- The task is to, again, identify the vegetation regions in the image

- Compare the result of using support vector classifiers to perform classification against results obtained from other models you have learned

- Which model is more appropriate for the general task of vegetation detection in aerial images (do not restrict yourself to which model performs better on just these two datasets)?

- Which model is more appropriate for other types of image processing (hand-writing digit classification for example)

- Your comparison should be both qualitative and quantitative

Hint: For your analysis, it's vital to consider the differences between the ways in which each of these models perform classification.

These differences can be gauged by looking at the differences between the decision boundaries drawn by the models and how these boundaries are determined.

- Are there any obvious draw backs to support vector classifiers as we have presented them to you?

- What might be some intuitive ways to address these draw backs?

Solution:

1. Load data and make a data scatter for all vegetation data sets
2. Visualize decision boundaries for all models and all data (training for new data)
3. For support vector machine:
   1. Support vectors identification and data fitting
   2. Parameter tuning and fitting test data
4. Quantitative comparison between SVM and other linear models (accuracy, false positives, false negatives)
5. Comparison between different methods
   1. Data characteristics: Number of features, class labels, availability of training data, categorical data, decision boundary shape, class separability
   2. Suitability of each model: Number of features limitations, size of data, Training cost, prediction cost, class separability, classification cost, number of labels, handling of categorical predictors, interpretability
6. Shortcomings of SVM
7. Test with polynomial kernels

Statement:

In this problem we revisit applications of classification, with the purpose of comparing the performance of support vector classifiers with other classifiers we have learned. We'll begin with the aeriel vegetation detection problem from Homework #7.

The data is contained in dataset\_1.txt and dataset\_2.txt (you are encouraged to use the datasets from Homework #7 as well). The first two columns of the data contains the latitude and longitudes of randomly sampled locations in the satellite image, and the last column contains a label indicating whether the location contains vegetation (1 denotes the presence of vegetation and 0 denotes otherwise). The task is to, again, identify the vegetation regions in the image.

* Compare the result of using support vector classifiers to perform classification against results obtained from other models you have learned. Which model is more appropriate for the general task of vegetation detection in aerial images (do not restrict yourself to which model performs better on just these two datasets)? Which model is more appropriate for other types of image processing (hand-writting digit classification for example) Your comparison should be both **qualitative** and quantitative.

**Hint:** For your analysis, it's vital to consider the differences between the ways in which each of these models perform classification. These differences can be gauged by looking at the differences between the decision boundaries drawn by the models and **how** these boundaries are determined.

* Are there any obvious draw backs to support vector classifiers as we have presented them to you? What might be some intuitive ways to address these draw backs?

Again, we provide you with a function plot\_decision\_boundary to visualize the decision boundary of a classifier.

Properties to be considered:

* Number of features
* Prediction cost
* Training cost
* Learning convergence
* Class separability
* Dataset size
* Decision boundary shape
* Number of labels and nature
* Training sets availability
* Categorical predictors
* In class distribution