Raster Classification with ArcGIS Desktop

Rebecca Richman
Andy Shoemaker
Raster Classification

• What is it?
  - Classifying imagery into different land use/land cover classes based on the pixel values of imagery bands

• Why do it?
  - Allows for analysis of land use/land cover change, identification of features, suitability analysis
Raster Classification

- What we will be talking about today
  - Workflow in ArcMap and ArcGIS Pro
  - Decisions that need to be made prior to classification
  - The different algorithms available in ArcGIS Desktop
  - Accuracy assessment
Workflows

How to do it
Workflow in ArcMap

1. Simplify data with Segment Mean Shift (Optional)
2. Train Classifiers
3. Run Classifiers
4. Analyze Samples against Accuracy Assessment points
Workflow in ArcGIS Pro

- Create Training Samples and Generate Classification Schema if desired
- Image Classification Wizard
  - Segment Mean Shift
  - Train Classifiers
  - Classify your Data
  - Merge Classes
- Do Accuracy Assessment
Prior to classification

Things to Consider
Unsupervised vs supervised
Leave the kids home alone….

- **Unsupervised**
  - Classification is based on the software analysis of an image without the user providing sample classes
  - Users decide on the number of classes and number of iterations used for the classification.
  - Easier, but less accurate
Supervised vs unsupervised
….or with a Babysitter?

- **Supervised**
  - Classification is based on user provided training samples that are representative of the different land use/land cover classes.
  - Training samples are selected based on user’s knowledge of the area represented by the image.
  - Users choose training sample, algorithm (Maximum Likelihood, Random Trees, Vector Support Machine)
  - More intensive, but better accuracy
Pixel based vs Object based

- **Pixel based classification**
  - Every pixel is classified separately

- **Object based classification**
  - Pixels are grouped into objects (Segmentation Mean Shift) and then classified
Determine Classes
Keepin’ it Classy

- What classes make sense for the analysis?
- Use a Preexisting Scheme or create a custom?
  - Anderson Land Use/Land Cover Schema
- Do classes need to be split for machine training?

<table>
<thead>
<tr>
<th>Level I</th>
<th>Level II</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Urban or Built-up Land</td>
<td>11 Residential</td>
</tr>
<tr>
<td>12 Commercial and Services</td>
<td></td>
</tr>
<tr>
<td>13 Industrial</td>
<td></td>
</tr>
<tr>
<td>14 Transportation, Communications, Utilities</td>
<td></td>
</tr>
<tr>
<td>15 Industrial and Commercial Complexes</td>
<td></td>
</tr>
<tr>
<td>16 Mixed Urban or Built-up Land</td>
<td></td>
</tr>
<tr>
<td>17 Other Urban or Built-up Land</td>
<td></td>
</tr>
<tr>
<td>2 Agricultural Land</td>
<td>21 Cropland and Pasture</td>
</tr>
<tr>
<td>22 Orchards, Groves, Vineyards, Nurseries, and Ornamental Horticultural Areas</td>
<td></td>
</tr>
<tr>
<td>23 Confined Feeding Operations</td>
<td></td>
</tr>
<tr>
<td>24 Other Agricultural Land</td>
<td></td>
</tr>
<tr>
<td>3 Rangeland</td>
<td>31 Herbaceous Rangeland</td>
</tr>
<tr>
<td>32 Shrub and Brush Rangeland</td>
<td></td>
</tr>
<tr>
<td>33 Mixed Rangeland</td>
<td></td>
</tr>
</tbody>
</table>
How to Train Your Data
Let’s think like machines!
How Many Classes Do You See?
I spy with my little eye.....

• Asphalt
• Buildings
• Planes
• Grass
• Trees
How to Teach an Algorithm to ‘See’
What Makes a Tree a Tree?

• How do humans identify classes in imagery?
  - Color
  - Size
  - Shape
  - Texture

• How do we teach an algorithm to identify classes?
  - Look at image as a machine does
    - Pixel by Pixel
    - Band by Band
    - Reflectance
Translate our Classes to a Computer’s

- Asphalt
- Buildings
- Planes
- Grass
- Trees
Translate our Classes to a Computer’s
01000001 01110010 01100011 01000111 01001001 01010011

- Asphalt
- Buildings
  - Grey Roofs
  - White Roofs
- Planes
- Grass
- Trees
- Shadows
How to train your data
AKA How to think like a machine

• General guidelines:

- 20-30 samples per class

- As evenly distributed across the image as possible

- Train, classify, adjust classes, repeat until classification is satisfactory, try different algorithms
Algorithms
All sorts of crazy maths
The Basic Equation
ISO Cluster
ISO Love Clusters

• **Unsupervised**
  - Iterative Self Organizing
  - Initially, means are placed on a 45 degree line in the multiband space and then each pixel is assigned to the closest mean.
  - New means are calculated.
  - The next iteration then assigns the pixels to the closest new mean.

• User selects number of classes and number of iterations

• Easy

• Not as accurate
Maximum Likelihood

• Supervised
  - Uses training sample variance and co-variance to create a multiband class signature. Assigns pixels to class based on the maximum likelihood of that they belong to that class
  - Assumes normal distribution of training data in multiband space
Support Vector

• Supervised
  - Similar to Maximum Likelihood
  - Places pixels in multiband space
  - Instead of determining the likelihood of a pixel belonging a class, pixels are assigned to classes so that the gaps between classes are as large as possible
Random Trees
An Entmoot?

- Supervised
  - Creates decision trees for random sub-samples of the training data
  - Each pixel is then classified by each tree
  - The class that is most often selected by the trees is assigned to the pixel

```
Band 4 > 0.3
  Band 7 > 0.4  Soil
  Band 7 < 0.4  Vegetation
Band 4 < 0.3
  Water
```
Accuracy Assessment
Assess the Accuracy
Accuracy Assessment

- Done to provide an idea of how well the Reclassification worked
- “Real Values” (Ground Truth/Testing data) can be collected in the field or visually from Satellite Images
- 3 Main tools
  - Create Accuracy Assessment Points
  - Update Points
  - Create Confusion Matrix

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Asphalt</th>
<th>Concrete</th>
<th>Grass</th>
<th>Tree</th>
<th>Building</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asphalt</td>
<td>2385</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>2394</td>
</tr>
<tr>
<td>Concrete</td>
<td>0</td>
<td>332</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>333</td>
</tr>
<tr>
<td>Grass</td>
<td>0</td>
<td>1</td>
<td>908</td>
<td>8</td>
<td>0</td>
<td>917</td>
</tr>
<tr>
<td>Tree</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1084</td>
<td>9</td>
<td>1093</td>
</tr>
<tr>
<td>Building</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>2053</td>
<td>2071</td>
</tr>
<tr>
<td>Total</td>
<td>2397</td>
<td>337</td>
<td>908</td>
<td>1099</td>
<td>2067</td>
<td>6808</td>
</tr>
</tbody>
</table>
Confusion Matrix
Will you take the Red Cell or the Blue Cell?

• Generates a Kappa Index of Agreement between classified raster and ground truth data
• Index is based on how well the Classified Raster reflects the Ground Truth Points
• Kappa Index is expressed as a value between 0 and 1
  - The closer to 1 the value is, the more accurate the reclassification was
• This tool is best used when comparing different Algorithms or Methods
  - This is because the Kappa Index assesses each raster independently

GARBAGE IN, GARBAGE OUT
So, Why do We Do This?
Becomes

This...
Questions?