Soft classification techniques

(previously using IDRISI GIS as example)

E. Baltsavias

Hard vs. Soft Classifiers

Hard

- Each classification unit (pixel, segment) belongs to ONE class

Soft

- Can belong to multiple classes
- For each class membership degree (%, or value 0-1)
- Relation to mixed pixels
- Subpixel classification
- Relation to probability, evidence
- Combine results of classification with GIS layers for additional info and evidence (if done, usually as post processing)
- Final decision with hardener possible

In-Process Classification Assessment

(used with IDRISI but could be used generally also as preprocessing step)

Part of classification with aim its improvement

- Combine hard and corresponding soft classifier
- Soft classifiers provide classification uncertainty image
- Problem cases -> those with high uncertainty

High uncertainty usually when

- Pixel appears as mixture of classes. This can be due to 2 reasons
  - poor training data: two classes can not be separated
  - mixture really exist at the spatial resolution of analysis ; for a given spatial resolution, mixture also depends on landcover (e.g. snow vs. urban), also (semi-)transparency (e.g. gaps vegetation, shallow water) on the pixel level
- Pixel belongs to unknown class (thus, with unknown spectral signature)

  -> create new class, get training/spectral library data for this class

Soft classifiers

In principle as hard classifiers, but

- **Output not a single thematic map/image (one code/colour per class), but**
- **One map per class with degree of membership to this class for each pixel AND**
- **Usually, also one map with degree of uncertainty for the decision above (uncertainty = quality measure)**

Examples of classifiers in Idrisi and other software packages (using different class membership metrics)

- **Based on Bayesian probability theory**
- **Dempster-Shafer theory**
- **Mahalanobis distance**
- **Fuzzy set theory**
- **Linear mixture model (for unmixing mixed pixels)**

All above are fuzzy (fuzzy here, synonymous to soft) measures

- **Relation to measurement errors (i.e. input image data are not error free)**
- **Relation to class definition errors (e.g. errors in training samples definition for each class)**

Soft classifiers

- Includes classification uncertainty image, which
  measures to what degree a class membership is superior to the rest.
- Classification uncertainty definition for Bayesian classification (in IDRISI)

$$\text{ClassificationUncertainty} = 1 - \frac{\max - \frac{\text{sum}}{n}}{1 - \frac{1}{n}}$$

where

max	=	the maximum set membership value for that pixel
sum	=	the sum of the set membership values for that pixel
n	=	the number of classes (signatures) considered

- Second term in above equation shows degree of commitment to a class relative to max possible commitment
- Other measures of uncertainty are possible (think about it!). Also consider that class membership is also uncertain!

Examples of classification uncertainty

Assuming a case where 3 classes are evaluated, consider those with the following allocations of membership:

(0.0 0.0 0.0)	Classification Uncertainty = 1.00
(0.0 0.0 0.1)	Classification Uncertainty = 0.90
(0.1 0.1 0.1)	Classification Uncertainty = 1.00
(0.3 0.3 0.3)	Classification Uncertainty = 1.00
(0.6 0.3 0.0)	Classification Uncertainty = 0.55
(0.6 0.3 0.1)	Classification Uncertainty = 0.60
(0.9 0.1 0.0)	Classification Uncertainty = 0.15
(0.9 0.05 0.05)	Classification Uncertainty = 0.15
(1.0 0.0 0.0)	Classification Uncertainty = 0.00

Note: the membership values are from 0 to 1. Each membership value should be non-negative.

Importance of uncertainty

-> collect new evidence (data) when uncertainty high -> gain more new information

ETH Eidgenössische Technische Hochschule Zürich

BAYESIAN CLASSIFICATION

- Similar to maximum likehihood classifier (which in reality is also soft)
- Method generates separate image for each class to express posterior probability of belonging to each class according to Bayes‘ theorem

$$p(h|e) = \frac{p(e|h) \cdot p(h)}{\sum_i p(e|h_i) \cdot p(h_i)}$$

where

$p(h|e)$ = the probability of the hypothesis being true given the evidence (posterior probability)
$p(e|h)$ = the probability of finding that evidence given the hypothesis being true
$p(h)$ = the probability of the hypothesis being true regardless of the evidence (prior probability)

Prior probabilities come from existing knowledge (even if not perfect), e.g. thematic maps

BAYESIAN CLASSIFICATION

- p(e/h) derived from variance covariance matrix from training data
- p(h/e) = same quantity that maximum likelihood evaluates
- Assumption: considered classes are the only possible classes
- That means it makes assignment to a class even with little evidence, if no support for belonging to other classes
- Used for sub-pixel classification / mixture analysis
- Posterior probabilities treated as evidence of class membership
- Assumptions:
  - classes are exhaustive, no other possible class exists
  - p(e/h) do not overlap in case of pure pixels
- These assumptions difficult to meet in practice

Fuzzy sets

- Sets (classes) without sharp (crisp) boundaries
- No binary decision
- Characterised by a fuzzy membership grade (called possibility), indicating a continuous increase from nonmembership (0) to complete membership (1)

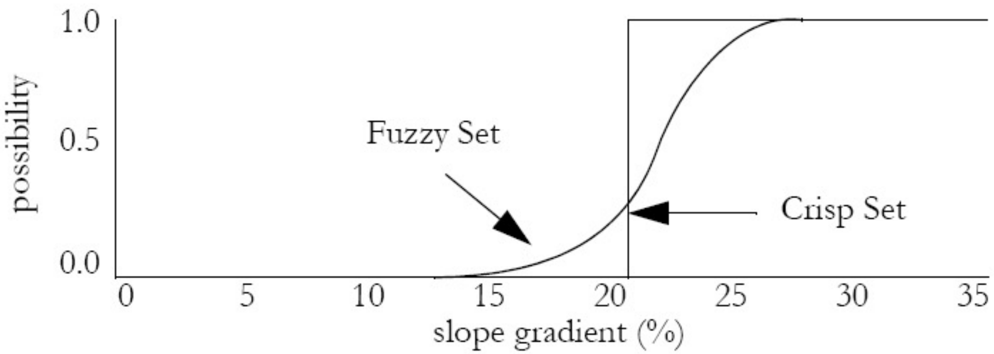

Figure 13-6 Fuzzy vs. Crisp Set Membership Functions

- Used in constructing decision rules in criteria evaluation and combination. NOT just classification

Fuzzy sets – Examples of membership functions

- Sigmoidal (most commonly used)

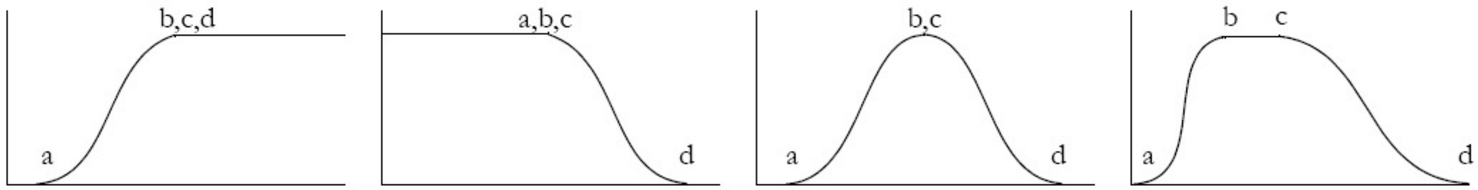

Figure13-7 Sigmoidal Membership Function

- J-shaped

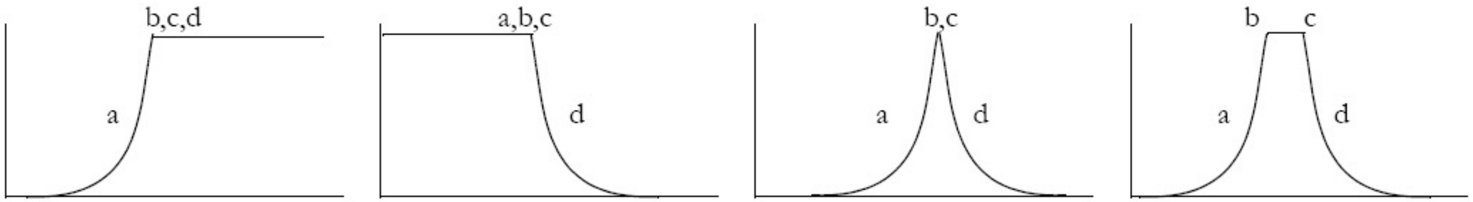

Figure 13-8 J-Shaped Membership Function

Fuzzy sets – Types of membership functions in Idrisi

- **Linear**

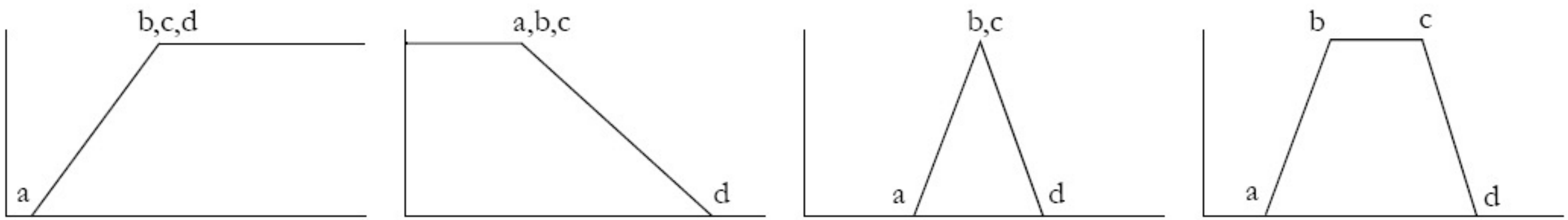

Figure 13-9 Linear Membership Function

- **User defined**

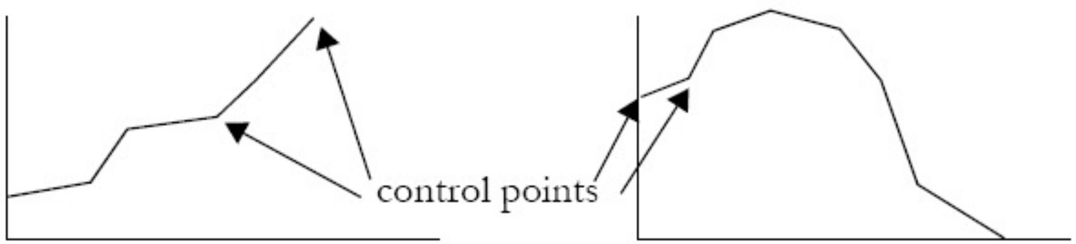

Figure 13-10 User-Defined Membership Function

Hardeners

Re-evaluate soft classification results to produce a hard classification

Practically one hardener for each of the respective soft classifiers

Examples:

- **Assign pixel to class with max posterior probability**
- **Up to N outputs possible, i.e. N highest probabilities**