Introduction to Image Classification

IDEA MAP SUDAN

Adapted from presentation prepared by
Charlotte Flasse
Université Libre de Bruxelles

overview

- Introduction
- Supervised classification
- Classification algorithms
- Accuracy assessment

IDEAMAP SUDAN

- Introduction

Image classification process

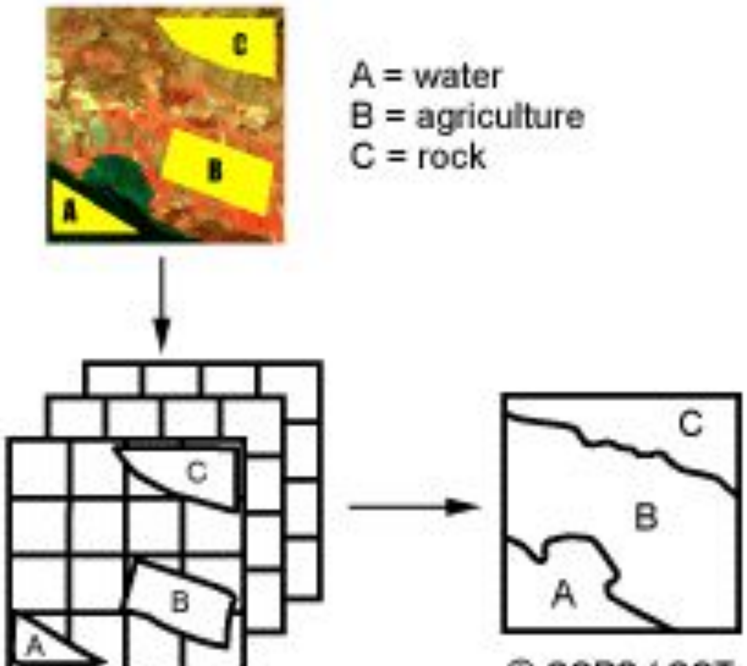
1. Select images
2. Define clusters in feature space
   - ❑ Unsupervised e.g. ISODATA, k-means
   - ❑ **Supervised** e.g. rule-based / **provide training data**
3. Select classification algorithm
   - ❑ Box classifier
   - ❑ **Minimum distance to means**
   - ❑ **Machine learning (Maximum likelihood, RF, SVM)**
   - ❑ **Deep learning (CNN)**
4. Run classification
5. Validation of the result

- Supervised classification

  a. With training data

Supervised classification

- **User must provide training data (a priori knowledge)**
- Classification algorithm learns from training data to find patterns and translate this into classes

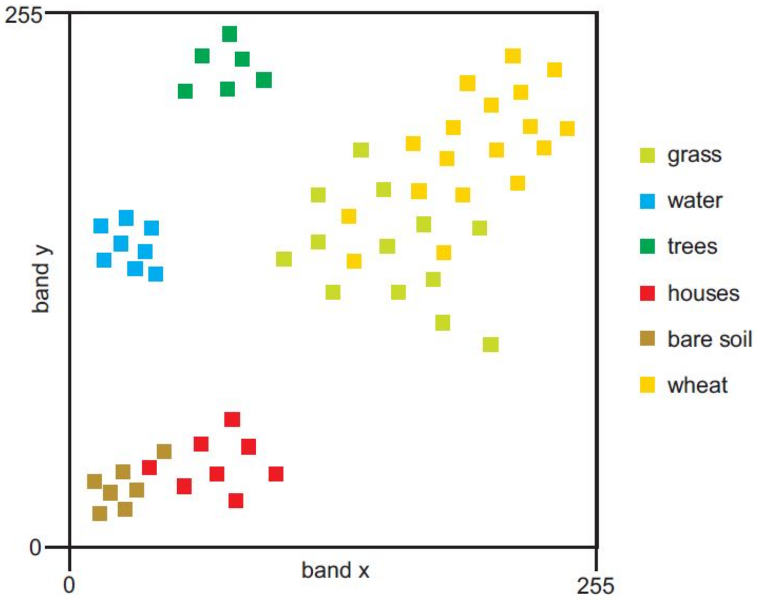


© CCRS / CCT

Provide training samples

Training samples should form clusters that:

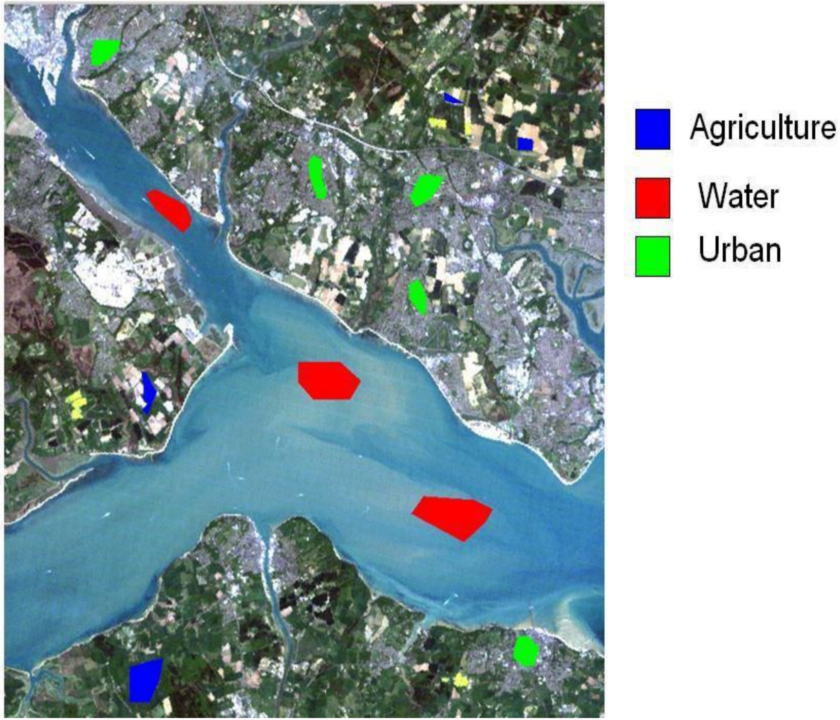
- Represent the variability within each class
- Have enough samples per class
- Do not overlap with other clusters



© Tempfli et al

Provide training samples

- Obtained directly in the field, or independent data, or interpreting an image.
- Taken over small zones of the image
- Training sample quality impacts classification



© University of Southampton

Training samples

- Advantages:
  - Analyst controls information classes
    - Specific purpose; change detection
- Disadvantages
  - No “natural” spectral classes (imposed)
  - Spectral class may be heterogeneous
  - Training data may not be representative

Classification algorithms

IDEAMAP SUDAN

Classification algorithms

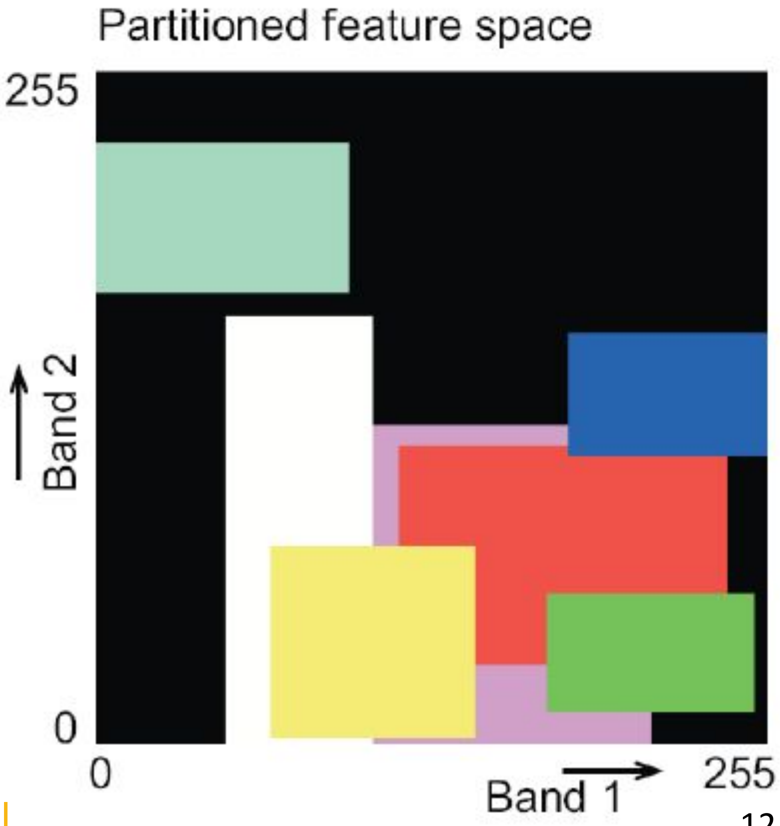
- Parametric (assumes normal distribution of data)
  - ❑ Minimum distance to means (MDM)
  - ❑ Maximum likelihood
- Non-parametric (doesn’t assume normal distribution)
  - ❑ Box classifier (parallelepiped)
  - ❑ Random forest (RF)
  - ❑ Support Vector Model (SVM)
  - ❑ Artificial Neural Networks (ANN)

Box classifier

Also known as the parallelepiped classifier

Advantages

- Computationally efficient
- Mathematically simple



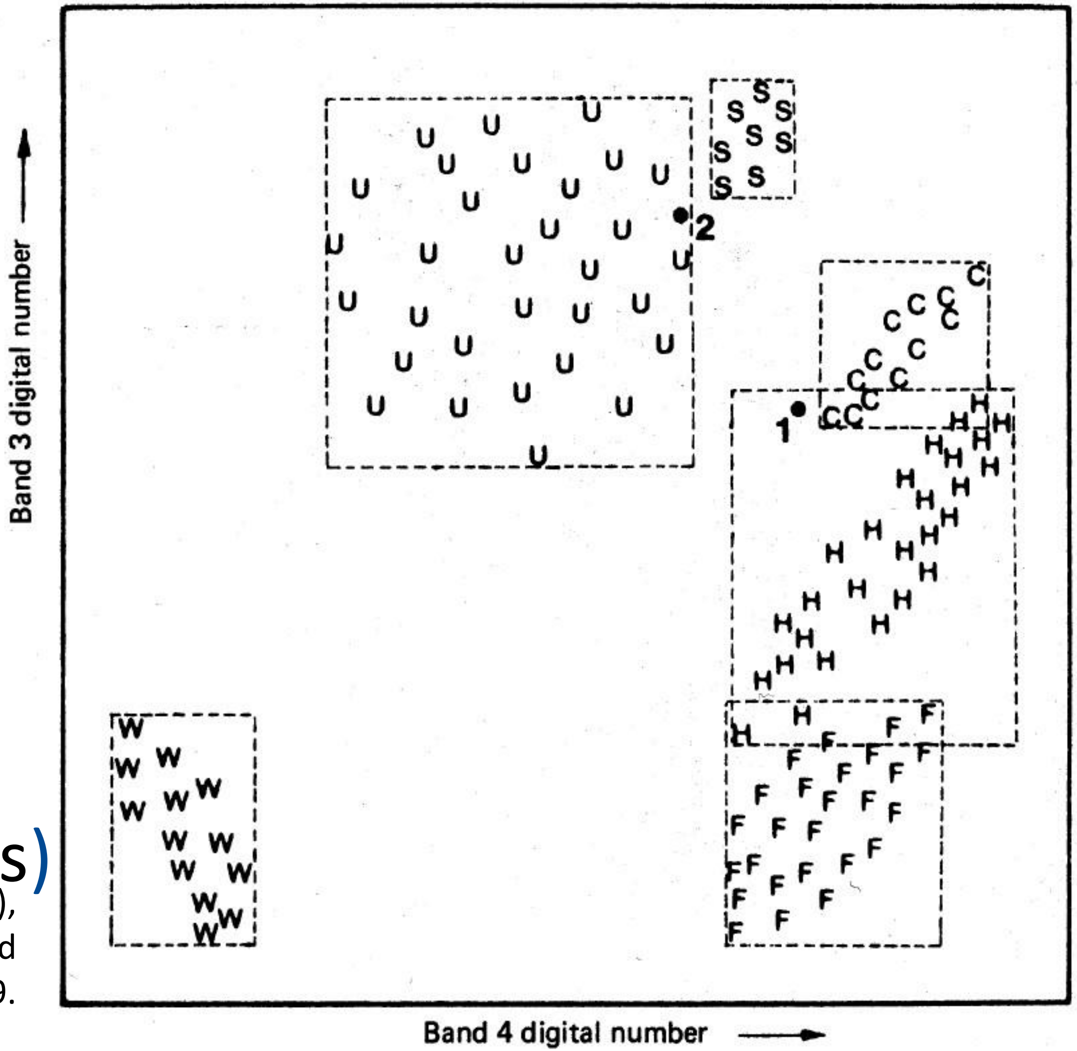
© Tempfli et al

Box

In which cluster would point 1 go?

Disadvantages

- Class overlap
- Insensitivity to covariance (shape & orientation of classes)



Source : Lillesand T., Kiefer R. W., Chipman J. (2015), Remote Sensing and Image Interpretation, Wiley and sons, 736p, ISBN: 978-1-118-34328-9.

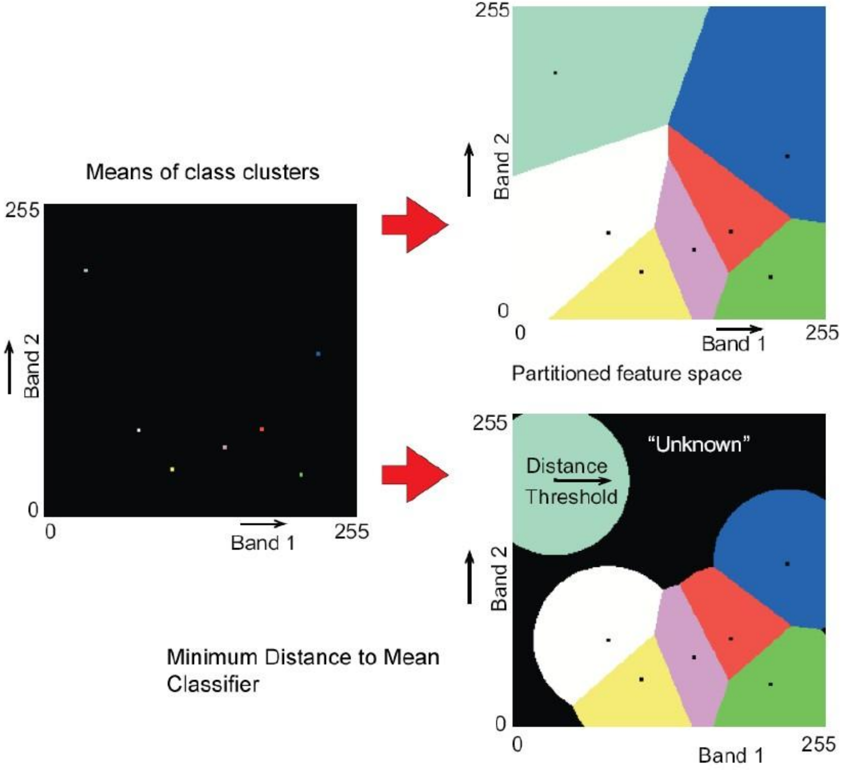
Minimum distance to means (MDM)

Advantages

- Computationally efficient
- Mathematically simple

Disadvantages

- Cluster mean may be far
- Insensitive to class variability



Minimum Distance to Mean Classifier

Without threshold distance

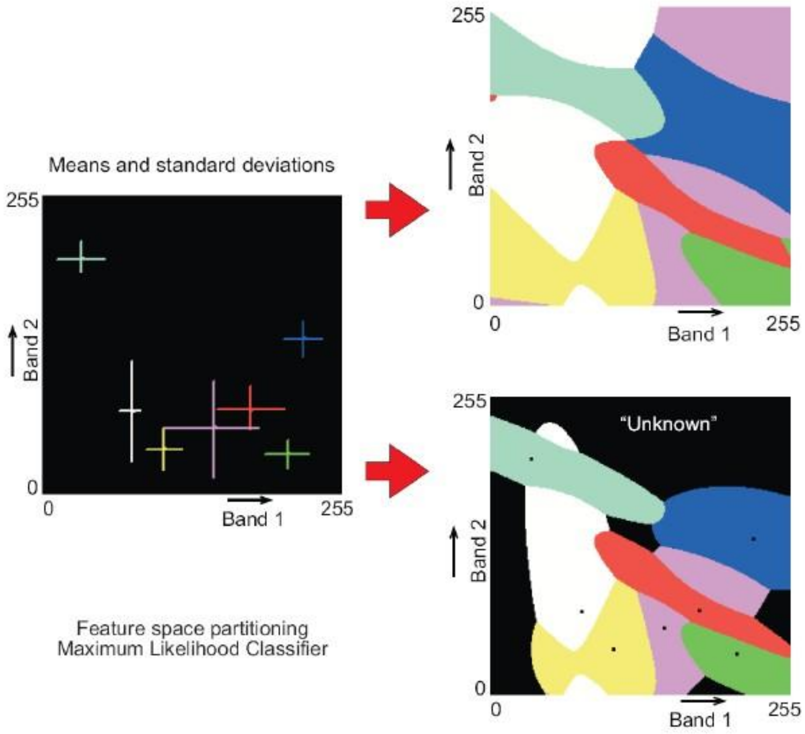
With threshold distance

© Tempfli et al

Maximum likelihood

- Sensitivity to variance and covariance:
  - Assumption: training data = Gaussian distribution
    - Generally reasonable for common spectral response distributions
  - Mean values and covariance matrix
  - Probability density function
- Probability of x to belong to each category
- Highest probability (most *likely* class) $\longrightarrow$ assigned to that category

Maximum likelihood


Means and standard deviations
255
0
Band 2
Band 1
0
255
Feature space partitioning
Maximum Likelihood Classifier
255
0
Band 2
Band 1
0
255
255
"Unknown"
0
Band 2
Band 1
0
255

Without threshold

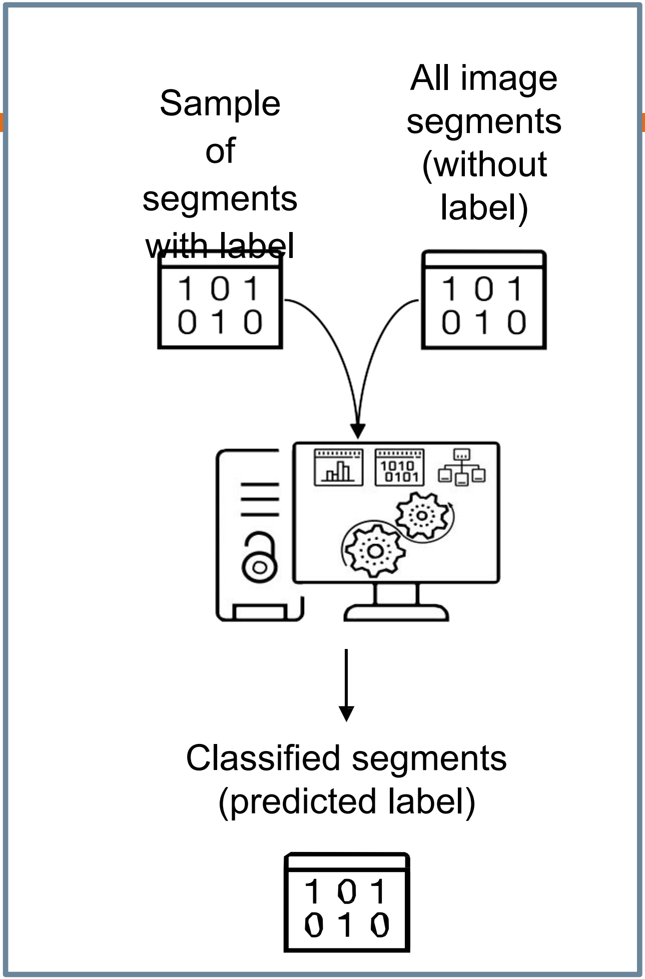
With threshold

© Tempfli et al

Other machine learning

Machine learning tools that are more robust with non-normal distributions:

- Random forest (RF)
- XGBoost (Extreme Gradient Boosting)
- Support vector machines (SVM)



Sample of segments with label

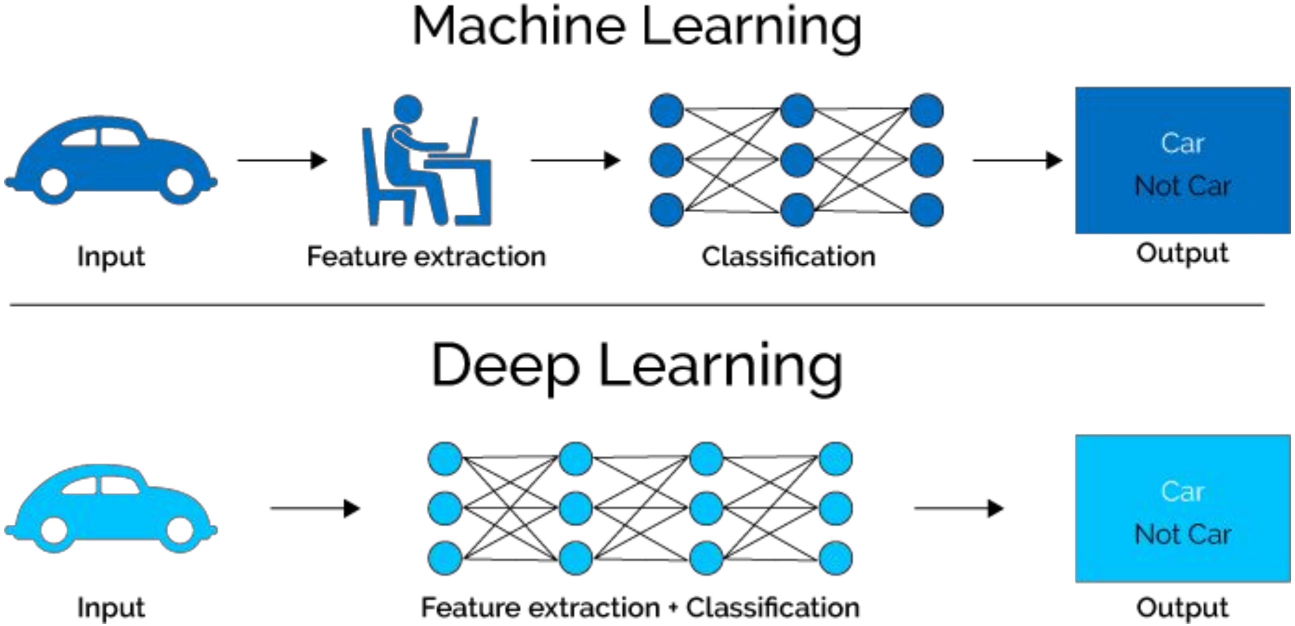
All image segments (without label)

Classified segments (predicted label)

© T. Grippa (ULB)

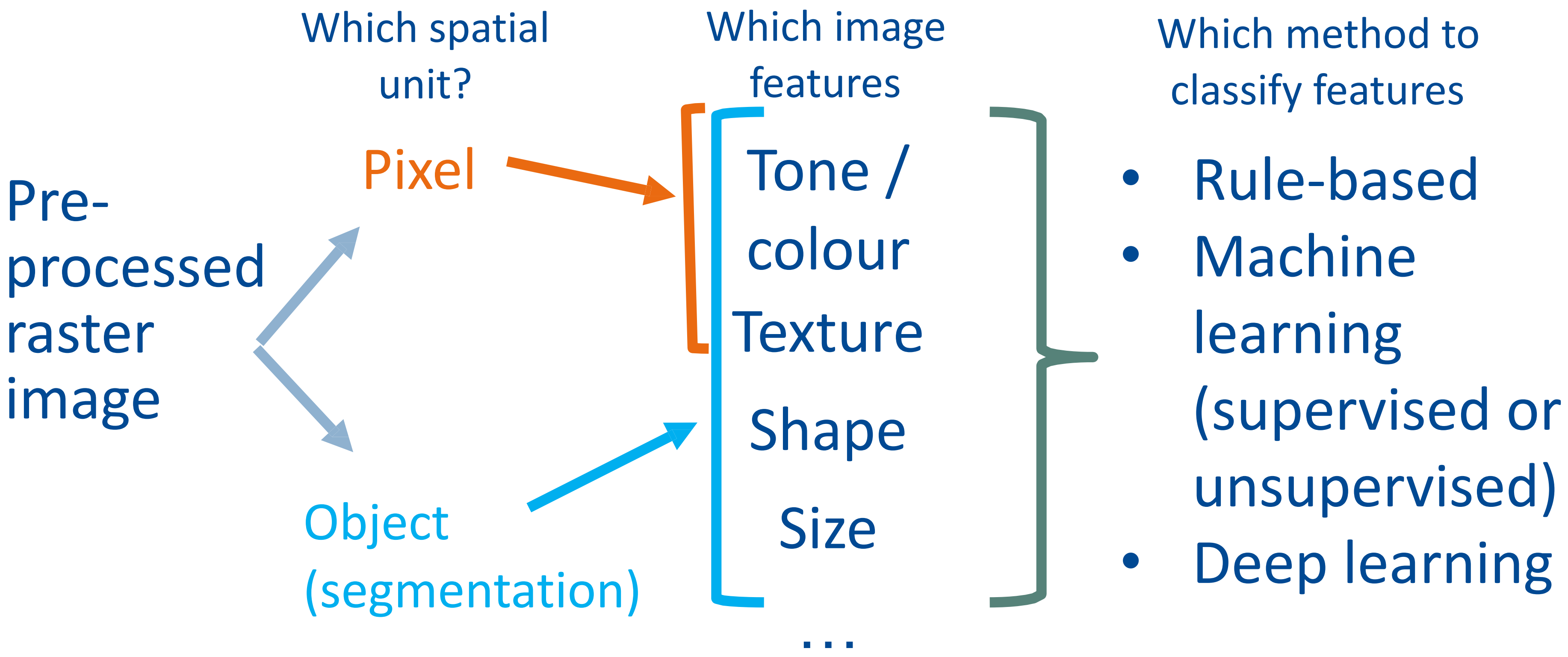
Machine learning vs. Deep learning

- Machine Learning >< Deep Learning



© Dey (2018) - TowardDatascience

Spatial unit <> Features <> Classification method



Pre-processed raster image

Which spatial unit?

Pixel

Object (segmentation)

Which image features

Tone / colour

Texture

Shape

Size

...

Which method to classify features

- Rule-based
- Machine learning (supervised or unsupervised)
- Deep learning

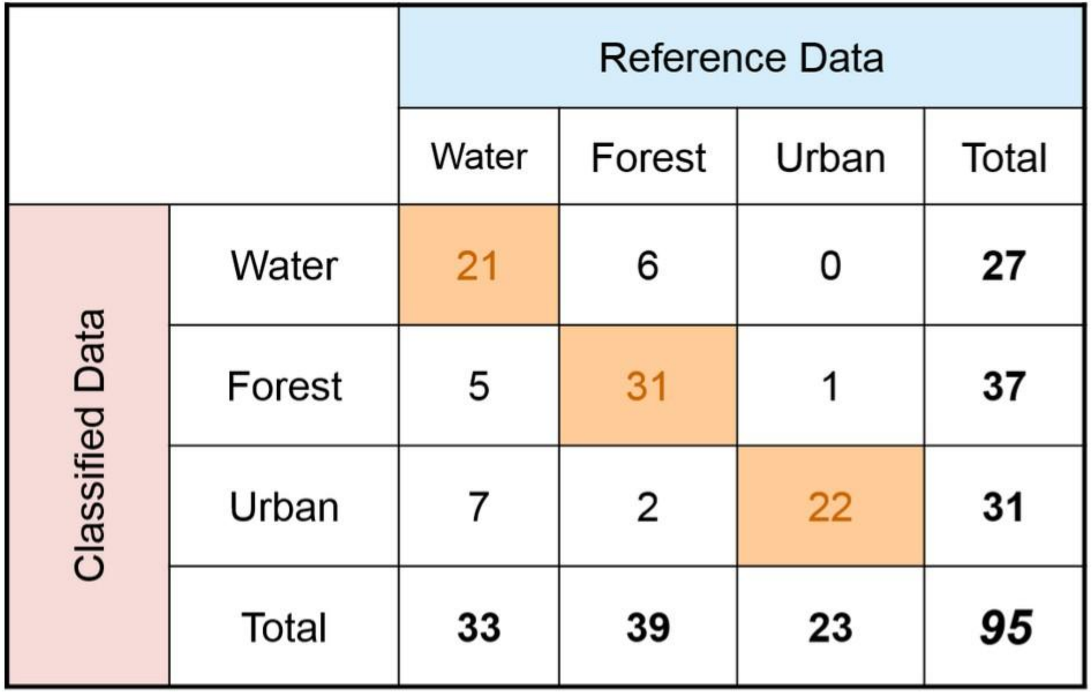
© PennState

Accuracy Assessment

Accuracy assessment

- Sources of errors
  - Mixels, pre-processing, classification, human
  - NOT distributed randomly
- **Essential !**
- Reference data
  - Field, high resolution imagery, …
  - As good as possible
  - Might be imperfect
- Comparison: classification vs. reference data
  - Accuracy assessment

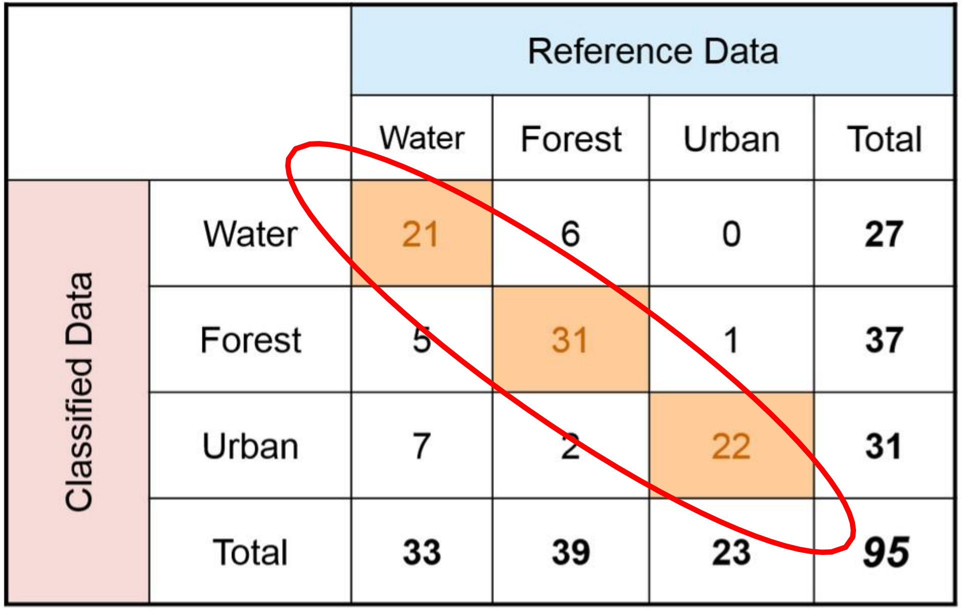
Error matrix



		Reference Data			
		Water	Forest	Urban	Total
Classified Data	Water	21	6	0	**27**
	Forest	5	31	1	**37**
	Urban	7	2	22	**31**
	Total	**33**	**39**	**23**	***95***

- 95 sample reference points in total
- Compare reference data with classified data for these samples

Error matrix: Overall accuracy (OA)

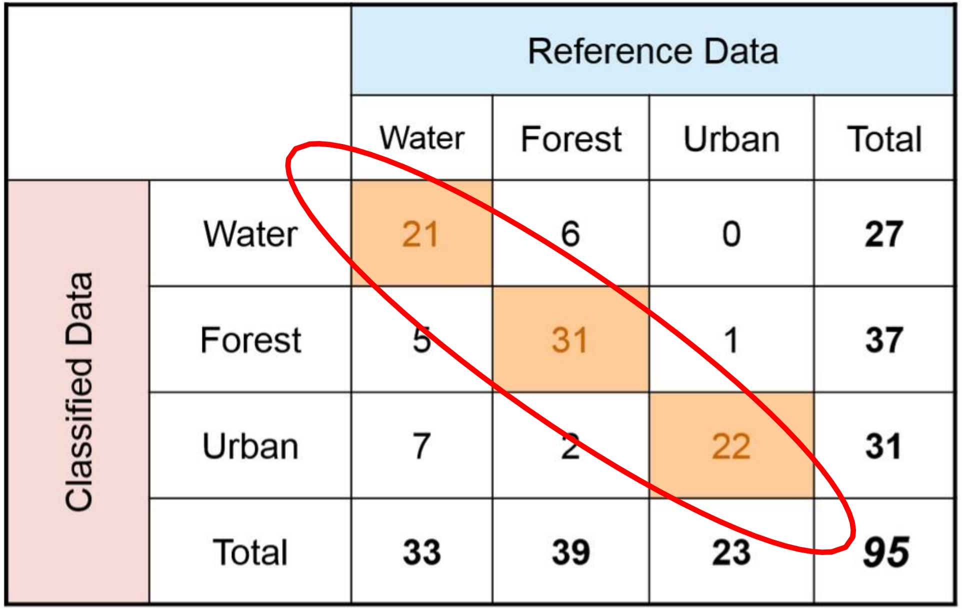


		Reference Data			
		Water	Forest	Urban	Total
Classified Data	Water	21	6	0	27
	Forest	5	31	1	37
	Urban	7	2	22	31
	Total	33	39	23	95

© Humboldt

- Diagonal = correctly classified samples
- OA = total of correctly classified pixels / total samples

Error matrix: Overall accuracy (OA)



		Reference Data			
		Water	Forest	Urban	Total
Classified Data	Water	21	6	0	**27**
	Forest	5	31	1	**37**
	Urban	7	2	22	**31**
	Total	**33**	**39**	**23**	***95***

© Humboldt

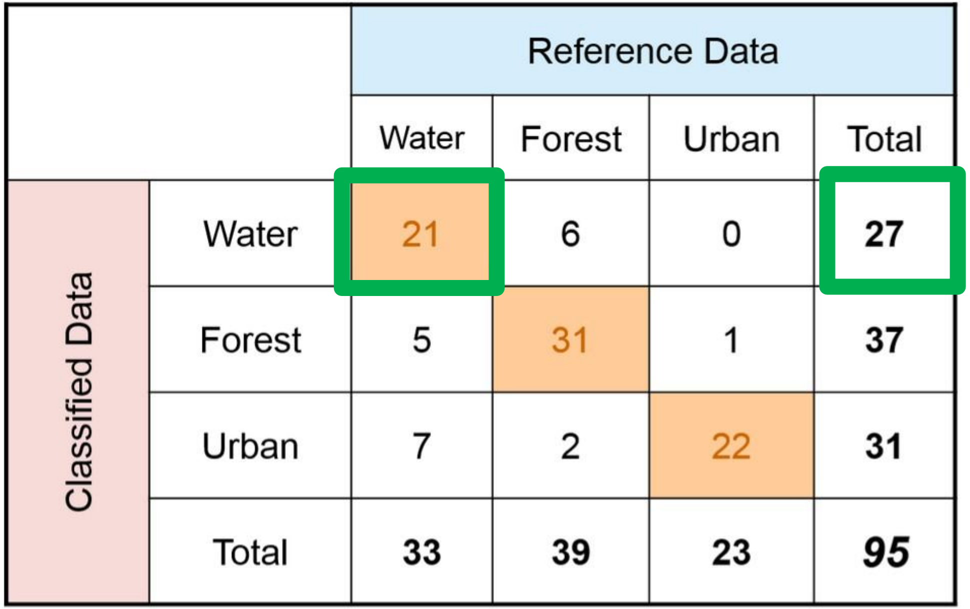
- Diagonal = correctly classified samples
- OA = total of correctly classified pixels / total samples

$$OA = \frac{21 + 31 + 22}{95} = 77.9\%$$

Error matrix

- Overall accuracy = GLOBAL accuracy measure
  BUT
  - Usually the errors in a classification are not random
  - Interesting to have per-class accuracy measures:
    - User's accuracy
      - If classified as grass in the image, how likely is it that it is really grass on the ground?
    - Producer's accuracy
      - If grass on ground, is it correctly classified as grass in the

User’s accuracy



		Reference Data			
		Water	Forest	Urban	Total
Classified Data	Water	21	6	0	**27**
	Forest	5	31	1	**37**
	Urban	7	2	22	**31**
	Total	**33**	**39**	**23**	***95***

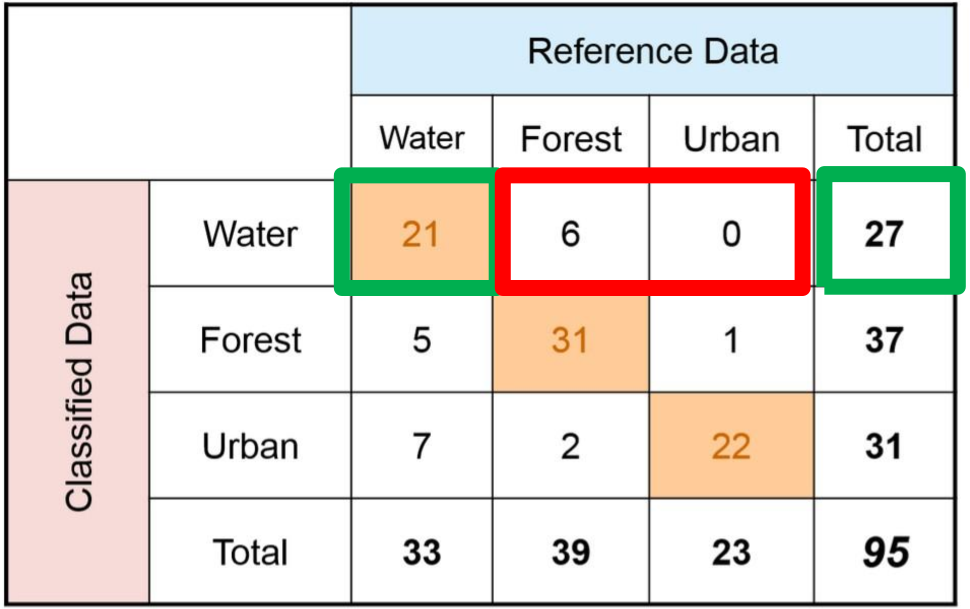
© Humboldt

For water class:

- UA = total correctly classified pixels / row total

User accuracy for Water class $= \dfrac{21}{27} = 78\%$

User’s accuracy + Commission error



		Reference Data			
		Water	Forest	Urban	Total
Classified Data	Water	21	6	0	**27**
	Forest	5	31	1	**37**
	Urban	7	2	22	**31**
	Total	**33**	**39**	**23**	***95***

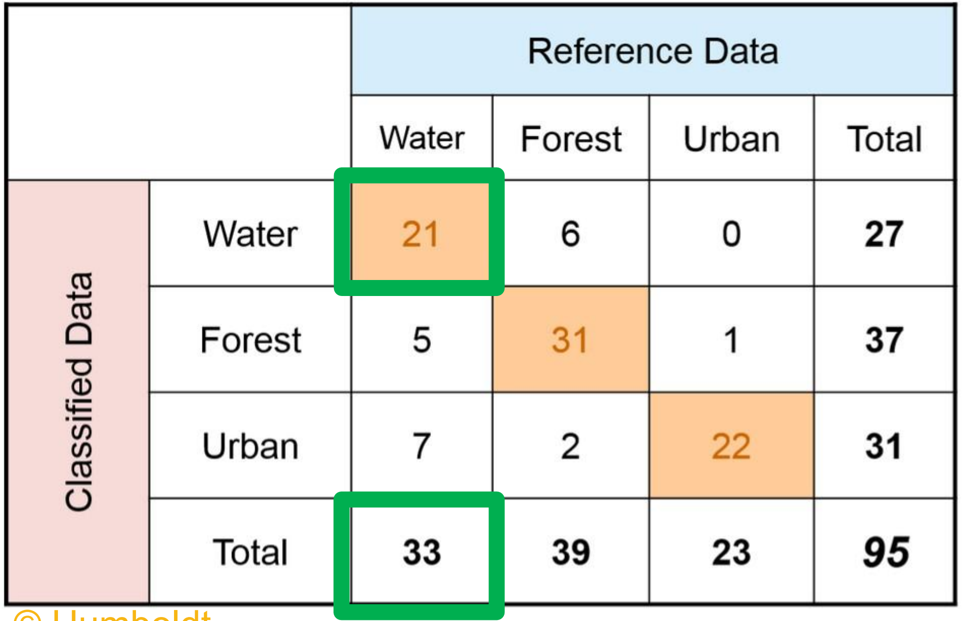
© Humboldt

For water class:

- Commission error = 100 – UA = 100 – 78 = 22%

User accuracy for Water class = $\frac{21}{27}$ = 78%

Producer’s accuracy



		Reference Data			
		Water	Forest	Urban	Total
Classified Data	Water	21	6	0	**27**
	Forest	5	31	1	**37**
	Urban	7	2	22	**31**
	Total	**33**	**39**	**23**	***95***

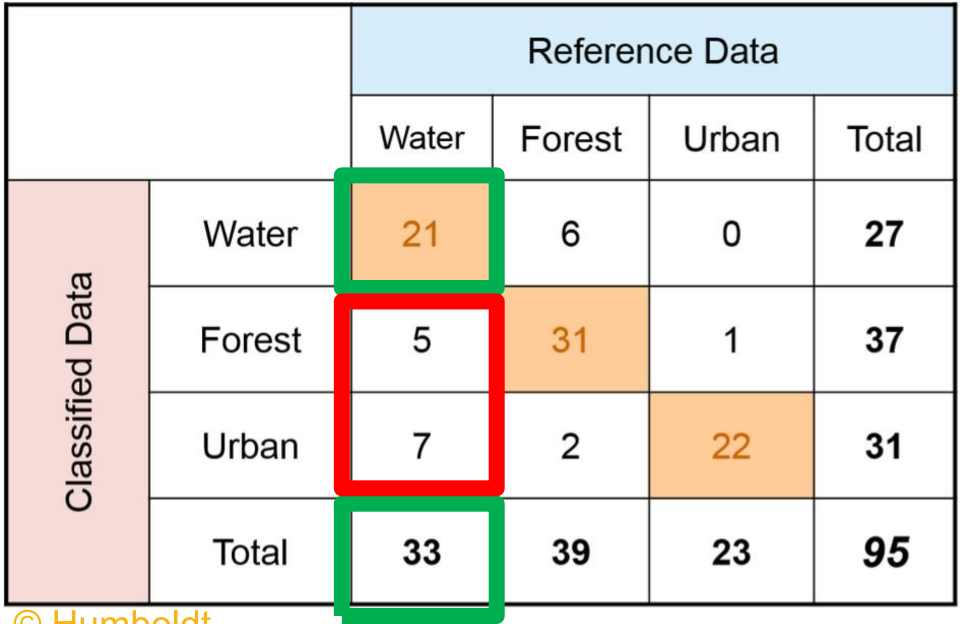
© Humboldt

For water class:

- PA = total correctly classified pixels / column total

Producer accuracy for Water class $= \frac{21}{33} = 64\%$

Producer’s accuracy + Omission error



		Reference Data			
		Water	Forest	Urban	Total
Classified Data	Water	21	6	0	**27**
	Forest	5	31	1	**37**
	Urban	7	2	22	**31**
	Total	**33**	**39**	**23**	***95***

© Humboldt

For water class:

- Omission error = 100 – PA = 100 – 64 = 36%

User accuracy for Water class = $\frac{21}{33}$ = 64%