**Date:** 20 October, 2015

**Experiment No. 11**

**Aim:** To obtain Fisher’s discriminant function.

**Experiment:**  In a study examining the possible difference in general proficiency of class 10 students, the class is divided into two groups and each group is given a test. The mean score in English (X1), history (X2), geography (X3), general science (X4) and mathematics (X5) of 98 students of group 1 and group 2 is obtained as follows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mean scores | X1 | X2 | X3 | X4 | X5 |
| X(1) group 1 | 62.1 | 65.24 | 61.31 | 70.31 | 75.25 |
| X(2) group 2 | 36.24 | 58.33 | 59.26 | 48.21 | 45.09 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 70.553 | 24.869 | 35.07 | 36.207 | 40.531 |
| 24.869 | 19.942 | 21.224 | 19.642 | 21.421 |
| 35.07 | 21.224 | 27.751 | 24.181 | 29.031 |
| 36.207 | 19.642 | 24.181 | 33.116 | 27.471 |
| 40.531 | 21.421 | 29.031 | 27.471 | 39.361 |

S =

1. Obtain Fisher’s discriminant function for the given data.
2. Assign the mean score vector.

**Theory:**

**Fisher Linear Discriminant Analysis**

Under the same setup assume that, 𝑋 | 𝜋1 ~ (𝜇1,Σ) where 𝜇1 is the mean vector of the 1st population and Σ is the covariance matrix for the 1st population, and 𝑋 | 𝜋2 ~ (𝜇2,Σ) where 𝜇2 is the mean vector of the 2nd population and Σ is the common covariance matrix for both the populations.

Further change 𝜋1 and 𝜋2 into two univariate populations by changing 𝑋 to some 𝑙′𝑋 by means of “some” 𝑙.

𝑋 | 𝜋1 ~ (𝜇1,Σ)⇒ 𝑙′𝑋 | 𝜋1 ~ (𝑙′𝜇1,𝑙′𝛴𝑙)

𝑋 | 𝜋2 ~ (𝜇2,Σ)⇒ 𝑙′𝑋 | 𝜋2 ~ (𝑙′𝜇2,𝑙′𝛴𝑙)

**Discrimination:** We are interested in finding or choosing 𝑙 such that the separation between the two univariate populations is maximum, i.e. maximization of statistical distance between 𝜋1 and 𝜋2 with respect to 𝑙.

A measure of statistical distance between the two populations is given by,

(𝑙′𝜇1−𝑙′𝜇2)2/(𝑙′Σ𝑙) = (𝑙′(𝜇1−𝜇2)) 2/(𝑙′Σ𝑙)

We want to obtain, 𝑙=argmax(𝑙′(𝜇1−𝜇2))2 / (𝑙′Σ𝑙)

Assuming that Σ is positive definite and defining 𝑎′=𝑙′Σ1/2 . We have, (𝑎′Σ−1/2(𝜇1−𝜇2)) 2/ (𝑎′𝑎)…(∗)

Using Cauchy Schwartz Inequality,

(𝑎′Σ−1/2(𝜇1−𝜇2)) 2 /(𝑎′𝑎)≤((𝑎′𝑎)((𝜇1−𝜇2)′Σ−1(𝜇1−𝜇2)))/ (𝑎′𝑎) =(𝜇1−𝜇2)′Σ−1(𝜇1−𝜇2): Mahalanobis Distance

Hence we have the distance between two populations is always less than or equal to the Mahalanobis Distance. (𝑙′(𝜇1−𝜇2)) 2 𝑙′Σ𝑙 is maximum when,

𝑎′=(𝜇1−𝜇2)′Σ−1/2⇒𝑙′=(𝜇1−𝜇2)′Σ−1∶ optimal 𝑙

Thus we have the optimal 𝑙 which provides the maximum separation (discrimination) between the two populations.

The quantity 𝑙′𝑋=(𝜇1−𝜇2)′Σ−1𝑋 is called the **Fisher Linear Discriminant Function (LDF),** which is an optimal separation between the two populations.

**Classification:** Let x0 be mean vector to be assigned tone of the two groups.

Define A1= (x0- X(1))’S-1(x0- X(1)) and A2= (x0- X(2))’S-1(x0- X(2)).

If A1 A2, assign x0 to group I else to group II.

**Algorithm:**

1. Open the file “in11.txt” to read the data and “out11.txt” to write the results using pointers.
2. Calculate inverse of the dispersion matrix.
3. Multiply it with the vector d̲’= (X(1) - X(2))’.
4. Then obtain A1 and A2 given and check for the possible assignment of the new mean score vector in the required group.
5. Results are expected in the file “out11.txt”.

**Results:**

The value of A1 is 63.912983.

The value of A2 is 78.831367.

**Conclusion:**

Since, A1 is less than A2, therefore the given mean score vector is assigned to group I.