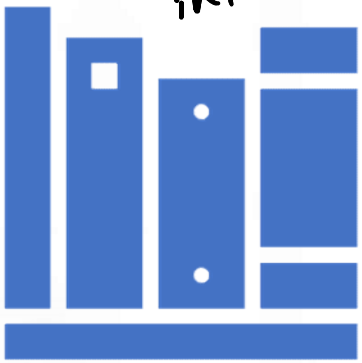
classification —— Prediction → Predicting some value

classify the data into diff. categories.

Predictive Modelling



categorical

continuous value

i.e. $-\infty < x < \infty$

LINEAR DISCRIMINANT ANALYSIS

We want to know whether somebody has lung cancer.

Hence, we wish to predict a Yes or No outcome. → dep. variable

ind. variable

Possible predictor variables: number of cigarettes smoked a day, coughing frequency and intensity etc.

We want to know whether a soap product is Good or Bad based on several measurements on the product such as weight, volume, people's preferential score, smell, color contrast etc. the object here is soap, the class category or the group ("good" and "bad") is what we are looking for (it is also called dependent variable)

$$\begin{bmatrix} \; \end{bmatrix}^{100\times100}_{n\times n} \longrightarrow \begin{bmatrix} \; \end{bmatrix}^{80\times80}_{n-m\times n-m}$$

Face Recognition: In the field of Computer Vision, face recognition is a very popular application in which each face is represented by a very large number of pixel values.

Linear discriminant analysis (LDA) is used here to reduce the number of features to a more manageable number before the process of classification.

variable

$y_1 \to \text{Yes} - \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $\quad \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix}$

$y_2 \to \text{No} - \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

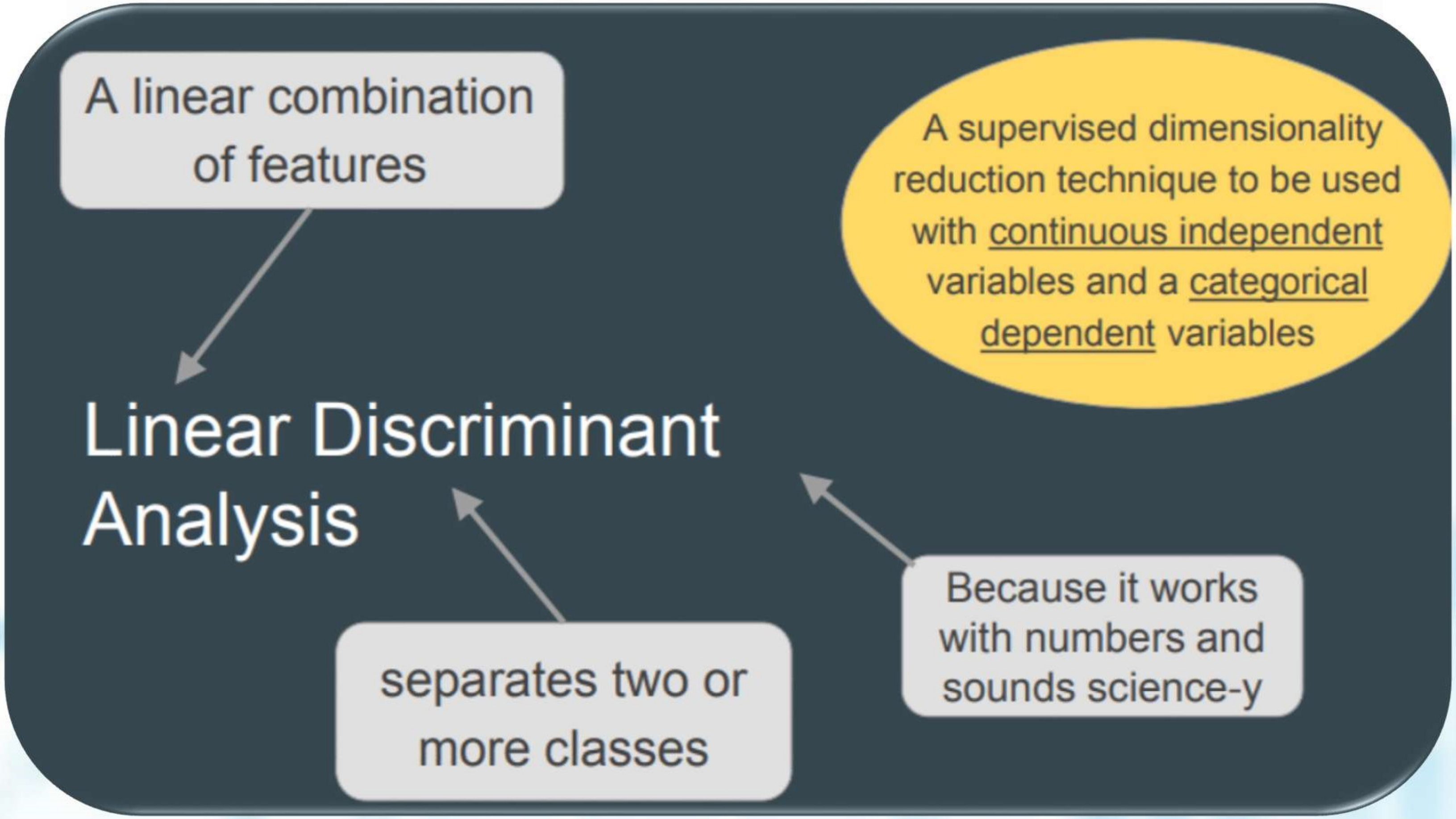
$[y] =$ $\begin{array}{cccc} x_1 & x_2 & \cdots & x_p \end{array}$

Medical: In this field, Linear discriminant analysis (LDA) is used to classify the patient disease state as Mild, Moderate or Severe based upon the patient various parameters and the medical treatment he is going through.

This helps the doctors to intensify or reduce the pace of their treatment.

Market Basket Analysis

Customer Identification: Suppose we want to identify the type of customers which are most likely to buy a particular product in a shopping mall. By doing a simple question and answers survey, we can gather all the features of the customers. Here, Linear discriminant analysis will help us to identify and select the features which can describe the characteristics of the group of customers that are most likely to buy that particular product in the shopping mall.


A linear combination of features
A supervised dimensionality reduction technique to be used with continuous independent variables and a categorical dependent variables
Linear Discriminant Analysis
separates two or more classes
Because it works with numbers and sounds science-y

Introduction

- Discriminant analysis is the appropriate statistical techniques when the dependent variable is a categorical (nominal or nonmetric) variable and the independent variables are metric variables.

- In many cases, the dependent variable consists of two groups or classifications, for example, male versus female or high versus low.

- In other instances, more than two groups are involved, such as low, medium, and high classifications.

Introduction

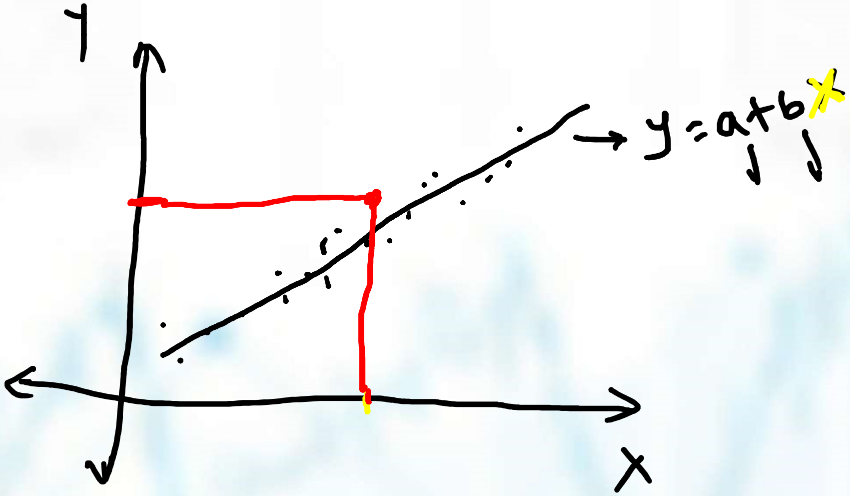
- Discriminant analysis is capable of handling either two groups or multiple (three or more) groups.

- When the criterion variable has two categories, the technique is known as two-group discriminant analysis.

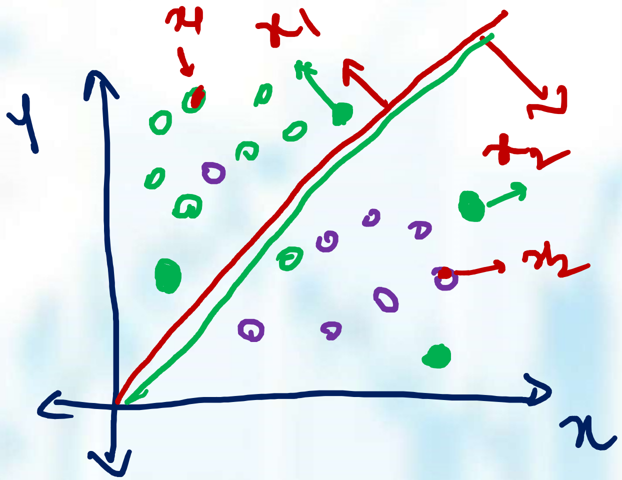
- When three or more categories are involved, the technique is referred to as multiple discriminant analysis.

Introduction

- If we can assume that the groups are linearly separable, we can use linear discriminant model (LDA).
- Linearly separable suggests that the groups can be separated by a linear combination of features that describe the objects.

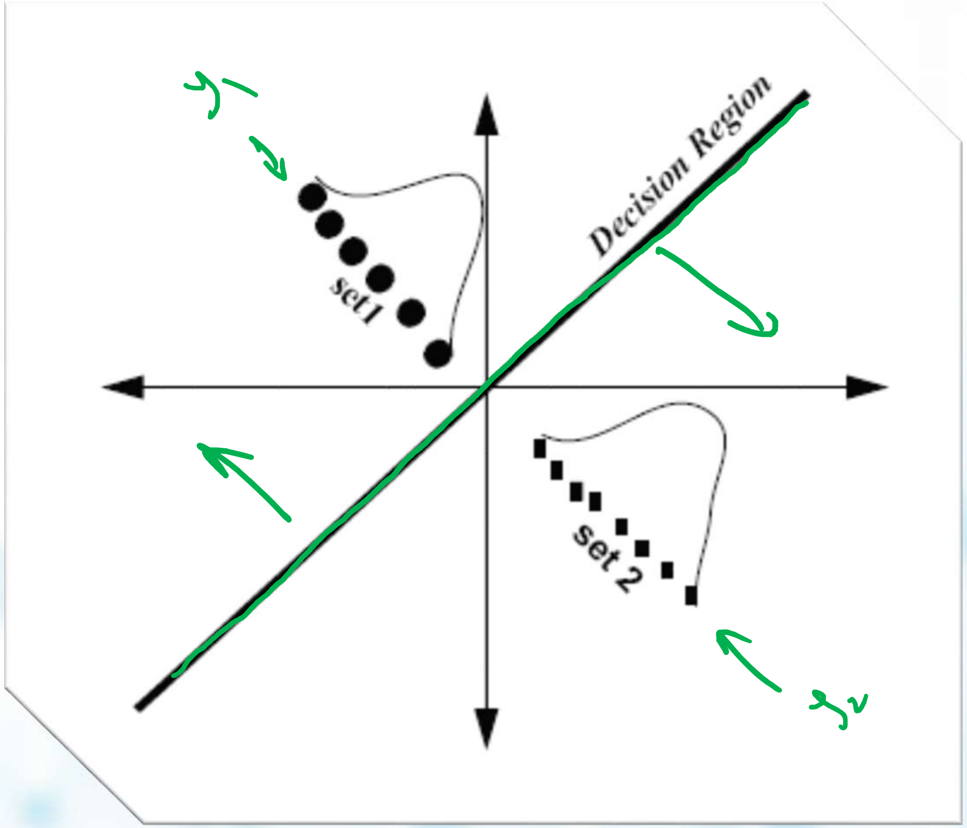


MLR

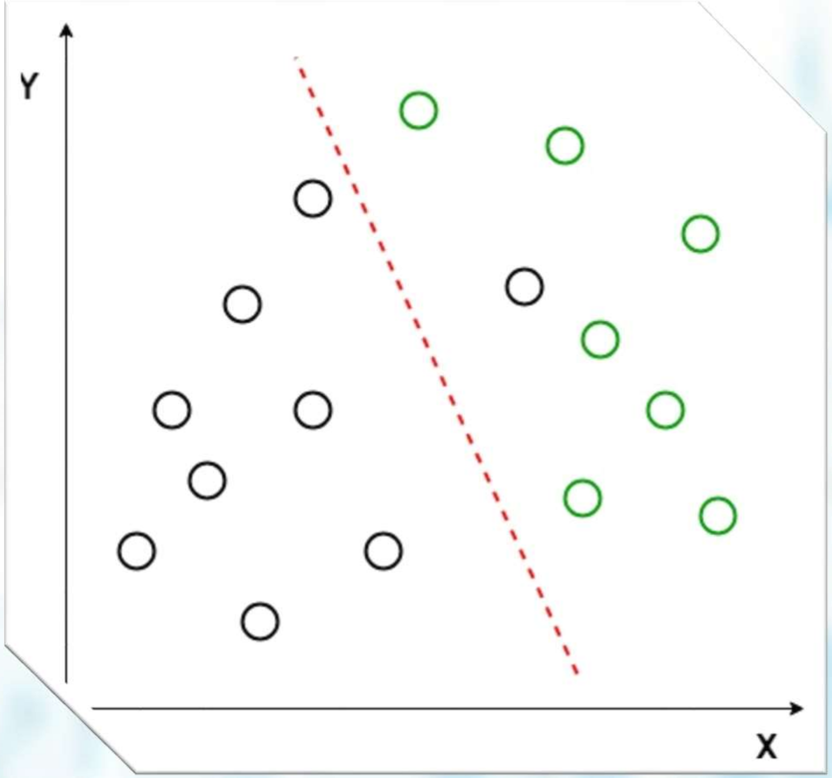


Visualisation (Two Outcomes)

If only two independent variables, the separators between objects group will become lines.



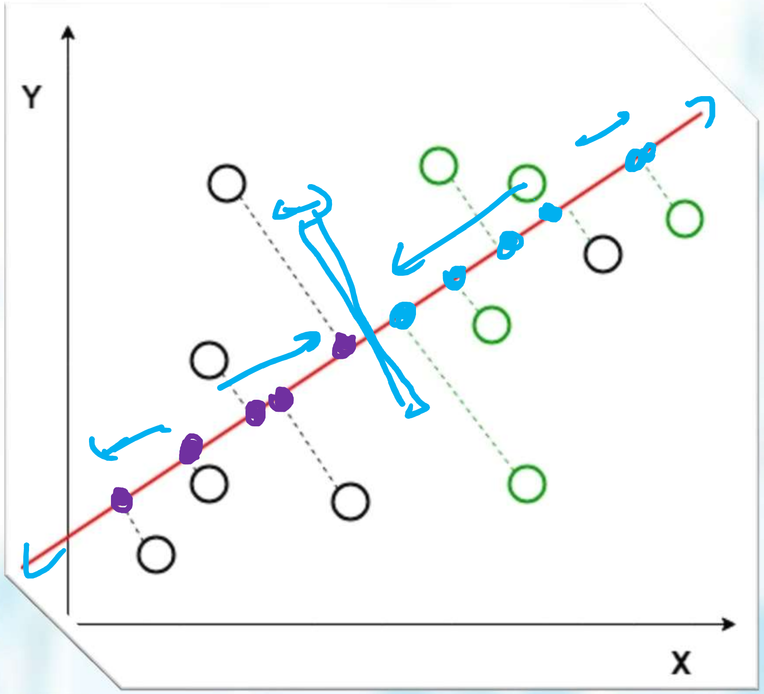
- Suppose we have two sets of data points belonging to two different classes that we want to classify.
- As shown in the given 2D graph, when the data points are plotted on the 2D plane, there's no straight line that can separate the two classes of the data points completely.
- Hence, in this case, LDA (Linear Discriminant Analysis) is used which reduces the 2D graph into a 1D graph in order to maximize the seperability between the two classes.



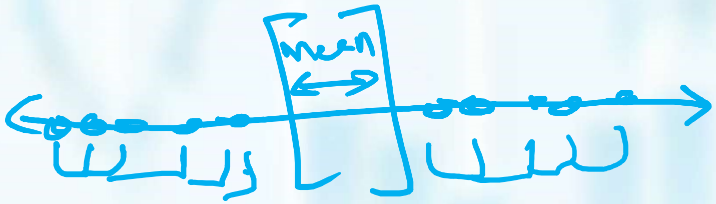
- Here, Linear Discriminant Analysis uses both the axes (X and Y) to create a new axis and projects data onto a new axis in a way to maximize the separation of the two categories and hence, reducing the 2D graph into a 1D graph.

- Two criteria are used by LDA to create a new axis:

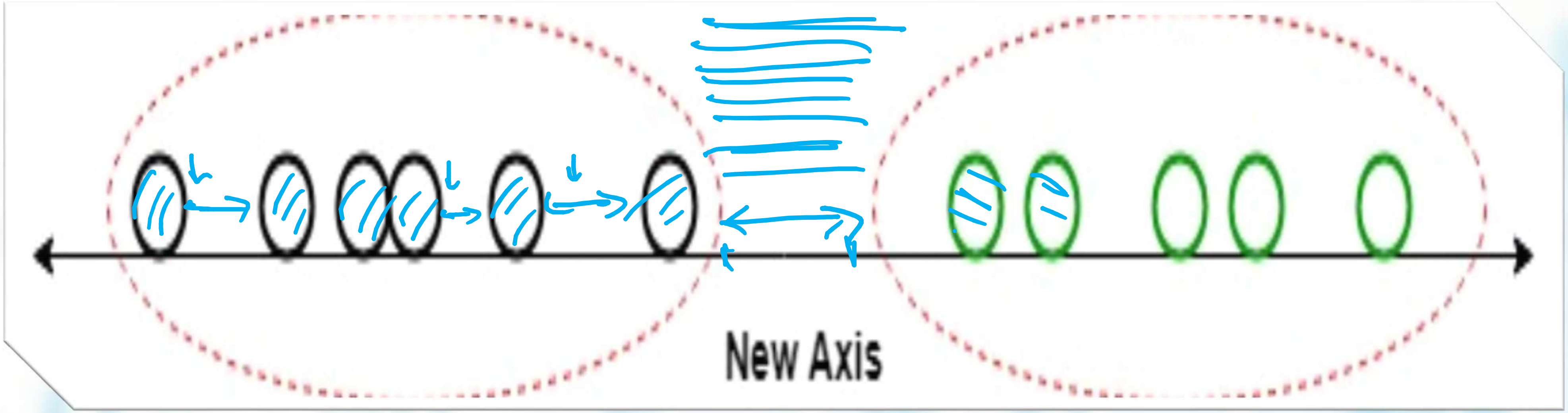
1. Maximize the distance between means of the two classes.
2. Minimize the variation within each class.



variance ↓



After generating this new axis using the above-mentioned criteria, all the data points of the classes are plotted on this new axis and are shown in the figure given below.



- The discriminant analysis model involves linear combinations of the following form:

$$D = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_kX_k$$

where

- D = discriminant score
- b 's = discriminant coefficient or weight
- X 's = predictor or independent variable
- The coefficients, or weights (b), are estimated so that the groups differ as much as possible on the values of the discriminant function.
- This occurs when the ratio of between-group sum of squares to within-group sum of squares for the discriminant scores is at a maximum.

Conducting Discriminant Analysis

Formulate the Problem

↓

Estimate the Discriminant Fun^c coeff.

↓

Determine the significance of the Discriminant function

↓

Intepret the results

↓

Assess validity of Discriminant Analysis

Linear Discriminant Analysis

1. Class Means:

$$\mu_1 = \frac{1}{N_1} \sum_{x \in \omega_1} x$$

$$\mu_2 = \frac{1}{N_2} \sum_{x \in \omega_2} x$$

2. Covariance Matrices:

$$S_1 = \sum_{x \in \omega_1} (x - \mu_1)(x - \mu_1)^T$$

n-1

$$S_2 = \sum_{x \in \omega_2} (x - \mu_2)(x - \mu_2)^T$$

n-1

3. **Within-class scatter matrix:**

$$S_w = S_1 + S_2$$

4. **The LDA projection is**

$$w^* = S_W^{-1}(\mu_1 - \mu_2)$$

Examples:

Compute the Linear Discriminant projection for the following two dimensional dataset.

Good / Yes →

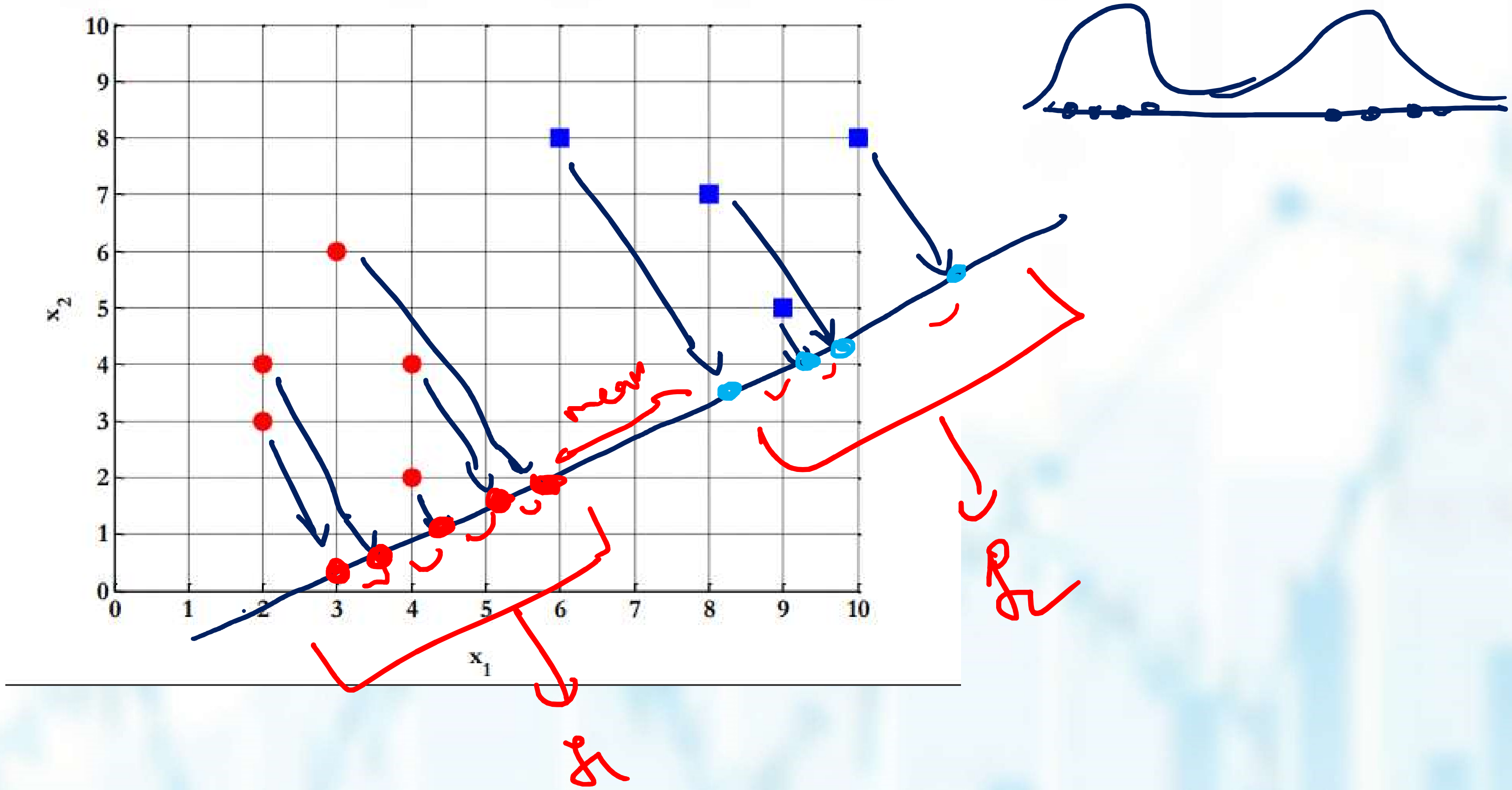
```
% samples for class 1
X1 = [4,2;
      2,4;
      2,3;
      3,6;
      4,4];
```

$n_1 = 5$

Bad / No →

```
% samples for class 2
X2 = [9,10;
      6,8;
      9,5;
      8,7;
      10,8];
```

$n_2 = 5$



x2
x1
mean

$$\mu_1 = \frac{\Sigma x_i}{n_1} = \frac{1}{5}\left[\begin{pmatrix}4\\2\end{pmatrix}_{2\times1} + \begin{pmatrix}2\\4\end{pmatrix}_{2\times1} + \begin{pmatrix}2\\3\end{pmatrix} + \begin{pmatrix}3\\6\end{pmatrix} + \begin{pmatrix}4\\4\end{pmatrix}\right] = \begin{bmatrix}3\\3.8\end{bmatrix}$$

$$\mu_2 = \frac{\Sigma x_i}{n_2} = \frac{1}{5}\left[\begin{pmatrix}9\\10\end{pmatrix} + \begin{pmatrix}6\\8\end{pmatrix} + \begin{pmatrix}9\\5\end{pmatrix} + \begin{pmatrix}8\\7\end{pmatrix} + \begin{pmatrix}10\\8\end{pmatrix}\right] = \begin{bmatrix}8.4\\7.6\end{bmatrix}$$

$$S_1 = \Sigma(x-\mu_1)(x-\mu_1)' = \begin{pmatrix}4-3\\2-3.8\end{pmatrix}\begin{pmatrix}4-3 & 2-3.8\end{pmatrix} + \begin{pmatrix}2-3\\4-3.8\end{pmatrix}\begin{pmatrix}2-3 & 4-3.8\end{pmatrix}$$

$$+ \cdots + \begin{pmatrix}4-3\\4-3.8\end{pmatrix}\begin{pmatrix}4-3 & 4-3.8\end{pmatrix}$$

$$= \begin{bmatrix}4 & -1\\-1 & 8.8\end{bmatrix}$$

$$S_1 = \begin{bmatrix} 4/4 & -1/4 \\ -1/4 & 8.8/4 \end{bmatrix} = \begin{bmatrix} 1 & -0.25 \\ -0.25 & 2.2 \end{bmatrix} \longrightarrow S_1$$

$$S_2 = \frac{1}{n-1} \Sigma (x-\mu_2)(x-\mu_2)' = \begin{bmatrix} 9.2/4 & -0.2/4 \\ -0.2/4 & 13.2/4 \end{bmatrix} = \begin{bmatrix} 2.3 & -0.05 \\ -0.05 & 3.3 \end{bmatrix} \longrightarrow S_2$$

$$S_w = S_1 + S_2 = \begin{bmatrix} 1 & -0.25 \\ -0.25 & 2.2 \end{bmatrix} + \begin{bmatrix} 2.3 & -0.05 \\ -0.05 & 3.3 \end{bmatrix}$$

$$S_w = \begin{bmatrix} 3.3 & -0.3 \\ -0.3 & 5.5 \end{bmatrix}$$

$$w^* = S_w^{-1}(\mu_1 - \mu_2)$$

$$= \begin{pmatrix} 3.3 & -0.3 \\ -0.3 & 5.5 \end{pmatrix}^{-1} \left[\begin{pmatrix} 3 \\ 3.8 \end{pmatrix} - \begin{pmatrix} 8.4 \\ 7.6 \end{pmatrix} \right]$$

$$w^* = \frac{1}{18.06} \begin{bmatrix} 5.5 & 0.3 \\ 0.3 & 3.3 \end{bmatrix} \begin{bmatrix} -5.4 \\ -3.8 \end{bmatrix}$$

$$= \begin{bmatrix} 0.304 & 0.0166 \\ 0.0166 & 0.1827 \end{bmatrix} \begin{bmatrix} -5.4 \\ -3.8 \end{bmatrix}$$

$$w^* = \begin{bmatrix} -1.707 \\ -0.782 \end{bmatrix}$$

$$z = W^{*T} x$$

Algorithm:

1. Compute the global mean (M) using the samples.
2. Compute the statistics like Mean Vector and Covariance Matrix for samples.
3. Compute within-class scatter matrix C. (Termed as pooled within group matrix)
4. Create Discriminant Functions (F1 and F2) by using formula Discriminant function is

$$f_i = \mu_i C^{-1} x_k^T - \frac{1}{2} \mu_i C^{-1} \mu_i^T + \ln P_i$$

Algorithm:

step 1 :- μ_1 , μ_2 $\Rightarrow$ Global Mean = μ

step 2 :- corrected data

$$x_i^0 = x_i - \mu$$

step 3 :- $C_1 = \frac{1}{n_1} x_1^T x_1$

$$C_2 = \frac{1}{n_2} x_2^T x_2$$

step 4 :- $C = \frac{n_1}{n_1+n_2} C_1 + \frac{n_2}{n_1+n_2} C_2 \rightarrow C^{-1}$

step 5 :- P = prior prob. vector $= \begin{bmatrix} n_1/n_1+n_2 \\ n_2/n_1+n_2 \end{bmatrix}$

Step 6 :- Discriminant func^n

$$f_1 = \mu_1 C^{-1} x_k^{T} - \frac{1}{2} \mu_1 C^{-1} \mu_1^T + \ln P_1$$

Given

Examples:

Factory "ABC" produces very expensive and high quality chip rings that their qualities are measured in term of curvature and diameter. Result of quality control by experts is given in the table below.

As a consultant to the factory, you get a task to set up the criteria for automatic quality control. Then, the manager of the factory also wants to test your criteria upon new type of chip rings that even the human experts are argued to each other. The new chip rings have curvature 2.81 and diameter 5.46.

Can you solve this problem by employing Discriminant Analysis?

Curvature	Diameter	Quality Control Result	
2.95	6.63	Passed	
2.53	7.79	Passed	x_1
3.57	5.65	Passed	
3.16	5.47	Passed	
2.58	4.46	Not Passed	
2.16	6.22	Not Passed	x_2
3.27	3.52	Not Passed	

$$x_1 = \begin{bmatrix} 2.95 & 6.63 \\ 2.53 & 7.79 \\ 3.57 & 5.65 \\ 3.16 & 5.47 \end{bmatrix} \qquad x_2 = \begin{bmatrix} 2.58 & 4.46 \\ 2.16 & 6.22 \\ 3.27 & 3.52 \end{bmatrix}$$

step I :- $\mu_1 = \begin{bmatrix} 3.05 & 6.38 \end{bmatrix}$ $\quad \mu_2 = \begin{bmatrix} 2.67 & 4.73 \end{bmatrix}$

GM = $\quad \mu = \begin{bmatrix} 2.88 & 5.676 \end{bmatrix}$

step 2:- corrected data : $x_i^{\circ} = x_i - \mu$

$$x_1^{\circ} = \begin{bmatrix} 0.07 & 0.954 \\ -0.35 & 2.114 \\ 0.69 & -0.026 \\ 0.28 & 0.206 \end{bmatrix} \qquad x_2^{\circ} = \begin{bmatrix} -0.3 & 1.216 \\ -0.72 & 0.544 \\ 0.39 & -2.156 \end{bmatrix}$$

step 3:- cov matrix for group 1

$$C_1 = \frac{1}{n_1} x_1^{T} x_1$$

$$C_1 = \frac{1}{4} \underset{2\times4}{\begin{bmatrix} 0.07 & -0.35 & 0.69 & 0.28 \\ 0.954 & 2.114 & -0.026 & 0.206 \end{bmatrix}} \underset{4\times2}{\begin{bmatrix} 0.07 & 0.954 \\ -0.35 & 2.114 \\ 0.69 & -0.026 \\ 0.28 & 0.206 \end{bmatrix}} = \begin{bmatrix} 0.166 & -0.192 \\ -0.192 & 1.349 \end{bmatrix}$$

III^{ly}, cov. matrix of group 2.

$$C_2 = \frac{1}{n_2} X_2^T X_2 = \begin{bmatrix} 0.259 & -0.286 \\ -0.286 & 2.142 \end{bmatrix}$$

Step 4 :- Within class scatter matrix

$$C = \frac{n_1}{n_1+n_2} C_1 + \frac{n_2}{n_1+n_2} C_2$$

$$= \frac{4}{7}\begin{bmatrix} 0.166 & -0.192 \\ -0.192 & 1.349 \end{bmatrix} + \frac{3}{7}\begin{bmatrix} 0.259 & -0.286 \\ -0.286 & 2.142 \end{bmatrix}$$

$$C = \begin{bmatrix} 0.206 & -0.233 \\ -0.233 & 1.089 \end{bmatrix}$$

step v :- C^{-1}

$$C^{-1} = \begin{bmatrix} 5.745 & 0.791 \\ 0.791 & 0.701 \end{bmatrix}$$

step vi :- P = prior probability vector

$$P = \begin{bmatrix} n_1/n_1+n_2 \\ n_2/n_1+n_2 \end{bmatrix} = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} 4/7 \\ 3/7 \end{bmatrix}$$

step 7 :- $f_i = \mu_i C^{-1} x_k' - \frac{1}{2} \mu_i C^{-1} \mu_i' + \ln P_i$

$$\begin{array}{cccc} 1 & 2 & \cdots & n \end{array}$$

$x_k = [2.81 \quad 5.46]$

HW

$f_1 = \mu_1 C^{-1} x_k' - \frac{1}{2} \mu_1 C^{-1} \mu_1' + \ln P_1 = 44.07 \rightarrow$ Passed

$f_2 = \mu_2 C^{-1} x_k' - \frac{1}{2} \mu_2 C^{-1} \mu_2' + \ln P_2 = 44.10 \rightarrow$ not Passed

$f_2 > f_1$

$44.10 > 44.07$

Examples:

Table lists the ratings of the new mixer on these two characteristics (at a specified price) by a panel of 10 potential purchasers. In rating the food mixer, each panel member is implicitly comparing it with products already on the market. After the product was evaluated, the evaluators were asked to state their buying intentions ("would purchase" or "would not purchase"). Five stated that they would purchase the new mixer and five said they would not. Can you solve this problem employing LDA?

Table 7.1 Kitchenade Survey Results for the Evaluation of a New Consumer Product

Groups Based on Purchase Intention	Evaluation of New Product* X_1 Durability	X_2 Performance
Group 1: Would purchase		
Subject 1	8	9
Subject 2	6	7
Subject 3	10	6
Subject 4	9	4
Subject 5	4	8
Group mean $= \mu_1$	7.4	6.8
Group 2: Would not purchase		
Subject 6	5	4
Subject 7	3	7
Subject 8	4	5
Subject 9	2	4
Subject 10	2	2
Group mean $= \mu_2$	3.2	4.4
Difference between group means	4.2	2.4

*Evaluations are made on a 10-point scale (1 = very poor to 10 = excellent).

$$x_1 = \begin{bmatrix} 8 & 9 \\ 6 & 7 \\ 10 & 6 \\ 9 & 4 \\ 4 & 8 \end{bmatrix} \qquad x_2 = \begin{bmatrix} 5 & 4 \\ 3 & 7 \\ 4 & 5 \\ 2 & 4 \\ 2 & 2 \end{bmatrix}$$

$$\mu_1 = \begin{bmatrix} 7.4 & 6.8 \end{bmatrix}$$

$$\mu_2 = \begin{bmatrix} 3.2 & 4.4 \end{bmatrix}$$

$$\mu = \begin{bmatrix} 5.3 & 5.6 \end{bmatrix}$$

$$x_1^0 = \begin{bmatrix} 2.7 & 3.4 \\ 0.7 & 1.4 \\ 4.7 & 0.4 \\ 3.7 & -1.6 \\ -1.3 & 2.4 \end{bmatrix} \qquad x_2^0 = \begin{bmatrix} -0.3 & -1.6 \\ -2.3 & 1.4 \\ -1.3 & -0.6 \\ -3.3 & -1.6 \\ -3.3 & -3.6 \end{bmatrix}$$

$$C_1 = \frac{1}{n} x_i x_i^T$$

$$= \frac{1}{5}\left\{\begin{bmatrix}2.7\\3.4\end{bmatrix}\begin{bmatrix}2.7 & 3.4\end{bmatrix} + \begin{bmatrix}0.7\\1.4\end{bmatrix}\begin{bmatrix}0.7 & 1.4\end{bmatrix} + \begin{bmatrix}4.7\\0.4\end{bmatrix}\begin{bmatrix}4.7 & 0.4\end{bmatrix}\right.$$

$$\left. + \begin{bmatrix}3.7\\-1.6\end{bmatrix}\begin{bmatrix}3.7 & -1.6\end{bmatrix} + \begin{bmatrix}-1.3\\2.4\end{bmatrix}\begin{bmatrix}-1.3 & 2.4\end{bmatrix}\right\}$$

$$= \frac{1}{5}\left\{\begin{bmatrix}7.29 & 9.18\\9.18 & 11.52\end{bmatrix} + \begin{bmatrix}0.49 & 0.98\\0.98 & 1.96\end{bmatrix} + \begin{bmatrix}22.09 & 1.88\\1.88 & 0.16\end{bmatrix} +\right.$$

$$\left.\begin{bmatrix}13.69 & -5.92\\-5.92 & 2.56\end{bmatrix} + \begin{bmatrix}1.69 & -3.12\\-3.12 & 5.76\end{bmatrix}\right\}$$

$$C_1 = \begin{bmatrix} 9.05 & 0.6 \\ 0.6 & 4.392 \end{bmatrix} \quad ; \quad |||ly, \quad C_2 = \begin{bmatrix} 5.77 & 3.04 \\ 3.04 & 4.08 \end{bmatrix}$$

Within scatter Matrix

$$C = \frac{n_1}{n_1+n_2} C_1 + \frac{n_2}{n_1+n_2} C_2$$

$$= \frac{1}{2}\begin{bmatrix} 9.05 & 0.6 \\ 0.6 & 4.392 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 5.77 & 3.04 \\ 3.04 & 4.08 \end{bmatrix}$$

$$C = \begin{bmatrix} 7.41 & 1.82 \\ 1.82 & 4.236 \end{bmatrix}$$

$$C^{-1} = \begin{bmatrix} 0.1508 & -0.064 \\ -0.064 & 0.2637 \end{bmatrix}$$

$$P = \begin{bmatrix} 5/10 \\ 5/10 \end{bmatrix} = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix}$$

$$f_1 = \mu_1 C^{-1} x_k' - \frac{1}{2} \mu_1 C^{-1} \mu_1' + \ln P_1$$

$$[\quad][\quad] x_k' - \frac{1}{2} [\quad][\quad][\quad] + \ln P_1$$

$$f_2 = \mu_2 C^{-1} x_k' - \frac{1}{2} \mu_2 C^{-1} \mu_2' + \ln P_2$$