Lab assignment #3: Detecting change in terrestrial environments with supervised methods

1. Logistics

Date assigned: February 26th, 2016
Date due: March 9th, 2016 (in conjunction with lab 2)
Points: 100 points for lab 3, 100 points for unsupervised change detection lab 2

2. Introduction

Now that you are familiar with a few of the unsupervised change detection procedures that are available, it is time to explore a completely different change detection approach: supervised multi-date composite change detection. This technique is becoming increasingly common, in part due to the availability of advanced, supervised machine-learning classification algorithms (support vector machines, decision trees, etc.) that can handle noisy data, missing features, and complex patterns in the training datasets.

The main goal of this lab is for you to detect change using two dates of satellite imagery, a sample of training data for stable and changed land cover types, and a support vector machine algorithm that is available within ENVI.

3. Supervised multi-date composite change detection

a. **Open ENVI and display the images**.

In addition to color composites, you may also wish to display your maps from the unsupervised change detection part of the lab. You should also open the band stack you used for the unsupervised k-means method (or recreate this if needed).

b. **Create a class scheme.**

Since you are now familiar with your study area, you should already have a pretty good idea of what classes you are going to map. Make sure to think about all 'stable' classes – those areas that are one land cover in the first time point and remain that land cover throughout the period – as well as the 'changed' areas. Determine the level of thematic precision you would like as well. Characterizing a broad class such as 'agricultural expansion 2000-2010' is going to be quite different than 'forest converted to agriculture, 2000-2010', 'desert converted to agriculture, 2000-2010', etc.

c. **Draw training sites for all classes.**

After making a detailed list of your classes, open the ROI tool and start digitizing exemplars of each class.

Remember to collect examples of each class across a range of locations throughout the image. Save your ROI file often, and include as much information in the name of the file as possible (e.g. egypt_training_data_15feb2015.roi).

d. **Run the support vector machine algorithm in ENVI.**

The SVM method is located on the ENVI main menu at

Classification > Supervised > Support Vector Machine

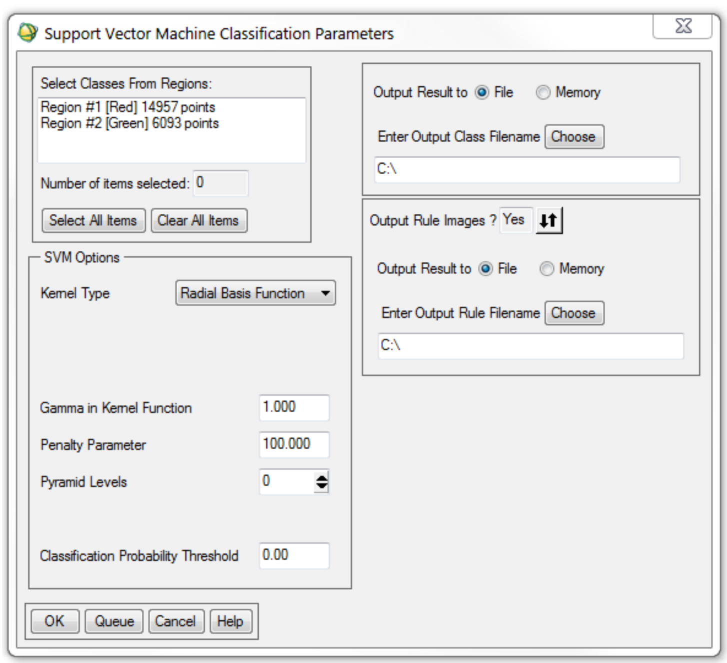
Select your input file and click ***OK***. The ***Support Vector Machine Classification Parameters*** dialogue appears with the following parameters:

Select Classes from Regions: Select the ROIs you wish to use for this classification.

SVM Options – Kernal Type: The SVM algorithm is built on the idea of using mathematical transformations ('kernals') applied to the training data to find the most optimal way to split the data into two classes. In ENVI, there are four options for the type of kernal you may apply during SVM classification: linear, polynomial, radial basis function, and sigmoid (for the equations associated with each, simply click on the help button at the bottom of the dialogue box). Depending on which option you choose, different fields will appear. For this lab, it is recommended to follow the remote sensing literature, which suggests that a radial basis function will typically provide the best results.

Gamma in Kernal Function: This refers to the value of gamma used within the kernal. Leave this at the default value, which is calculated as the inverse of the number of bands used in the input data.

Penalty Parameter: This optional parameters allows you to specify the level of misclassification allowed. A high value forces the classifier to have stricter margins for the boundaries of the hyperplane, and fewer misclassified pixels will be allowed on each side of the hyperplane. The default is 100, which permits some misclassification to occur without creating strict rules that might lead to over-fitting of the algorithm to the training data (remember that once the rules are created from the training data, the rest of the image data will be classified using these rules; if the rules fit the training data too closely, then the noisy, less perfect image data will not be classified correctly).



Pyramid Levels: SVM can be a time-consuming algorithm to run, so using the pyramid levels option allows ENVI to resample the data to a coarser resolution for classification. You might try setting this option to a value greater than zero, however, for best results, use the default value of zero.

Classification Probability Threshold: This optional parameter allows you to specify the probability required to classify a pixel. If set to zero, no threshold is applied, and a pixel is classified according to whatever the highest probability is across all the classes. If you specify a value for the threshold, a pixel must have a probability greater than the threshold to be labeled as that class (i.e. a value greater than zero will generate stricter rules for classification, and more pixels may be left unclassified). For your initial runs, a value of zero is recommended.

Output Result to File: Specify your directory and provide a name for your classified map.

Output Rule Images: This option will produce a multi-band image, where each band corresponds to one of your land cover classes. The values in each band (ranging between 0 and 100) indicate the probability of a given pixel belonging to that class.

If you choose to experiment with the optional parameters or different kernal types, remember to <u>change only one parameter at a time</u>. If you change multiple parameters for a given classification, it might be impossible to determine the effect of individual parameters.

e. **Assess the results**.

Take a moment to visually assess the map. Where did the classifier perform well? Where did it fail?

f. **Augment the training data and rerun the classification.**

While the SVM algorithm is optimized to perform well with small amounts of training data, you will likely find that your class accuracies increase when more training data are provided as input. Therefore, you may wish to add new sites to your classes, especially in areas where the classifier did perform as expected. Due to the time-consuming nature of the SVM algorithm, you are not required to perform a large number of classification iterations to perfect your map. It is more important for you to understand where the method worked well, and where it did not, and to discuss the possible reasons for this in your final lab report.

4. Accuracy assessment using an independent, random set of test sites

a. **Restart ENVI.**

When you are ready to assess the accuracy of your supervised change detection results, shut down ENVI and restart it. This will close all your training sites so that they are do not become confused with your test sites.

b. **Open the images, maps, and test sites**.

Open the color composites for each date, as well as your final map from the supervised change detection. Then re-open your set of test sites from your unsupervised change detection. It is likely that the class scheme in your map and your test site classes do not align. Therefore, you will need to adjust one or the other. If you wish to revise your test sites, simply revisit each site and relabel it as one of the more detailed changed or stable classes in your supervised change detection. It is far less time consuming to simply 'cross-walk' your final map to match your original set of test sites, however, as shown in the following step.

c. **Revise the class scheme in your map**.

First, go to a file folder (outside of ENVI) and make a copy of both the map.bsq image file, AND the header file.

Rename each file so that they are easy to distinguish from your original map file, and make sure the names of the new header and image file are identical (e.g. ***revisedmap.bsq*** and ***revisedmap.hdr***). You will want to keep your original, multi-class map to show in the results section of your write-up, of course, but you will report the accuracy based on your smaller set of classes.

Open the new map in an ENVI window, and open the interactive class tool window. Take a minute to write down how you will remap the classes. For example:

stable agriculture → no change class

stable forest → no change class

stable water → no change class

bare ground to forest → change to forest class

agriculture to forest → change to forest class

forest to bare ground → change to bare ground class

Now remap the classes in ENVI using

Interactive Class Tool > Options > Merge classes…

Select a base class, and merge all appropriate classes into the base class. Hit ***OK***, and then save the class map file

Interactive Class Tool > File > Save Changes to File…

Repeat this step until you have reduced your class map down to the three-class scheme of your test sites.

d. **Estimate the confusion matrix.**

On the main menu bar, go to:

Classification > Post Classification > Confusion Matrix > Using Ground Truth ROIs

Select your final classified map from the image list, and then click ***OK***. Match your ground truth ROIs to the newly revised thematic classes in your final supervised change map, and then click ***OK***.

e. **Save the confusion matrix results and format in Excel or Word**.

Once you have exported your confusion matrix results to a text file, import the file into Excel and edit/delete the accuracy results for your final report. Please format and edit the data into easy-to-read tables (e.g. type in the full names of classes, make sure all columns and rows have labels, eliminate any extra information that is not necessary in the table, and give the table a caption).

5. Assignment

Your assignment is to produce a final report that describes the results of all three change detection approaches (the two unsupervised techniques in lab 2 and the supervised approach in lab 3), and provides a discussion on the efficacy of each method. The report must comply with acceptable standards of technical writing within the geoscience community. The most common organizational form of a scientific report is:

- Introduction (statement of problem)
  - set the stage – why should we care?
  - state the goal and research questions
  - state your approach to answering the research questions
- Study Area
  - describe the area that you are working on (no more than one paragraph)
- Methods
  - describe the three approaches used (a few paragraphs)
- Results
  - present pertinent data in the form of graphs, figures, and tables (each with caption); (this is evidence for the discussion that follows)
  - provide snapshots of your maps and accuracy assessment results
- Discussion
  - ***discuss*** and ***interpret*** the meaning of the results; why was each approach successful? why not? what do the results mean?
- Conclusions
  - sum up results and findings in a few sentences

Please submit your report electronically by sending it to **envst556@gmail.com**. You must put your name and 'labs 2-3' in the subject heading of your email. For student Bucky Badger, the subject line would read:

Bucky Badger – labs 2-3

The name of your *word document* or *pdf file* must also have this format:

Bucky_Badger_labs2-3.doc