THE NEW CONCEPT FOR INTER-CLASS SEPARABILITY AND FAST MAXIMUM LIKELIHOOD CALCULATION

Minoru AKIYAMA
Geographical Survey Institute of Japan
Kitasato-1, Tsukuba-shi, Ibaraki-ken 305
JAPAN
Commission III

1. Introduction

The per pixel classification method applied in remote sensing is to classify a pixel to the most likely class by comparing the statistics of classes with the gray levels of the pixel. In order to improve the classification performance, classes should be well separated each other. There are several indices to represent inter-class separability, however, satisfactory index has not yet been proposed.

Here, new index for inter-class separability corresponding the maximum likelihood classification will be proposed. This index corresponds to a rate of points whose probability to class A (likelihood) is greater than that of the point which shows the maximum likelihood among the misclassified points.

Furthermore, classification time could be reduced by calculating these inter-class separability in advance.

2. Inter-class separability

Figure-1 shows one dimensional likelihood functions. Here, all points of class A in $x_2<x<x_3$ are mis-classified to class B. Generally speaking, separability is considered to be high when S_{AB}^* the rate of points in this area, is small. However, it is difficult to calculate this value exactly. Therefore, we usually refer other indices as a substitute of separability.

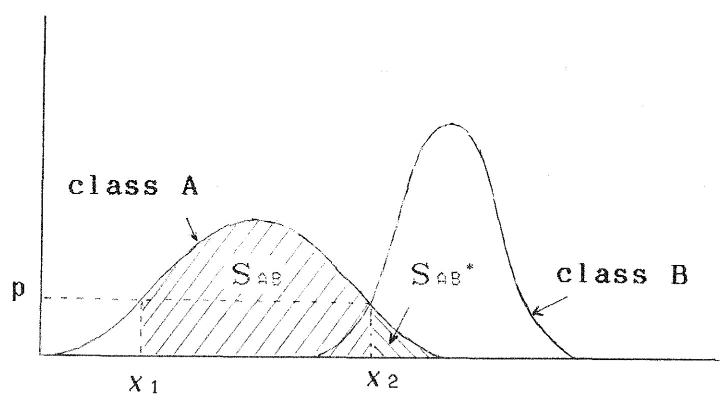

Figure-1 Probability functions

In this figure, likelihood functions of classes A and B intersect at points x_2 and x_3, where likelihood at x_2 is greater than that of point x_3. Let p be the likelihood at x_2 and S_{AB} be the area in $x_1<x<x_2$ which is the area of points whose likelihood is greater than p. Obviously $S_{AB} \neq 1 - S_{AB}^*$, however, the area corresponding the difference between S_{AB} and $(1 - S_{AB}^*)$

is the area of low likelihood and of little importance. Therefore, ignoring this difference, S_{AB} was adopted as a new index for inter-class separability.

In other words, the new concept for inter-class separability was defined as follows.

a) The separability of class A against class B is expressed by S_{AB}.
b) S_{AB} is the rate of points whose likelihood is greater than p,
c) where p is the maximum likelihood among the points misclassified to B.

Figure-2 shows rather extreme case. In this case, the area of points of class A which is misclassified to class B corresponds to S_{AB}^* of $x_1<x<x_2$. This is obviously the area of high likelihood of class A. Therefore, it is not appropriate to consider that the separability of class A against class B is high only if $1-S_{AB}^*$ is large. In other words it is not appropriate, in this case, to take S_{AB}^* as the separability of A against B.

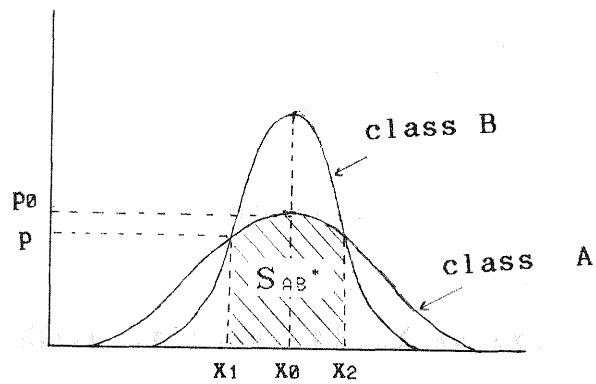

Figure-2 Probability functions (extreme case)

In accordance with the new concept, the point which shows the maximum likelihood among misclassified points is x_0, the mean of class A. And there is no points in class A whose likelihood is greater than p_0, the likelihood at point x_0. Therefore, S_{AB}, the separability of class A against class B is obtained to be zero.

On the contrary, S_{BA}, the separability of class B against class A is defined as the area of class B in $x_1<x<x_2$, which takes rather large value. In other words, in this case, it can be said that class B is well separated from class A while class A is not separated from class B.

3. Multi-dimensional inter-class separability

The likelihood that a point χ is belonged to class A in multi-dimensional feature space is expressed as follows.

$$P(\chi \mid A)=\frac{1}{(2\pi)^{n/2}\mid \Sigma_A \mid^{1/2}}\exp\left\{-\frac{1}{2}\,{}^t(\chi-\mu_A)\Sigma_A^{-1}(\chi-\mu_A)\right\} \qquad (1)$$

where $P(\chi \mid A)$; likelihood of point χ to class A
Σ_A ; variant matrix of class A
μ_A ; mean vector of class A

Usually, the following value is calculated and compared instead of (1).

$$g(\chi \mid A)=2 \log P(\chi \mid A)$$
$$=-n \log 2\pi -\log \mid \Sigma_A \mid -{}^t(\chi-\mu_A)\Sigma_A^{-1}(\chi-\mu_A)$$
$$=-C_A -{}^t(\chi-\mu_A)\Sigma_A^{-1}(\chi-\mu_A) \quad (2)$$

The intersection between likelihood functions of classes A and B is obtained as the points which satisfy the following condition.

$$g(\chi \mid A)=g(\chi \mid B) \quad (3)$$

The point which shows the maximum $g(\chi \mid A)$ among points of above mentioned intersection is the point of maximum likelihood among misclassified points. Let χ_0 be such point, then it satisfies

$$g(\chi_0 \mid A)-g(\chi_0 \mid B)=0$$

that is

$${}^t(\chi_0-\mu_A)\Sigma_A^{-1}(\chi_0-\mu_A)-{}^t(\chi_0-\mu_B)\Sigma_B^{-1}(\chi_0-\mu_B)=C_B-C_A \quad (4)$$

It also satisfies

$$d\, g(\chi_0 \mid A)=0$$

In other words,

$${}^t(\chi_0-\mu_A)\Sigma_A^{-1}d\chi=0 \quad (5)$$

for any $d\chi$ which satisfies the following.

$$d\,\{g(\chi_0 \mid A)-g(\chi_0 \mid B)\}=0$$

$$\{{}^t\chi_0(\Sigma_A^{-1}-\Sigma_B^{-1})-({}^t\mu_A\Sigma_A^{-1}-{}^t\mu_B\Sigma_B^{-1})\}\, d\chi=0 \quad (6)$$

Therefore, the following two vectors should be parallel.

$${}^t\chi_0(\Sigma_A^{-1}-\Sigma_B^{-1})-({}^t\mu_A\Sigma_A^{-1}-{}^t\mu_B\Sigma_B^{-1})$$
$${}^t(\chi_0-\mu_A)\Sigma_A^{-1}$$

This condition can be expressed as

$$(\Sigma_A^{-1}-\Sigma_B^{-1})\chi_0-(\Sigma_A^{-1}\mu_A-\Sigma_B^{-1}\mu_B)=k\Sigma_A^{-1}(\chi_0-\mu_A) \quad (7)$$

Or

$$\{(1-k)\Sigma_A^{-1}-\Sigma_B^{-1}\}\ \chi_0=(1-k)\Sigma_A^{-1}\mu_A-\Sigma_B^{-1}\mu_B \quad (8)$$

where k is a scalar. Here, let m=1-k, then we obtain

$$\chi_0=\{m\Sigma_A^{-1}-\Sigma_B^{-1}\}^{-1}\ \{m\Sigma_A^{-1}\mu_A-\Sigma_B^{-1}\mu_B\} \quad (9)$$

Finally, χ_0 can be obtained as a point to satisfy the equations (4) and (9) simultaneously.

Let P_0 be the likelihood at χ_0. P_0 is, of course, the likelihood of class A as well as of class B at point χ_0.

$$P_0 = P(\chi_0 \mid A) = P(\chi_0 \mid B)$$

$$= \frac{1}{(2\pi)^{n/2} \mid \Sigma_A \mid^{1/2}} \exp \{-\frac{1}{2} {}^t(\chi_0 - \mu_A) \Sigma_A^{-1} (\chi_0 - \mu_A)\} \tag{10}$$

$$= \frac{1}{(2\pi)^{n/2} \mid \Sigma_B \mid^{1/2}} \exp \{-\frac{1}{2} {}^t(\chi_0 - \mu_B) \Sigma_B^{-1} (\chi_0 - \mu_B)\} \tag{11}$$

S_{AB} , the separability of class A against class B, is obtained by the following.

$$S_{AB} = \int_{P \geqq P_0} P(\chi \mid A)\, d\chi \tag{12}$$

similarly, S_{BA} , the separability of class B against class A, is obtained by the following.

$$S_{BA} = \int_{P \geqq P_0} P(\chi \mid B)\, d\chi \tag{13}$$

Generally $S_{AB} \neq S_{BA}$

4. Fast maximum likelihood calculation

In maximum likelihood method, the probability for all classes should be calculated and compared to each other to find the most likely class. This makes the calculation time too long. However, if each classes are well separated to each other, probabilities to most classes are quite small, and only a few classes give eminent value for a point. In this regard, there is a possibility to reduce calculation time by omitting classes of obviously small probability for each point.

This can be accomplished by calculating inter-class separability proposed here for each combination of classes in advance. Let $P_0(AB)$ be the probability derived from (10) or (11) between classes A and B. If the probability of a point to class A is greater than $P_0(AB)$, as follows,

$$P(\chi \mid A) > P_0(AB)$$

then obviously

$$P(\chi \mid A) > P(\chi \mid B)$$

Therefore, it is clear that the point is never belonged to class B without calculating probability to class B.

The effect of time reduction depends on the algorithm to decide the order of classes to be calculated, however, the more classes the more reduction can be expected.