

LESSON 11: STATISTICAL TECHNIQUES FOR CLASSIFYING VEGETATION SPECIES



Aim of Lesson

Using statistical techniques to classify vegetation species with hyperspectral imagery

Objectives

1. Introduction to statistical techniques for classifying hyperspectral imagery
2. Using binary classifiers for multi-class problems
3. Reducing salt and pepper by smoothing classified images
4. Introduction to the Discrete Wavelet Transform

Background Information

This lesson relates to the Sections 9.1.2 and 9.1.4.2 of the HyperTeach Theory Syllabus. For further details about the techniques involved readers are recommended to read these sections. This lesson describes a method to classify vegetation species in the nature reserve 'De Westhoek' using AISA imagery. 'De Westhoek' is located at the West coast of Belgium near the French border.

SOFTWARE

A PC with ENVI® software is required to carry out the lesson.

IMAGE DATA

The AISA image with 32 bands is used (AISA_bands.img).

A pixel based classification image is also provided (AISA_class.img) and the output of a Discrete Wavelet Transform Transformation (AISA_dwt.img). Please refer to lesson 10 for more information of the data.

Lesson outline

1. Introduction to statistical techniques for classifying hyperspectral imagery

1.1 Fundamentals

Supervised classification, as opposed to unsupervised classification focuses on pre-defined classes. The classifier can be built with a training set and later be used for classifying new objects. Supervised classification requires ground truth data, for labelling the classes, needed for training. Unsupervised classification clusters the data without prior definition of the groups. This lesson will only focus on supervised classification.

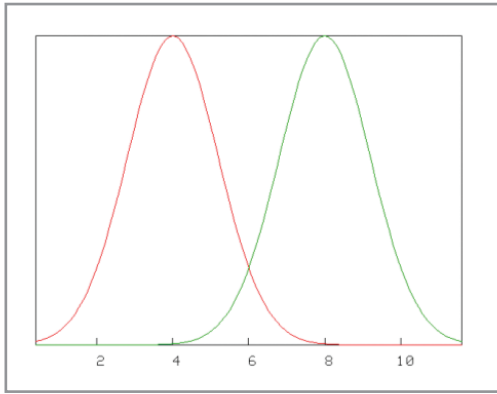
There are many methods for supervised classification. Among them, linear discriminant analysis (LDA) is very popular. Although relying on strong assumptions that are often not met in practice, this method has proven to be very useful.

- Discriminant analysis is based on an assumption of normality.
- In addition, linear discriminant analysis assumes that all the classes have the same variance.

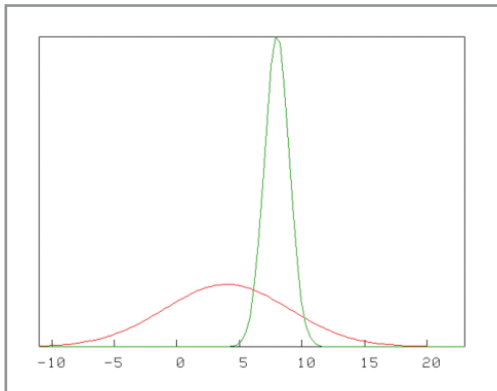
Question 11.1: The following distributions are all normal. Where would you define a classifier and how do the distributions differ? Are the assumptions for LDA met in all figures? What can you tell about the distribution in Figure 11.3?

How should we deal with prior knowledge about the class probabilities?

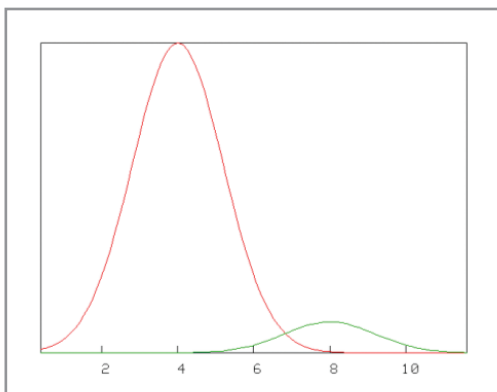
LESSON 11: STATISTICAL TECHNIQUES FOR CLASSIFYING VEGETATION SPECIES



>> Figure 11.1: Class probability distribution 1 for a given output variable x.



>> Figure 11.2: Class probability distribution 2 for a given output variable x



>> Figure 11.3: Class probability distribution 3 for a given output variable x

Question 11.2: The training and test set used for validation must be two different sets. Why?

Question 11.3: What is meant by over-training? Is there a relationship between the dimensionality of the variables and the number of training samples required? What does this mean for classification of hyperspectral data?

Question 11.4: Often, the available ground truth is split. Half of the data is used for training and half for validation. In case few ground truth is available, what are the options you can think of?

Linear Discriminant Analysis (LDA) uses the notion of scatter matrices⁵. Similar to covariance matrices⁵, they indicate the extend to which two random variables co-vary. By definition, the covariance matrix is symmetric. Also, the covariance of any component of the random variable with itself, is that component's variance. The magnitude of a covariance between component X_j and a component X_k depends on the standard deviations of both X_j and X_k . To obtain a more direct indication on how two components co-vary, we scale covariance to obtain correlation:

$$\rho_{jk} = \frac{\text{cov}(X_j, X_k)}{\sigma_j \sigma_k},$$

LESSON 11: STATISTICAL TECHNIQUES FOR CLASSIFYING VEGETATION SPECIES



where σ_j and σ_k are the standard deviations of X_j and X_k , respectively. By construction, the correlation is always between -1 and +1.

2. Using binary classifiers for multi-class problems

Fisher LDA provides a solution for both binary and multi-class problems. From the Theory Syllabus, it is clear that the projection vector is easier to compute for binary classes than for the multi-class problem. However, the multi-class problem can be solved by combining binary classifiers. Often, this method even turns out to outperform the direct multi-class solution.

Question 11.5: Suggest two methodologies to combine binary classifiers to solve the multi-class problem.

3. Reducing salt and pepper by smoothing classified images

Pixel based classification results are cursed with a noise referred to as salt and pepper. Individual pixels with label A are in contrast to homogeneous patches with label B from which they are surrounded. Not only, the effect is less "pleasing" to the end user, small isolated areas are less likely to occur in nature and tend to be misclassifications.

These artefacts can be reduced using post-classification methods. A quick and dirty method is based on a majority analysis, using a specific kernel (moving window). This method is included in the ENVI package.

Action 1: Open the classification image AISA_class.img in ENVI.

Perform a smoothing operation using the post classification method in ENVI

Hint: Menu: Classification->Post Classification->Majority/Minority Analysis, with a kernel size of 3x3. Select all classes, and specify an output filename. Compare the two classification images using Tools->Link->Link Display. Click on the image and check if single pixels are reduced. The same exercise can be performed using a kernel of size 5x5.

4. Introduction to the Discrete Wavelet Transform

Action 2: Open the multi band images AISA_bands.img and AISA_dwt.img in ENVI. This image corresponds to the 32 coefficients obtained from a Discrete Wavelet Transform performed on the AISA image. The filter used is the HAAR filter. Compare the spectra of the two images, using the Z-profile build-in function.

Question 11.6: Where are the low scale coefficients located, where are the high scale coefficients located?

Question 11.7: How many bands are used for each scale?

Tip: Think about the Heisenberg uncertainty and the downscaling after each filtering operation.

RELATED PAPERS

Duda, R.O., Hart P. E., and Stork D.G., Pattern Classification, 2nd, Edition. Wiley, 2001.

Answers to questions

Answer 11.1: The following distributions are all normal. Where would you define a classifier and how do the distributions differ? Are the assumptions for LDA met in all figures? What can you tell about the distribution in Figure11.3 ?

How should we deal with prior knowledge about the class probabilities?

A classifier can be defined with its threshold for decision located at the cross section of both distributions. In the second plot, two thresholds must be defined. This distribution does not meet the assumption of equal standard deviations. All three plots have identical means. The last plot shows an example of different prior probabilities⁶. Notice how the ideal thresholds differ in plot 11.1 and plot 11.3, though means and standard deviations are identical. Prior probabilities should be multiplied to the conditional distributions to obtain the posterior probabilities needed for classification (Bayes).

Answer 11.2: The training and test set used for validation must be two different sets. Why?

Training and test set used for validation must be two different sets, because of positive bias. No knowledge of validation data can be used during training, or the estimation of accuracy is too optimistic.