9. Classification Optimization

Objective: to simplify classification and/or improve classification accuracy.

There are three stages at which one has opportunity to improve the accuracy and efficiency of classification:

1. Measurement
   - The measurements (spectral bands, bandwidths, spatial resolution, quantization, polarization, look angle, etc.) can be optimized for specific targets and specific situations. By selecting those measurements that most completely and uniquely characterize the desired targets, one can simplify and/or improve classification.
   - If the targets of interest are very distinct (separable) from all other data in measurement space, a simple classifier may be sufficient and highly accurate.
   - Select the most appropriate data available (sensor, bands, time of observation, look angle). There is usually little control over the design of the imaging system.

2. Feature Extraction
   - Knowing the characteristics of the classifiers, select the most appropriate bands and adjust the feature space to fit the way the intended classifier models the data distribution.

3. Classification
   - Match the classifier to the problem; choose the classifier that best models the data distribution.
   - Select the simplest classifier that is capable of handling the problem. There is no need to use a parametric maximum-likelihood classifier if a minimum-distance classifier will suffice.

Feature Extraction

Goal: optimize classifier performance

1. Feature reduction - reduce the number of variables needed for classification
   - Principal component analysis
   - Select one variable from pairs of highly correlated variables (only one is likely to be necessary/useful).
   - Eliminate variables which are known to be ineffective.

2. maximize separability of target classes.
   - select variables which individually maximize the separability of the target classes.
   - define new variables (combining the original variables) to maximize separability.
Measures of separability

**Divergence**: a measure of separability based on the degree of overlap of two probability distribution functions.

\[ L_{ij}(x) = \frac{p(x|\omega_i)}{p(x|\omega_j)} \Rightarrow L'_{ij}(x) = \ln[p(x|\omega_i)] - \ln[p(x|\omega_j)] \]

\[ d_{ij} = \int_x L'_{ij}(x) = \int_x \int y L'_{ij}(x) + \int_x \int y L'_{ij}(x) \]

**DETERMINATION OF OPTIMAL FEATURES**

**Principal Components Analysis**

Compresses data into a minimal number of images containing the maximum variance. Minimizes the number of significant bands given that information can be equated with variance.

**Canonical Analysis**

In the illustration above (Figure A), a principal components analysis would not provide the optimal separation between the two classes since PC2 is not oriented along the direction of maximum separability. However, one can define an axis that optimizes the separability of the two classes using the criteria that the classes have the largest possible separation between their means when projected onto that axis, while in that same projection, each class should have the minimum variance. Conceptually, the goal is to maximize the ratio:

\[ \frac{\sigma_b^2}{\sigma_w^2} = \frac{\text{between class variance}}{\text{within class variance}} \]

Since the classes are treated as multinormal distributions, the estimate of variance must be derived from the covariance matrix. The average within class covariance matrix is defined as
The within class covariance matrix is given by:

\[ \sum_w = \frac{\sum_{i=1}^{M} (n_i - 1) \Sigma_i}{S_n} \]

where \( \Sigma_i \) is the covariance matrix of the data in training set for class \( i \), \( M \) is the total number of classes, \( n_i \) is the population of the \( i \)th class and \( S_n = \sum_{i=1}^{M} n_i \).

The among class covariance matrix is given by:

\[ \sum_A = \frac{1}{M} \sum_{i=1}^{M} (m_i - m_0)(m_i - m_0)^\dagger \]

where: \( m_0 = \frac{\sum_{i=1}^{M} n_i m_i}{S_n} \)

We have explored the use of several different classifiers. Generally the better the classifier models the multivariate distribution of the target data, the more precise the classification will be and the more complex (and costly) the classifier will be to run. Often the cost of operating the classifier can be reduced by reducing the number of variables (spectral bands). In some cases new variables can be defined which will either optimize the efficiency of a classifier or will allow the use of a simpler classifier.

Reducing the number of variables is sometimes simple. For example, if two spectral bands are very highly correlated, it is unlikely that both will be necessary for discriminating among most targets. Similarly, if variance can be equated with information, one may use principal component analysis to define a reduced set of variables without loss in discriminability.