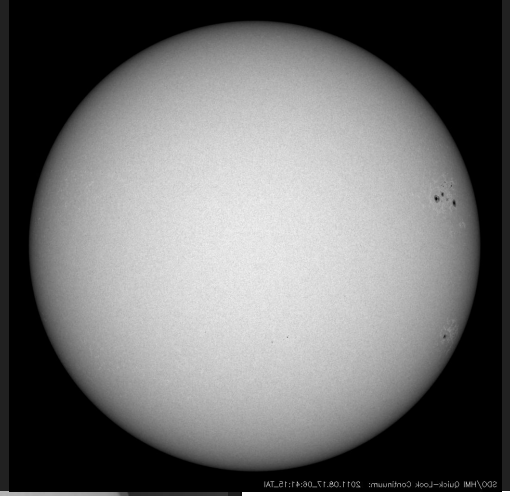
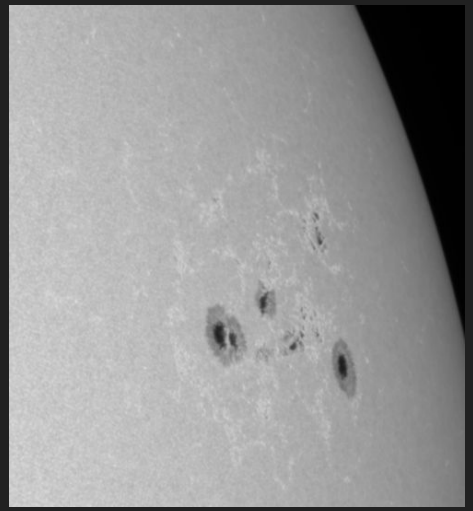
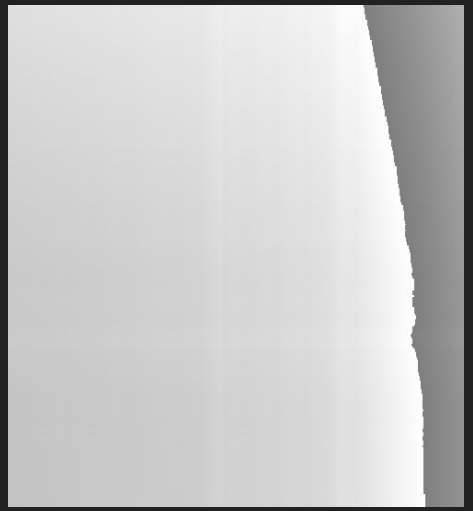


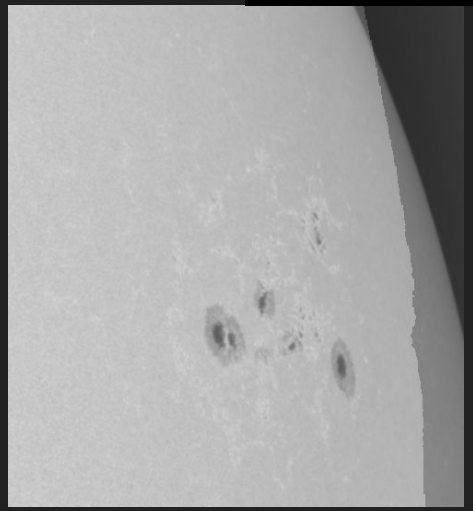
Image processing

- Correct Limb effect









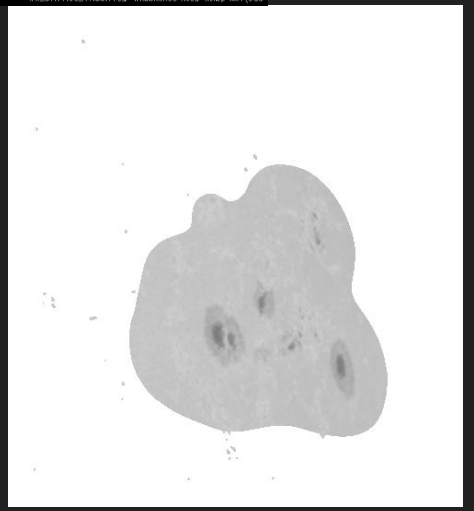


Image processing

- Find spots
  - Threshold image
  - Find contours of spots

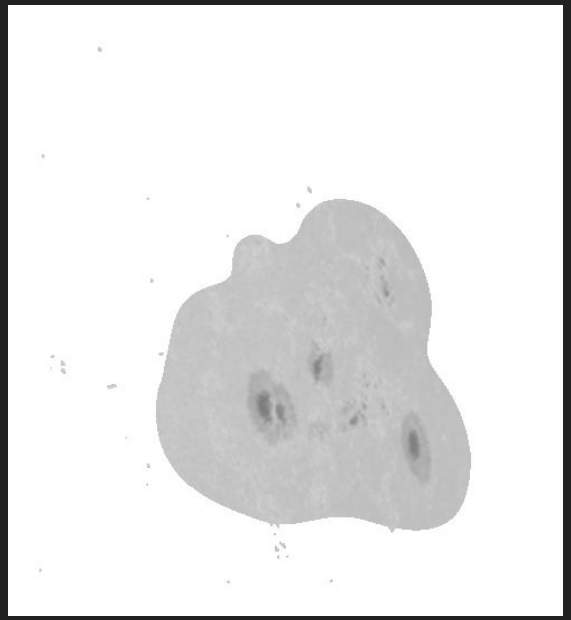


Image processing

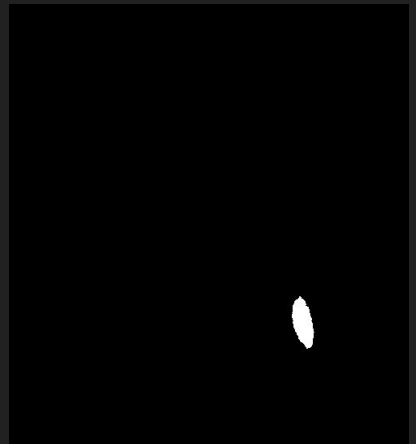
- Longitudinal extent
  - Find furthest left and right spots
  - Find edges of extreme spots
  - Find distance between furthest edges of extreme spots
  - Multiply by scale factor

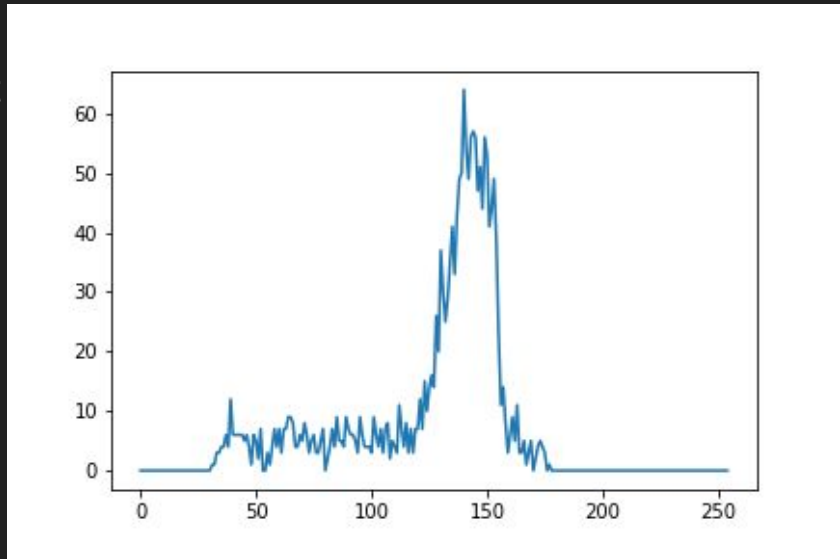
Image processing

- Polarity
  - Defined as bipolar if the greatest separation between spots is greater than 3 degrees
- Method:
  - Find furthest left and right spots
  - Find centers of extreme spots
  - Find distance between centers of extreme spots
  - Multiply by scale factor

Image processing

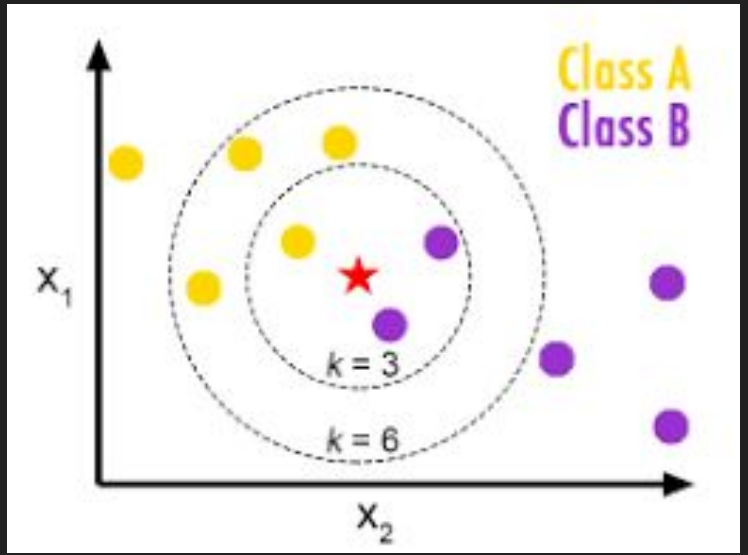
- Presence of Penumbra
  - Isolate one contour of a spot
  - Create a black image and draw contour in white (create mask)
  - Make a histogram of pixel value for spot
  - Analyze peaks of histogram
    - 2 peaks means penumbra is present
    - 1 peak means that no penumbra is present





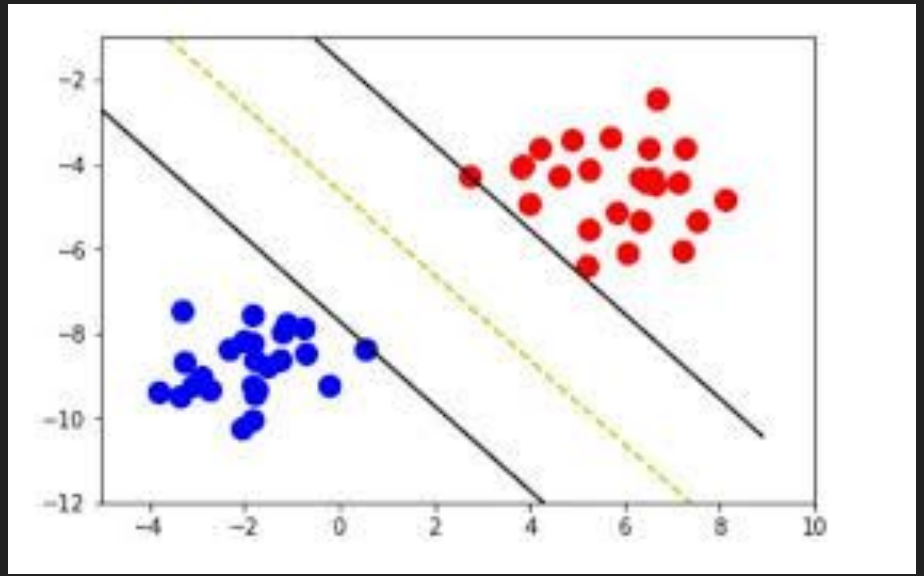
Machine learning techniques

- K Nearest Neighbor
  - Training
    - Graph all training points and their labels
  - Classifying
    - Graph point to be classified
    - Find k closest points
    - Between closest points, which label occurs the most



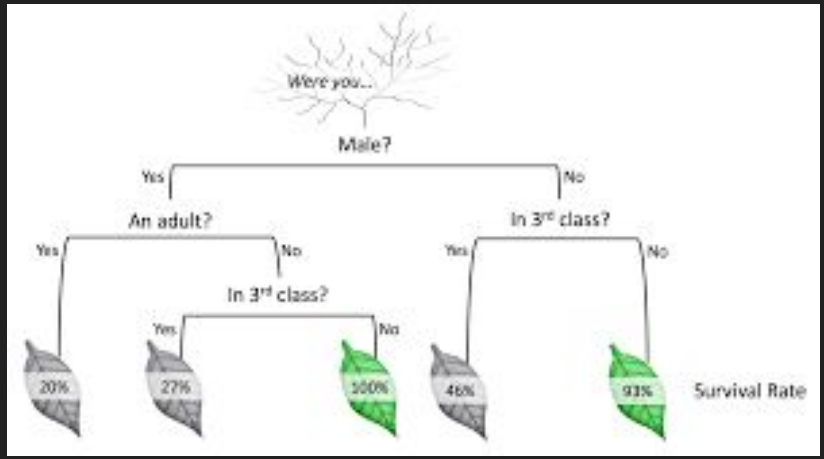
Machine learning techniques

- Support Vector Machine
  - Training
    - Graph all training points
    - Find line in between labels that is the farthest away from any point
  - Classifying
    - Graph point to be classified
    - Determine which side of the line it is on



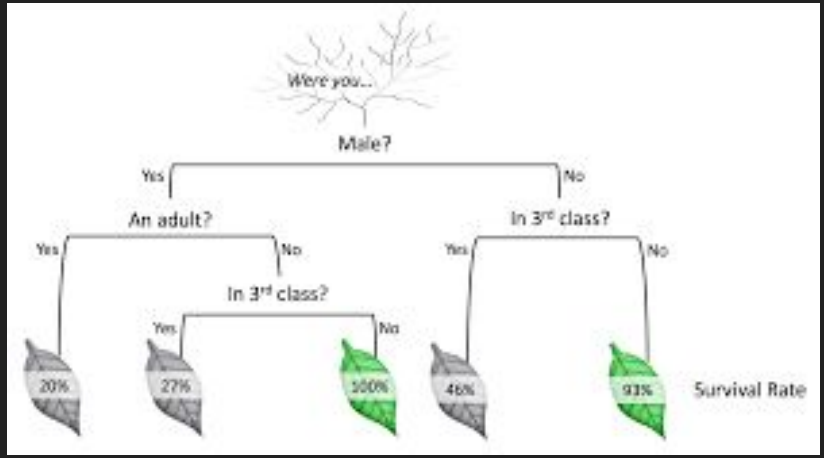
Machine learning techniques

- Random Forest
  - Training
    - Creates a random decision tree with a random subset of training points
    - Continues to make trees until the specified number is reached
  - Testing
    - Puts point to be classified through each decision tree
    - Averages the results of the trees to yield a classification



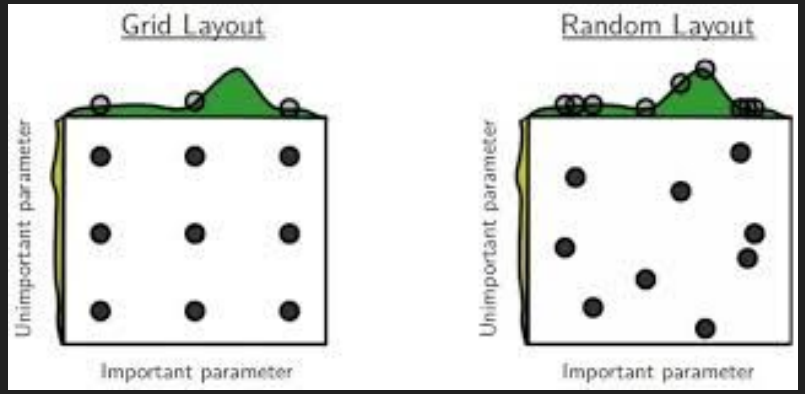
Machine learning techniques

- AdaBoost classifier
  - Training
    - Makes a random decision tree with random subset of data
    - Finds which training points the tree misclassified
    - Trains new decision tree with more weight on incorrectly classified points
    - Repeats until it has created the specified number of trees
  - Classifying
    - Finds result of point to be classified from each decision tree
    - Weighs the probabilities returned by each tree to yield a classification



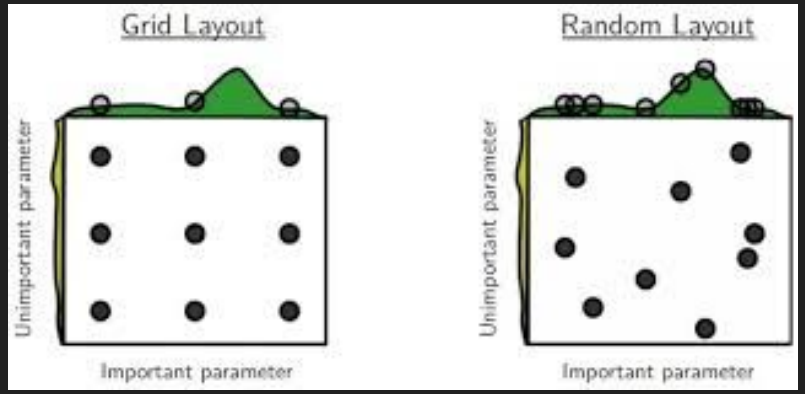
Randomized hyperparameter search

- Runs a given estimator (machine learning method) on given data a specified amount of times
  - examples
- Each time it randomizes the parameters
- Returns the parameters that yielded the best results.



Hyperparameter Grid Search

- Runs a given estimator with each possible combination of parameters that it is given
- Returns parameters that yielded the best results
- Slower than randomized parameter search



Performance of Classification using SVM

Overall score:
51.2% accurate

	A: Unipolar, no penumbra	B: Bipolar, no penumbra	C: Bipolar, penumbra on one side	D: Bipolar, penumbra on both sides, <10°	E: Bipolar, penumbra on both sides, >10°, <15°	F: Bipolar, penumbra on both sides, >15°	H: Unipolar, possesses penumbra
# of Samples	157	193	423	443	192	44	341
# Matched	58	56	225	285	85	5	204
Accuracy	36.9%	29.0%	53.1%	64.3%	44.3%	11.8%	59.8%

Performance using SVM, oversampling

	A: Unipolar, no penumbra	B: Bipolar, no penumbra	C: Bipolar, penumbra on one side	D: Bipolar, penumbra on both sides, <10°	E: Bipolar, penumbra on both sides, >10°, <15°	F: Bipolar, penumbra on both sides, >15°	H: Unipolar, possesses penumbra	Total:
# of Samples	157	193	423	443	192	44	341	
# Matched	86	75	183	259	102	10	206	
Accuracy	59.8%	38.9%	43.1%	61.2%	53.1%	20.4%	60.4%	51.4%

Comparison with literature

	A: Unipolar, no penumbra	B: Bipolar, no penumbra	C: Bipolar, penumbra on one side	D: Bipolar, penumbra on both sides, <10°	E: Bipolar, penumbra on both sides, >10°, <15°	F: Bipolar, penumbra on both sides, >15°	H: Unipolar, possesses penumbra	Total
# of Samples	157	193	423	443	192	44	341	
# Matched	86	75	183	259	102	10	206	
Accuracy	59.8%	38.9%	43.1%	61.2%	53.1%	20.4%	60.4%	51.4%
# of Samples	19	24	38	51	36	17	76	
# Matched	14	10	21	25	18	13	63	
Accuracy	73.7%	41.7%	55.3%	49.0%	50.0%	76.5%	82.9%	63%

Moving Forward

- Improve Z value
  - Find polarity using magnetogram
    - Current metric for polarity is 78% accurate
  - Improve contouring for penumbra
- Find c value
  - Number of spots between leader and follower
  - Presence of penumbra on spots between leader and follower
  - Maturity of penumbra on spots between leader and follower

Moving Forward

- Eliminate error in dataset
  - Eliminate pictures with too low of a range of pixel values
  - Limb effects
    - Write code to transform angled data to flat data
- Eliminate error in classifications
  - Comparison with mount wilson magnetic classification
  - Remove points by classifiers that are less accurate
  - Compare to flaring activity instead of classifications (McIntosh, P., 1990)
- Uses
  - Perform classification on all points in SHARP dataset to be used for research in the future

Acknowledgements

This research was supported by the National Science Foundation REU program, Award #165987 and the Solar Dynamics Observatory

Thanks to my menors Laura Sandoval and Andrew Jones

Thanks to the rest of the flares group: Wendy Carande, Tracy Moreland, Stephane Beland, Kim Kokkonen, Maxine Hartnett, and Justin Cai