Notes on Statistical Pattern Recognition

Compiled by D. Gueorguiev , 11/22/24

# Elementary Decision Theory

## Bayes’ Decision Rule for Minimum Error

Consider classes , , with *a priori* probabilities (the probabilities of each class occurring) , assumed known. If we wish to minimize the probability of making an error and we have no information regarding an object other than the class probability distribution, then we would assign an object to class if

,

This classifies all objects as belonging to one class: the class with the largest prior probability. For classes with equal prior probabilities, patterns are assigned arbitrarily between those classes.

However, we do have an *observation vector* or *measurement vector* and we wish to assign an object to one of the classes based on the measurements . A decision rule based on probabilities is to assign (here we refer to an object in terms of its measurement vector) to class if the probability of class given the observation , that is , is greatest over all classes . That is, assign to class if

, (1)

This decision rule partitions the measurement space into regions such that if then belongs to class . The regions may be disconnected.

The *a posteriori* probabilities may be expressed in terms of the *a priori* probabilities and the class conditional density functions using Bayes’ theorem as

and so the decision rule (1) may be written: assign to if

, (2)

This is known as Bayes’ rule for *minimum error*.

For two classes, the decision rule (1.2) may be written:

implies

The function is known as the *likelihood ratio*.

The Figure 1 below shows an example of two-class discrimination problem. Class is normally distributed with zero mean and unit variance, . Class is a *normal mixture* (a weighted sum of normal densities) .

A diagram of a function

Description automatically generatedFigure 1: for classes and : for in region , is assigned to class .

The function

A diagram of a function

Description automatically generatedFigure 2: the likelihood ratio against the ratios of the priors

Let us understand why decision rule (2) minimizes the misclassification error-

The probability of making a misclassification error is given with

(3)

where is the probability of misclassifying patterns from class . This is given by

(4)

the integral of the class-conditional density function over , the region of measurement space outside ( is the complement operator). Thus .

# References

[Statistical Pattern Recognition, Andrew Webb, Keith Copsey, 3rd Edition, 2011](https://github.com/dimitarpg13/image_processing/blob/main/literature/books/Statistical_Pattern_Recognition_Webb_Copsey_3rdEd.pdf)

[Pattern Recognition and Machine Learning, Christopher Bishop, 2006](https://github.com/dimitarpg13/image_processing/blob/main/literature/books/Bishop-Pattern_Recognition_and_Machine_Learning.pdf)

[Pattern Classification, Richard O. Duda, Peter E. Hart, David G. Stork, Second Edition, 2nd Edition, 2001](https://github.com/dimitarpg13/image_processing/blob/main/literature/books/Richard_Duda_Peter_Hart_David_Stork-Pattern_classification_2001.pdf)