GEOG 4110/5100
Advanced Remote Sensing
Lecture 13

• Classification (Supervised and Unsupervised)

• Spectral Unmixing
  Richards: 11.10
By the time of the exam, you will have become familiar with the mechanics of image processing, and will have tried a number of different techniques to give you a better feeling for the advantages and limitations of remote sensing data in “real-world” applications. Through your project you should demonstrate a proficiency in the analysis of digital imagery for the purposes of extracting geophysical information.

There are three required presentations. On March 23rd, you will give a five-minute presentation about your topic and the status of the data. On April 18th, you will give a seven-minute presentation on the project’s progress. The last week of class will be devoted to the final presentation of your results. Written reports are due on May 1st, and you should prepare a fifteen-minute presentation, which you will give on May 2nd or 4th.

**GEOG 5100**

Pick an area with which you have some familiarity, collect all the ancillary data, such as geologic maps, land-use maps, soils maps, census data, or anything else that might help you interpret the information contained in the imagery. Your report should focus on how the remote sensing data you work with can be used to address a particular problem or question, how the remote sensing data correlates with the ancillary data, and what advantages and limitations are of the data processed by the various techniques that you have learned.

Start now to choose an area to work on, and begin to search for data that will suit your purposes. You should be able to find data to work with for most applications via the internet, however, the ESOC library does contain some historic Landsat, AVIRIS, and radar imagery that you can also consider using for your project. After you have chosen a project, begin to collect the data you need immediately. A topic and outline for your term project are due **February 26th March 9th**.

**GEOG 4110**

You are encouraged to develop a project idea on your own, and are free to work with a partner. If you would prefer, I can provide you with some project ideas from which to choose. You must present me with the topic and an outline by **February 26th March 9th**. Each group member (or individual) will prepare a report on the project, and the group will give a fifteen-minute presentation during the last week of class.
Two Dimensional Multi-Spectral Classes

Two-dimensional multi-spectral space with the spectral classes represented by Gaussian probability distributions. (Fig. 3.8 from Richards and Jia, 2006)

Each class modeled by a normal distribution specified by a mean vector and covariance matrix, which determine location (mean), width, height, and robustness (covariance matrix) of detection boundaries.
Classification

Unsupervised classification:
• The assigning of pixels of an image to spectral classes without the knowledge of their existence and names.
• Performed using clusters. The methods determine the location and the number of classes in the data and the class of each pixel.
• Identifying the classes using a reference data (maps, field).
• It is useful in identifying the spectral classes of an image before further analysis (e.g. supervised).

Supervised classification
• A number of statistical and non-statistical methods are available.
• Statistical methods assume that each spectral class has a particular probability distribution (Gaussian) function in multispectral space.
• Consists of three broad phases: (1) Selection of training pixels (field data, maps, ...), (2) Compute the mean and covariance matrix, and (3) Assigning each pixel to a class using the highest probability.
Six Steps in Supervised Classification

1. Decide on set of ground cover types into which the image is classified.
2. Choose representative pixels or training data from each class.
   – Based on knowledge of the region acquired either through ancillary information, or interpretation of the imagery
3. Use the training data to estimate the parameters of the particular classifier algorithm to be used.
   – Properties that define a probability model
   – Equations that define partitions in multi-spectral space
   – Signature of that class
4. Use the trained sets to classify every pixel in the image into one of the information classes.
5. Produce tables or thematic maps that summarize the results of the classification.
6. Assess the accuracy of the classification using a testing dataset.
Maximum Likelihood Classification

• Most common supervised classification with remote sensing imagery.

• We define a vector \( \mathbf{x} \) that is the set of brightness values of a pixel in multi-spectral space.

\[
\begin{array}{c|ccccc}
\text{Band} & 1 & 2 & 3 & 4 & 5 \\
\hline
\text{Brightness} & 35 & 74 & 244 & 20 & 133 \\
\end{array}
\]

\[ x = \begin{bmatrix} 35 \\ 74 \\ 244 \\ 20 \\ 133 \end{bmatrix} \]

• This vector has a certain probability of being in one of \( M \) spectral classes \( \omega_i \) in an image

\[ p(\omega_i | \mathbf{x}), \ i = 1, 2, \ldots M \]

• \( \mathbf{x} \) is classified as follows

\[ \mathbf{x} \in \omega_i, \quad \text{if} \quad p(\omega_i | \mathbf{x}) > p(\omega_j | \mathbf{x}) \quad \text{for all} \ j \neq i \]

— i.e. the probability of a given pixel is greatest that it falls into class \( i \) rather than any other class [pixel in a class]
Mean Vector and Covariance

The mean vector \((\mathbf{m})\) is the vector average of the individual components of a vector.

The covariance between two real-valued random variables describes how one variable varies in relation to another.

\[
\text{Cov}(X,Y) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})
\]

\[
C_{XX} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^T
\]

Which has high covariance and which has a low?

Mean Vector and Covariance

The covariance matrix ($\Sigma_x$) is a matrix of covariance values that describes the scatter or spread between variables.

\[ \sum_x = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - m)(x_i - m)^t \]

Computation of Covariance Matrix
(Table 8.1 from Richards and Jia, 2006)

<table>
<thead>
<tr>
<th>$x$</th>
<th>$x - m$</th>
<th>$[x - m][x - m]^t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2.00</td>
<td>4.00 0.66</td>
</tr>
<tr>
<td>2</td>
<td>-0.33</td>
<td>0.66 0.11</td>
</tr>
<tr>
<td>2</td>
<td>-1.00</td>
<td>1.00 1.33</td>
</tr>
<tr>
<td>1</td>
<td>-1.33</td>
<td>1.33 1.77</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>1.00 -1.33</td>
</tr>
<tr>
<td>1</td>
<td>-1.33</td>
<td>-1.33 1.77</td>
</tr>
<tr>
<td>5</td>
<td>2.00</td>
<td>4.00 -0.66</td>
</tr>
<tr>
<td>2</td>
<td>-0.33</td>
<td>-0.66 0.11</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>1.00 1.67</td>
</tr>
<tr>
<td>4</td>
<td>1.67</td>
<td>1.67 2.79</td>
</tr>
<tr>
<td>2</td>
<td>-1.00</td>
<td>1.00 -1.67</td>
</tr>
<tr>
<td>4</td>
<td>1.67</td>
<td>-1.67 2.79</td>
</tr>
</tbody>
</table>

whereupon $\Sigma_x = \begin{bmatrix} 2.40 & 0 \\ 0 & 1.87 \end{bmatrix}$

Covariance of $x_1$ and $x_2$ is 0 (not correlated)

\[
\begin{align*}
m & = 3.00 \\
2.33 &
\end{align*}
\]
Mean Vector and Covariance

The covariance matrix ($\Sigma_x$) is a matrix of covariance values that describes the scatter or spread between variables.

\[
\sum_x = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - m)(x_i - m)^t
\]

Computation of Covariance Matrix
(Table 8.1 from Richards and Jia, 2006)

<table>
<thead>
<tr>
<th>$x$</th>
<th>$x - m$</th>
<th>$[x - m][x - m]^t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2.00</td>
<td>4.00 0.66</td>
</tr>
<tr>
<td>2</td>
<td>-0.33</td>
<td>0.66  0.11</td>
</tr>
<tr>
<td>2</td>
<td>-1.00</td>
<td>1.00  1.33</td>
</tr>
<tr>
<td>1</td>
<td>-1.33</td>
<td>1.33  1.77</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>1.00  -1.33</td>
</tr>
<tr>
<td>4</td>
<td>-1.33</td>
<td>-1.33 1.77</td>
</tr>
<tr>
<td>5</td>
<td>2.00</td>
<td>4.00  -0.66</td>
</tr>
<tr>
<td>2</td>
<td>-0.33</td>
<td>-0.66  0.11</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>1.00  1.67</td>
</tr>
<tr>
<td>4</td>
<td>1.67</td>
<td>1.67  2.79</td>
</tr>
<tr>
<td>2</td>
<td>-1.00</td>
<td>1.00  -1.67</td>
</tr>
<tr>
<td>4</td>
<td>1.67</td>
<td>-1.67 2.79</td>
</tr>
</tbody>
</table>

whereupon

\[
\Sigma_x = \begin{bmatrix} 2.40 & 0 \\ 0 & 1.87 \end{bmatrix}
\]

Variance of $x_1$ and $x_2$

\[
m = \begin{bmatrix} 3.00 \\ 2.33 \end{bmatrix}
\]
## Signature

Comprised of mean values in each band from a sample population

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>44.27</td>
</tr>
<tr>
<td></td>
<td>28.82</td>
</tr>
<tr>
<td></td>
<td>22.77</td>
</tr>
<tr>
<td></td>
<td>13.89</td>
</tr>
<tr>
<td>Fire Burn</td>
<td>42.85</td>
</tr>
<tr>
<td></td>
<td>35.02</td>
</tr>
<tr>
<td></td>
<td>35.96</td>
</tr>
<tr>
<td></td>
<td>29.04</td>
</tr>
<tr>
<td>Vegetation</td>
<td>40.46</td>
</tr>
<tr>
<td></td>
<td>30.92</td>
</tr>
<tr>
<td></td>
<td>57.50</td>
</tr>
<tr>
<td></td>
<td>57.68</td>
</tr>
<tr>
<td>Urban</td>
<td>63.14</td>
</tr>
<tr>
<td></td>
<td>60.44</td>
</tr>
<tr>
<td></td>
<td>81.84</td>
</tr>
<tr>
<td></td>
<td>72.25</td>
</tr>
</tbody>
</table>
Mean Vector and Covariance Signatures

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean Vector</th>
<th>Covariance Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Low variance: Homogeneous</td>
</tr>
<tr>
<td>Water</td>
<td>44.27</td>
<td>14.36</td>
</tr>
<tr>
<td></td>
<td>28.82</td>
<td>3.71</td>
</tr>
<tr>
<td></td>
<td>22.77</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td>13.89</td>
<td>4.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fire Burn</td>
<td>42.85</td>
<td>9.38</td>
</tr>
<tr>
<td></td>
<td>35.02</td>
<td>20.29</td>
</tr>
<tr>
<td></td>
<td>35.96</td>
<td>22.10</td>
</tr>
<tr>
<td></td>
<td>29.04</td>
<td>20.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.05</td>
</tr>
<tr>
<td>Vegetation</td>
<td>40.46</td>
<td>5.56</td>
</tr>
<tr>
<td></td>
<td>30.92</td>
<td>7.46</td>
</tr>
<tr>
<td></td>
<td>57.50</td>
<td>1.96</td>
</tr>
<tr>
<td></td>
<td>57.68</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19.71</td>
</tr>
<tr>
<td>Urban</td>
<td>63.14</td>
<td>43.58</td>
</tr>
<tr>
<td></td>
<td>60.44</td>
<td>60.57</td>
</tr>
<tr>
<td></td>
<td>81.84</td>
<td>17.38</td>
</tr>
<tr>
<td></td>
<td>72.25</td>
<td>-9.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-14.86</td>
</tr>
</tbody>
</table>

Class signatures generated from the training pixels (adapted from Richards, 2013 Table 8.1).
## Maximum Likelihood Classification

**Tabular summary of the thematic map (adapted from Richards and Jia, 2006)**

<table>
<thead>
<tr>
<th>Class</th>
<th>No. of Pixels</th>
<th>Area (ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>4830</td>
<td>2137</td>
</tr>
<tr>
<td>Fire Burn</td>
<td>14182</td>
<td>6274</td>
</tr>
<tr>
<td>Vegetation</td>
<td>28853</td>
<td>12765</td>
</tr>
<tr>
<td>Urban</td>
<td>22791</td>
<td>10083</td>
</tr>
</tbody>
</table>

*Classifications include: Water, Fire Burn, Vegetation, Urban.*
Maximum Likelihood Thresholds

Figures to the right illustrate poor classification for patterns lying near the tails of the distribution functions of all spectral classes. B. Use of a threshold to remove poor classification (Fig. 8.1 from Richards and Jia, 2006).

In 2 dimensions, width, height, and robustness of detection boundaries determined by covariance matrix.
Maximum Likelihood Thresholds

• The threshold can be determined from the statistics of the distribution

\[ T_i = 0.5[\chi^2] - (0.5)\ln |\Sigma_i| + \ln p(\omega_i) \]

Where

\[ \chi^2 = \text{the chi-squared distribution for N degrees of freedom and} \]
\[ |\Sigma_i| = \text{the determinant of the covariance matrix of the ith} \]
\[ \text{spectral class (e.g. fire, water, soil, vegetation)} \]

DOF = #bands - 1

• Example: if we want to eliminate 5% of the least likely pixels for a \( \chi^2 \) distribution for five bands, \( \chi^2 \) is 9.488, so for a given class:

\[ T_i = 4.744 - (0.5)\ln |\Sigma_i| + \ln p(\omega_i) \]

\[ \chi^2 \text{ Table} \]

<table>
<thead>
<tr>
<th>df</th>
<th>( P = 0.05 )</th>
<th>( P = 0.01 )</th>
<th>( P = 0.001 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.84</td>
<td>6.64</td>
<td>10.83</td>
</tr>
<tr>
<td>2</td>
<td>5.99</td>
<td>9.21</td>
<td>13.82</td>
</tr>
<tr>
<td>3</td>
<td>7.82</td>
<td>11.35</td>
<td>16.27</td>
</tr>
<tr>
<td>4</td>
<td>9.49</td>
<td>13.28</td>
<td>18.47</td>
</tr>
<tr>
<td>5</td>
<td>11.07</td>
<td>15.09</td>
<td>20.52</td>
</tr>
</tbody>
</table>
Calculating Determinants

Determinant of a 2x2 matrix

\[ |A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc. \]

Determinant of a 3x3 matrix

\[
|A| = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}
= aei + bfg + cdh - ceg - bdi - afh.
\]

Determinant of a 4x4 matrix

\[
\begin{vmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{vmatrix} = a \begin{vmatrix} f & g & h \\ j & k & l \\ n & o & p \end{vmatrix} - b \begin{vmatrix} e & g & h \\ i & k & l \\ m & o & p \end{vmatrix} + c \begin{vmatrix} e & f & h \\ i & j & l \\ m & n & p \end{vmatrix} - d \begin{vmatrix} e & f & g \\ i & j & k \\ m & n & o \end{vmatrix}.
\]
Calculating Determinants

For 4×4 Matrices and Higher

The pattern continues for 4×4 matrices:

- **plus a** times the determinant of the matrix that is **not** in **a**'s row or column,
- **minus b** times the determinant of the matrix that is **not** in **b**'s row or column,
- **plus c** times the determinant of the matrix that is **not** in **c**'s row or column,
- **minus d** times the determinant of the matrix that is **not** in **d**'s row or column,

\[
\begin{vmatrix}
  a & x \\
  f & g & h \\
  j & k & l \\
  n & o & p \\
\end{vmatrix}
\begin{vmatrix}
  b & x \\
  e & g & h \\
  i & k & l \\
  m & o & p \\
\end{vmatrix}
\begin{vmatrix}
  c & x \\
  e & f & h \\
  i & j & l \\
  m & n & p \\
\end{vmatrix}
\begin{vmatrix}
  d & x \\
  e & f & g \\
  i & j & k \\
  m & n & o \\
\end{vmatrix}
\]

As a formula:

\[
|A| = a \cdot \begin{vmatrix}
  f & g & h \\
  j & k & l \\
  n & o & p \\
\end{vmatrix}
- b \cdot \begin{vmatrix}
  e & g & h \\
  i & k & l \\
  m & o & p \\
\end{vmatrix}
+ c \cdot \begin{vmatrix}
  e & f & h \\
  i & j & l \\
  m & n & p \\
\end{vmatrix}
- d \cdot \begin{vmatrix}
  e & f & g \\
  i & j & k \\
  m & n & o \\
\end{vmatrix}
\]

Notice the + - + - pattern (+a... -b... +c... -d...). This is important to remember.

The pattern continues for 5×5 matrices and higher.
Minimum Distance Classification

• Large number of pixels needed for maximum likelihood in order to calculate mean vector and covariance matrix for each spectral class
  – Requires sufficient number of training pixels for each class
• Minimum Distance Classification does not use co-variance information
  – Relies instead on mean positions of the spectral classes
  – Can be performed when number of training pixels is limited
• Training data used to determine class means
• Classification performed by placing pixel in class of nearest mean

\[ x \in \omega_i \quad \text{if} \quad d(x, \mu_i)^2 < d(x, \mu_j)^2 \quad \text{for all } j \neq i \]

where: \( d(x, \mu_i)^2 = (x - \mu_i)^t (x - \mu_i) \)
Parallelepiped Classification

• Simple classification technique based on histograms of individual spectral components in training data

Histograms for the components of a 2-D set of training data corresponding to a single spectral class. (Fig. 8.5 from Richards and Jia, 2006)

The upper and lower boundaries of the histograms above define the edges of a 2-D parallelepiped.
Parallelepiped Classification

- Computationally efficient
- Many cases of dual classification
- Many cases of null classification

Distribution of class data in 2-D spectral space (left) and parallelepipeds associated with those distributions (right)
Context Classification

• Maximum Likelihood, Minimum Distance, and Parallelepiped classifiers are all pixel-specific

• Context Classification considers characteristics of neighboring behavior
  – Sensors acquire some energy from adjacent pixels
  – Ground cover variability is usually larger than pixel sizes
  – Greater probability that a pixel will be similar to its neighboring pixel than one far away
  – Can reduce misclassifications due to noisy data
Unsupervised Classification

• Unsupervised classification (clustering) is the portioning of remote sensing data into different spectral classes.
• Successful application of supervised classification, depends on how correctly we delineated the training classes.
• It is not easy to identify unimodal groups, therefore clustering methods are practical alternatives. They are applied to indentify the spectral classes of the data.
Unsupervised Classification Methods

Clustering

• Grouping of pixels in a multispectral space. Pixels in each cluster are spectrally similar.
• Euclidean distance can be used to measure the similarity between pixels.

$$d(x_1, x_2) = \left\{ \sum_{i=1}^{N} (x_{1i} - x_{2i})^2 \right\}^{\frac{1}{2}}$$

Where: $N = \text{number of spectral bands}$

Fig. 9.1. From Richards, 2013
Unsupervised Classification Methods

• Sum of squared error (SSE) can be used to measure the quality of clusters. It calculates the cumulative distance to the center for each cluster and the sums the distances of all clusters. The clusters are favorable when the distance is small.

• It requires calculating big numbers of SSE to evaluate all clusters.
Unsupervised Classification Methods

Migrating means (Iterative optimization, Isodata)

- Based on assigning the pixel vectors into candidate clusters.

\[
x = \begin{bmatrix}
35 \\
74 \\
244 \\
20 \\
133 \\
\end{bmatrix}
\]

- Moving the cluster centers until a minimum SSE is reached.
Steps for migrating means method

• Selecting N points in multispectral space to serve as cluster centers. The cluster means should be spaced uniformly over the data. As with the means, the number of clusters must be identified beforehand.

\[ \hat{m}_i, i = 1, 2, \ldots, N \]

• Based on Euclidean distance each pixel at a location \( x \) is assigned to the nearest cluster.

• New means are calculated from the clustering resulted from the previous step.

\[ m_i, i = 1, 2, \ldots, N \]

• If \( \hat{m}_i = m_i \) the procedure is terminated, otherwise \( \hat{m}_i \) is redefined as \( m_i \).
Once the clustering is completed. The clusters can be examined for:

- Any clusters contain very few points will not be useful as training classes for supervised classification.
- Some clusters are so close they need to be merged.

Clustering by Isodata method (Fig. 9.2. from Richards, 2013)
Hybrid Supervised and unsupervised Classification

- The major shortcoming of the supervised classification is the need to delineate unimodal spectral classes in advance. Unsupervised or clustering can be used to identify the training classes.

Steps for hybrid

1. Use unsupervised classification to determine the spectral classes of the image (representative subset of the data).
2. Using field data or other reference data to associate the spectral classes with information classes (average 2-3 spectral classes).
3. Evaluating if all bands needed for classification.
4. Using supervised classification to classify the entire image.
5. Label the ground cover types.
Training classes

• Major step of supervised classification is the selection of the training pixels. The information about training classes may be obtained from the field, maps, areal photos (generally for major land cover types), etc.

• It is important that the data used to select the training pixels be collected at the same time of the data to be classified.

• It is important to at least identify training classes for all classes of interest, preferably all classes in the image.

• Use of thresholds result in more accurate results.

• Refinement of the training classes and reclassification may be necessary.