Raster Classification with ArcGIS Desktop

Rebecca Richman

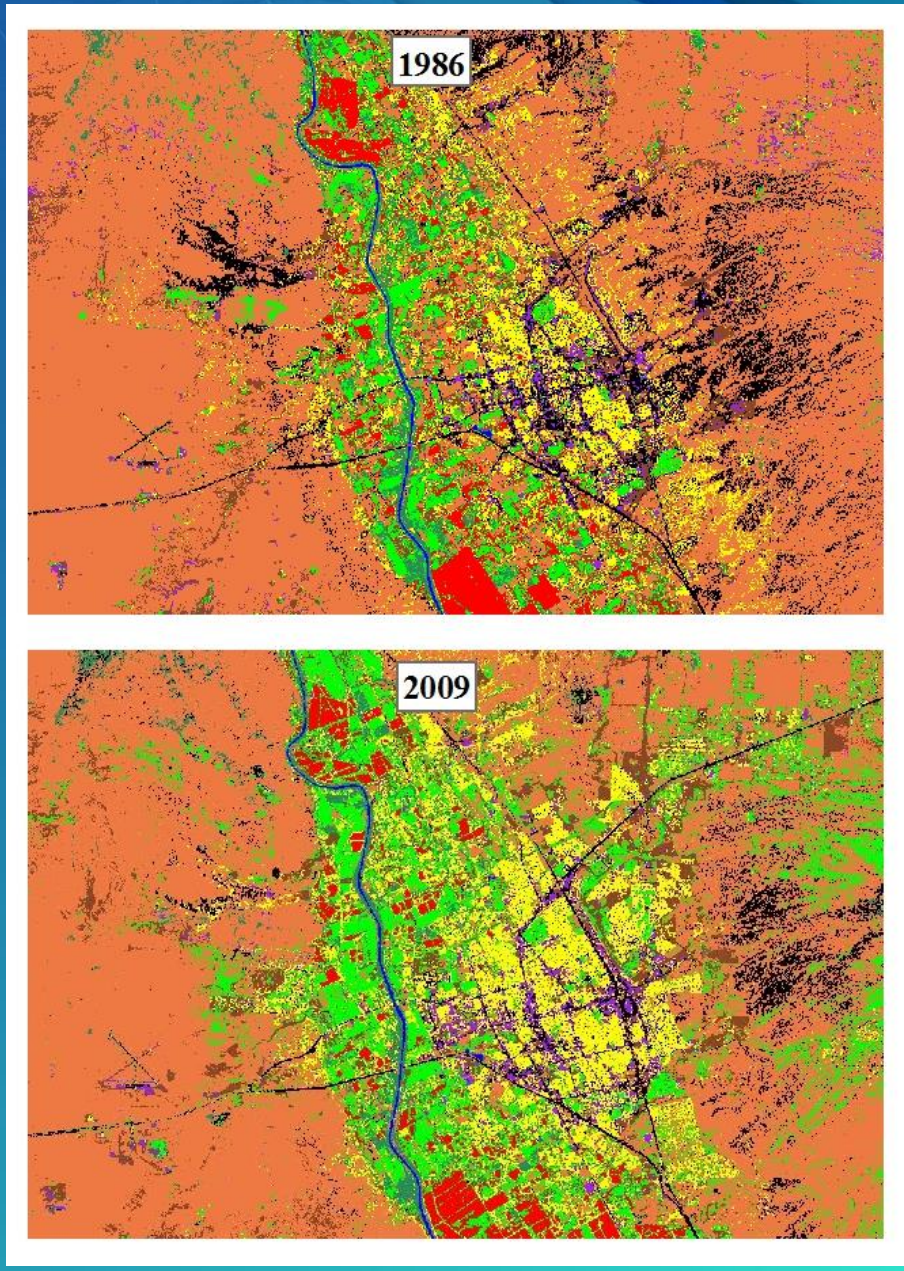
Andy Shoemaker

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 classification
  - The different algorithms available in ArcGIS Desktop
  - Accuracy assessment

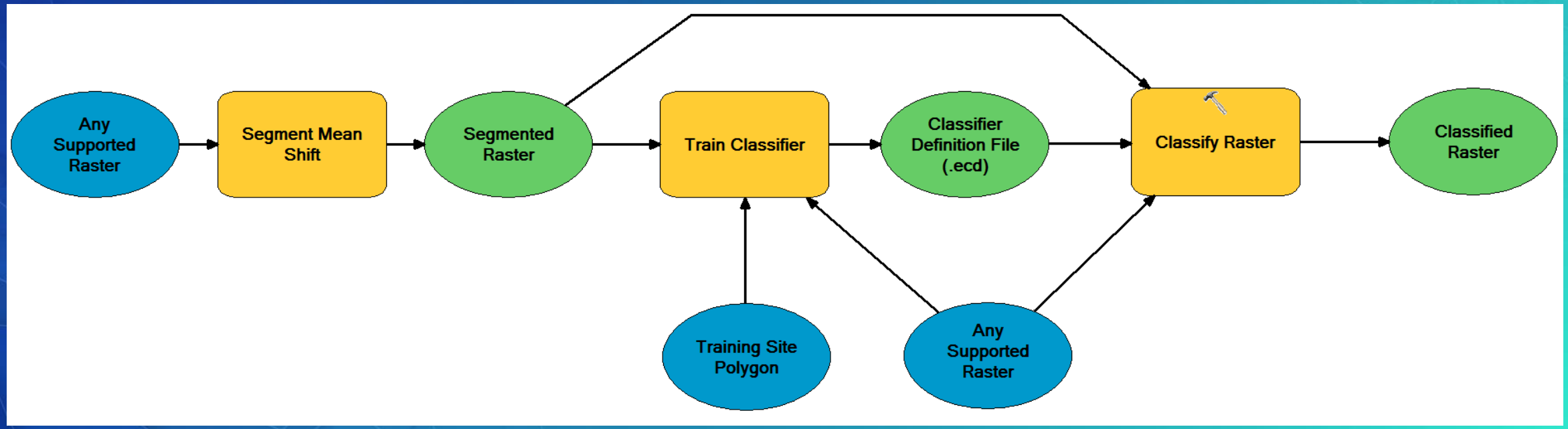


Workflows

How to do it

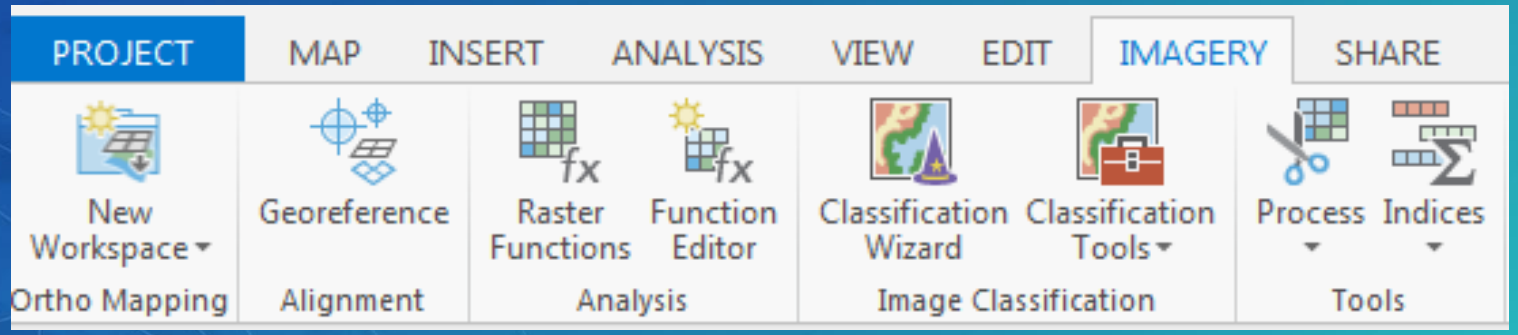
Workflow in ArcMap

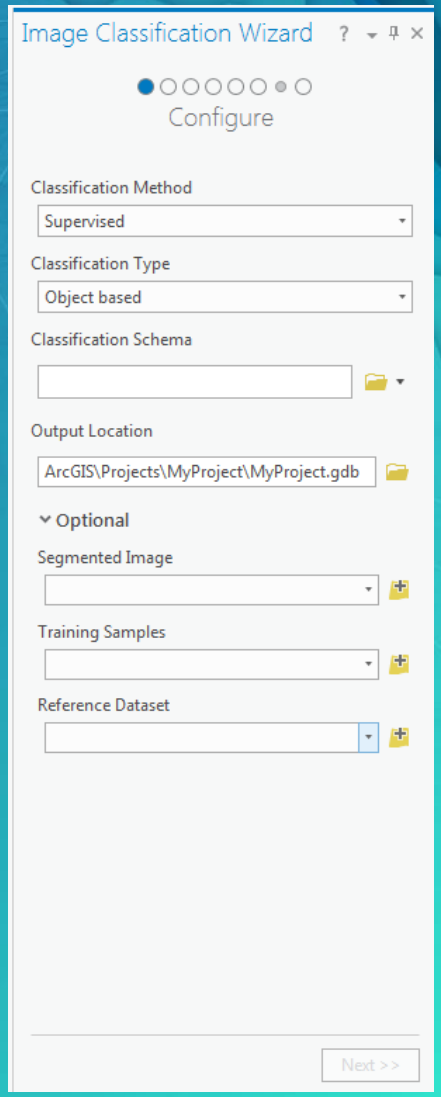
1. Simplify data with Segment Mean Shift (Optional)
2. Train Classifiers
3. Run Classifiers
4. Analyze Samples against Accuracy Assessment points



Workflow in ArcGIS Pro

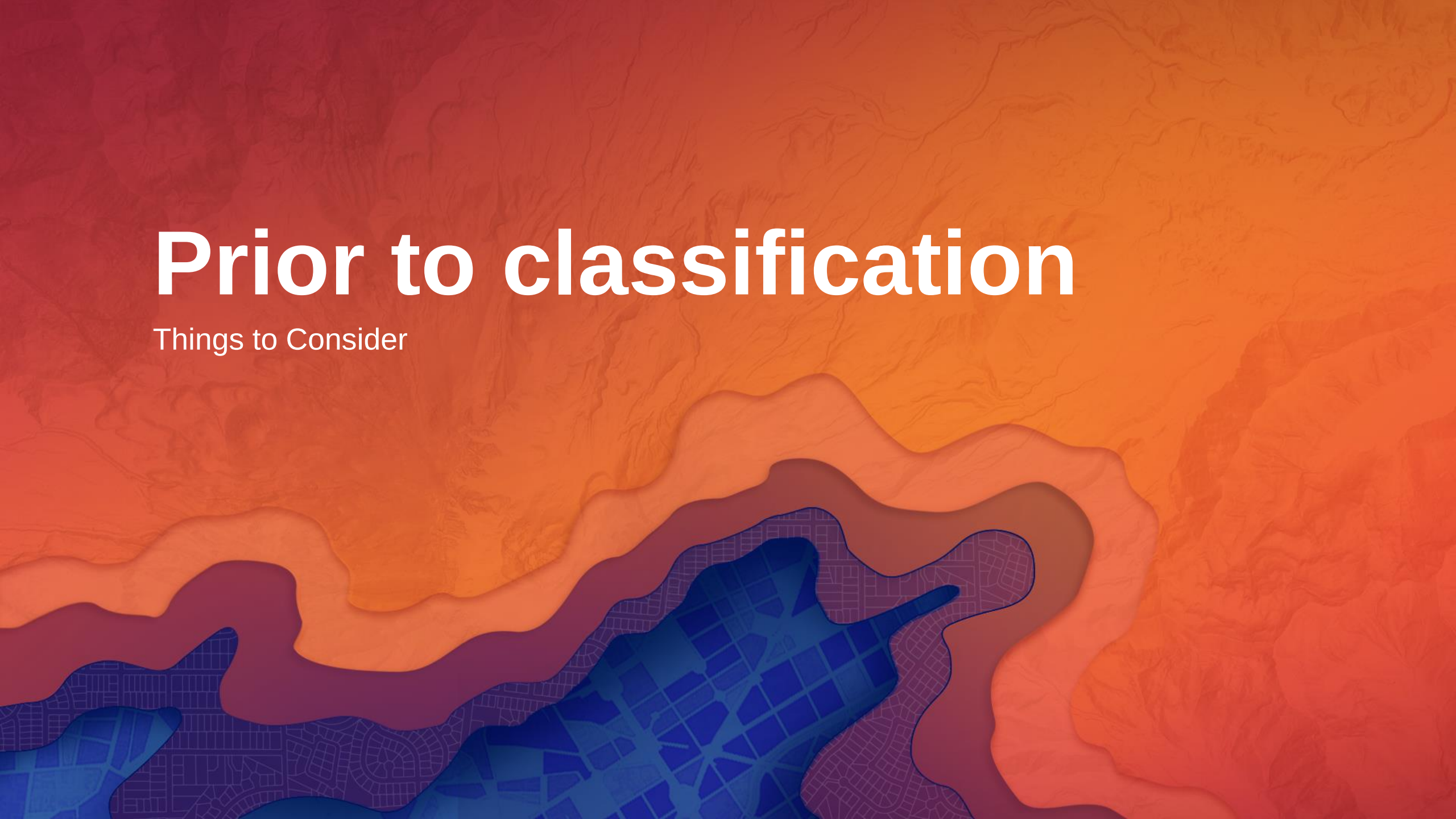
- Create Training Samples and Generate Classification Schema if desired
- Image Classification Wizard
  - Segment Mean Shift
  - Train Classifiers
  - Classify your Data
  - Merge Classes
- Do Accuracy Assessment





Prior to classification

Things to Consider



Unsupervised vs supervised

Leave the kids home alone….

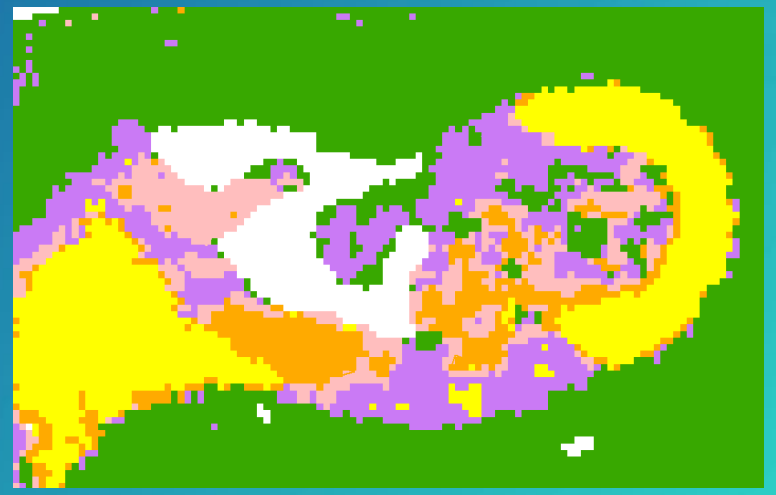
- **Unsupervised**
  - Classification is based on the software analysis of an image without the user providing sample classes
  - Users decide on the number of classes and number of iterations used for the classification.
  - Easier, but less accurate

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

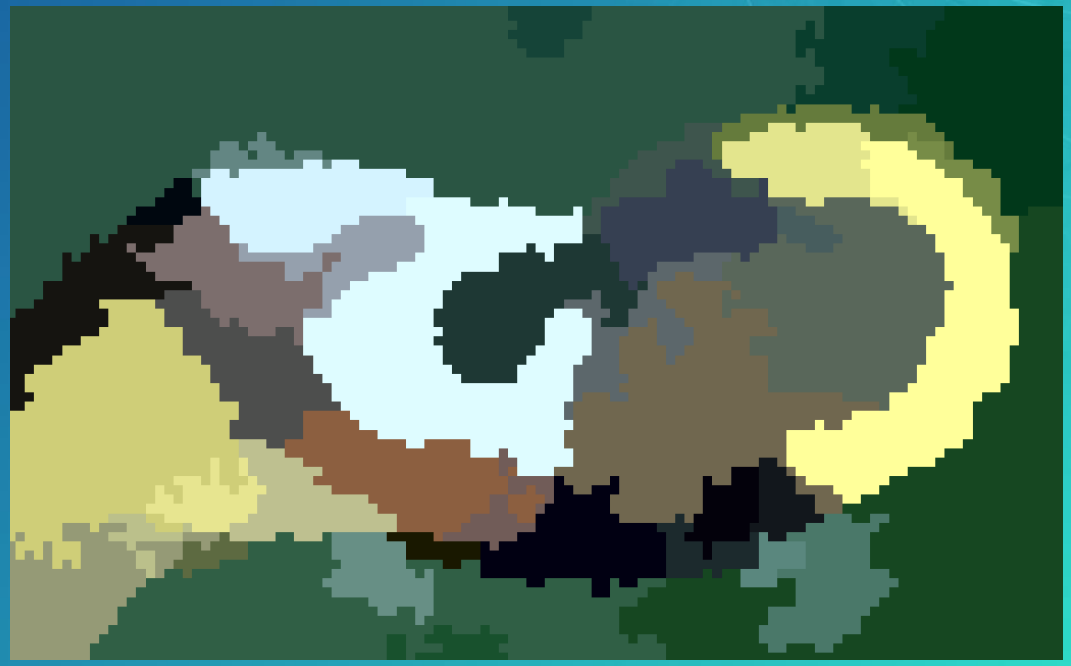
ID	Class Name	Value	Color	Count
1	Yellow	1		72
2	Pink	13		11
3	White	16		48
4	Orange	30		18
5	Purple	36		10
6	Grass	40		108



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?

Level I	Level II
1 Urban or Built-up Land	11 Residential
	12 Commercial and Services
	13 Industrial
	14 Transportation, Communications, Utilities
	15 Industrial and Commercial Complexes
	16 Mixed Urban or Built-up Land
	17 Other Urban or Built-up Land
2 Agricultural Land	21 Cropland and Pasture
	22 Orchards, Groves, Vineyards, Nurseries, and Ornamental Horticultureal Areas
	23 Confined Feeding Operations
	24 Other Agricultural Land
3 Rangeland	31 Herbacious Rangeland
	32 Shrub and Brush Rangeleand
	33 Mixed Rangeland

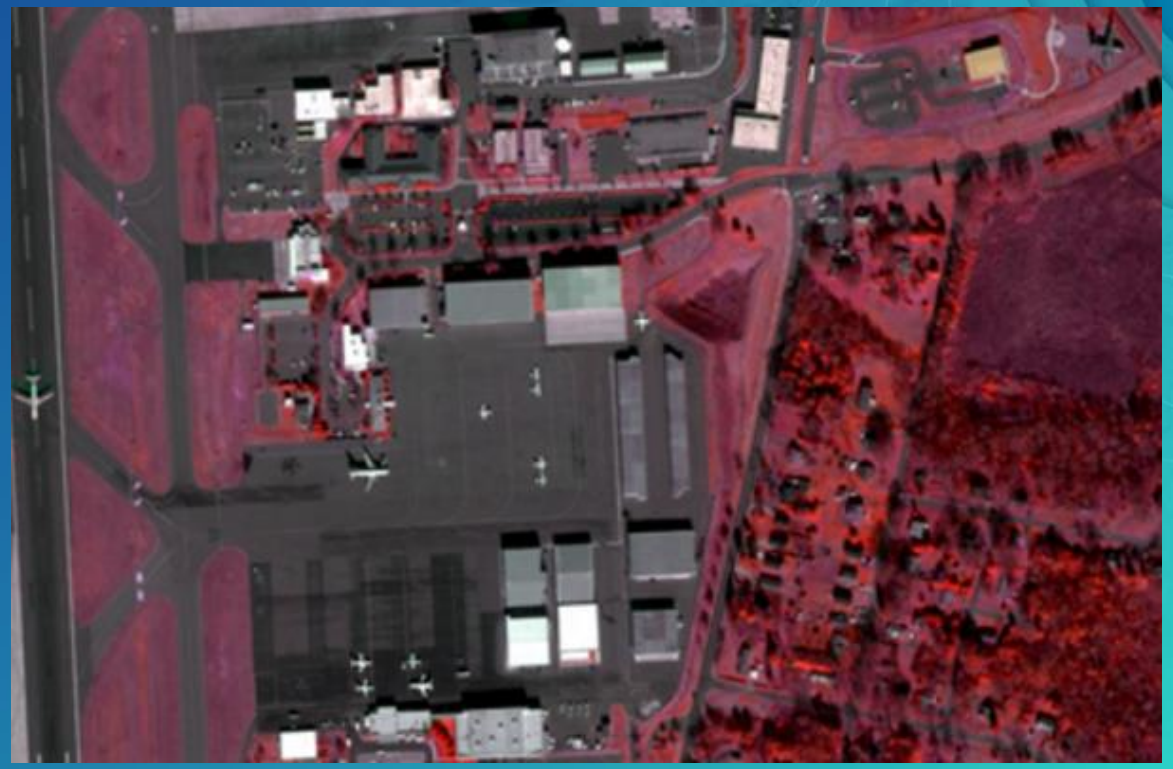
How to Train Your Data

Let’s think like machines!

How Many Classes Do You See?

I spy with my little eye…..

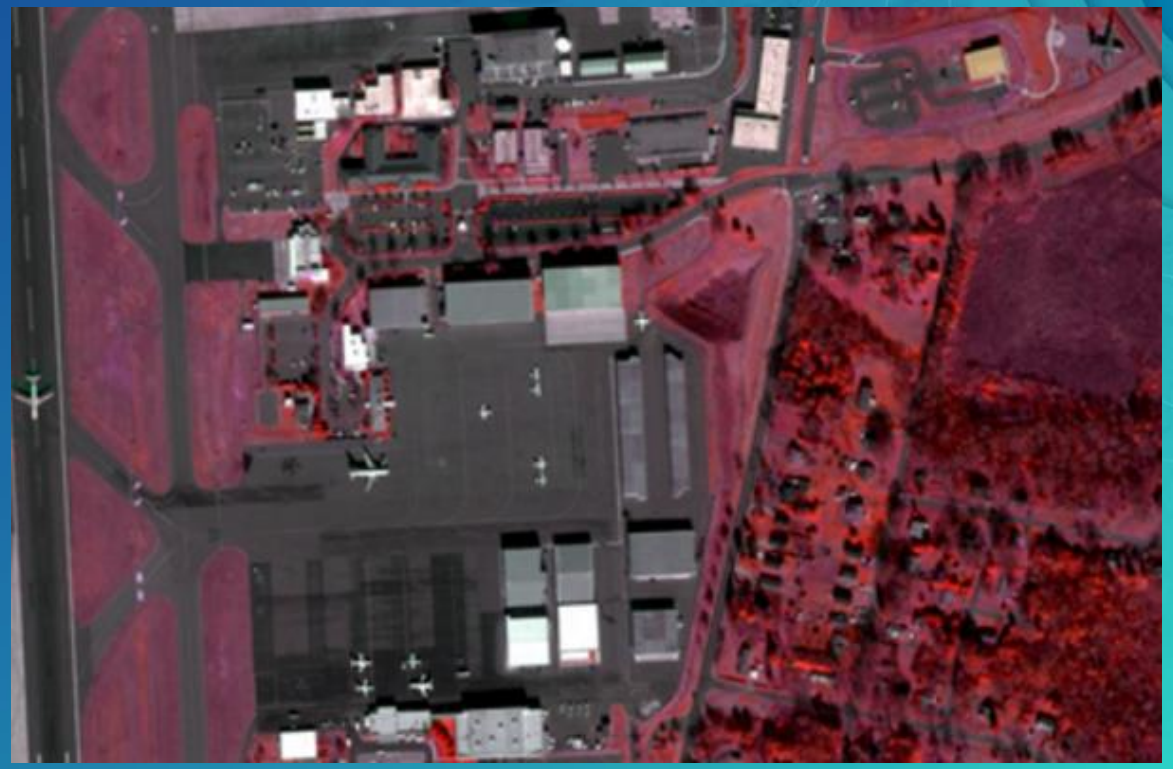
- Asphalt
- Buildings
- Planes
- Grass
- Trees



How to Teach an Algorithm to ‘See’

What Makes a Tree a Tree?

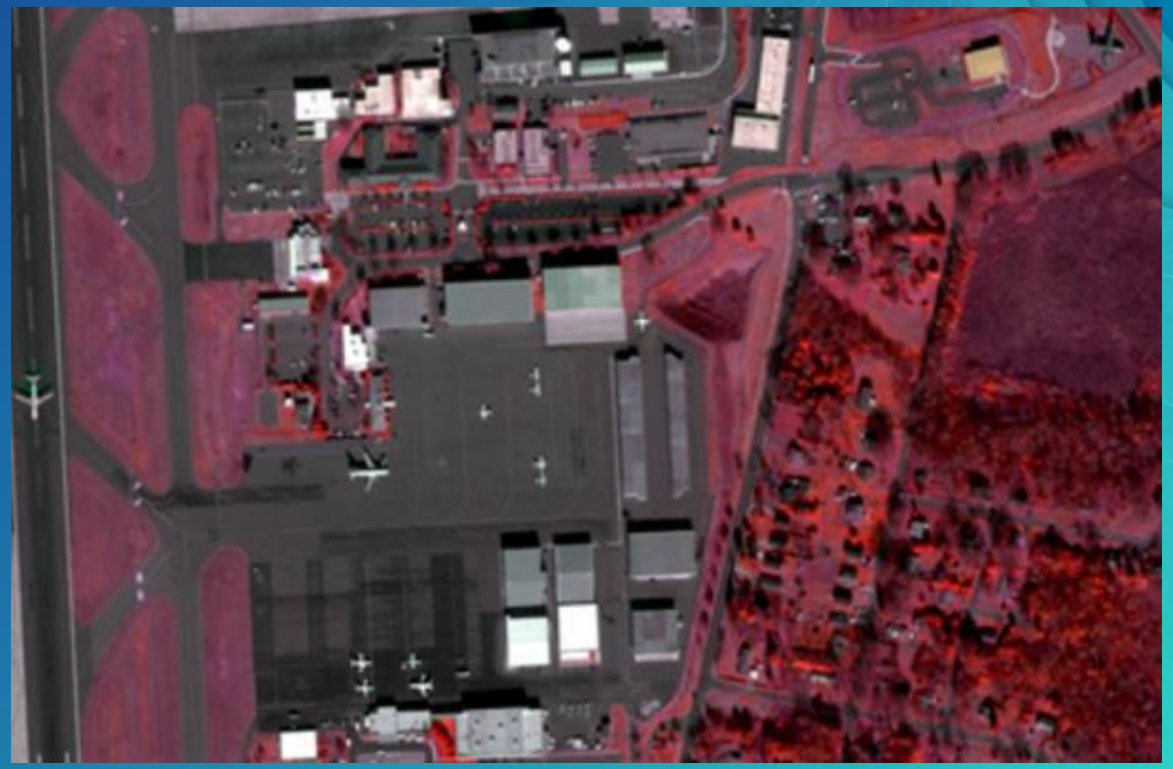
- How do humans identify classes in imagery?
  - Color
  - Size
  - Shape
  - Texture
- How do we teach an algorithm to identify classes?
  - Look at image as a machine does
    - Pixel by Pixel
    - Band by Band
    - Reflectance



Translate our Classes to a Computer's

01000101 01010011 01010010 01001001

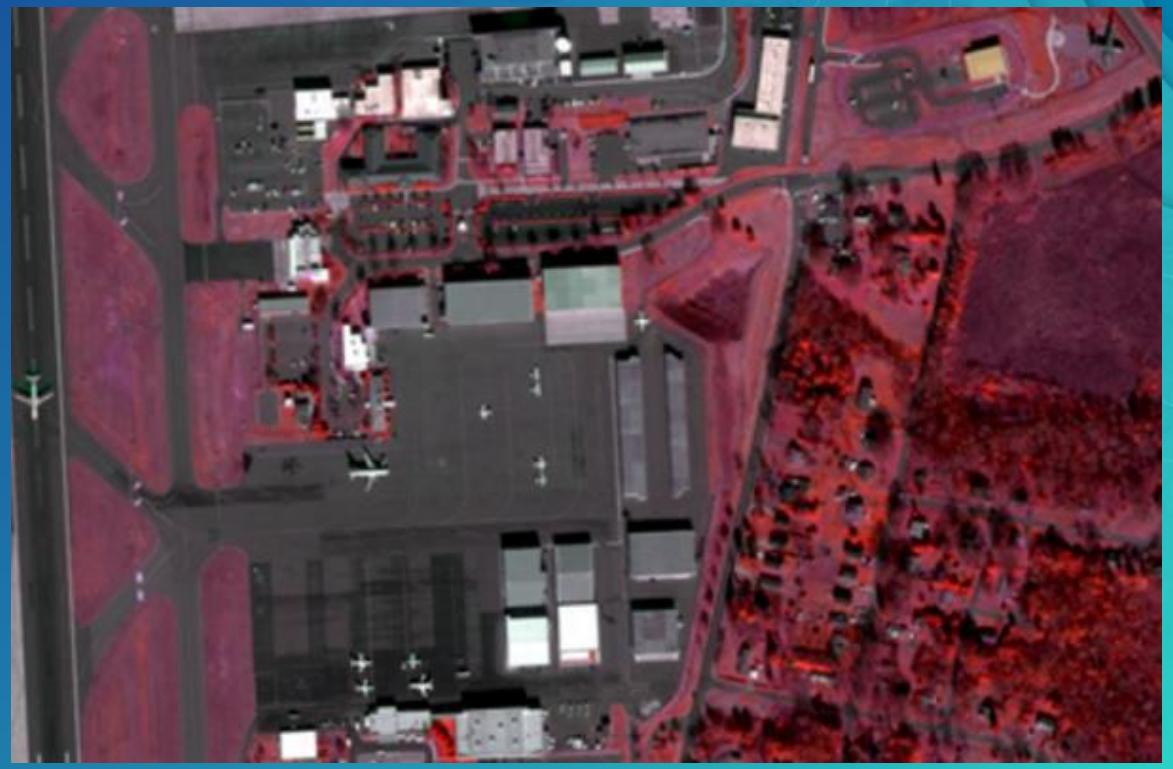
- Asphalt
- Buildings
- Planes
- Grass
- Trees



Translate our Classes to a Computer's

01000001 01110010 01100011 01000111 01001001 01010011

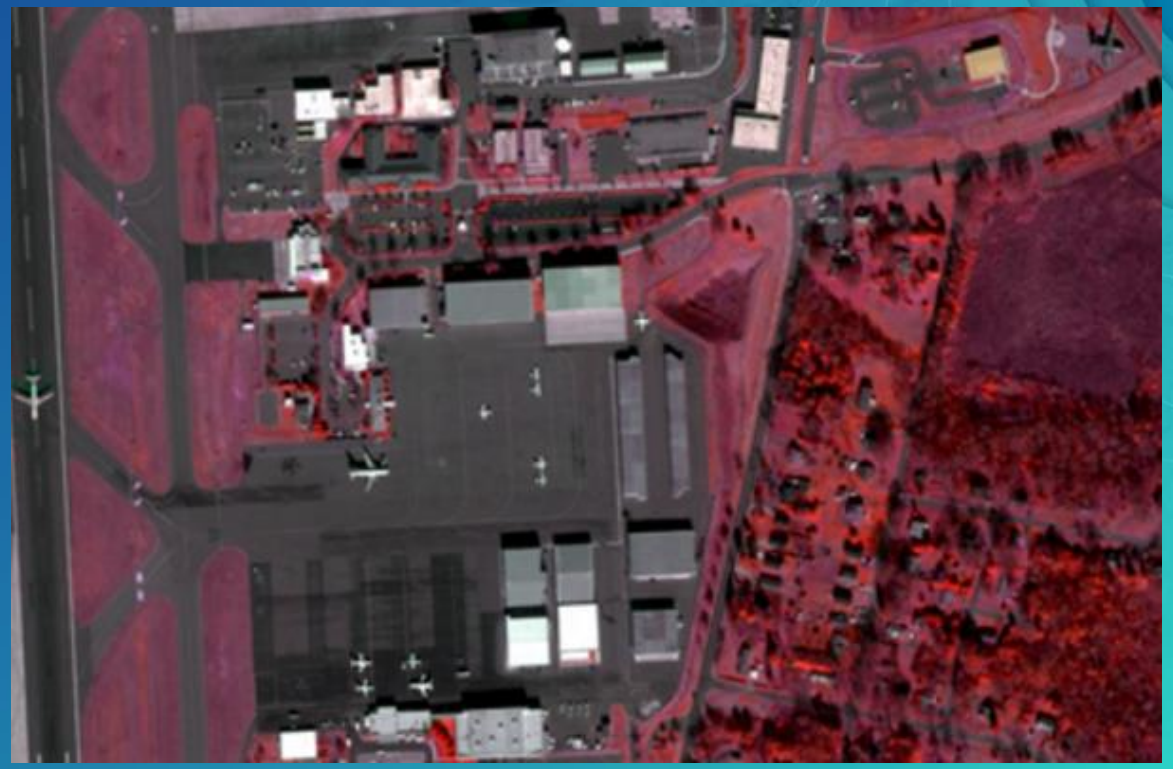
- Asphalt
- Buildings
  - Grey Roofs
  - White Roofs
- Planes
- Grass
- Trees
- Shadows

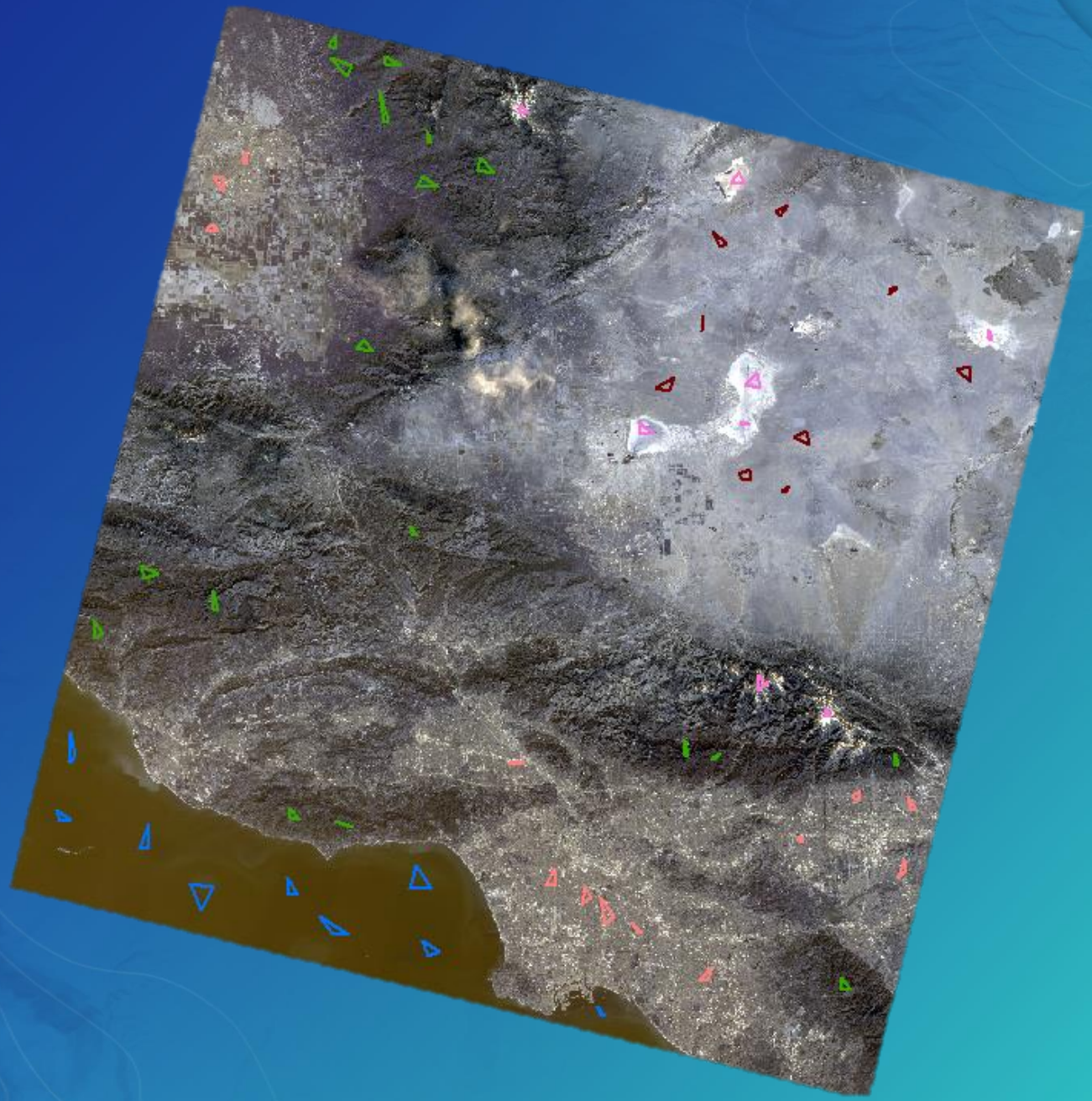


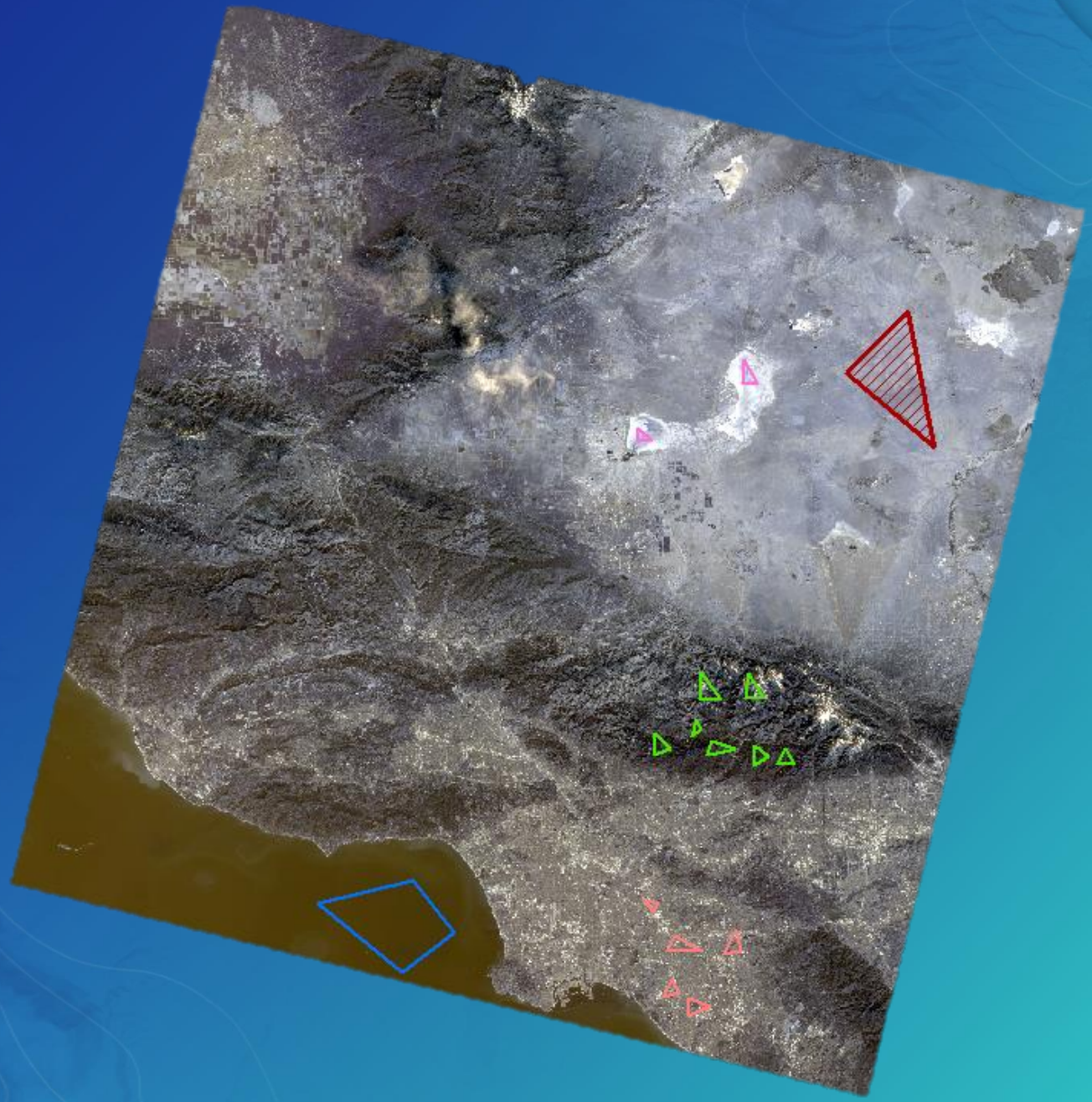
How to train your data

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

$$\frac{}{A \vdash A}\ \text{init} \qquad \frac{\Gamma_1 \vdash A, \Sigma_1 \quad \Gamma_2, A \vdash \Sigma_2}{\Gamma_1, \Gamma_2 \vdash \Sigma_1, \Sigma_2}\ \text{cut}$$

$$\frac{\Gamma \vdash \Sigma}{\Gamma, !A \vdash \Sigma}\ \text{weak}_L \qquad \frac{\Gamma \vdash \Sigma}{\Gamma \vdash ?A, \Sigma}\ \text{weak}_R \qquad \frac{\Gamma, !A, !A \vdash \Sigma}{\Gamma, !A \vdash \Sigma}\ \text{contr}_L \qquad \frac{\Gamma \vdash ?A, ?A, \Sigma}{\Gamma \vdash ?A, \Sigma}\ \text{contr}_R$$

$$\frac{\Gamma, A, B \vdash \Sigma}{\Gamma, A \otimes B \vdash \Sigma}\ \otimes_L \qquad \frac{\Gamma_1 \vdash A, \Sigma_1 \quad \Gamma_2 \vdash B, \Sigma_2}{\Gamma_1, \Gamma_2 \vdash A \otimes B, \Sigma_1, \Sigma_2}\ \otimes_R \qquad \frac{\Gamma \vdash \Sigma}{\Gamma, \mathbf{1} \vdash \Sigma}\ \mathbf{1}_L \qquad \frac{}{\vdash \mathbf{1}}\ \mathbf{1}_R$$

$$\frac{\Gamma, A \vdash \Sigma}{\Gamma, A \,\&\, B \vdash \Sigma}\ \&_{L1} \qquad \frac{\Gamma, B \vdash \Sigma}{\Gamma, A \,\&\, B \vdash \Sigma}\ \&_{L2} \qquad \frac{\Gamma \vdash A, \Sigma \quad \Gamma \vdash B, \Sigma}{\Gamma \vdash A \,\&\, B, \Sigma}\ \&_R \qquad \text{no } \top_L \qquad \frac{}{\Gamma \vdash \top, \Sigma}\ \top_R$$

$$\frac{\Gamma_1, A \vdash \Sigma_1 \quad \Gamma_2, B \vdash \Sigma_2}{\Gamma_1, \Gamma_2, A \parr B \vdash \Sigma_1, \Sigma_2}\ \parr_L \qquad \frac{\Gamma \vdash A, B, \Sigma}{\Gamma \vdash A \parr B, \Sigma}\ \parr_R \qquad \frac{}{\bot \vdash}\ \bot_L \qquad \frac{\Gamma \vdash \Sigma}{\Gamma \vdash \bot, \Sigma}\ \bot_R$$

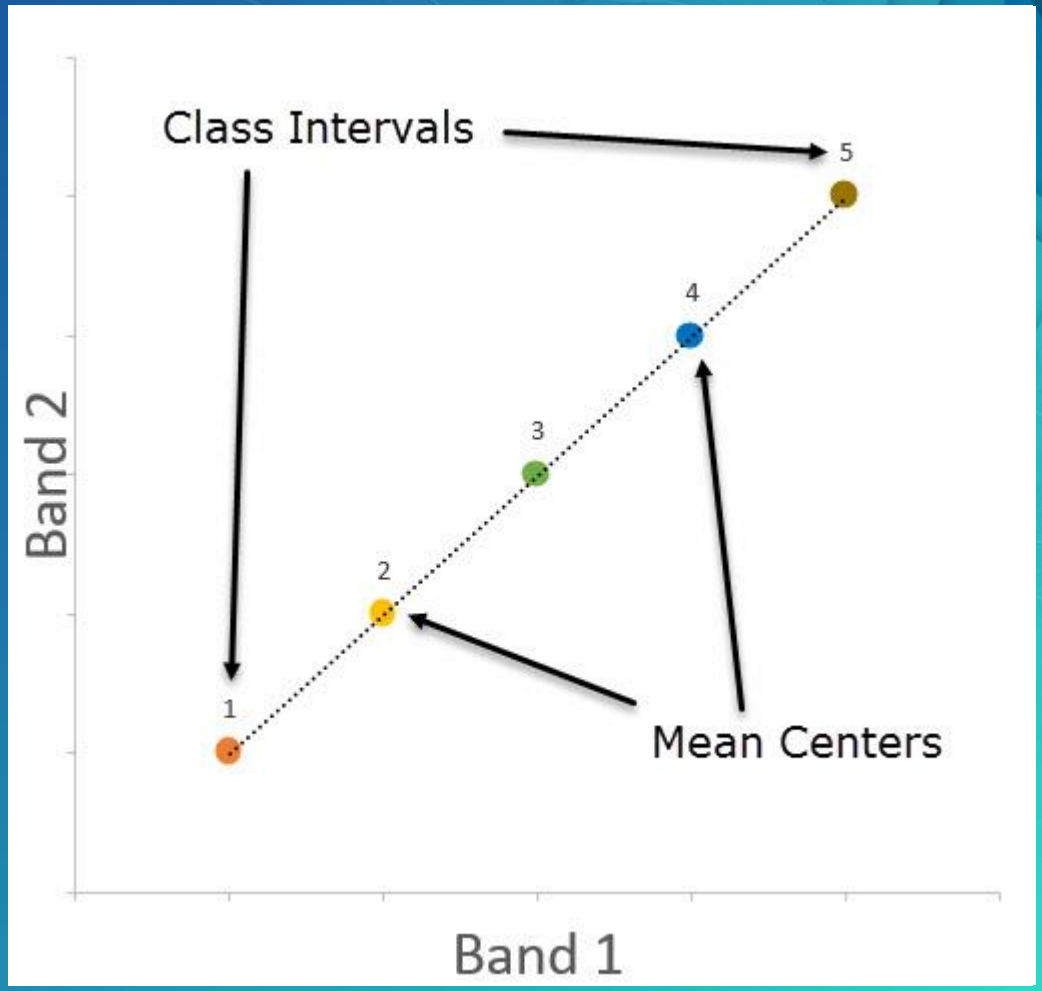
$$\frac{\Gamma, A \vdash \Sigma \quad \Gamma, B \vdash \Sigma}{\Gamma, A \oplus B \vdash \Sigma}\ \oplus_L \qquad \frac{\Gamma \vdash A, \Sigma}{\Gamma \vdash A \oplus B, \Sigma}\ \oplus_{R1} \qquad \frac{\Gamma \vdash B, \Sigma}{\Gamma \vdash A \oplus B, \Sigma}\ \oplus_{R2} \qquad \frac{}{\Gamma, \mathbf{0} \vdash \Sigma}\ \mathbf{0}_L \qquad \text{no } \mathbf{0}_R$$

$$\frac{\Gamma, A \vdash \Sigma}{\Gamma, !A \vdash \Sigma}\ !_L \qquad \frac{!\Gamma \vdash A, ?\Sigma}{!\Gamma \vdash !A, ?\Sigma}\ !_R \qquad \frac{!\Gamma, A \vdash ?\Sigma}{!\Gamma, ?A \vdash ?\Sigma}\ ?_L \qquad \frac{\Gamma \vdash A, \Sigma}{\Gamma \vdash ?A, \Sigma}\ ?_R$$

ISO Cluster

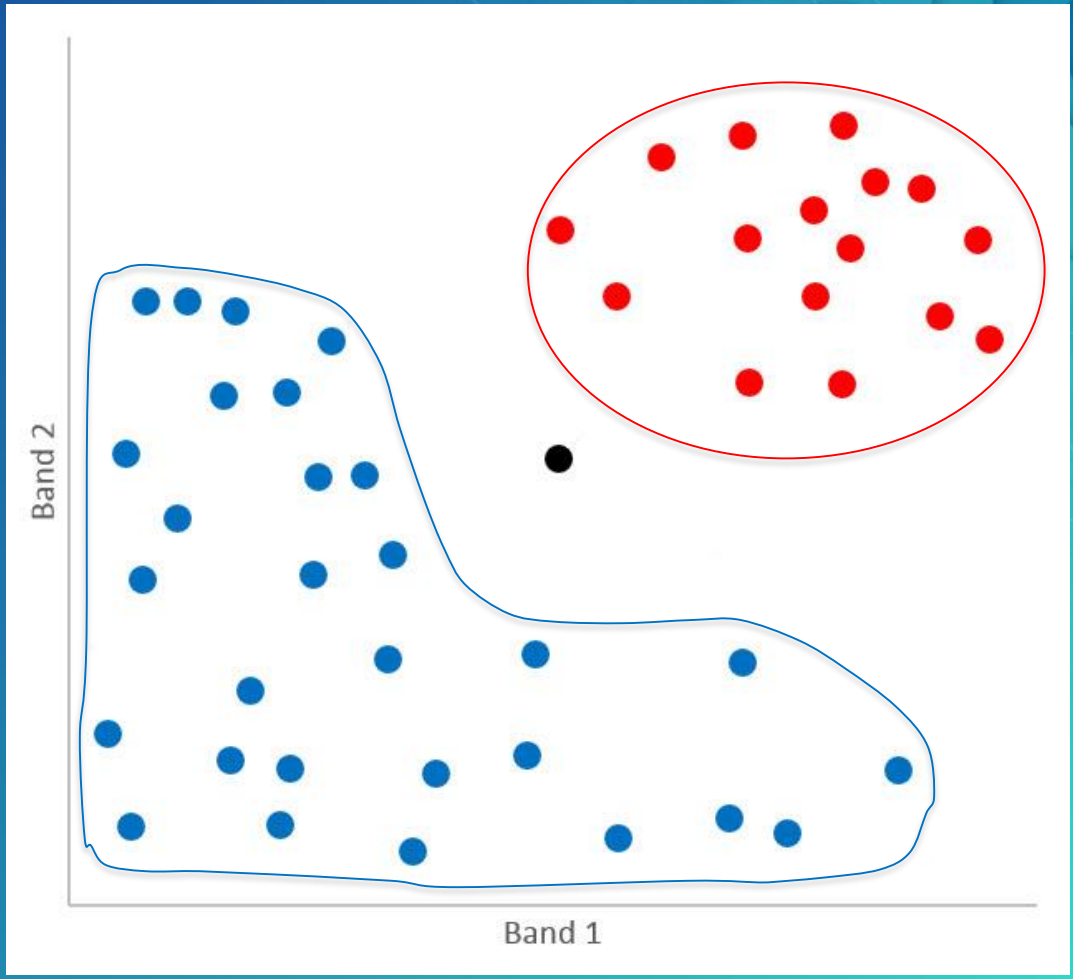
I SO Love Clusters

- Unsupervised
  - Iterative Self Organizing
  - Initially, means are placed on a 45 degree line 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



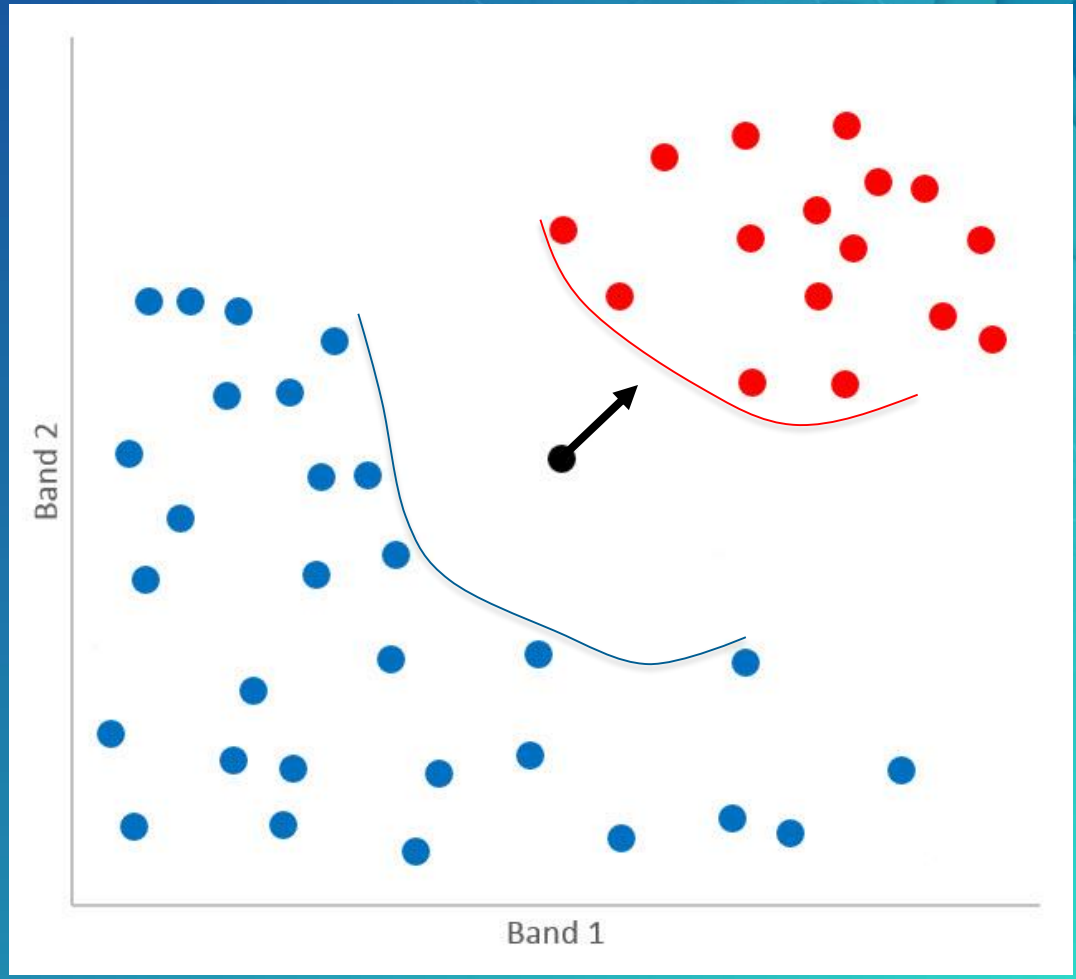
Maximum Likelihood

- 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

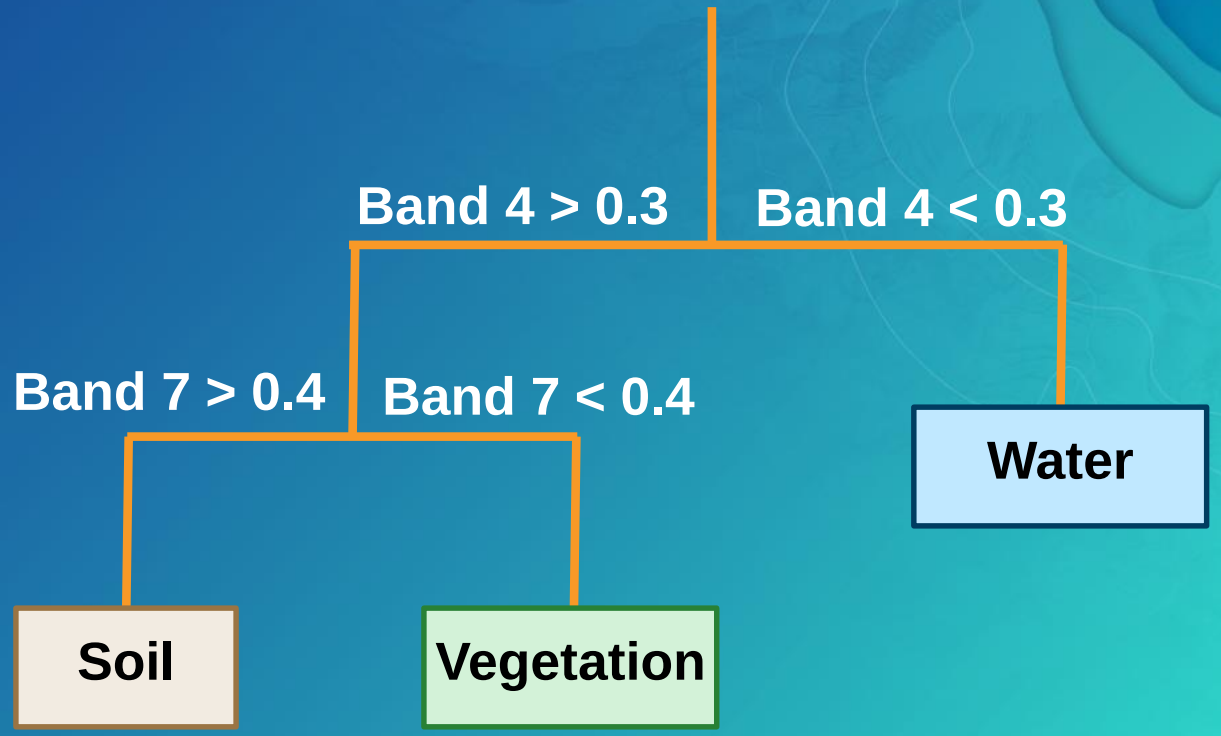
- Supervised
  - Similar to Maximum Likelihood
  - Places pixels in multiband space
  - Instead of determining the likelihood of a pixel belonging a class, pixels are assigned to classes so that the gaps between classes are as large as possible



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



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

		Truth					
		Asphalt	Concrete	Grass	Tree	Building	Total
Predicted	Asphalt	2385	4	0	1	4	2394
	Concrete	0	332	0	0	1	333
	Grass	0	1	908	8	0	917
	Tree	0	0	0	1084	9	1093
	Building	12	0	0	6	2053	2071
	Total	2397	337	908	1099	2067	6808

Confusion Matrix

Will you take the Red Cell or the Blue Cell?

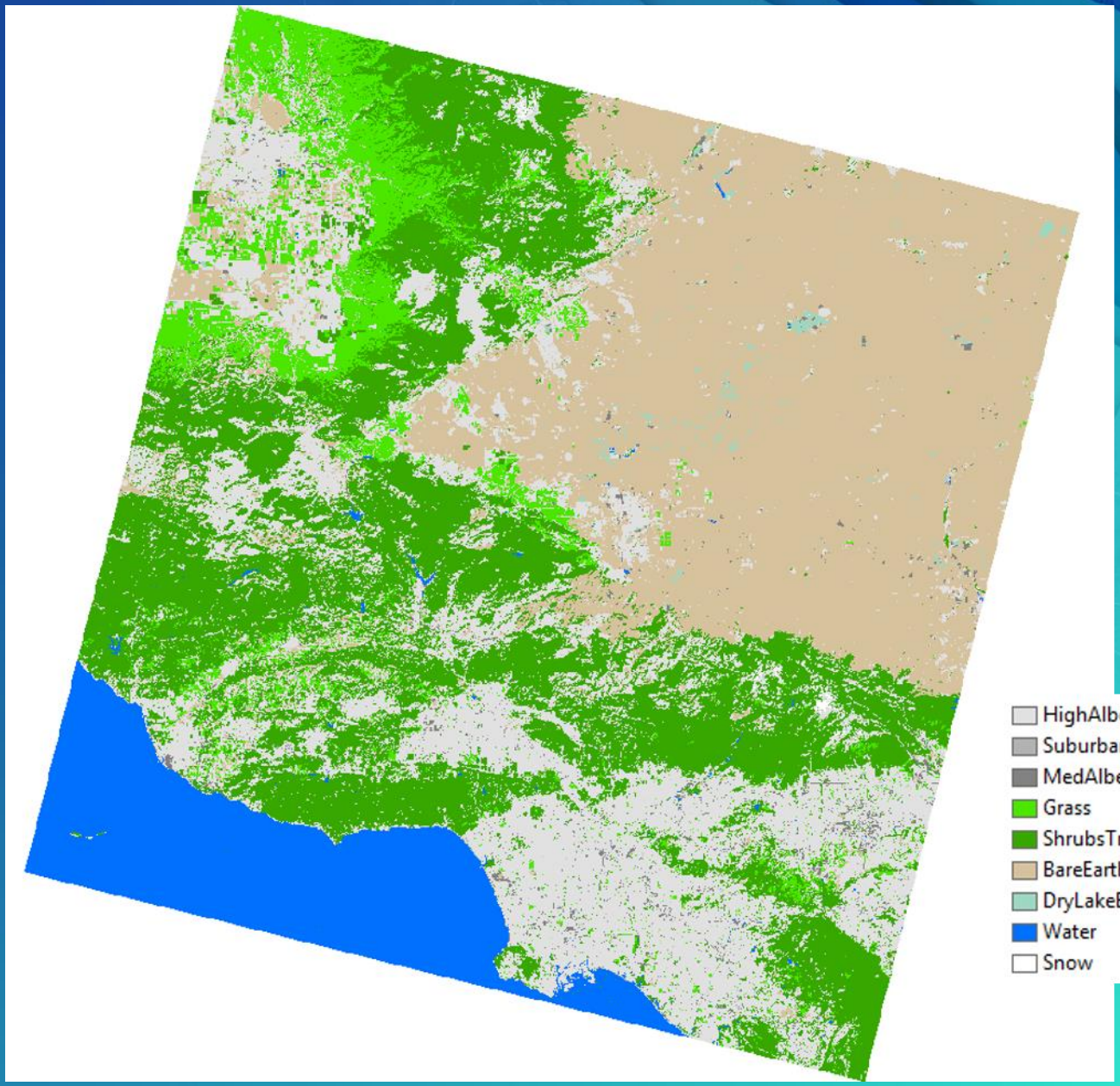
- Generates a Kappa Index of Agreement between classified raster and ground truth data
- Index is based on how well the Classified Raster reflects the Ground Truth Points
- Kappa Index is expressed as a value between 0 and 1
  - The closer to 1 the value is, the more accurate the reclassification was
- 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?

THIS…

Becomes This…




Questions?


esri
THE
SCIENCE
OF
WHERE