Machine Learning

CS F464

BITS Pilani
Dubai Campus

Dr. Pranav M. Pawar

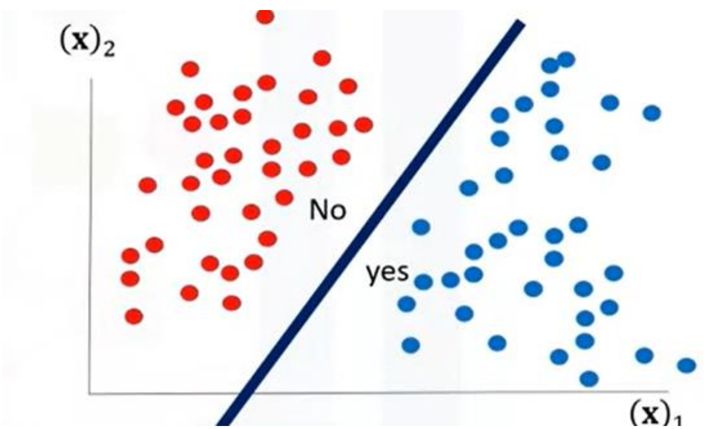
Contents

- Discriminant Functions
  - Two classes
  - Multiple classes
- Least square for classification
- Fisher Linear Discriminant (Linear Discriminant)

Introduction (1)

- Goal of classification
  - To take input vector *x* and to assign it to one of *K* discrete classes C_k where $k = 1,\ldots,K$.
- The input space is divided into *decision regions* and its boundary called *decision boundaries* or *decision surfaces*.
- Linear models for classification =>

  Decision surfaces are linear functions of input vector x, hence are defined using (D-1) dimensional hyperplane within the D-dimensional input.



Introduction (2)

- Linearly separable data sets => If classes can be separated exactly by linear decision surface.
- The generalised linear model for classification is,

$$y(\mathbf{x}) = f\left(\mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0\right)$$

  - *f*(·) is a fixed non-linear function (activation function)
    - E.g.

$$f(u) = \begin{cases} 1 \text{ if } u \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

    - Decision boundary between classes will be linear function of *x.*

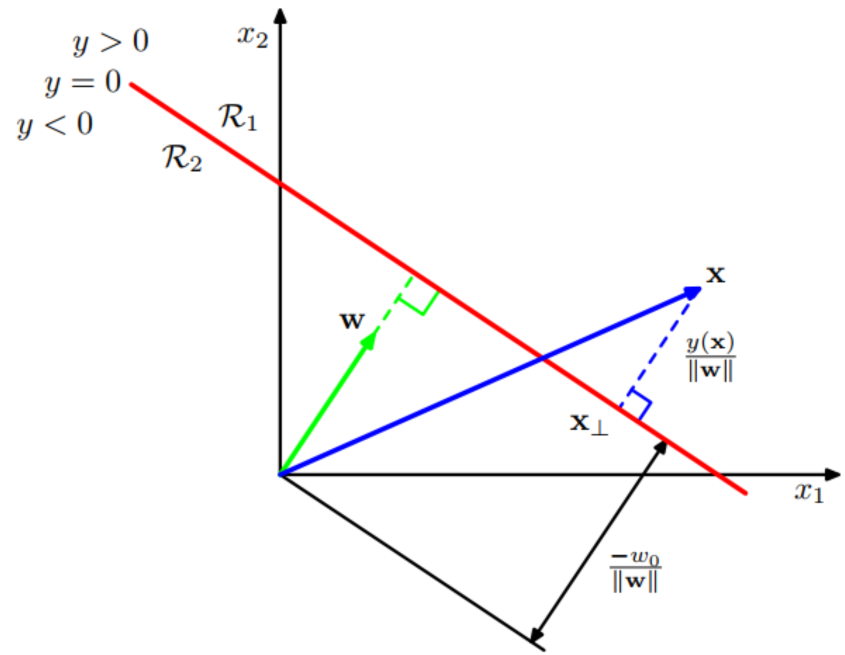
Discriminant Function (1)

- Function which takes an input vector *x* and assigns it to one of *K* classes.
- Two class
  - 2 class problem,
  - Simple linear discri $t_i \in \{0, 1\}$ function is,

$$y(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0$$

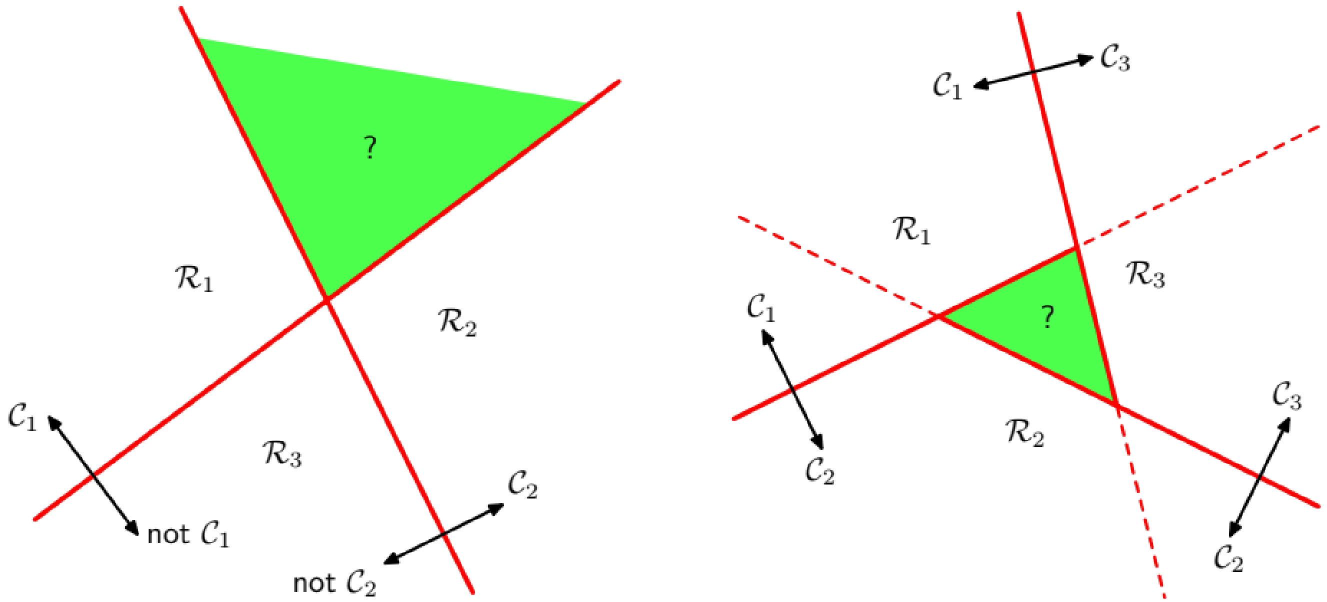
  - Use of threshold for predicting class. Threshold is negative of bias.
  - Input vector *x* belongs to class C_1 if $y(x) >= 0$ and to class C_2 otherwise. Corresponding decision boundary is $y(x) = 0$.
  - Perpendicular distance of *x* in *w* direction is

$$\frac{\boldsymbol{w}^T\boldsymbol{x}}{||\boldsymbol{w}||}$$



Discriminant Function (2)

- Multiple classes



- A linear discriminant between two classes separates with a hyperplane
- One-versus-the-rest method: build K - 1 classifiers, between Ck and all others
- One-versus-one method: build $K(K$ - 1)/2 classifiers, between all pairs

Discriminant Function (3)

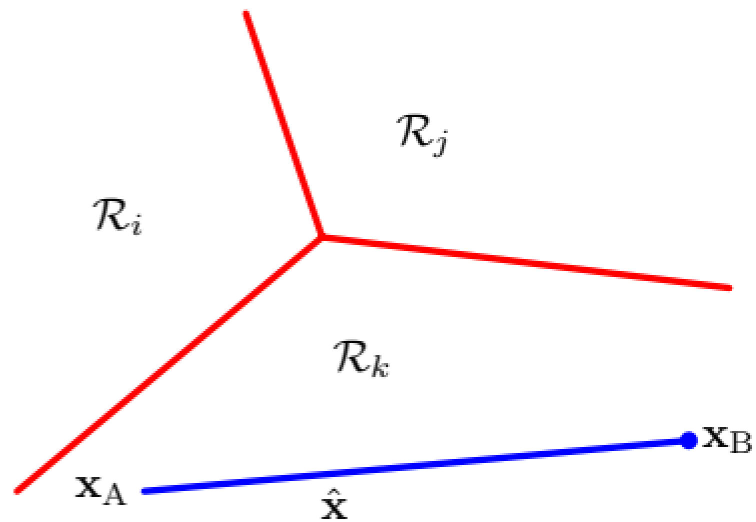
- A solution is to build *K* linear function,

$$y_k(\mathbf{x}) = \mathbf{w}_k^{\mathrm{T}}\mathbf{x} + w_{k0}$$

- Assign point x to class C_k if $y_k(x) > y_j(x)$ for all *j* not equal to *k.*
- Decision boundary between class C_k and C_j is given by $y_k(x) = y_j(x)$ and *(D-1)* dimensional hyperplane is defined as,

$$(\mathbf{w}_k - \mathbf{w}_j)^{\mathrm{T}}\mathbf{x} + (w_{k0} - w_{j0}) = 0.$$

- Decision of such a discriminant is always singly connected and convex.

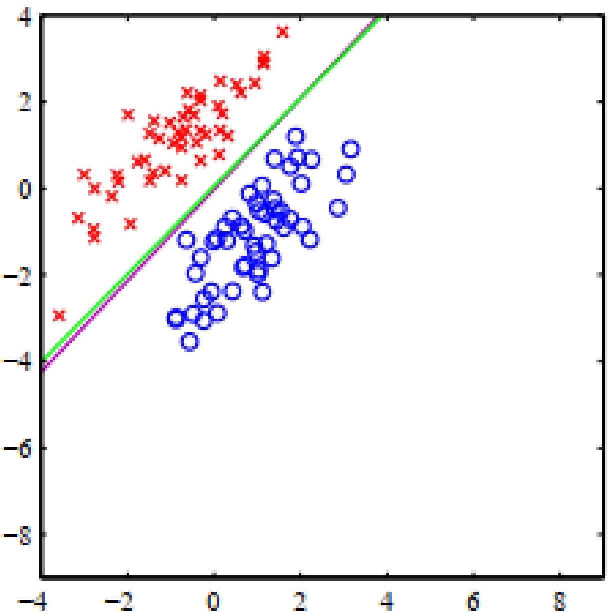


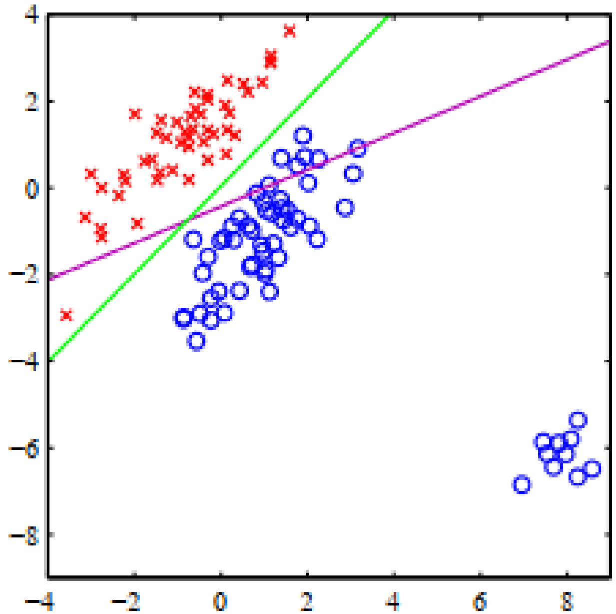
Least squares for classification

- How do we learn the decision boundaries ($\boldsymbol{w}_k$; w_{k0})?
- One approach is to use least squares.
- Find $\boldsymbol{W}$ to minimize squared error over all examples and all components of the label vector:

$$E(\boldsymbol{W}) = \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K}(y_k(\boldsymbol{x}_n) - t_{nk})^2$$

Problems with Least Square (1)

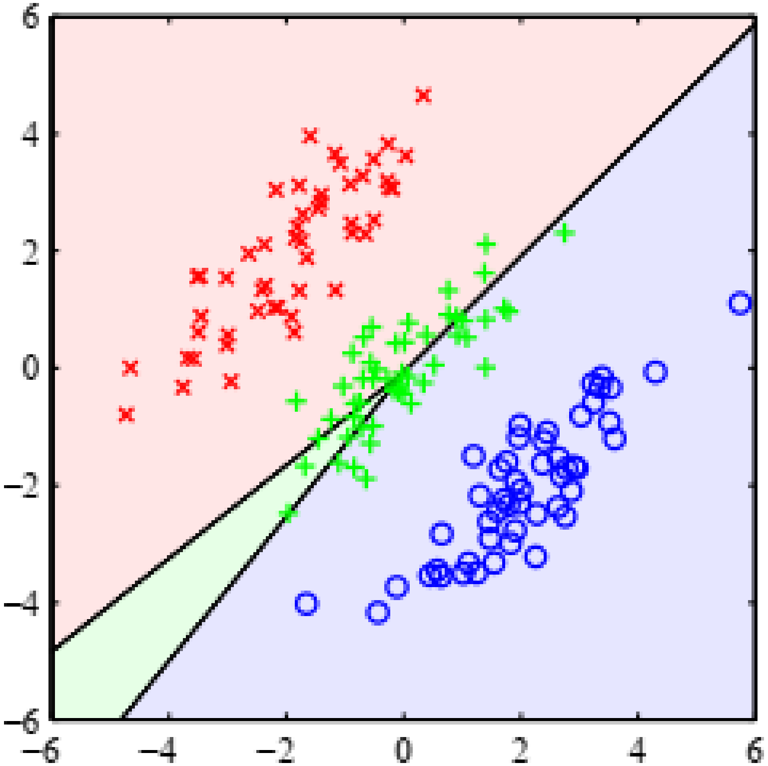


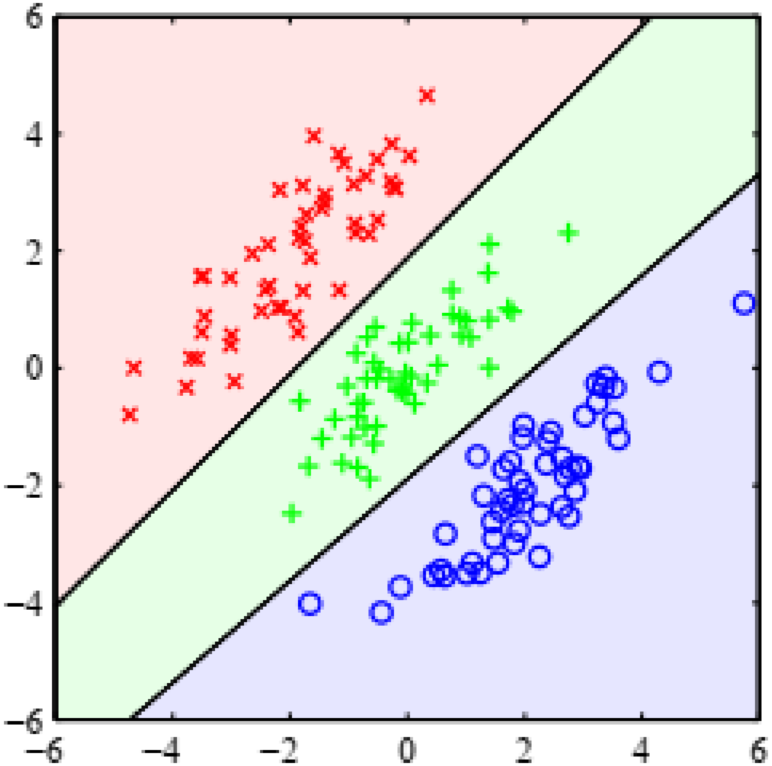


- Looks okay... least squares decision boundary
  - Similar to logistic regression decision boundary (more later)

- Gets worse by adding easy points?!
- Why?
  - If target value is 1, points far from boundary will have high value, say 10; this is a large error so the boundary is moved

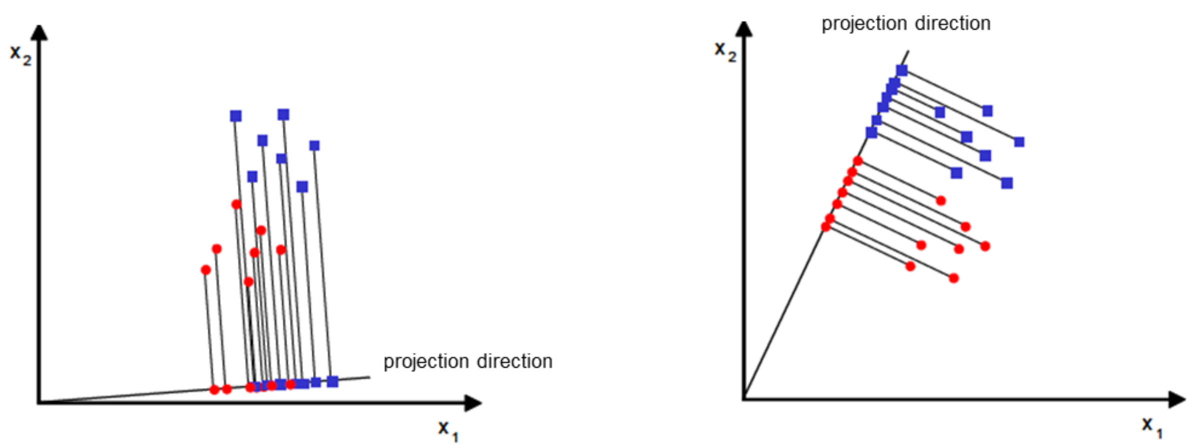
Problems with Least Square (2)





Fisher Linear Discriminant (Linear Discriminant Analysis (LDA))

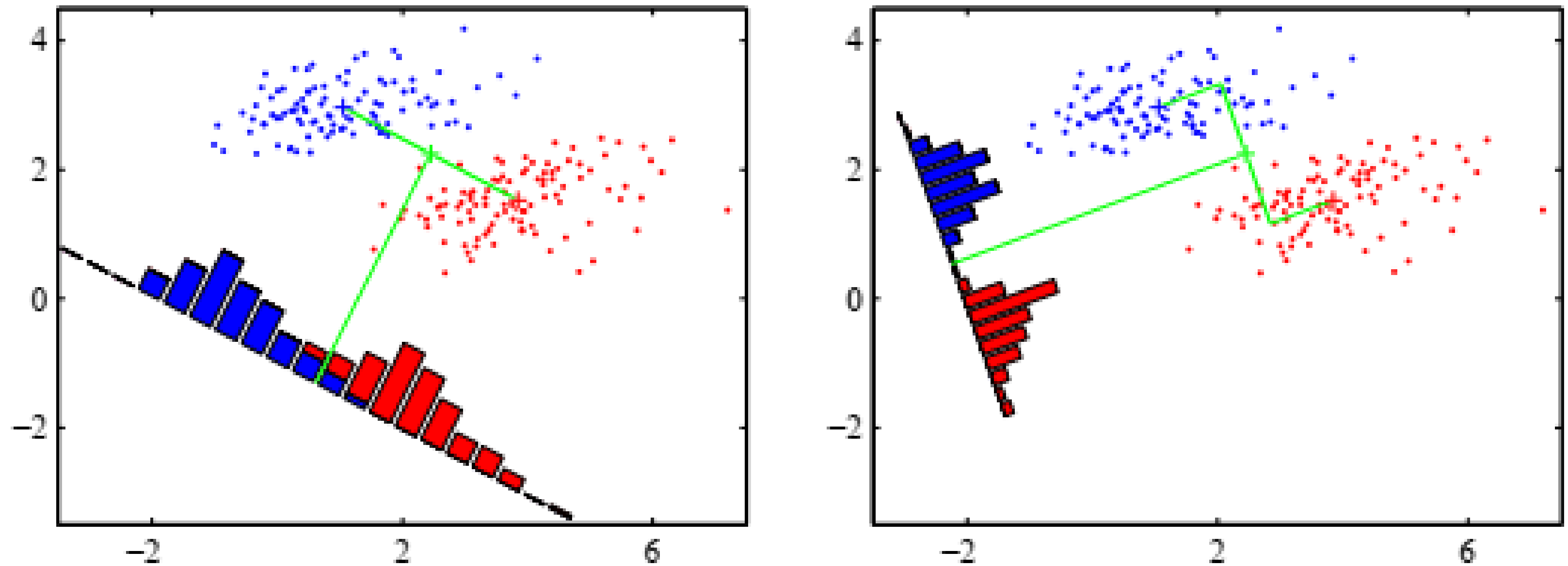
- LDA maximizes the separation between multiple classes.
- LDA seeks a projection that **best discriminates** the data.
- Goal
  - Seeks to find directions along which the classes are best separated (i.e., increase discriminatory information).
  - It takes into consideration the scatter (i.e., variance) within-classes and between-classes.



Bad separability Good separability

  - LDA is also used as dimensionality reduction technique as pre-processing step for ML application.

Fisher Linear Discriminant (Linear Discriminant Analysis (LDA))



- A natural idea would be to project in the direction of the line connecting class means
- However, problematic if classes have variance in this direction
- Fisher criterion: maximize ratio of inter-class separation (between) to intra-class variance (inside)

Linear Discriminant Analysis (LDA)

- Let us assume C classes with each class containing M_i samples, i=1,2,..,C and M the total number of samples:

$$M = \sum_{i=1}^{C} M_i$$

- Let μ_i is the mean of the i-th class, i=1,2,…,C and μ is the mean of the whole dataset:

$$\boldsymbol{\mu} = \frac{1}{C}\sum_{i=1}^{C} \boldsymbol{\mu}_i$$

Within-class scatter matrix

$$S_w = \sum_{i=1}^{C}\sum_{j=1}^{M_i} (\mathbf{x}_{ij} - \boldsymbol{\mu}_i)(\mathbf{x}_{ij} - \boldsymbol{\mu}_i)^T$$

Between-class scatter matrix

$$S_b = \sum_{i=1}^{C} (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T$$

Linear Discriminant Analysis (LDA)

- Suppose the desired projection transformation is:

$$\mathbf{y} = U^T \mathbf{x}$$

- Suppose the scatter matrices of the projected data $\mathbf{y}$ are:

$$\tilde{S}_b, \tilde{S}_w$$

- LDA seeks transformations that maximize the between-class scatter and minimize the within-class scatter:

$$\max \frac{|\tilde{S}_b|}{|\tilde{S}_w|} \quad \text{or} \quad \max \frac{|U^T S_b U|}{|U^T S_w U|}$$

Example

Find the linear discriminant projection vector and classify the for following data samples,

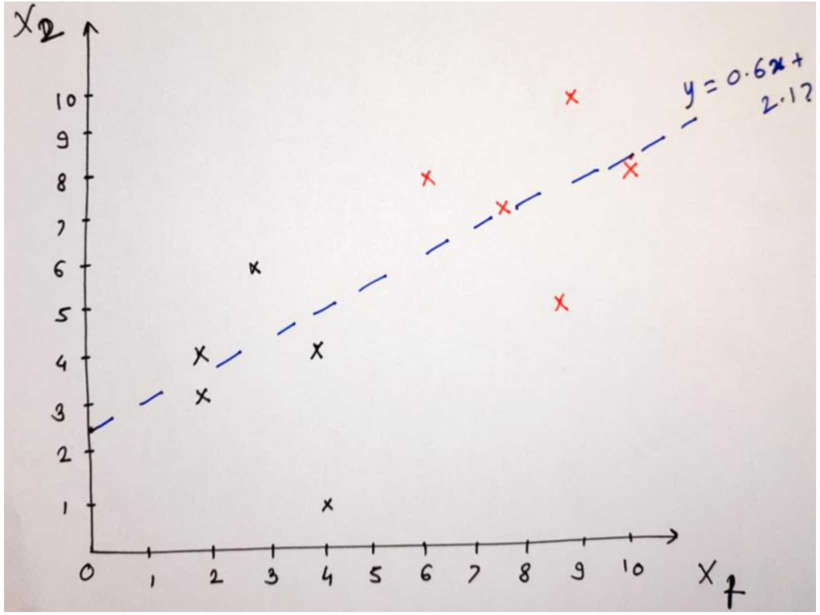
Class 1 =>

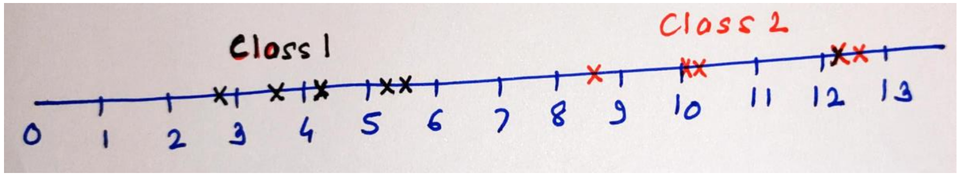
X1 = (x1,x2) = {(4,2), (2,4), (2,3), (3,6), (4,4)}

Class 2 =>

X2 = (x1,x2) = {(9,10), (6,8), (9,5) ,(8,7), (10,8)}

Solution:





Sources

- Chapter 4, Christopher M Bishop: Pattern Recognition & Machine Leaning, 2006 Springer.
- http://vda.univie.ac.at/Teaching/ML/15s/LectureNotes/04_classification.pdf
- Chapter 9, Christopher M. Bhisop, Pattern Recognition & Machine Learning, Springer, 2006.
- Chapter 6 and 14, Marsland Stephen, Machine Learning – An Algorithmic Perspective, 2e, CRC Press,2015.
- Chapter 6 and 7, Alpaydin Ethem. Introduction to Machine Learning, 3e, PHI, 2014.
- http://www.facweb.iitkgp.ac.in/~sudeshna/courses/ml08
- https://www.cse.unr.edu/~bebis/CS479/Lectures
- https://cse.iitkgp.ac.in/~dsamanta/courses/da/resources
- http://www.cvip.louisville.edu/wordpress/wp-content/uploads/2010/01/LDA-Tutorial-1.pdf

innovate achieve lead

BITS Pilani
Dubai Campus

Thank You!