

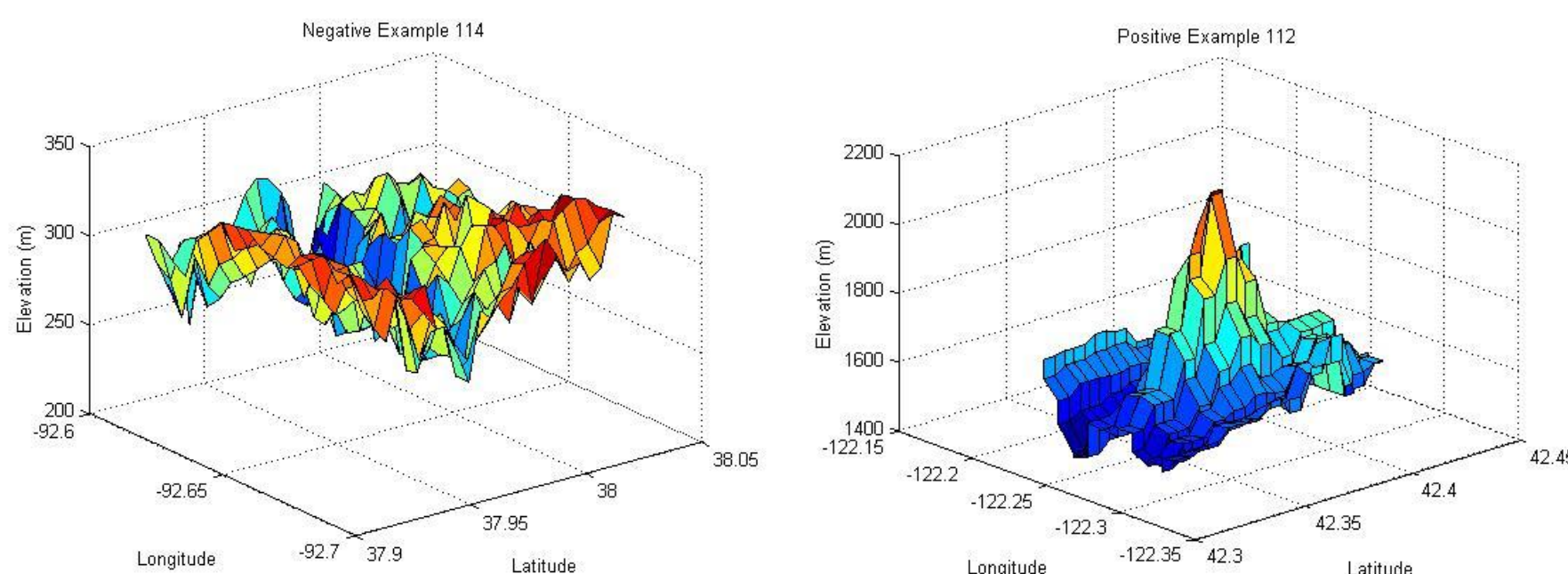


IDENTIFYING VOLCANOES FROM ELEVATION PROFILE

ELIZABETH CULBERTSON

DATA

Elevation data was acquired for 115 volcanoes and 119 non-volcano regions using the USGS Bulk Point Query Service (BPQS). The latitude and longitude of the center of each region were run through a script to generate a 21x21 grid of coordinates centered around the input point and spanning a width of .1°. This grid was then submitted to the BPQS to fill in the elevation at each coordinate. Examples of the positive and negative examples are below.



In general, qualitatively, the positive examples tended to look more sharply peaked and radially symmetric, while negative examples tended to be either relatively flat, or only be sloped along a particular direction. These observations motivated the feature selection.

FEATURES

Two approaches were taken to extract features from the data. One approach took the 441 coordinate-elevation pairs as the features for a training example (denoted "all" in table). The other, in an attempt to capture key characteristics of a volcano, extracted two "scores" from the data to use as features (denoted "scores" in table):

$$\text{elevation score} = \max \text{elevation} - \min \text{elevation}$$

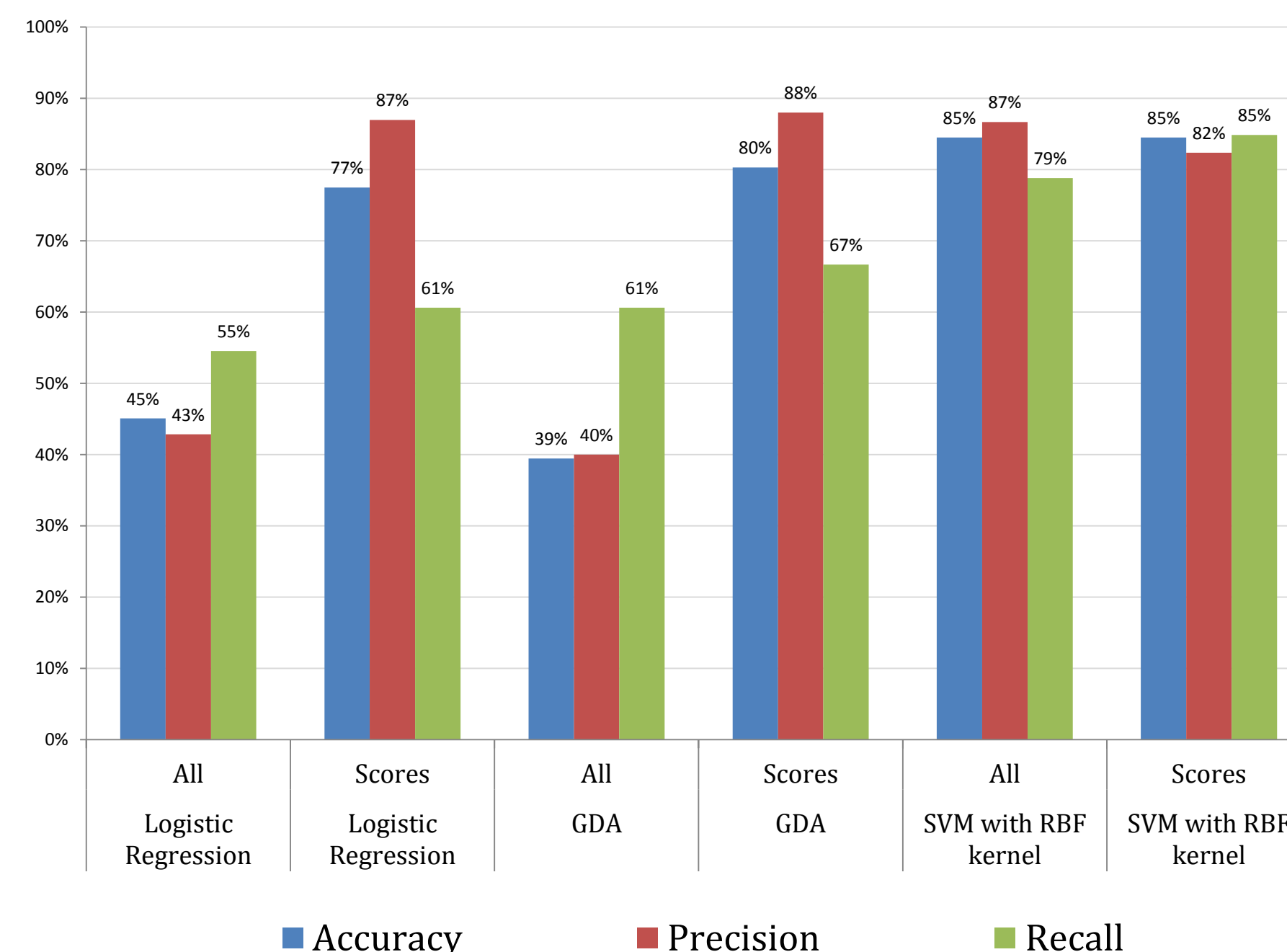
$$\text{rotational symmetry score} = x \cdot \text{flipped}(x)$$

MODEL COMPARISON

Performance of each model on each type of extracted features is compared in the table and graph below.

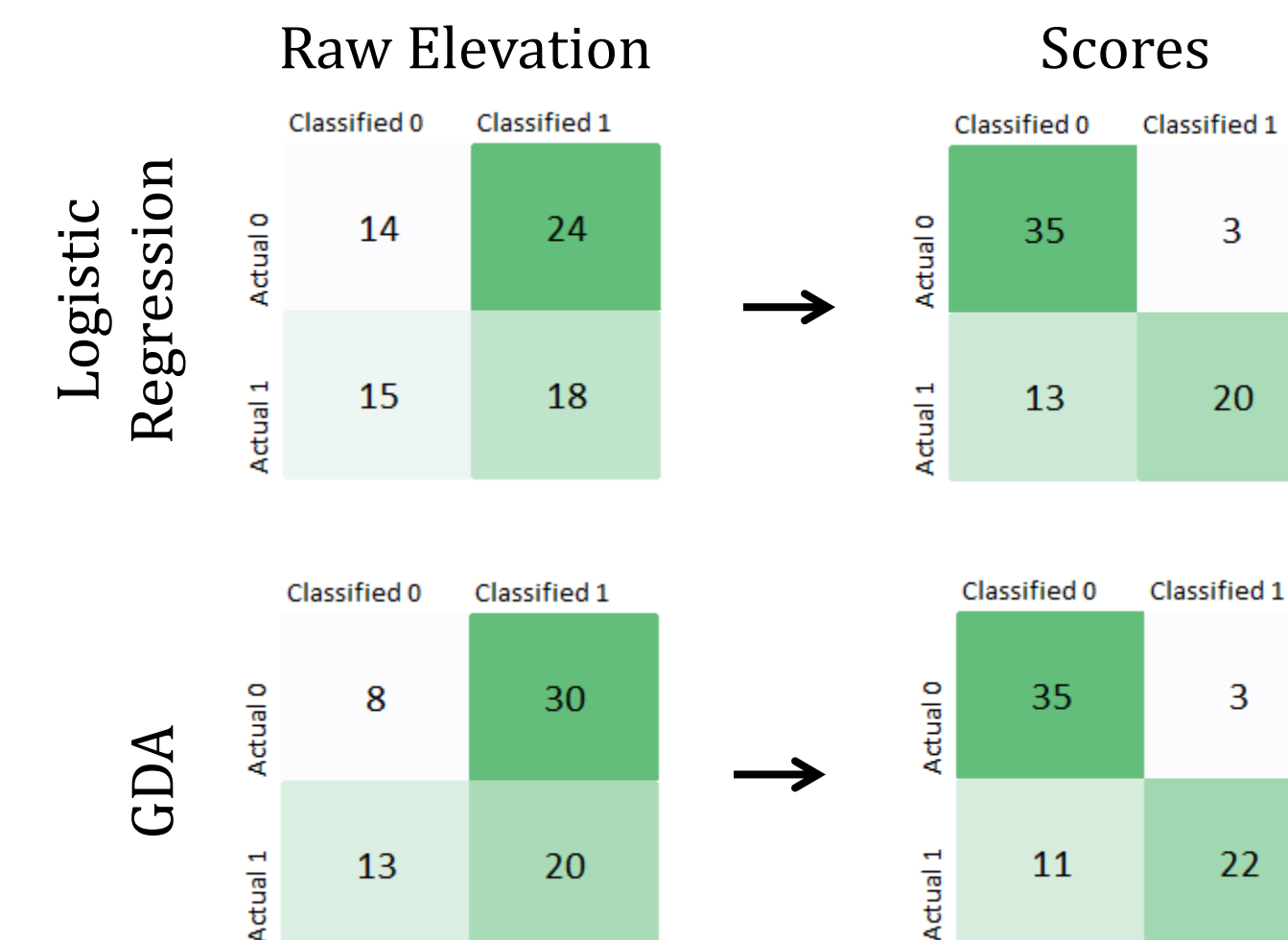
Model	Features	Accuracy	Precision	Recall
Logistic Regression	All	45%	43%	55%
Logistic Regression	Scores	77%	87%	61%
GDA	All	39%	40%	61%
GDA	Scores	80%	88%	67%
SVM w/ RBF Kernel	All	85%	87%	79%
SVM w/ RBF Kernel	Scores	85%	82%	85%

Algorithm Performance Comparison



RAW ELEVATION OR SCORES?

Using the raw elevation values as features rather than the contrived scores resulted in a tendency towards false positives. e.g., see confusion matrices for logistic regression and GDA using raw elevations vs. scores:



REGULARIZATION OF SVM

Optimized box constraints for the SVM were found to be 1.75 and 0.75 for the 2-feature and 441-feature models.

