ECEM/EAML 2004

Bled, Slovenia 2004

Using wavelets for classification of hyperspectral images

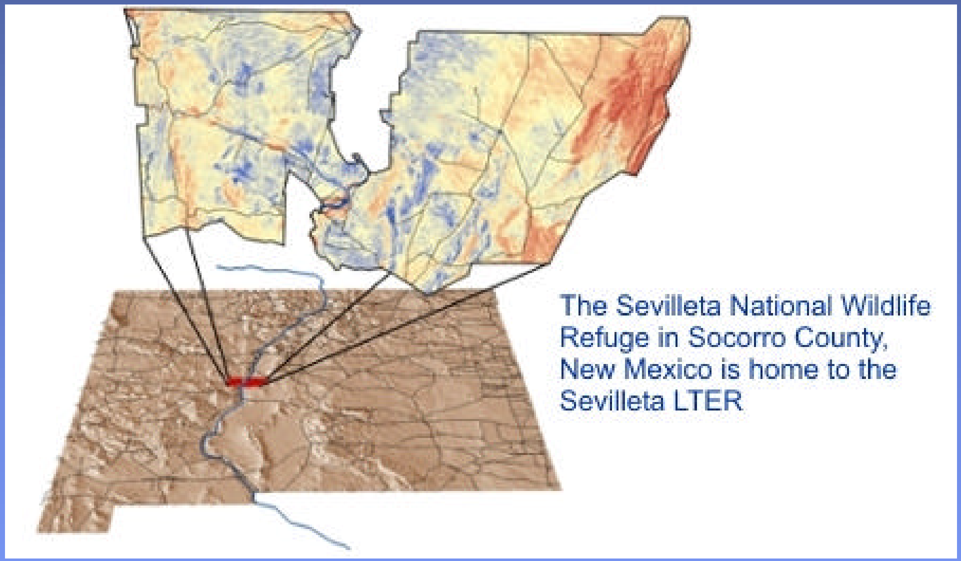
Hector Jasso, Peter Shin, Tony Fountain

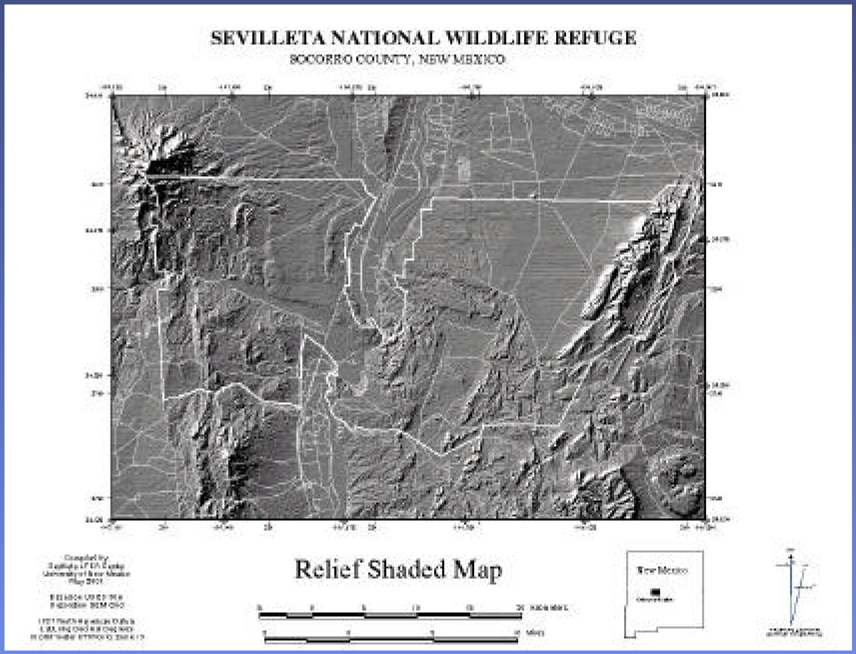
San Diego Supercomputer Center, University of California San Diego (UCSD)

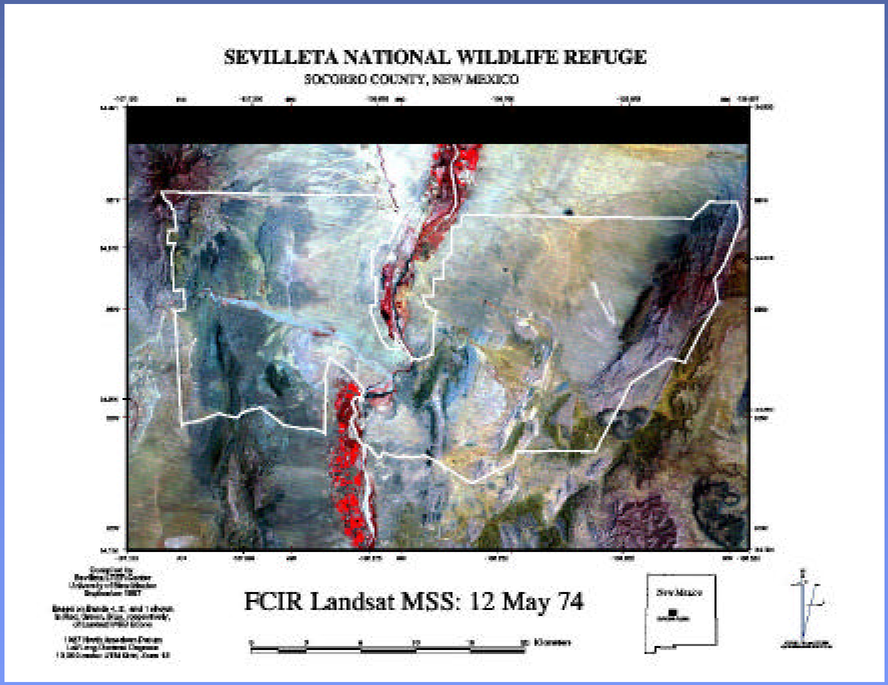
Deana Pennington

Long Term Ecological Research (LTER), University of Albuquerque, New Mexico

New Mexico Sevilleta National Wildlife Refuge

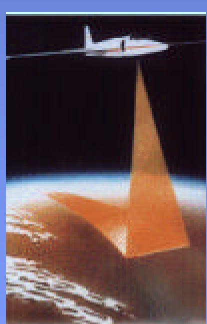


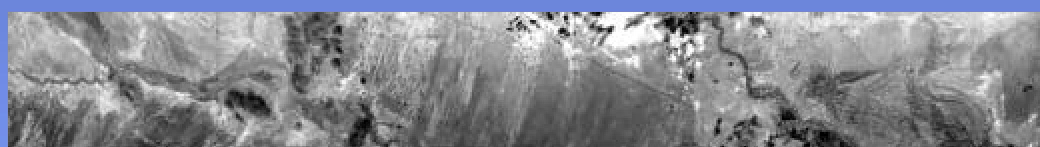

SEVILLETA NATIONAL WILDLIFE REFUGE
SOCORRO COUNTY, NEW MEXICO
Relief Shaded Map
New Mexico

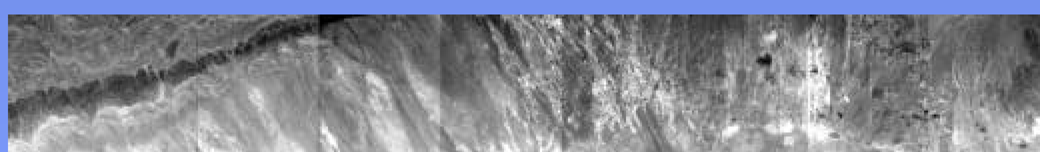

SEVILLETA NATIONAL WILDLIFE REFUGE
SOCORRO COUNTY, NEW MEXICO
FCIR Landsat MSS: 12 May 74
New Mexico

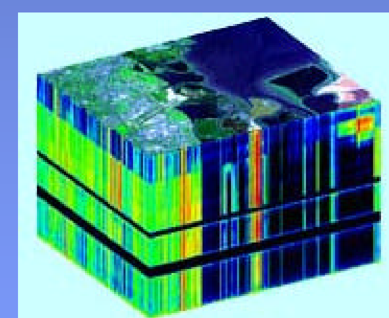
NASA's JPL (Jet Propulsion Lab) Aviris (Airborne Visible/Infrared Imaging Spectrometer) scans

- Scanned from an altitude of 20km, 10km flightline
- 614x512 pixel image with 20m resolution, subsetted to 300x300
- 224 bands of electromagnetic information, infrared to ultraviolet
- 24 water absorption bands were removed from the data, resulting in 201 bands
- Preprocessing steps for radiometric and atmospheric corrections (sun angle and atmospheric scatter)
- 20 million pieces of information, over an area of $6km^2$

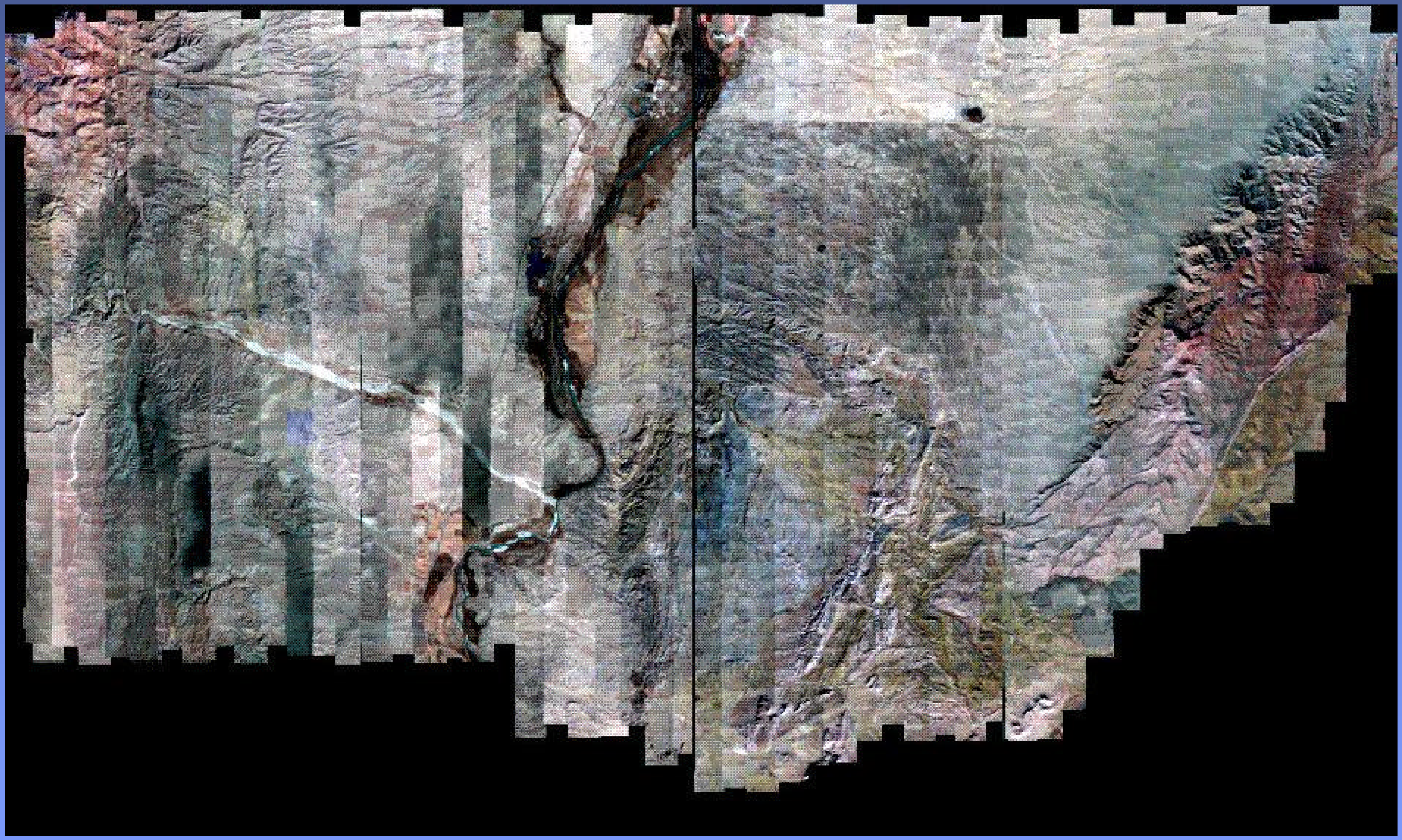






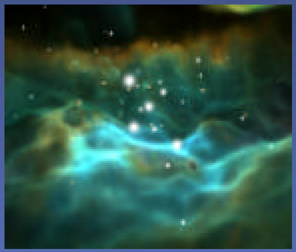


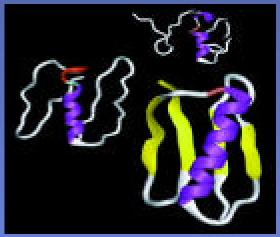
Complete Aviris scan of the Sevilleta Wildlife refuge

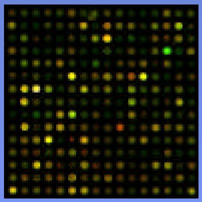


A paradigm shift is occurring in scientific data analysis

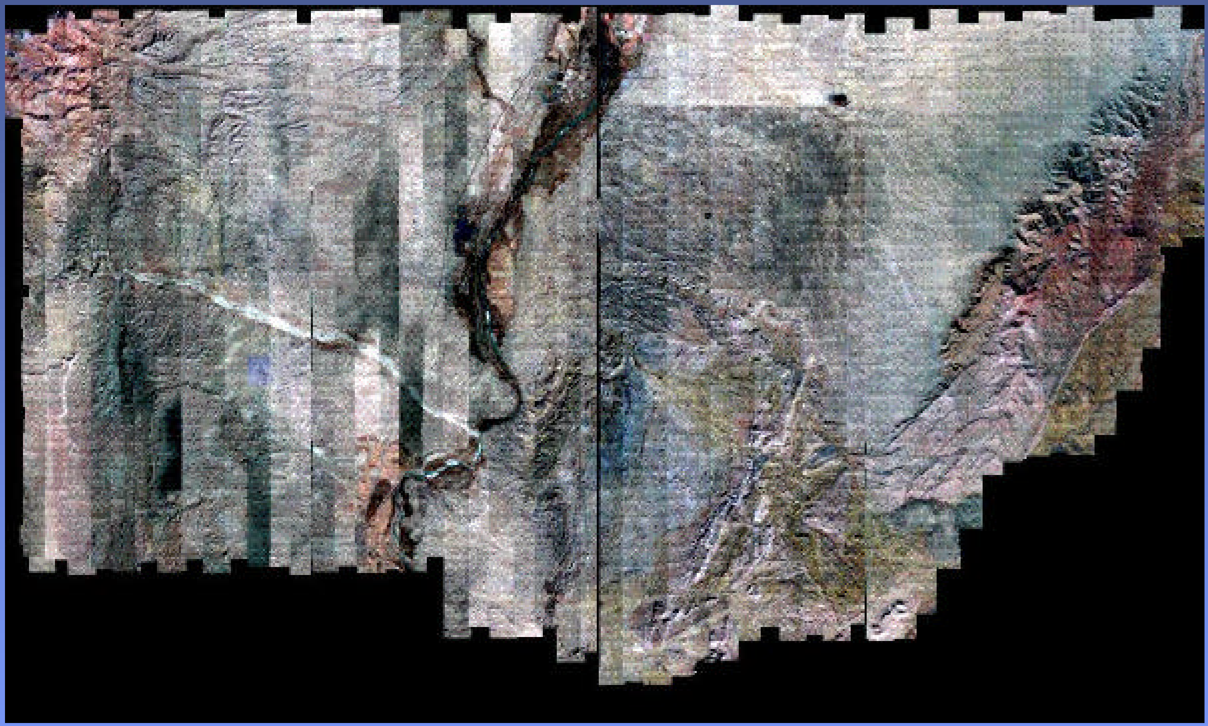
- **Deluge from new sources**
  - **Remote sensing**
  - **Microarray processing**
  - **Wireless communication**
  - **Simulation models**
  - **Instrumentation – microscopes, telescopes**
  - **Digital publishing**
  - **Federation of collections**
- **"5 exabytes (5 million terabytes) of new information was created in 2002" (source: UC Berkeley researchers Peter Lyman and Hal Varian)**
- **This is the result of a recent paradigm shift: from hypothesis-driven data collection to data mining**







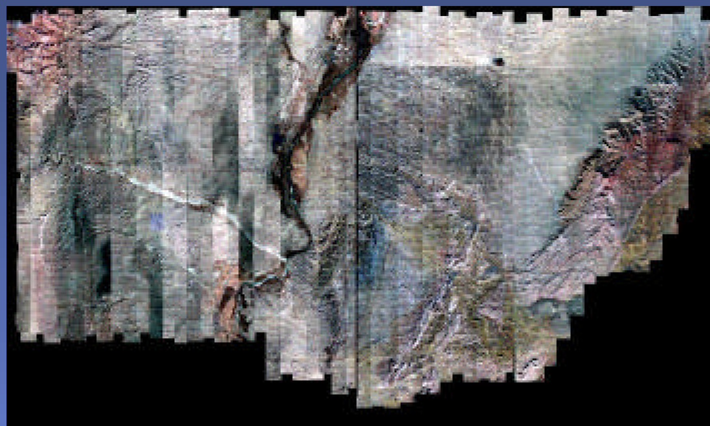
Objective: ????



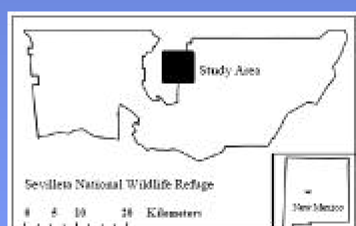
National Partnership for Advanced Computational Infrastructure

San Diego Supercomputer Center

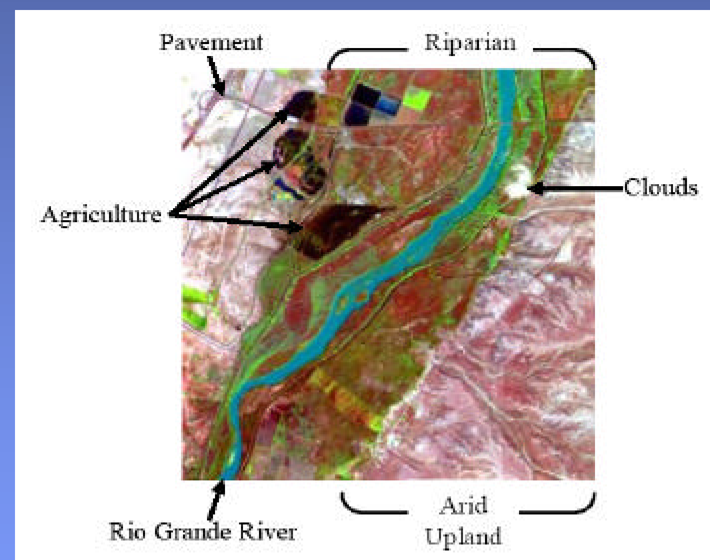
Pixel-by-pixel classification of land cover



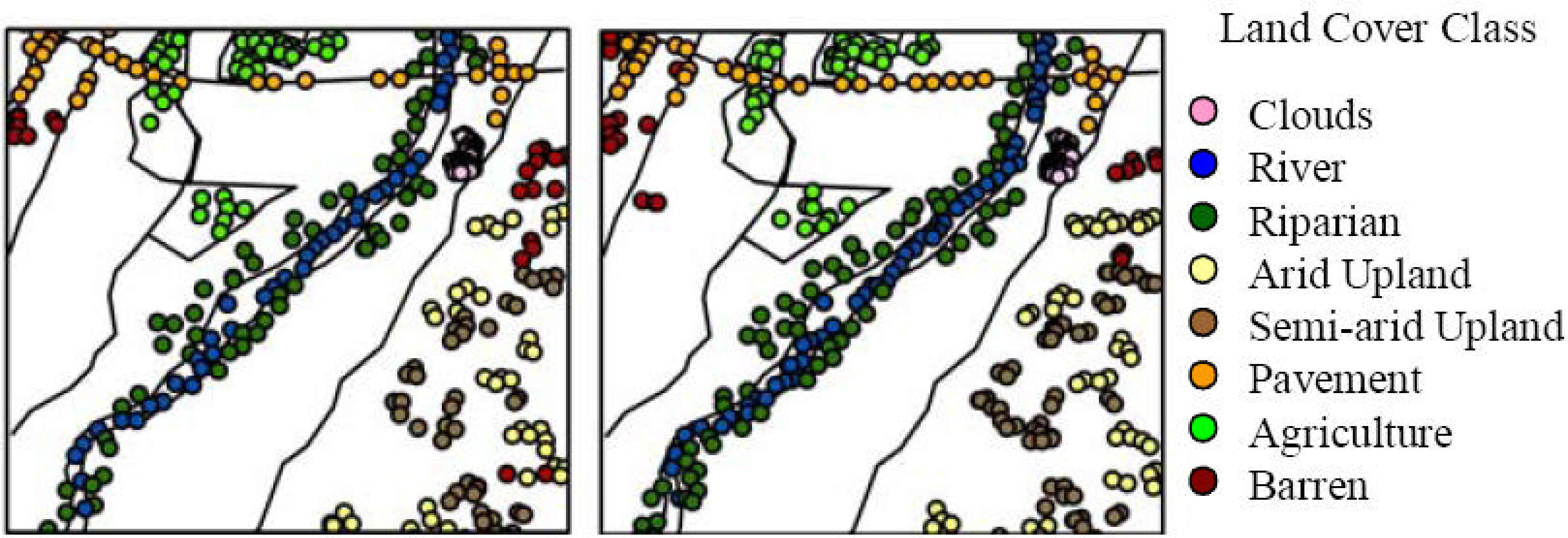
Complete Aviris scan of the Sevilleta Wildlife refuge



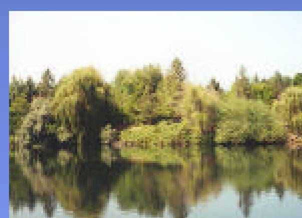
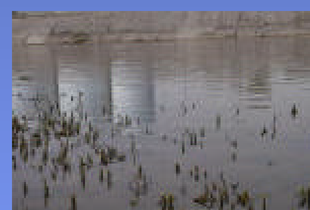
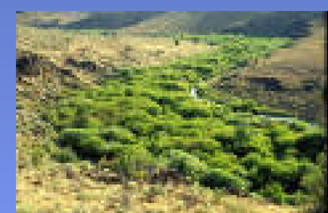
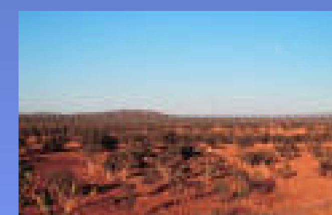
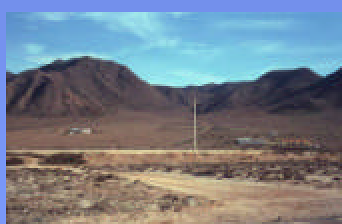
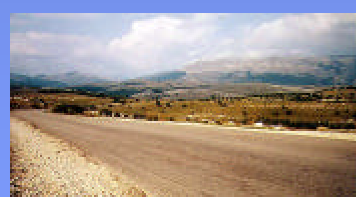
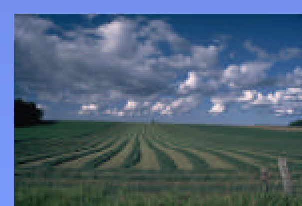
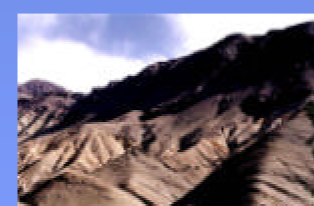
Study area



False color composite of study area



Training data　　Test data

Data set

Class	*Train*	*Test*
Clouds	24	26
River	42	60
Riparian	59	73
Arid Upland	44	49
Semi-arid Up.	41	52
Pavement	35	36
Agriculture	43	42
Barren	25	24
Total	**313**	**362**

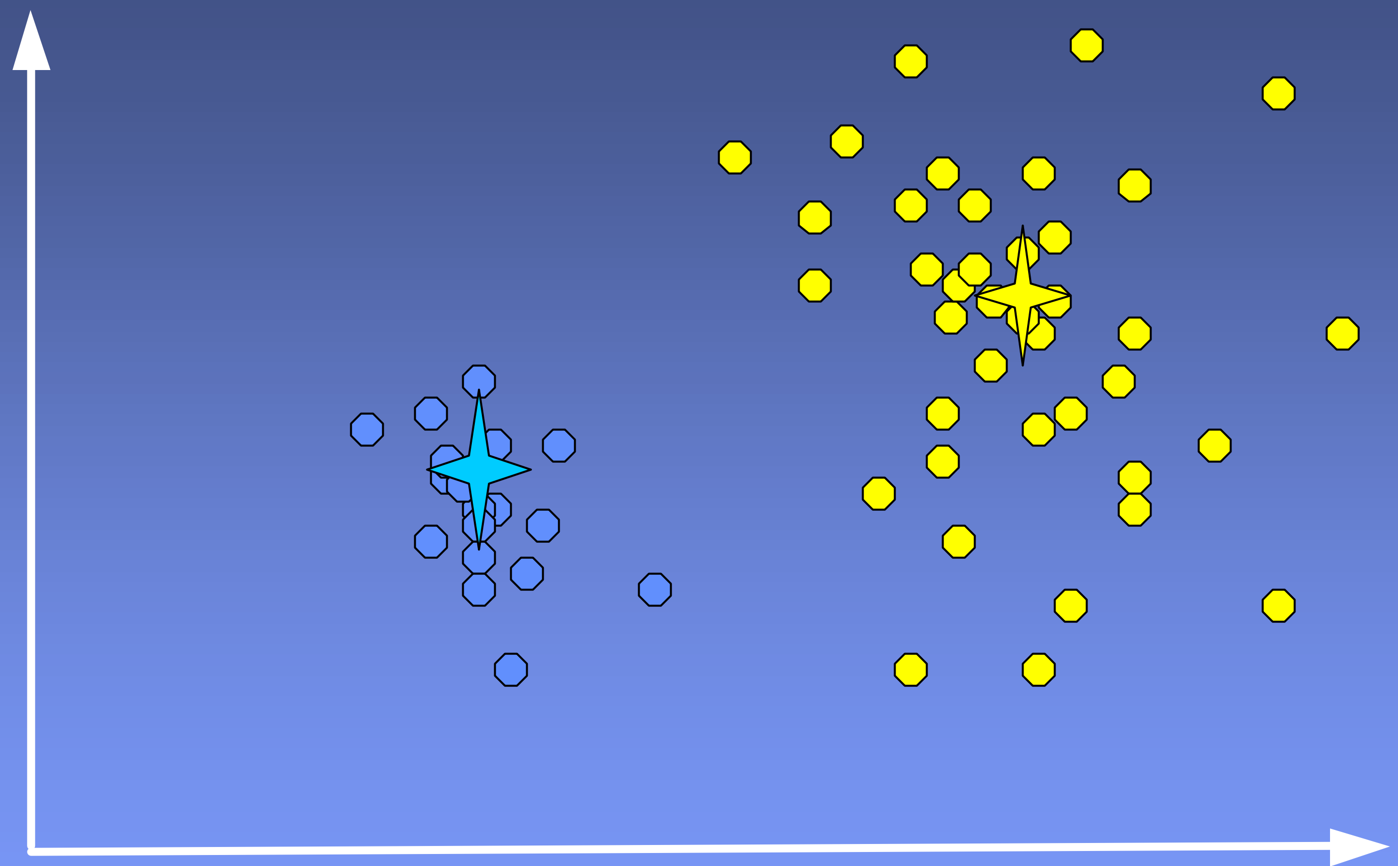
Objective

- **Automatically classify the whole image using:**
  - Minimum distance
  - Naïve Bayesian classifier
  - Support Vector Machines (SVMs), “winner-take-all” scheme
- **Improve accuracy by using wavelet transformation to pre-process data**
- **Improve accuracy by using an ensemble method: “one-against-one” scheme for SVM**

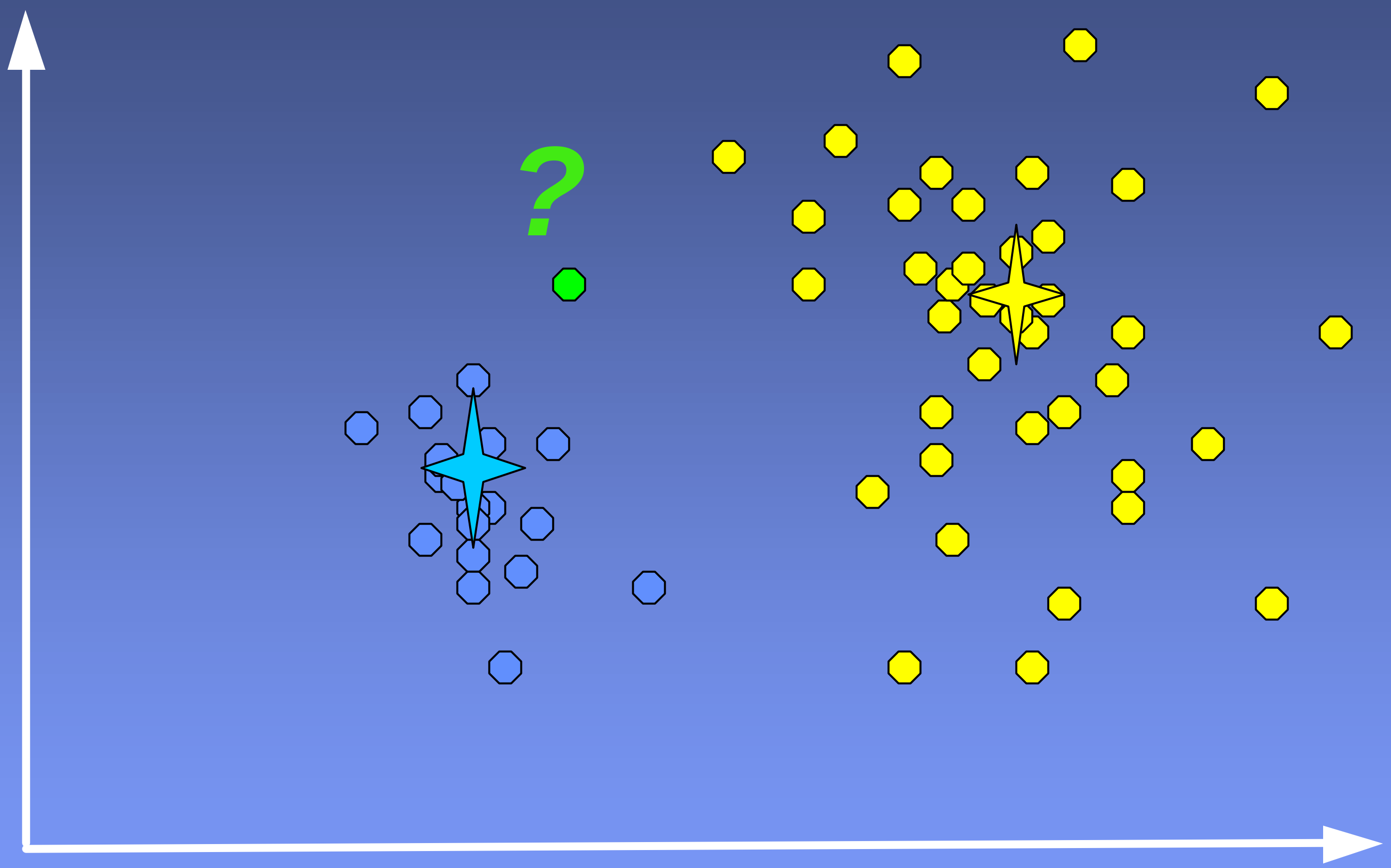
Challenges of Classifying Hyperspectral Data

- **High dimensionality**
- **Low number of samples**
- **Within-class variability (e.g. many types of agriculture)**
- **Between-class mixtures (e.g. riparian and river)**
- **Voluminous data**

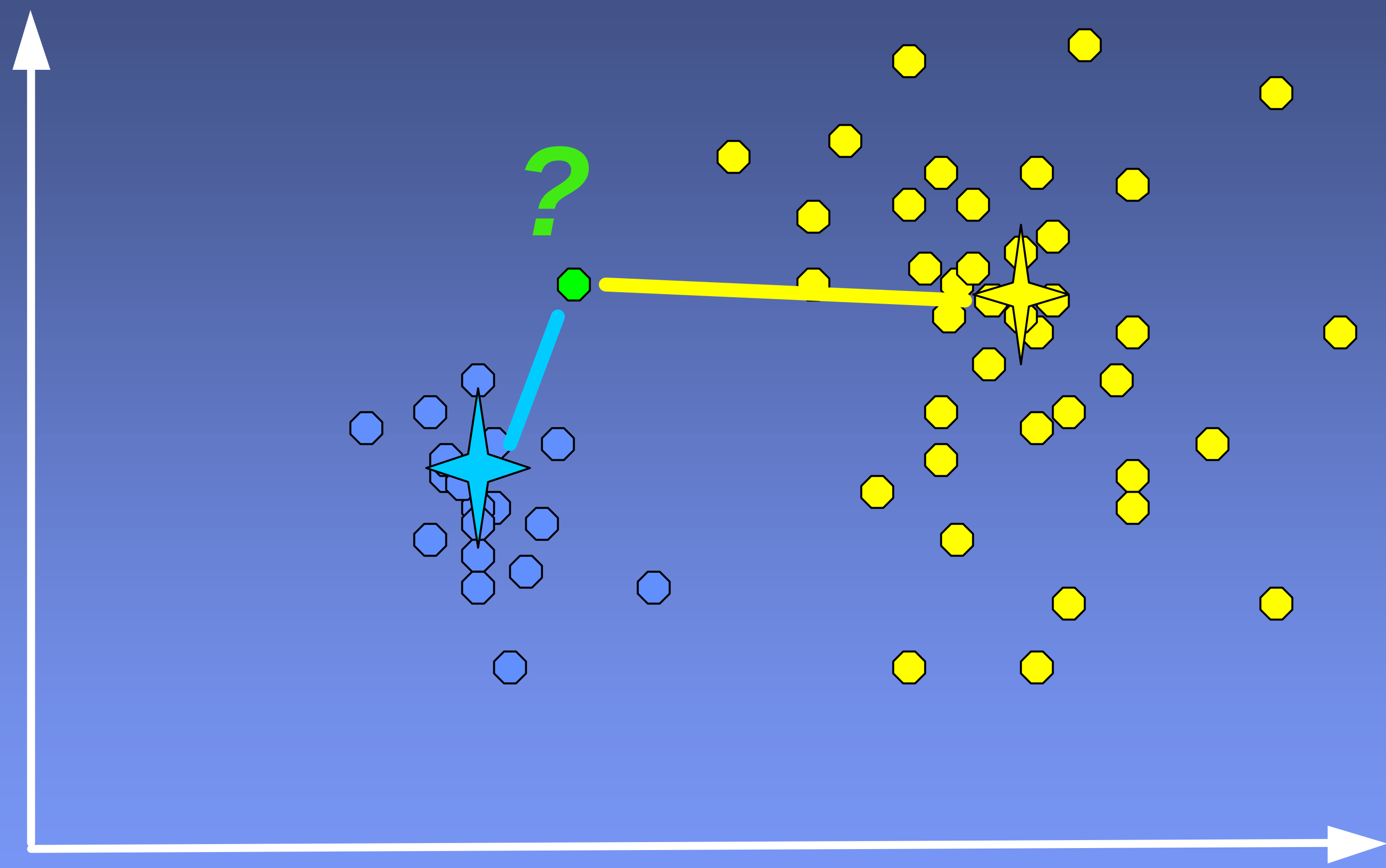
Minimum Distance



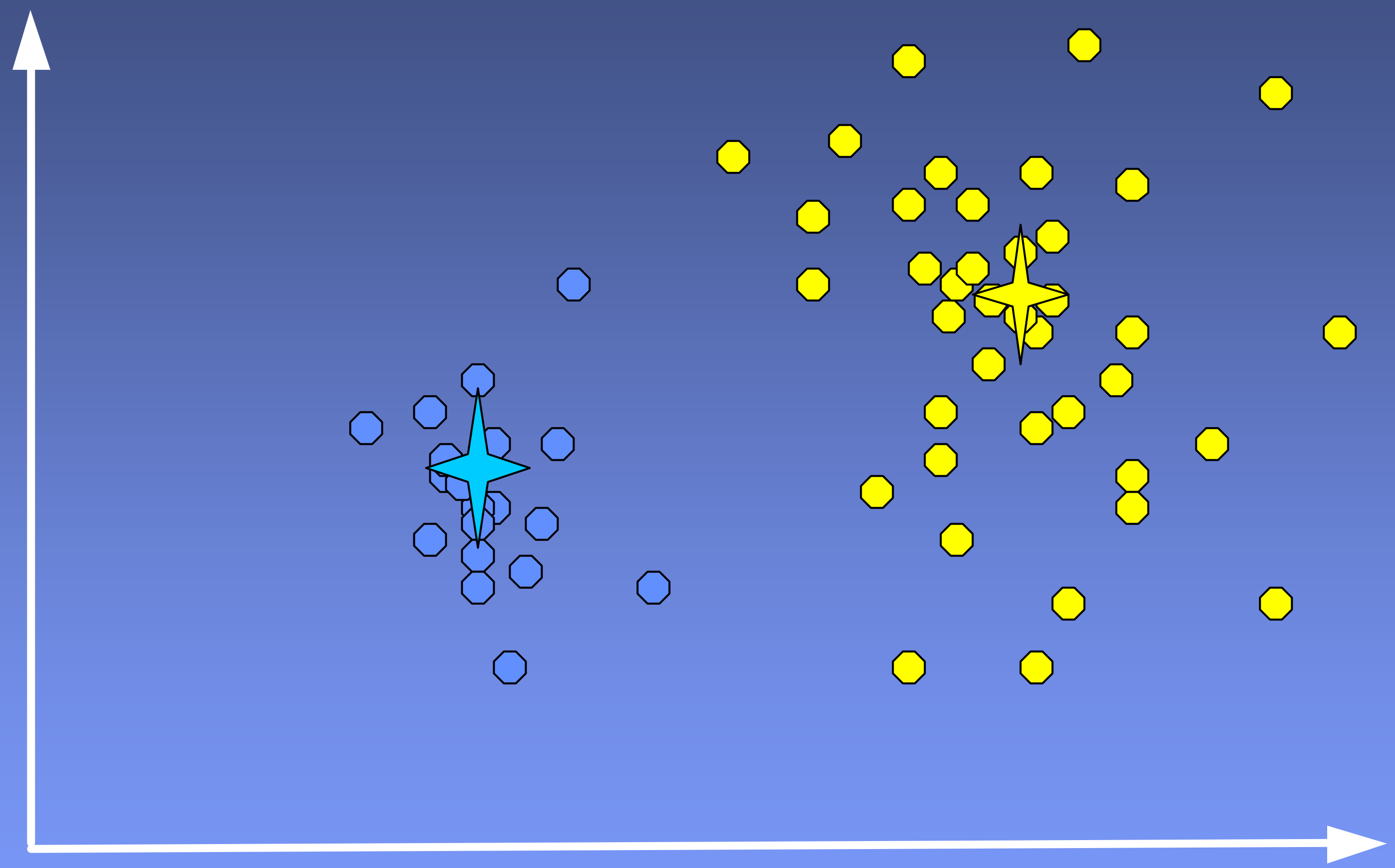
Minimum Distance



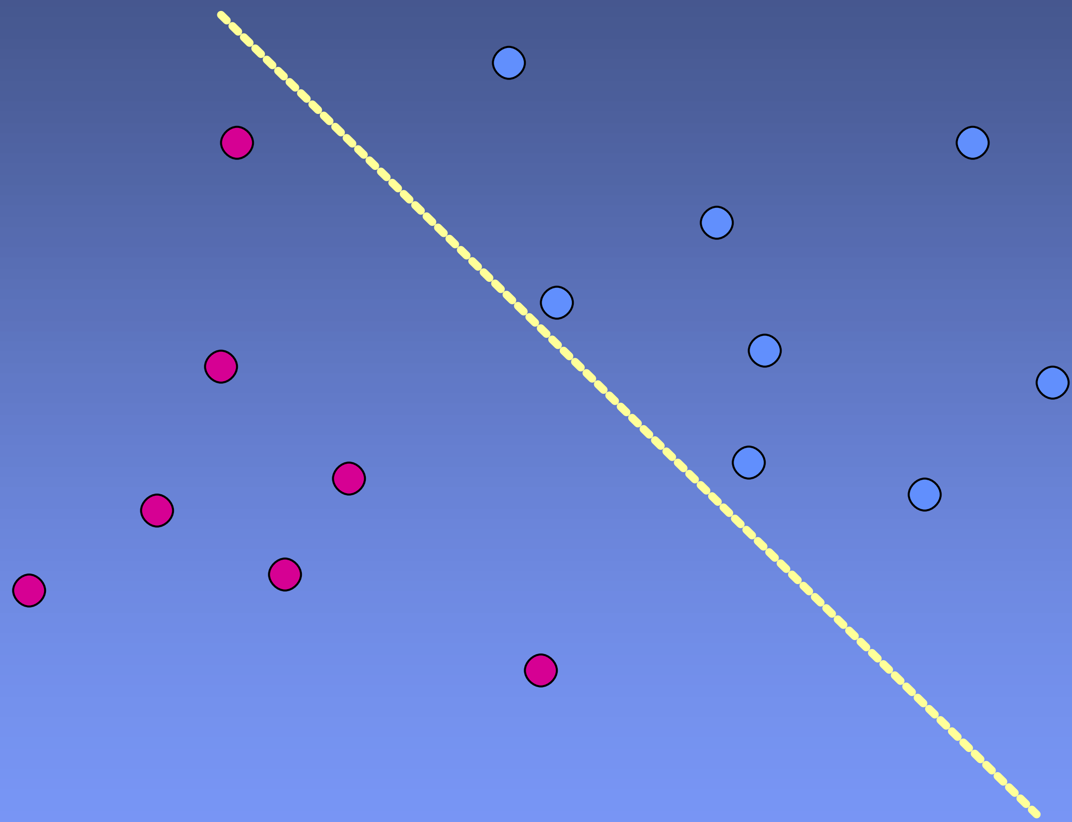
Minimum Distance



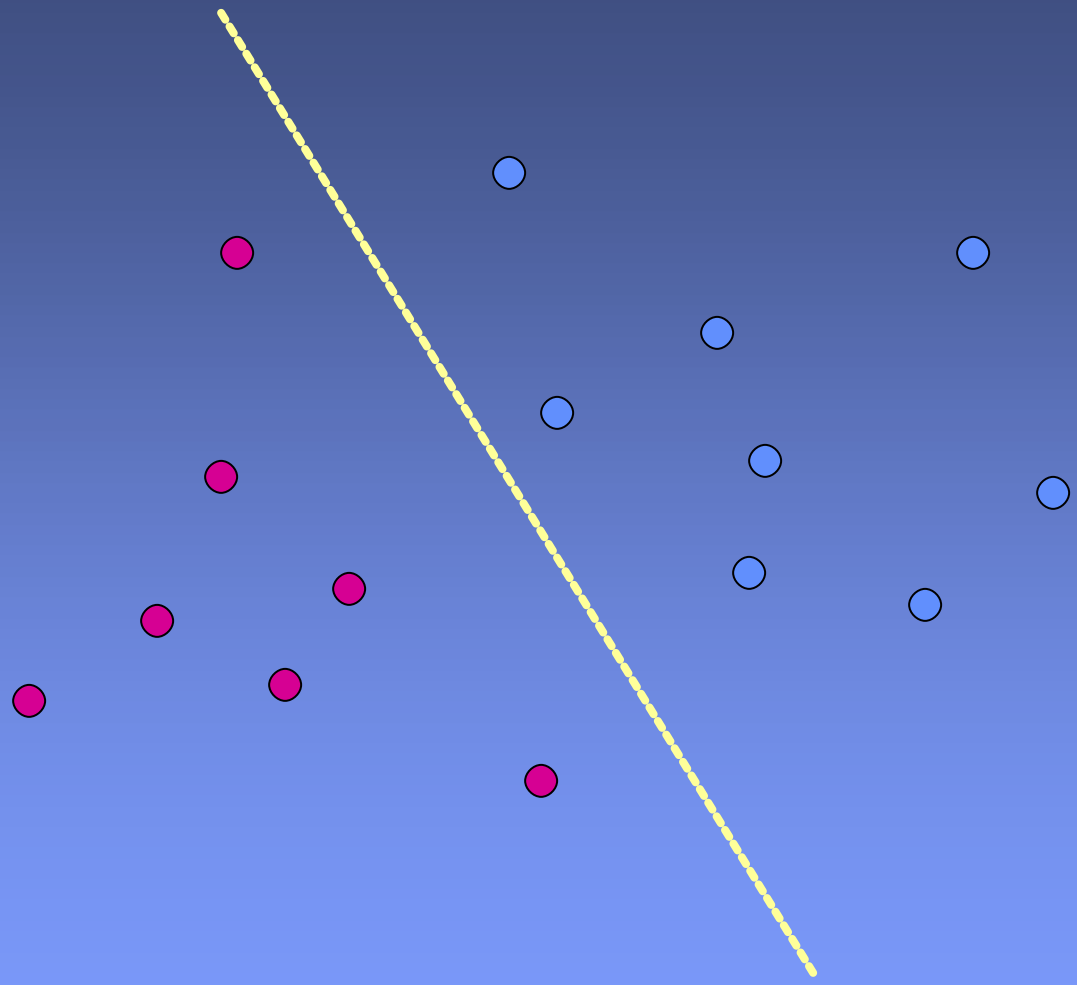
Minimum Distance



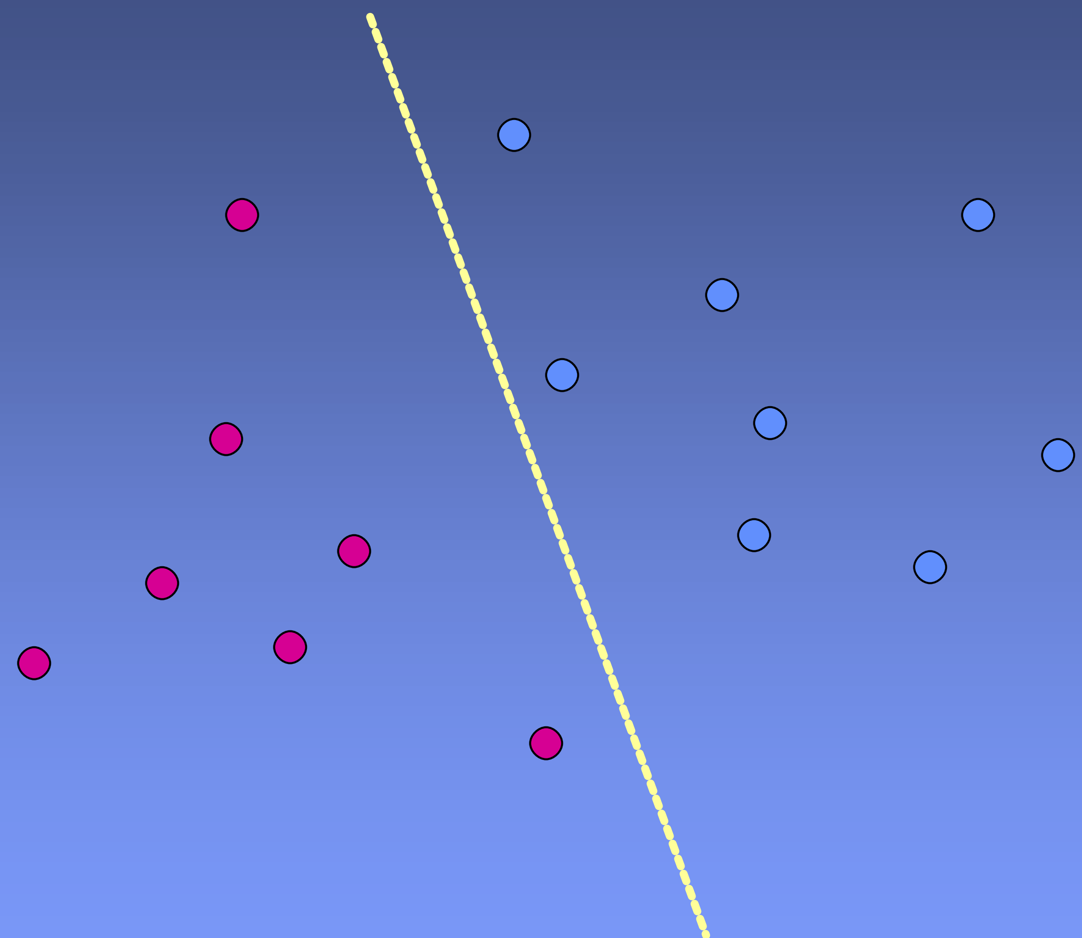
How Support Vector Machines (SVMs) work



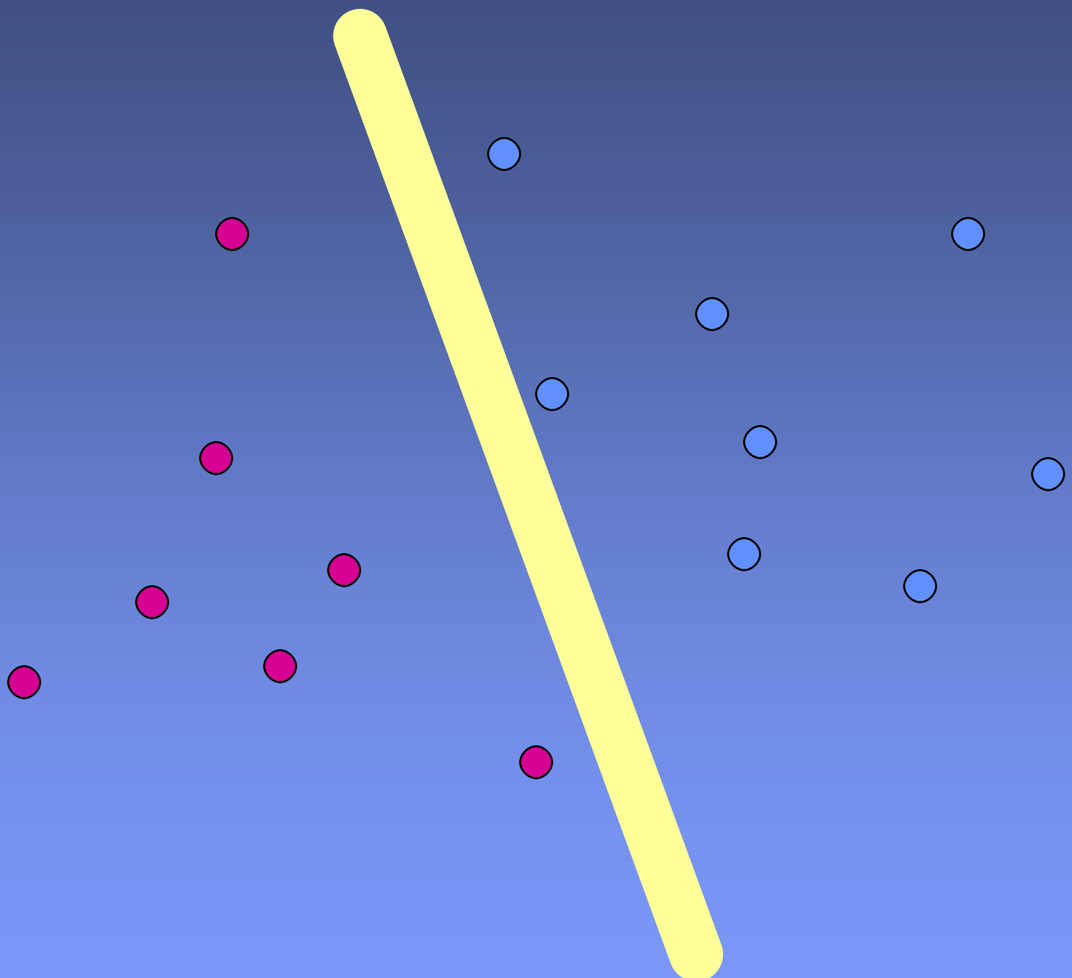
How Support Vector Machines (SVMs) work



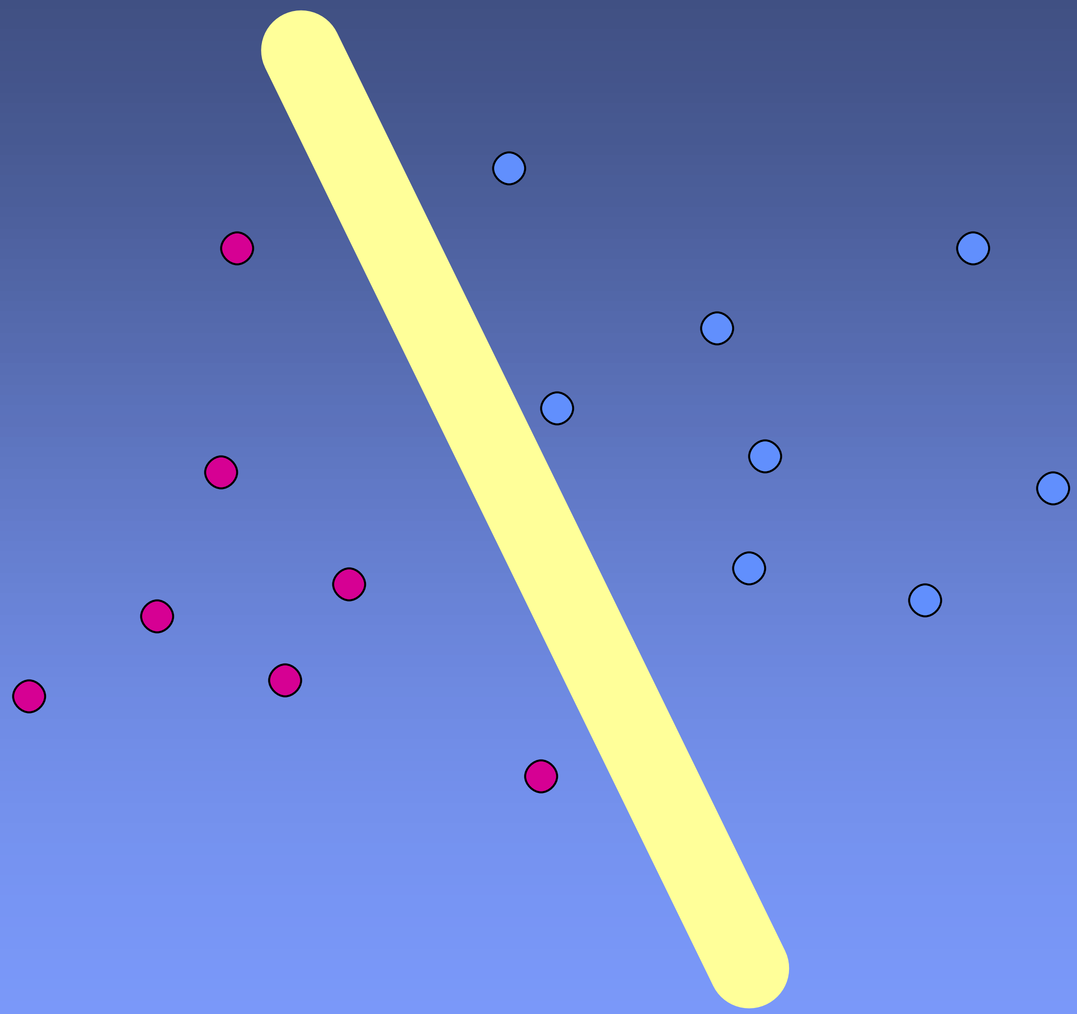
How Support Vector Machines (SVMs) work



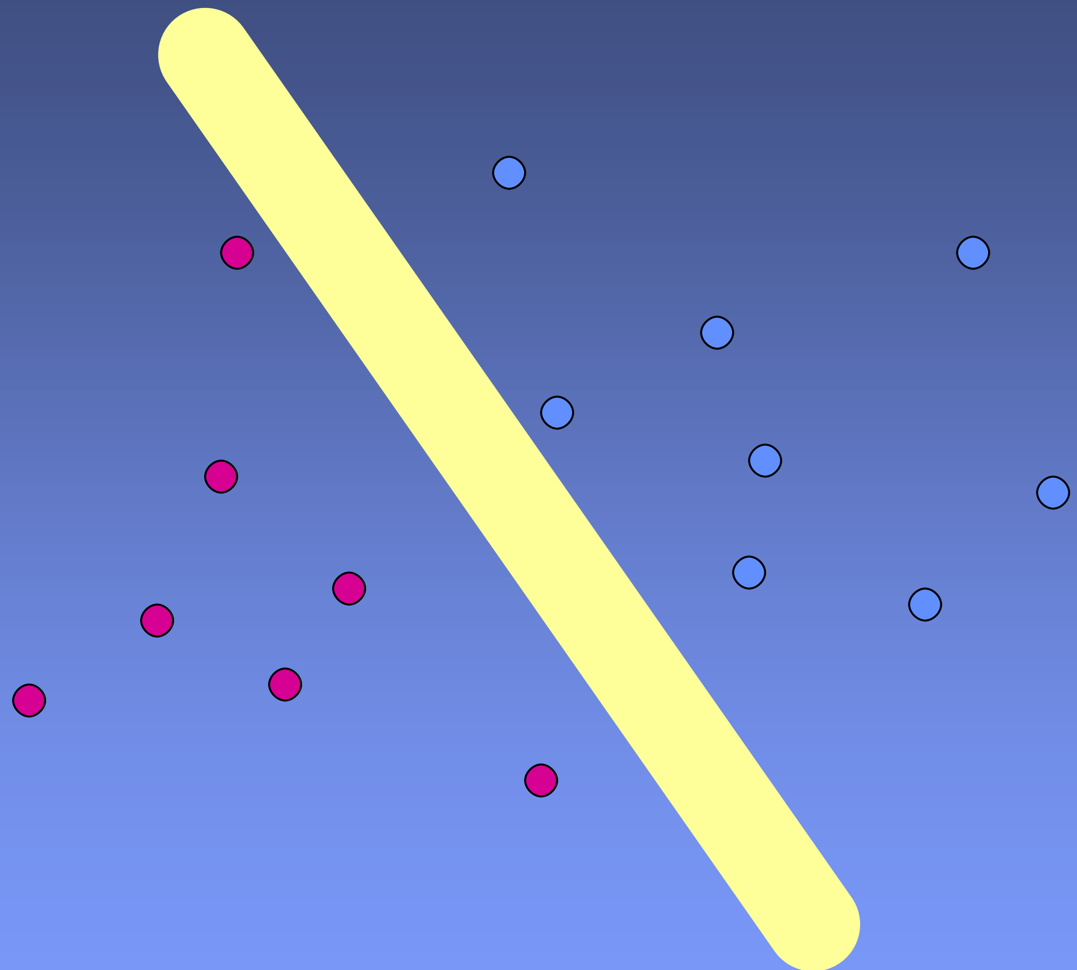
How Support Vector Machines (SVMs) work



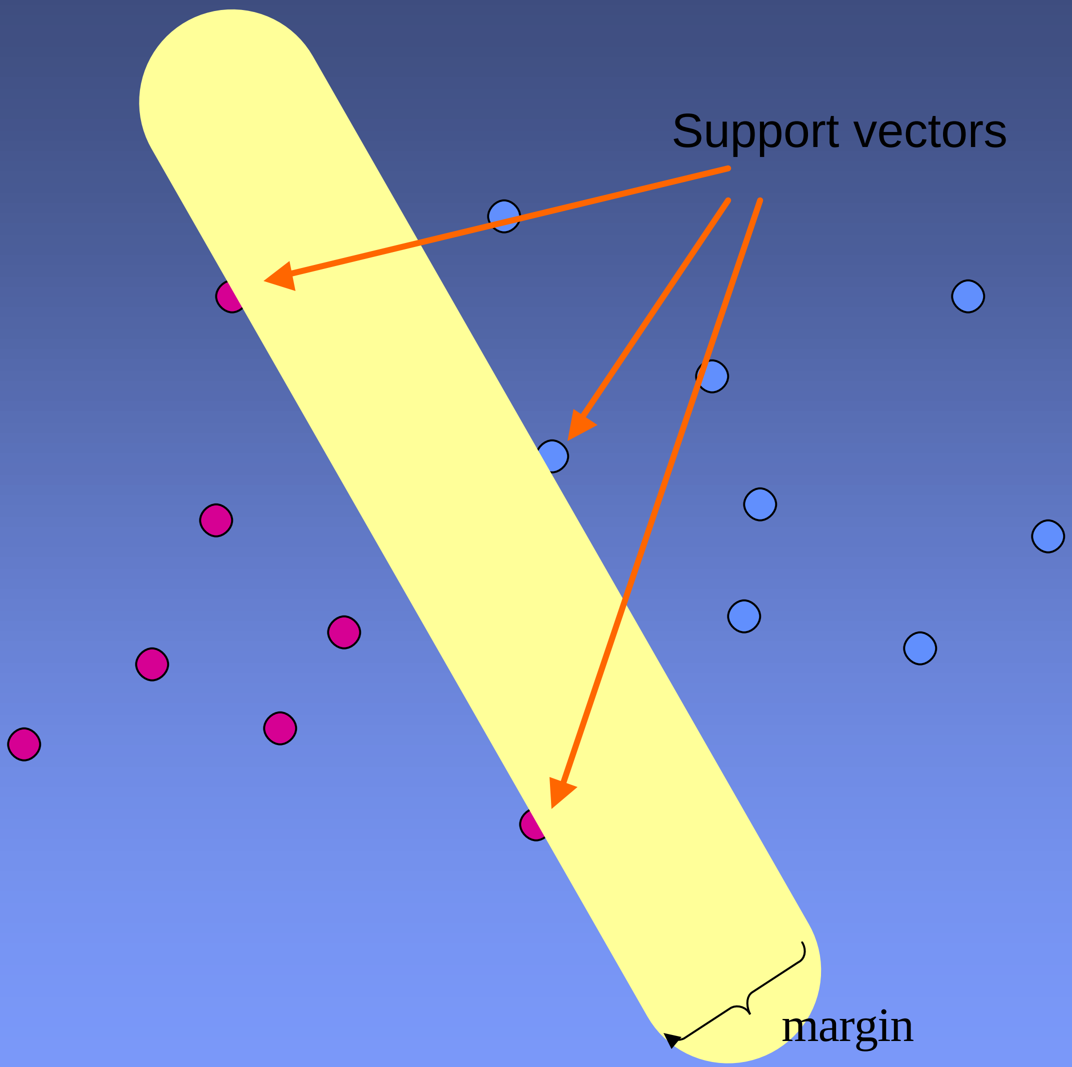
How Support Vector Machines (SVMs) work



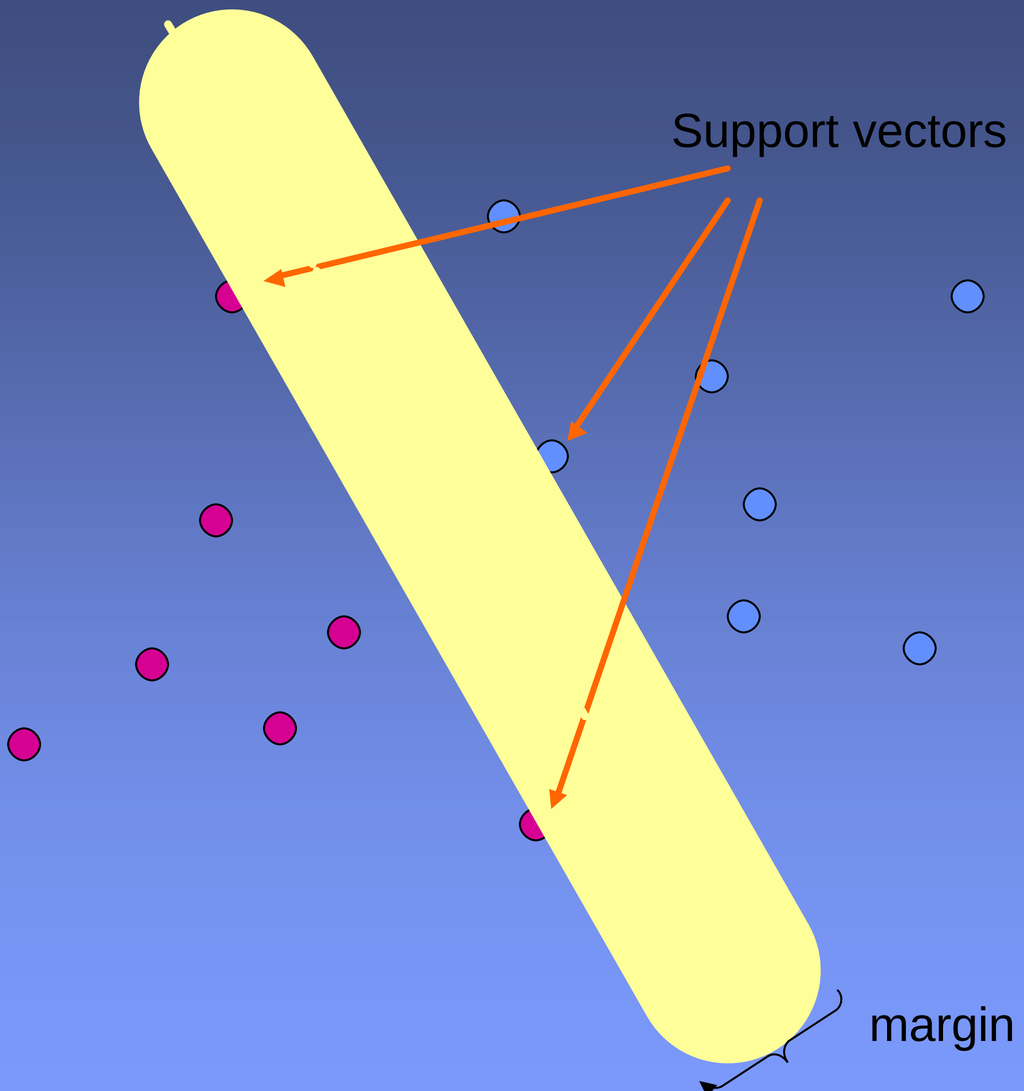
How Support Vector Machines (SVMs) work



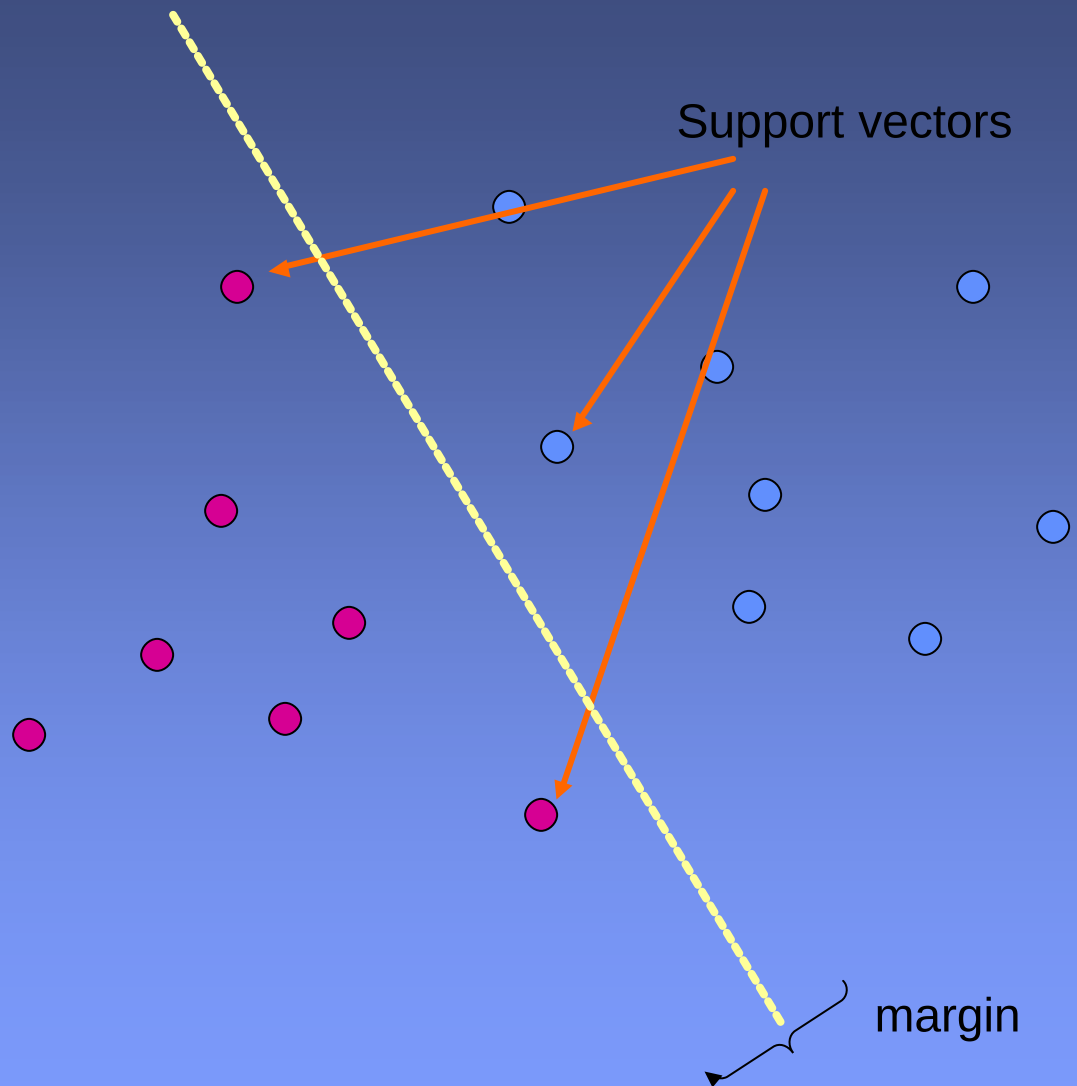
How Support Vector Machines (SVMs) work



How Support Vector Machines (SVMs) work



How Support Vector Machines (SVMs) work



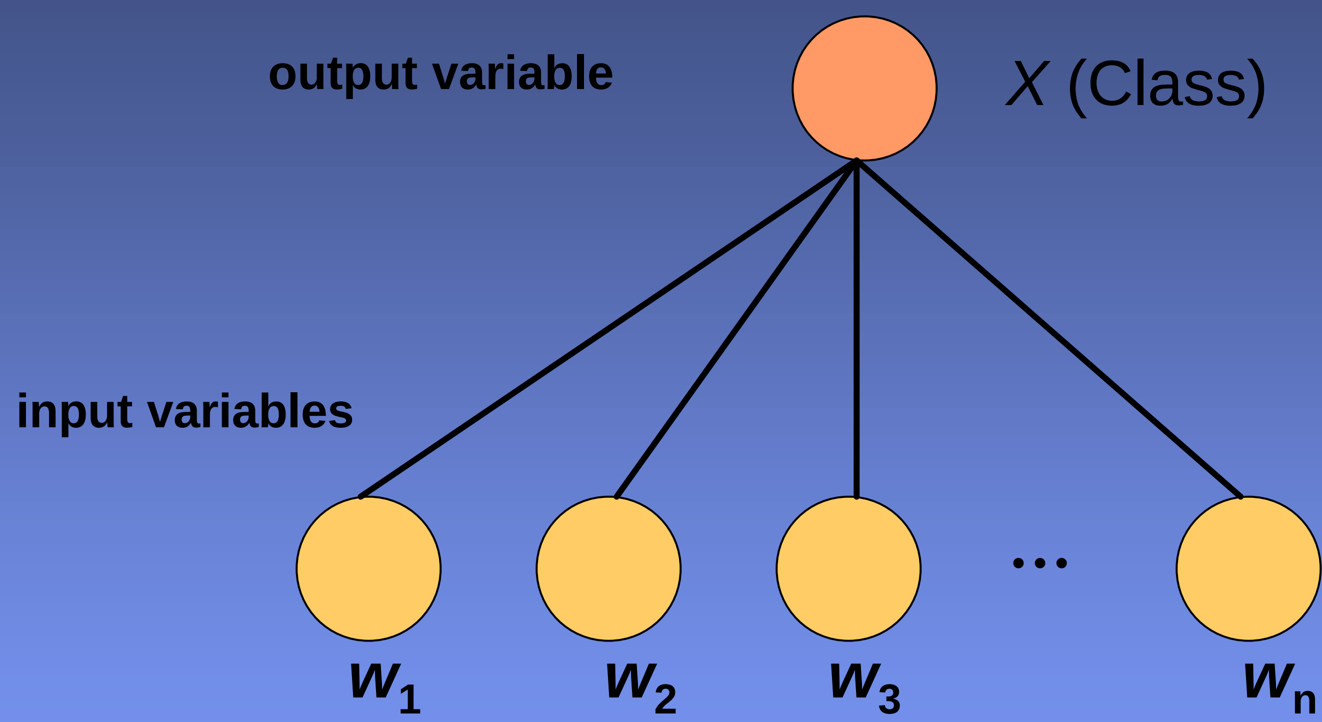
SVM winner-take-all

- **SVM is a binary classifier, but we have 8 classes!**
- **Solution:**
  - Train 8 SVMs, one per class: label correct class as +1, all other classes as -1
  - Run a new data point on all SVMs. The one with the highest positive value wins

Characteristics of SVMs

- Scales well to high-dimensional problems
- Fast convergence to solution
- Has well-defined statistical properties

Naïve Bayesian Classifiers



$$p(X, w_1, w_2, \ldots, w_n) = p(X) \prod_{i=1}^{n} p(w_i \mid X)$$

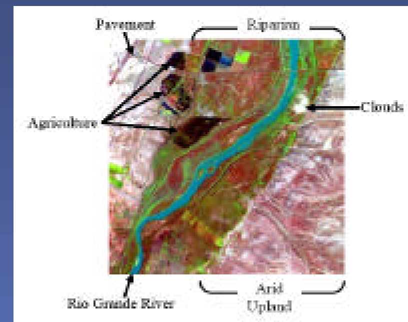
Characteristics of Naïve Bayesian Classifiers

- Scales well to high-dimensional problems
- Fast to compute
- Based on Bayesian probability theory

Results

Independent test set

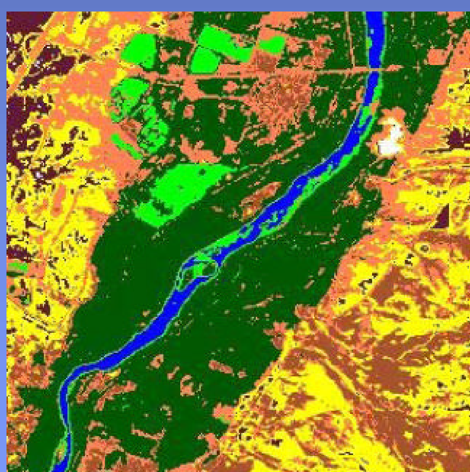
Classifier	Accuracy
Naïve Bayesian Classifier	90.9%
Minimum Distance	88.4%
SVM winner-take-all	72.9%



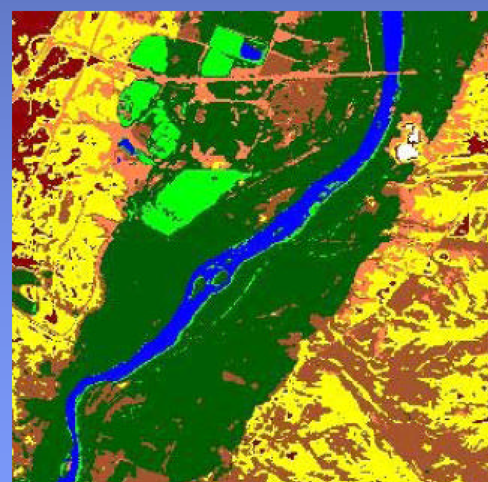
- Clouds
- River
- Riparian
- Agriculture
- Arid Upland
- Semi-arid Upland
- Barren
- Pavement

Full image

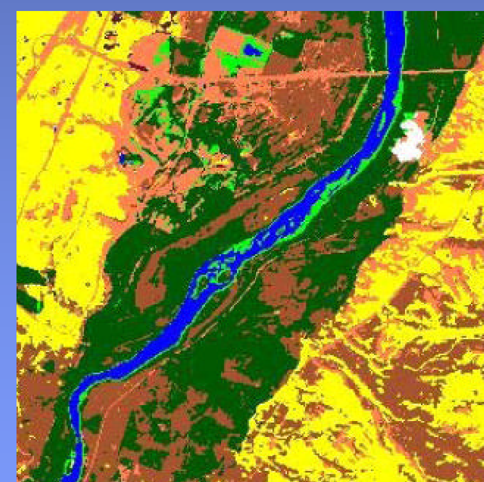
Naïve Bayesian Classifier



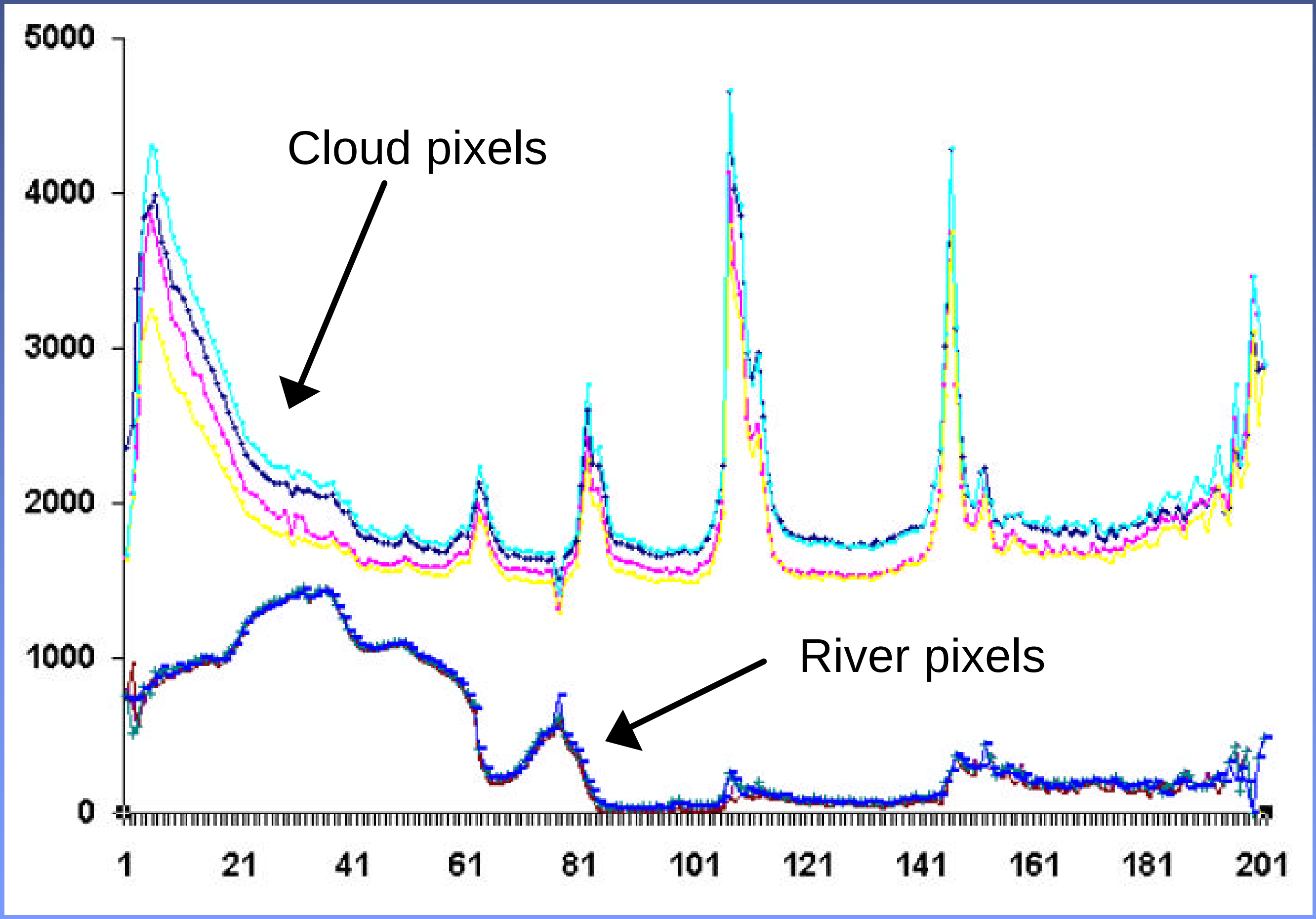
Minimum Distance



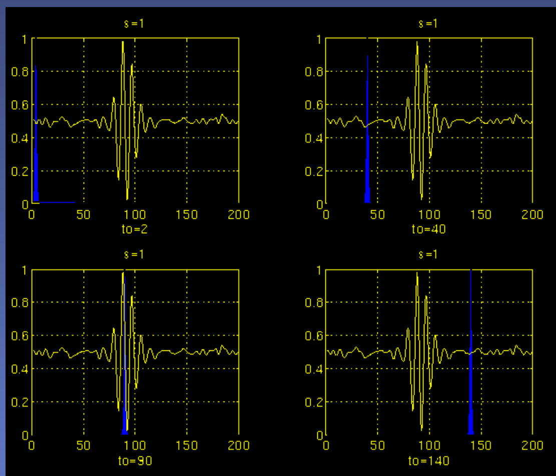
SVM winner-take-all

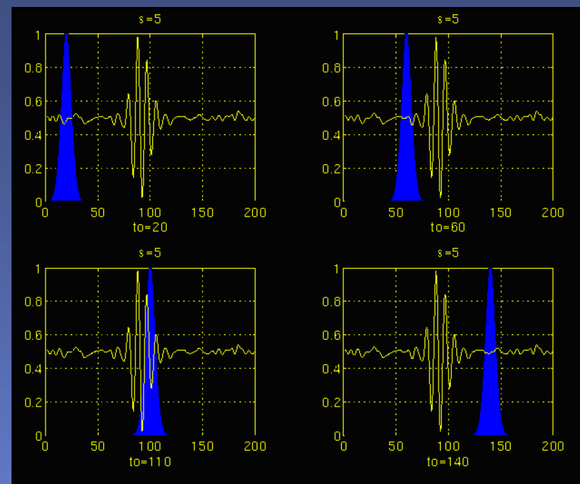


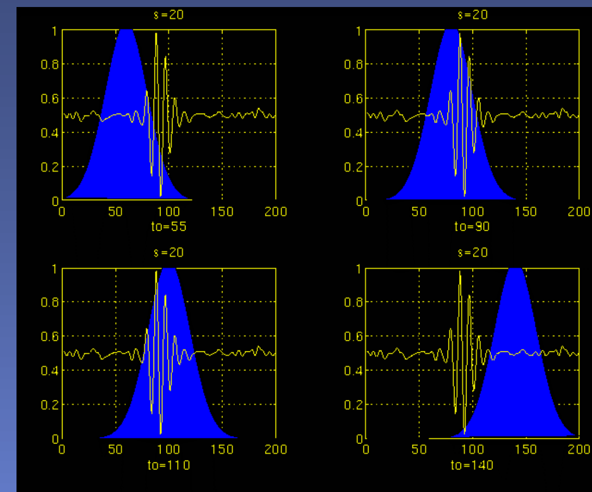
“Frequency component” of Hyperspectral data

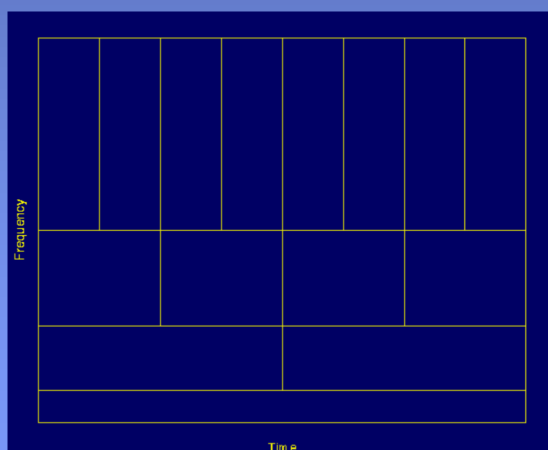


Wavelets

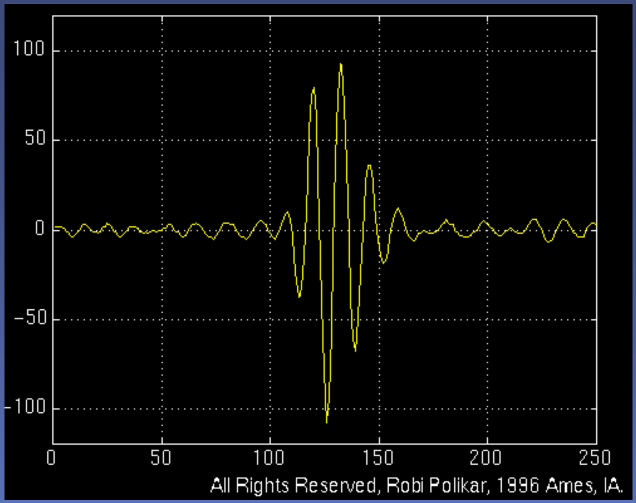


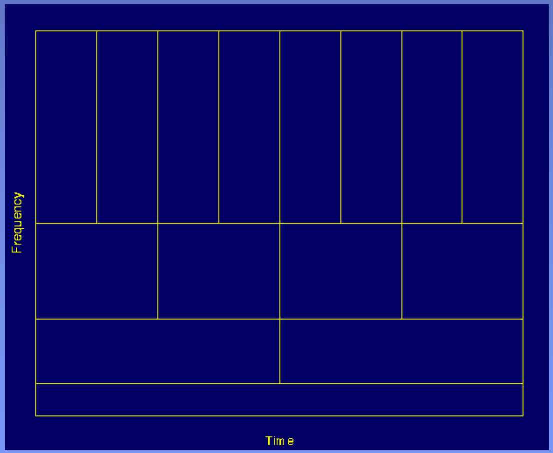


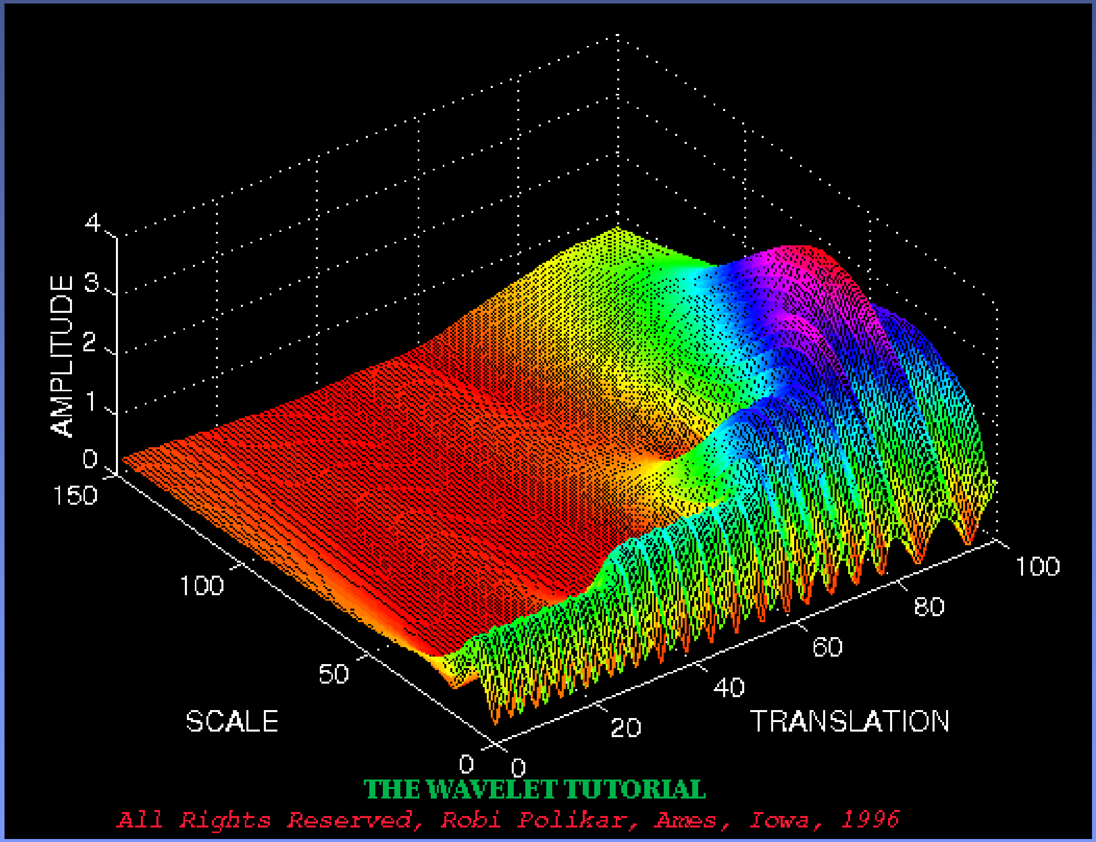




From Wavelet Tutorial for Engineers, by Robi Polikar

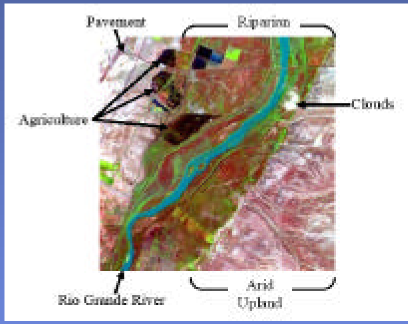

100
50
0
-50
-100
0
50
100
150
200
250
All Rights Reserved, Robi Polikar, 1996 Ames, IA.


Frequency
Time

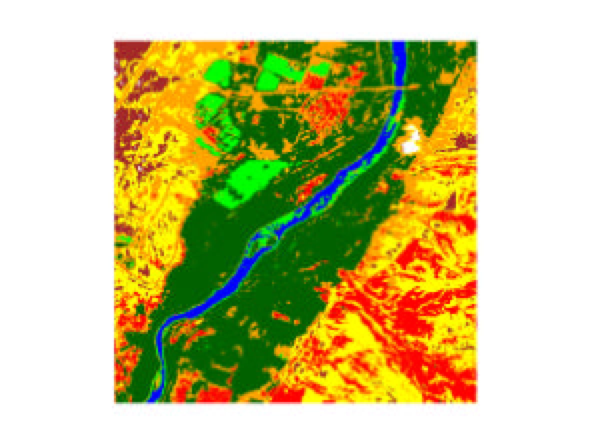

AMPLITUDE
4
3
2
1
0
150
100
50
0
SCALE
0
20
40
60
80
100
TRANSLATION
THE WAVELET TUTORIAL
All Rights Reserved, Robi Polikar, Ames, Iowa, 1996

Results with wavelets

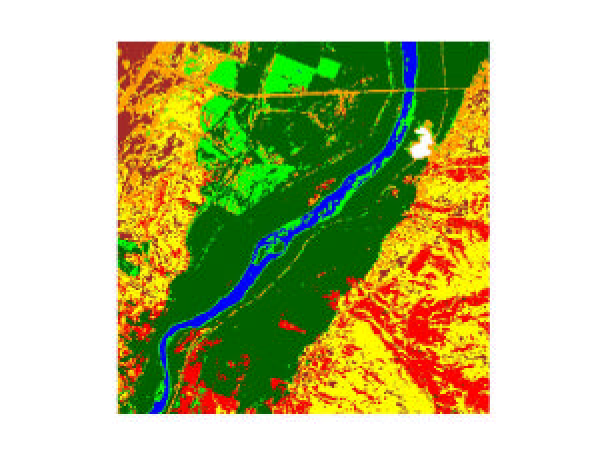
- **Daubechies transformation (db1)**
- **SVM winner-take-all improved, from 77.6% to 90.1%**
- **Naïve Bayesian Classifier improved from 90.9% to 94.2%**



Naïve Bayesian Classifier, raw data

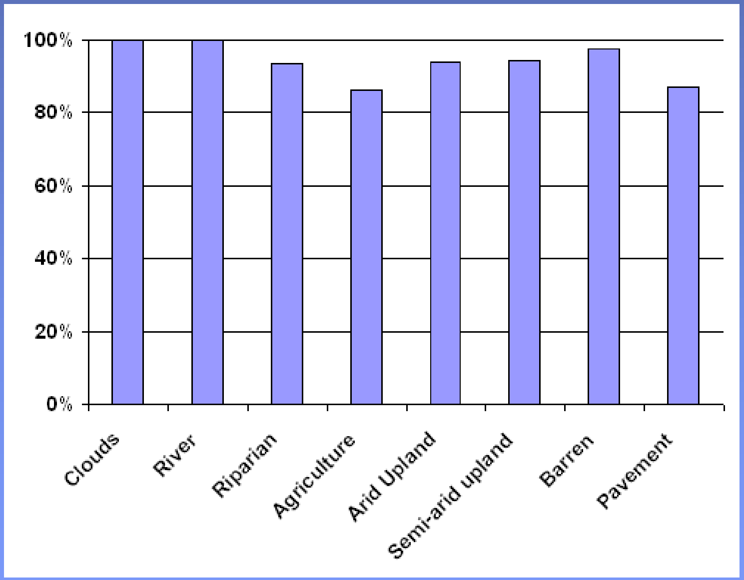


Naïve Bayesian Classifier, wavelet transformation



Naïve Bayesian Classifier, Wavelet transformation

	Clouds	River	Riparian	Agriculture	Arid Upland	Semi-arid upland	Barren	Pavement
Clouds	26							
River		60						
Riparian			71			2	3	
Agriculture				44	3			4
Arid Upland			2	1	48			
Semi-arid upland				1	1	33		
Barren						1	39	
Pavement				3				20

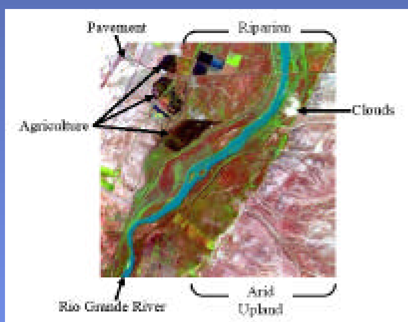


SVM One-against-one

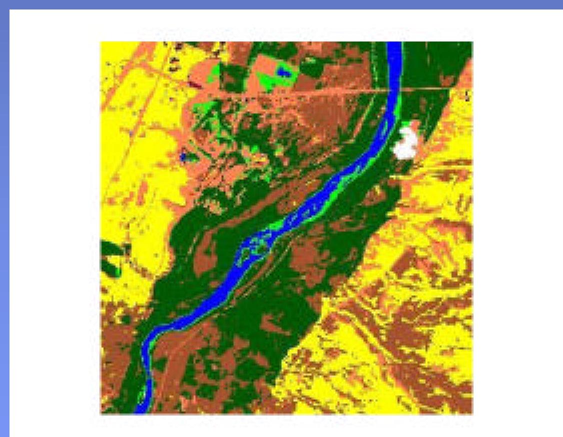
- **Winner-take-all: the SVM with the highest positive value wins**
- **One-against-one:**
  - Train as many SVMs as possible groupings of classes into new positive and negative classes
  - On a new data point, run all SVMs and use voting to determine the winner

Results with one-against-one

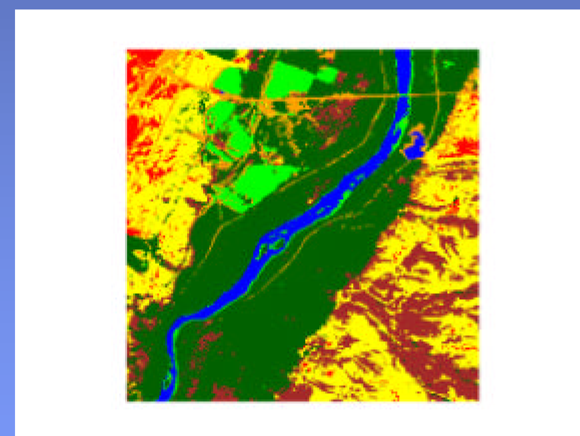
- **Improves, from 90.1% to 97.1%**



SVM, winner-take-all, raw data



SVM, one-against-one, wavelet transformation



Conclusions

- Paradigm shift in scientific data analysis occurring, evident in hyperspectral data and microarray analysis
- High-dimensional data means either 'go back basic algorithms', or 'try more sophisticated algorithms that work well in high-dimensions'
- Ensemble schemes and wavelet preprocessing help improve algorithms

Next steps

- Monitor landscape dynamics
- Make this a Grid service
  - **Globus - Grid**
  - **Kepler - scientific data flow**
  - **Resource nodes to store data**
- Improve accuracy
  - **Tentative classification of neighboring pixels should be used**
  - **Use other ensemble algorithms**
  - **Use other wavelets**
  - **Use other wavelet-like preprocessing**

Thanks!