GEO/EVS 425/525 Unit 10
Supervised Classification

This exercise continues our investigations of satellite-images. Today we will be doing a supervised classification. Classification is a process for extracting information from imagery using statistical tools of various sorts. It identifies patterns – not spatial patterns, but rather patterns of similarity of digital numbers in various bands. Pattern recognition is the craft of finding *meaningful* patterns in raw data which can be extracted to form an image that has more meaning to the user than the original raw image had. We are accustomed to doing this. By spatially and spectrally enhancing an image, the human brain automatically sorts textures and colors received from the eye into categories. Classification in a GIS is a more scientific (although perhaps less romantic) process than what happens between the eye and the brain. Statistics are derived from the spectral characteristics of all of the pixels in an image. Then the pixels are sorted on the basis of well understood mathematical criteria.

There are two parts to the classification process: training and classifying. The training process recognizes patterns in the data. These patterns may be controlled by the analyst, as in *supervised classification*, or they may be controlled by the raw data, as in *unsupervised classification*. In this laboratory, we shall concentrate on supervised classification, as it is a bit simpler than unsupervised classification, and its biases are clearer.

Supervised classification operates via an estimate of the likelihood each pixel in the image belongs to a particular class. The classes must be identified *a priori*, and the constellation of digital numbers that is thought to belong to each class must be determined, based on *training areas*, or type areas from the image thought to represent typical areas from each class.

At its most straightforward, supervised classification requires that you [1] identify the classes you wish to consider, [2] identify training areas that you can use to typify each class, [3] create areas of interest on the image for each training area, so that you identify the range of digital numbers from each band corresponding to each class, and [4] apply the algorithm that associates each pixel to a class, based on an estimator of the likelihood that the pixel belongs to a class. In practice, there are several other operations that you might want to carry out to refine your characterization of the classes.

Put the TM quadrangle image you rectified in Unit 9 into a viewer. Open the Area-of-Interest Tool palette by clicking on AOI-Tools in the viewer menu bar. An Area of Interest (AOI) is simply a subset of an image which you define as having special interest to you. The nature of your interest in this exercise will be that each AOI will represent the training areas for each class. You will define the AOIs in two ways: by polygons and by growing an area from a seed.

Now click on AOI-Seed Properties. A dialog will open that will allow you to determine the characteristics by which you will generate an AOI from seed. You need to choose several parameters. First, choose a "Neighborhood." As several times before in this course, you can choose the 4-way neighborhood (i.e. no diagonals; this is the default), or you can choose the 8-way neighborhood. It should make little difference in this exercise. You must also choose an area to limit AOI growth. The default is 1000 pixels, although you can choose any number you want, and the units can be in number of pixels, hectares, acres, or square miles. 1000 pixels is the default, and it is appropriate for most instances. Finally (and most important), you must identify a spectral Euclidian distance. This parameter determines how far the digital number (DN) of a given pixel can be from the seed pixel and still be accepted as a member of the AOI. The pixels that are accepted will be within this spectral distance from the mean of the seed pixel. The default is 1 DN unit. You should raise it to 5 at first, but you might find that this is too low – or too high. As you gain experience, you may very well want to change it to a different number if you think it appropriate. You should not raise it too high, however, because this will make your training-site definitions less precise.

Let's begin. First, make a list of the classes you want to consider. Examples are water, grass, lowdensity residential, medium-density residential, high-density residential, commercial, industrial, pavement, forest, agriculture, etc. Your area may lend itself to using these classes; it may lend itself to using fewer classes or different classes. The choice is yours.

Now let's begin to identify training areas and create AOIs for them. Our first training editor will be defined by using the mouse as a digitizer. Find an area of relatively uniform DN distribution, where you know the class to which it belongs. Water is usually a good choice for this first AOI. From the AOI Tool palette, select the "Create Polygon AOI" icon. As you move your mouse into the viewer, the cursor turns into a crosshair. Move the cursor to the edge of your training area, and click with the left mouse button. Every left-click will digitize a point that will eventually become a polygon representing the AOI for the training area. When you have digitized as much of your AOI as you want, double-click the left mouse button to close the polygon. This is how you will create new polygon AOIs using your mouse to digitize.

Now go to the Signature Editor. Select the "Create New Signature(s) from AOI" icon. This will take the image area defined by the AOI and add its spectral information to the Signature Editor. By default, this signature will be called "Class 1." Edit the text field and enter a suitable name (e.g. Water) as the class name for the signature.

Find a second training area. We will use the seed-growing method to define the signature for this area. Again, the area should be relatively uniform, and you should know what it is. Select the "Region Grow AOI" icon. Again, the curser turns into crosshairs when you move it into the viewer. Move the cursor to the center of your chosen training area and click the left mouse button. What happens? Zoom in so that you can see the AOI created by this process. Change the spectral Euclidian distance to 10 (or 2), and click on "Redo." Change from the 4-way neighborhood rule to the 8-way neighborhood rule. Do you see a difference? What combinations give results which are most intuitively meaningful?

Again go to the Signature Editor when you are satisfied with your AOI and select the "Create New Signature(s) from AOI" icon. This will take the image area defined by the AOI and add its spectral information to the Signature Editor. Edit the text field and enter a suitable name (e.g. Grass) as the class name for the signature.

Complete your signature set by defining AOIs for training areas for each class you wish to consider. Which training areas lend themselves to digitizing? To using the region-grow technique? Note that you can digitize several areas for each class and then merge the digital information for each class into a single aggregate class. Be sure you know how to do this.

When you have finished your signature set, save it by selecting File-Save from the Signature Editor menu bar. Give your signature set a suitable name.

We could go, at this point, to the actual classification of the image. More often, we will want to evaluate the signatures created by the technique used so far. To do this, we use the Signature Editor. Select two classes whose spectral responses are very different. You might select forest and high-density residential, for example. Select the rows so that they are highlighted in yellow. Click on the "Display Histograms Window" icon. This will bring up the Histogram Plot Control Panel. Click on the "All Selected Signatures" and "All Bands" radio buttons and then click on "Plot." Reposition the windows so that you can view them simultaneously. The screen should show a series of plot windows showing two histograms in each. The color of the histograms will reflect the color of the corresponding signature. How different are these two signatures? Close the Histogram Plot Control Panel.

Now view the signature statistics, so that you can gauge which bands contain significant overlap. Within the Signature Editor, make sure that your indicator (>) is selecting one of the signatures you plotted using the Histograms tool. You can move the location of the indicator by simply clicking on the appropriate row within the indicator column. Click on the "Display Statistics Windows" icon, and a Statistics Plot dialog appears. Your statistics can be toggled by moving the Signature Editor indicator back and forth between your selected classes.

You may find that two particular signatures contain only a small amount of unwanted overlap. To eliminate this overlap, you don't need to remake the signature by defining a new AOI. You can adjust the signature parameters by selecting from the Signature Editor-Edit Parallelepiped Limits. Enter new Lower and Upper values in the text fields, or click Set to look at other options. Close the Statistics dialog.

Choose 4 classes from within the Signature Editor so that the rows are highlighted in yellow. Choose rather different features, so that they are not likely to be spectrally similar. Click on the "Display Mean Plot Window" icon on the Signature Editor tool bar, and the Signature Mean Plot dialog will display. To view all of the signatures you have chosen, click on the "Switch Between Single and Multiple Signature Mode" icon. You will see the means for each signature for each band of data. Move your cursor into the Mean plot. In the Status field below, you will see the layer and mean values automatically updated, reflecting the position of the cursor.

Consider all of your signatures. Each of them should fall into one of the following classes:

Bi-Modal	Spectrally confused training sample; should probably be redone
Normal distribution	Ideal training sample, spectrally separated
Very narrow range	Too small a sample, not representative of all of the feature
Very broad range	Possibly heterogeneous training area; signature confused with other features

Based on your plots within the Signature Editor, are any of them confused? That is, do they overlap so much with other signatures that they will not be able to differentiate different classes? Consider what you can do to revise the signatures: [1] delete the signature and remake it by redefining its training area's AOI, [2] edit the signature's parallelepiped limits, or [3] decide that the classes are inseparable and that they must be combined or remade using a different technique.

When you are satisfied with your signature set, you are ready to run the classification. If the Signature Editor is open, you can select Classify-Supervised from the menu bar. Otherwise, you can select Classifier-Supervised Classification from the main Imagine Control Panel. Your input image is your rectified TM quadrangle image; your input signature set is the signature set you have just created. Give your output a suitable name. You should also choose to output a distance file. This will be a raster image showing how far each pixel lies from the training set to which it was assigned. Give it a suitable name, such as XYZDistance.IMG, if your output image was XYZ.IMG. For the parametric rule, choose "Maximum Likelihood." Make sure that either none or all of the signatures in your signature set are selected. If only some signatures are selected, then the classification will be based only on the selected signatures. Use the remaining defaults, and click OK.

Load your new output image into a viewer. Make sure that it loads as a pseudocolor image. Adjust the colors so that they are representative and easy to interpret. Compare this image with your original TM image and with the image you created with your unsupervised classification in Unit 10. **The raw image generated by the supervised classification, with a suitable legend, should be included in your portfolio for this unit.**

You may note that your image contains a number of very small zones of one class in areas characterized by a different class. For example, you may have some individual pixels of "forest" within areas of high-density residential housing. This is a phenomenon known as "salt and pepper" and is commonly thought of as being inappropriate.

To eliminate the small zones, click on Interpreter-GIS Analysis-Clump. Your input file is the supervised classification you have just produced. Give your output file a suitable name. Choose 4-way for your connected-neighbors. Click OK. As when you ran Clump before, this algorithm will generate a series of clumps. It is the small ones you want to get rid of, by merging them with their nearest neighbors.

Click on Interpreter-GIS Analysis-Eliminate. The Eliminate dialog box will appear. The input file is the file you just produced using the Clump algorithm. Give your output file a suitable name. In the "Minimum" section, select 10 pixels as the threshold limit. This will eliminate all zones of 10 pixels or less and merge them with their nearest neighbors. Click on OK.

Load the eliminated file into a viewer, and compare it with your raw supervised image. To see the former in color, start the raster attribute editor for each file. Click on the column heading for colors in the raw supervised image to select it, then right-click it to invoke hidden functions. Choose "Copy." Now click on the column heading for colors in the eliminated file, and right-click to invoke hidden functions. Choose "Paste." The two files should now have the same color schemes. **This image, with a suitable legend, should be included in your portfolio for this unit.**

Questions to Consider

1. Which training areas lend themselves to digitizing as the most appropriate method of defining AOIs? Which lend themselves to using the region-grow technique as the most appropriate method of defining AOIs? Do you think that your answer will apply generally to all images of urban areas?

2. What is meant by Parallelepiped limits? Are they the most useful way to present limits?

3. Which spectral bands in your image show the greatest convergence and divergence in the plots? Where does this convergence or divergence come from, and what are its implications for your classification?

4. How do you use the distance file created in the generation of the classified image?

5. Do you agree that "salt and pepper" patterns are inappropriate? Where did they come from? When should you eliminate them?

Portfolio

1. The raw supervised classification, with suitable legend.

2. The supervised classification with the "salt and pepper" removed using Clump-Eliminate.