Pattern Classification

02. Training Patterns

AbdElMoniem Bayoumi, PhD

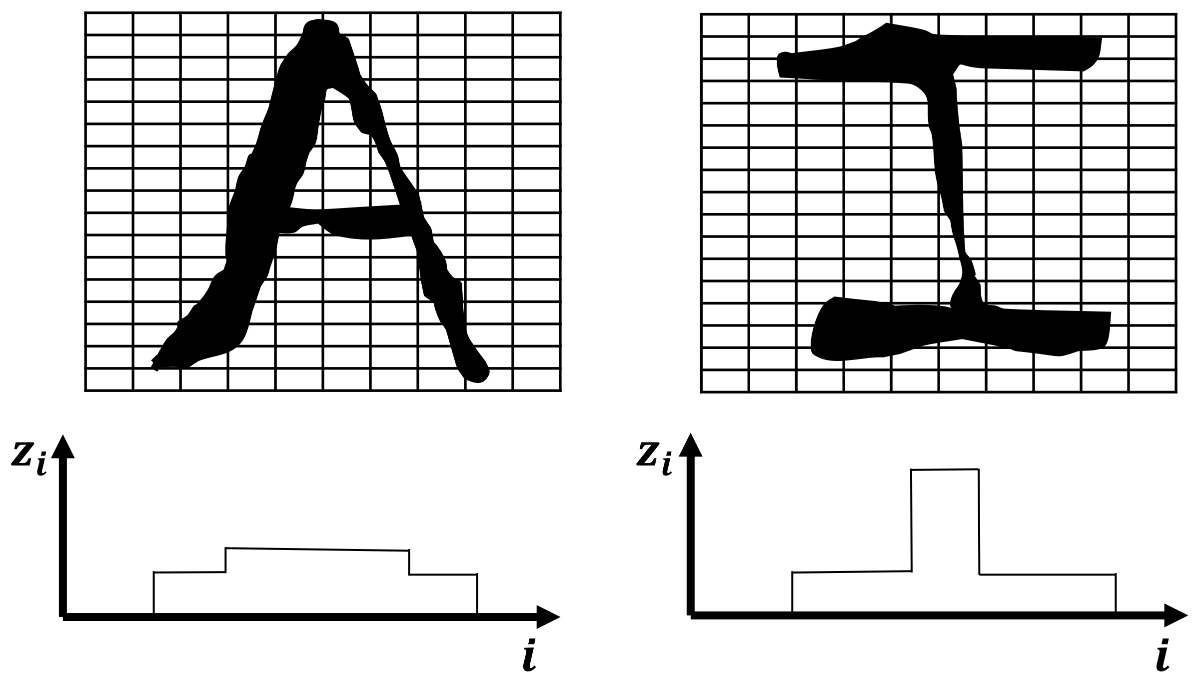
Fall 2021

Acknowledgment

- These slides have been created relying on lecture notes of Prof. Dr. Amir Atiya

Recap: OCR

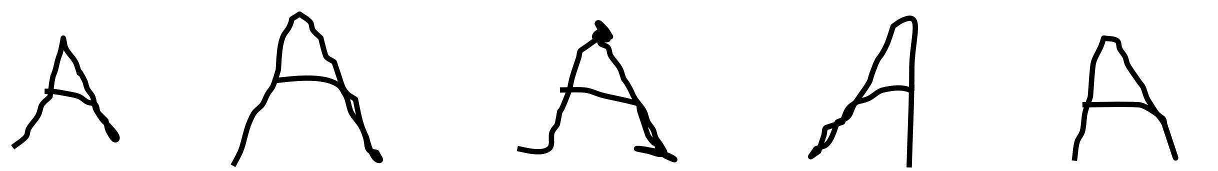
- z_i = sum of black pixels along column i



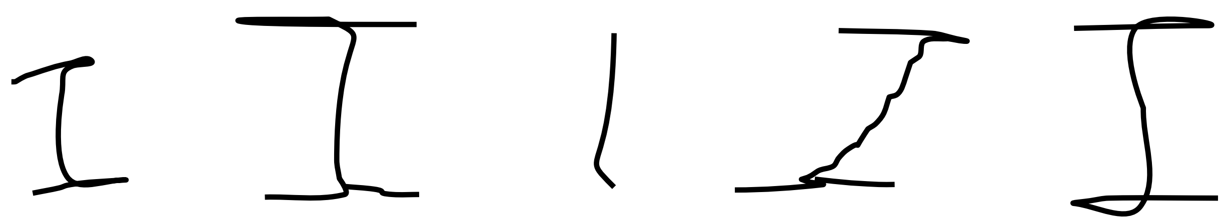
- Feature: $\mathbf{z} = max_i(z_i)$
  - z = 3 for 'A'
  - z = 10 for 'I'

Recap: OCR

- Construct a histogram based on the feature values of the training patterns



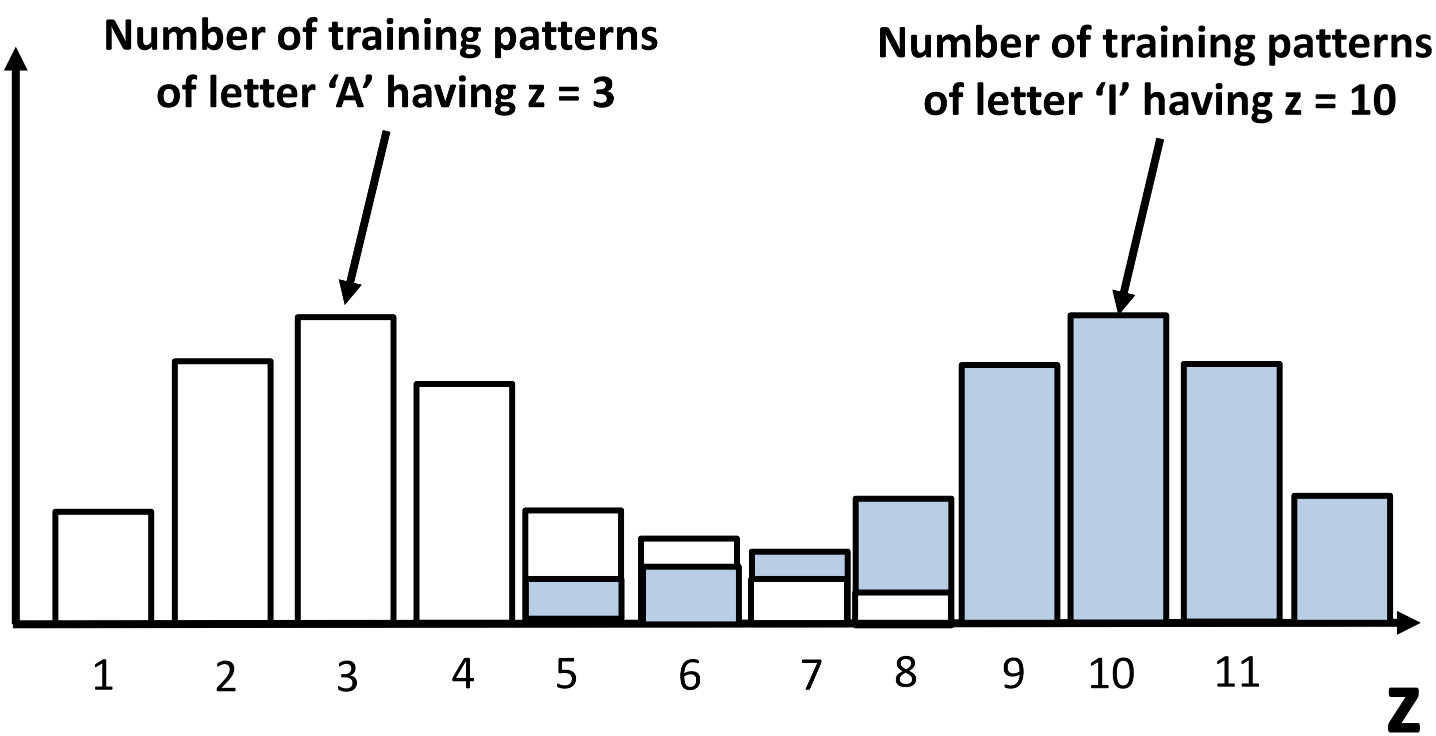
$z =$	2	3	1	4	3



$z =$	8	10	9	5	7

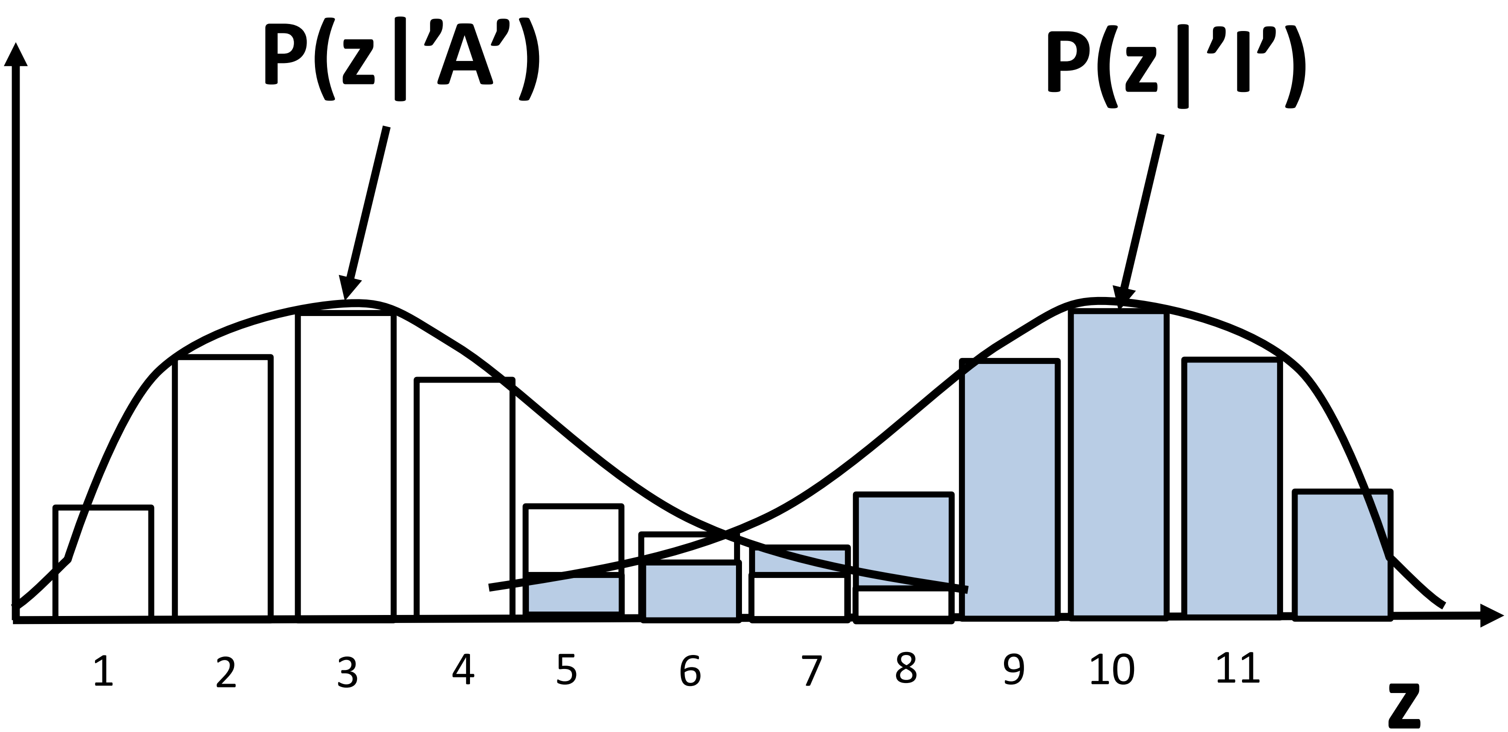
Recap: OCR

- Construct a histogram based on the feature values of the training patterns



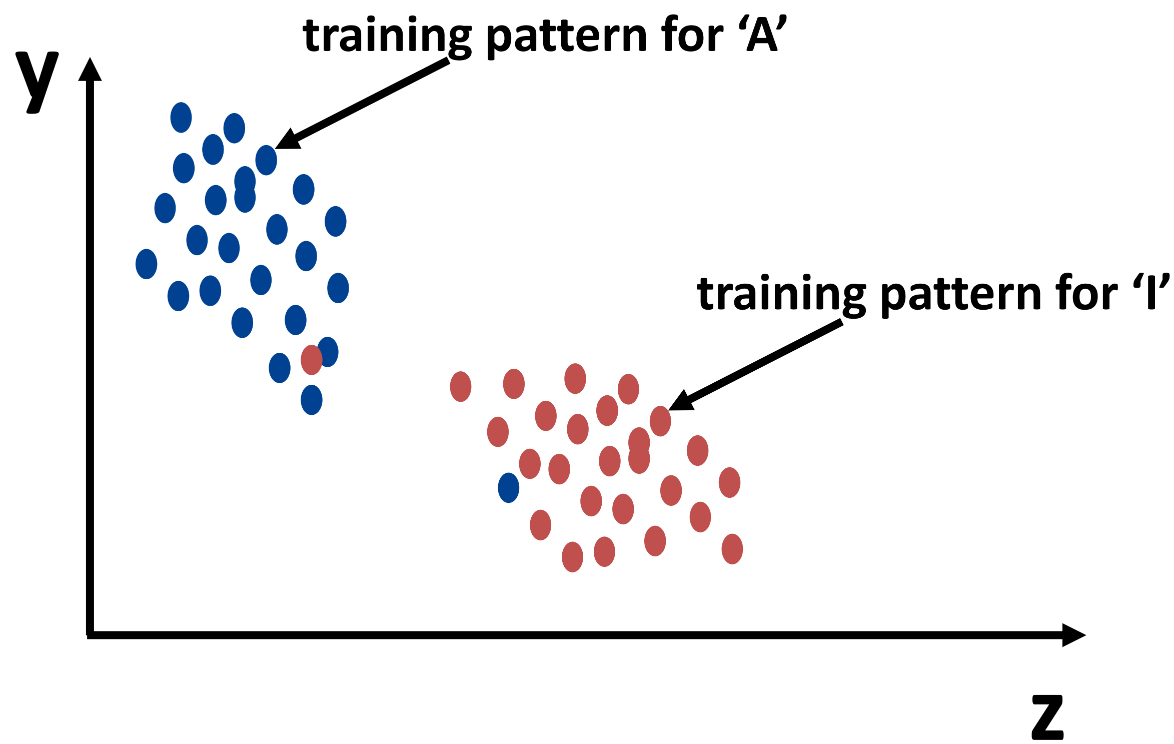
Recap: OCR

- Estimate density functions



Recap: Feature Space

- Multiple features reduce the classification error

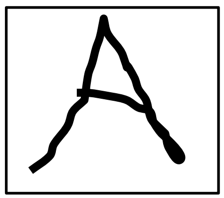
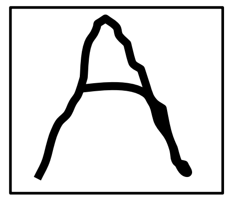
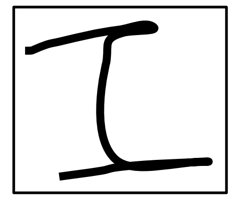


2d feature space for features y & z

Feature Vector

- Let $\underline{\boldsymbol{X}}(\boldsymbol{m}) = \begin{bmatrix} \boldsymbol{X}_1(\boldsymbol{m}) \\ \boldsymbol{X}_2(\boldsymbol{m}) \\ \vdots \\ \boldsymbol{X}_N(\boldsymbol{m}) \end{bmatrix}$ be the feature vector of the m^{th} training pattern
- N is the number of features, i.e., the dimension of $\underline{X}(m)$
- M is the number of the training patterns

Example

#1	#2	#3
		
z = 3	z = 4	z = 9
y = 6	y = 7	y = 3
$\underline{X}(1) = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$	$\underline{X}(2) = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$	$\underline{X}(3) = \begin{bmatrix} 9 \\ 3 \end{bmatrix}$

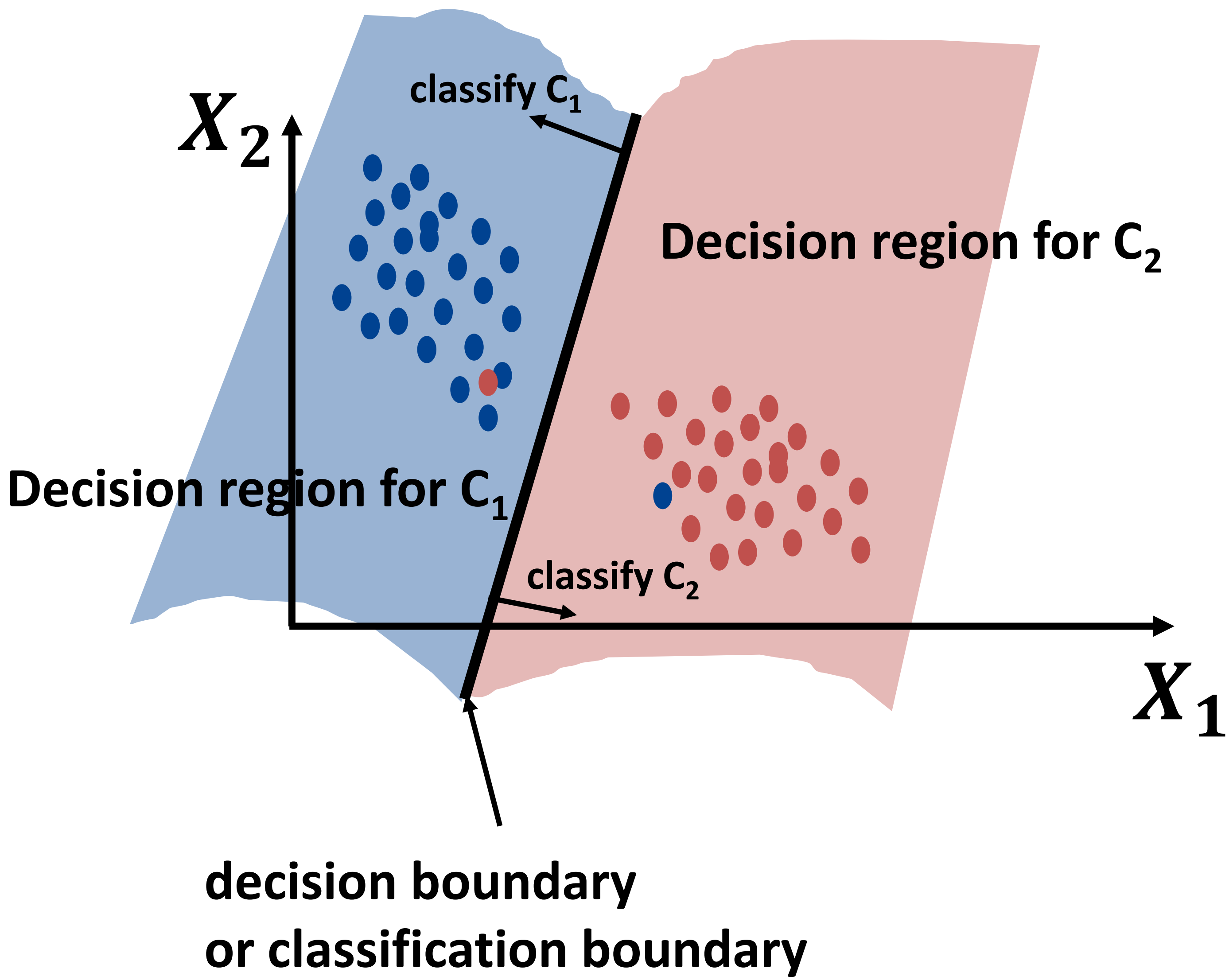
Decision Regions

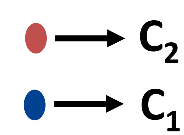
- Features from patterns from the same class tend to be similar

- The data points (patterns) of each class occur in groupings or clusters in the feature space plot

- Utilize this fact to design the classifier by detecting the regions where the patterns of each class are grouped

Decision Regions





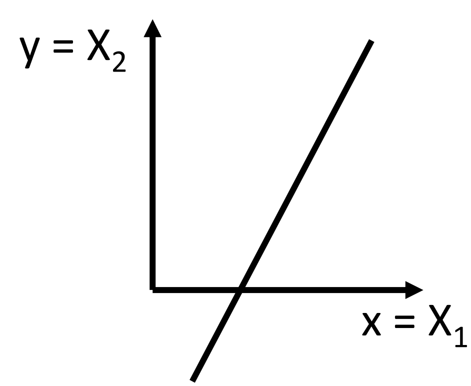
Decision Boundary

- If the decision boundary is linear (a line in 2D, a plane in 3D, or a hyperplane in more than 3D), then its equation follows:

$$\boldsymbol{W_0 + W_1 X_1 + \cdots + W_N X_N = 0}$$

- $\boldsymbol{W_0}$ and $\boldsymbol{W_i}$ are the constants that determine the position of the hyperplane

- Example:



$$y = mx + b$$

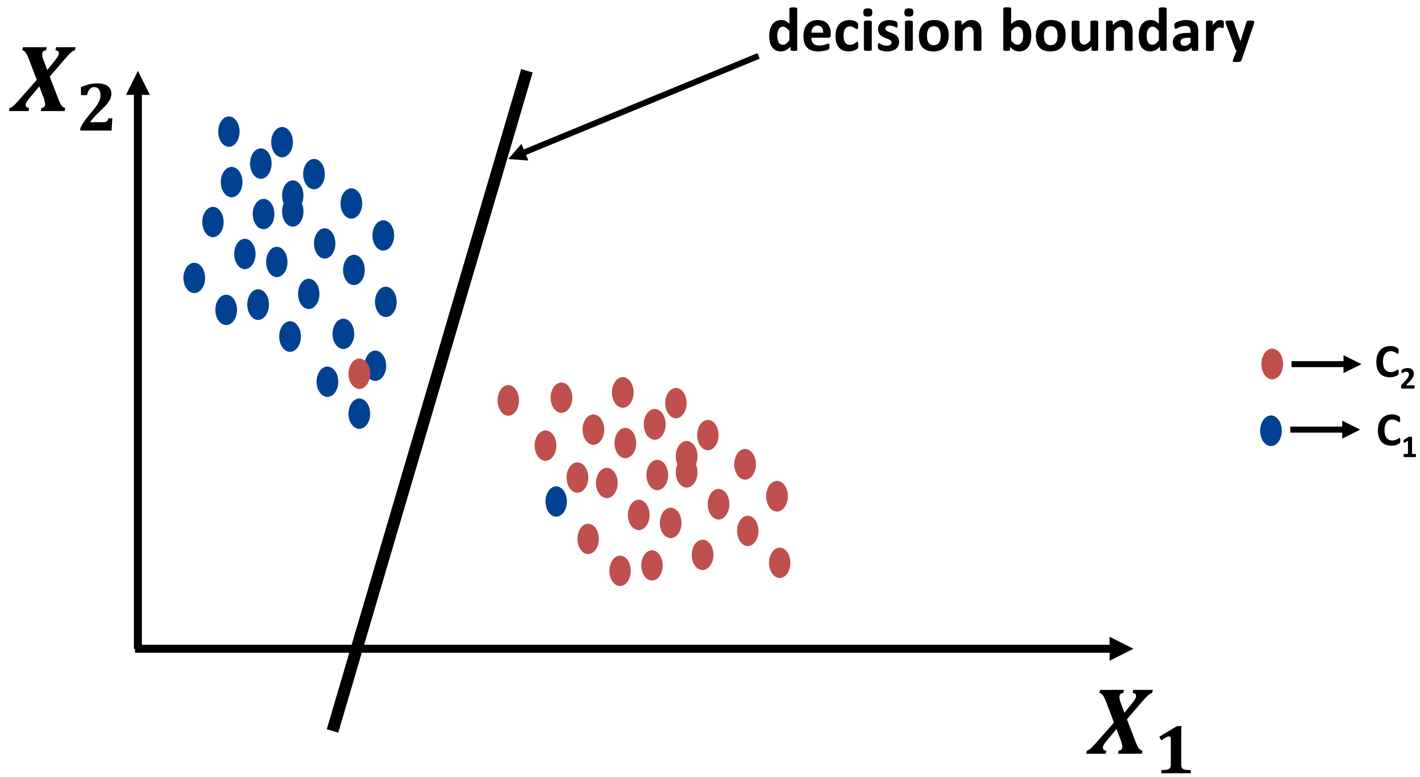
or

$$W_1 x + W_2 y + W_0 = 0$$

or

$$W_1 X_1 + W_2 X_2 + W_0 = 0$$

Decision Boundary



$$W_0 + \sum_{i=1}^{N} W_i X_i(m) \begin{cases} > 0 & \text{classification region for C1} \\ = 0 & \text{on the decision boundary} \\ < 0 & \text{classification region for C2} \end{cases}$$

Decision Boundary

- Let $\underline{X}(m)$ be the feature vector from the training set

- We want to compute:

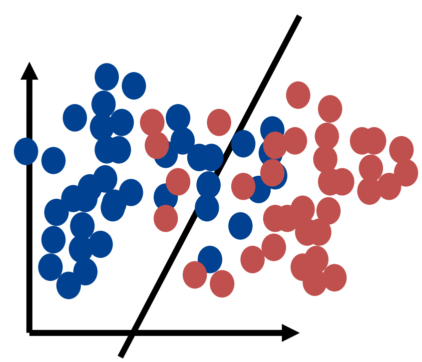
$$\boldsymbol{W_0} + \underline{\boldsymbol{W}}^{\boldsymbol{T}}\underline{\boldsymbol{X}}(\boldsymbol{m}) \begin{cases} > \mathbf{0} & \text{for most } \underline{X}(m) \text{ of } C_1 \\ < \mathbf{0} & \text{for most } \underline{X}(m) \text{ of } C_2 \end{cases}$$

- Where $\underline{\boldsymbol{W}} = \begin{bmatrix} \boldsymbol{W_1} \\ \boldsymbol{W_2} \\ \vdots \\ \boldsymbol{W_N} \end{bmatrix}$ and $\underline{\boldsymbol{W}}^{\boldsymbol{T}}\underline{\boldsymbol{X}}(\boldsymbol{m}) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{W_i X_i}(\boldsymbol{m})$

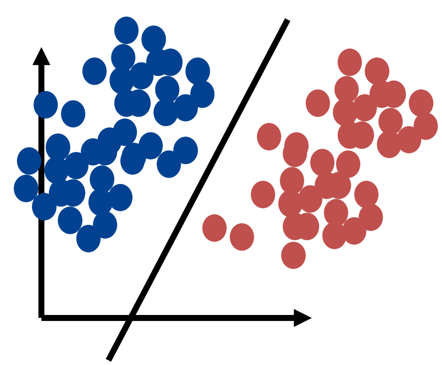
Types of Problems

- A problem is said to be ***linearly separable*** if there is a hyperplane that can separate the training data points of class C_1 from those of C_2

- Otherwise it is said to be ***not linearly separable***

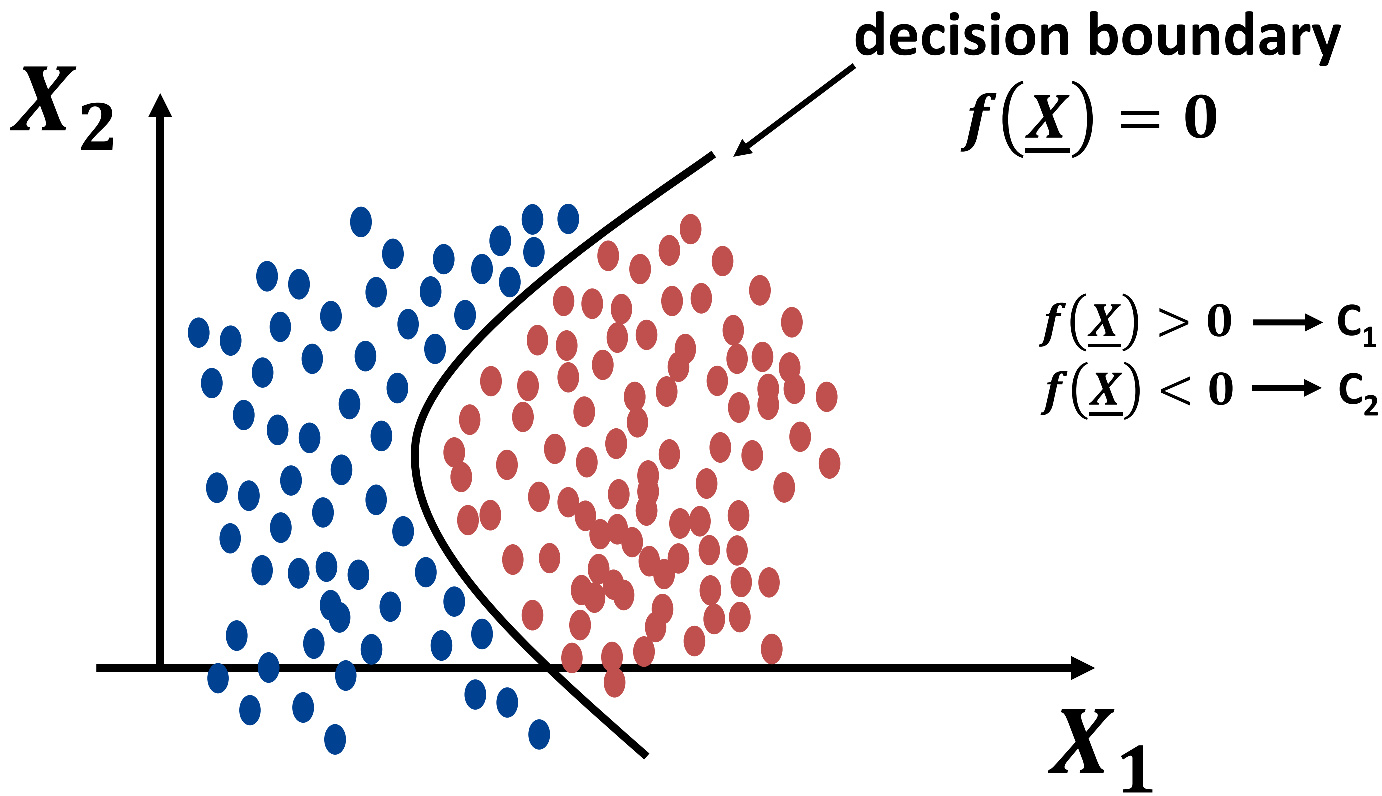


Not linearly separable



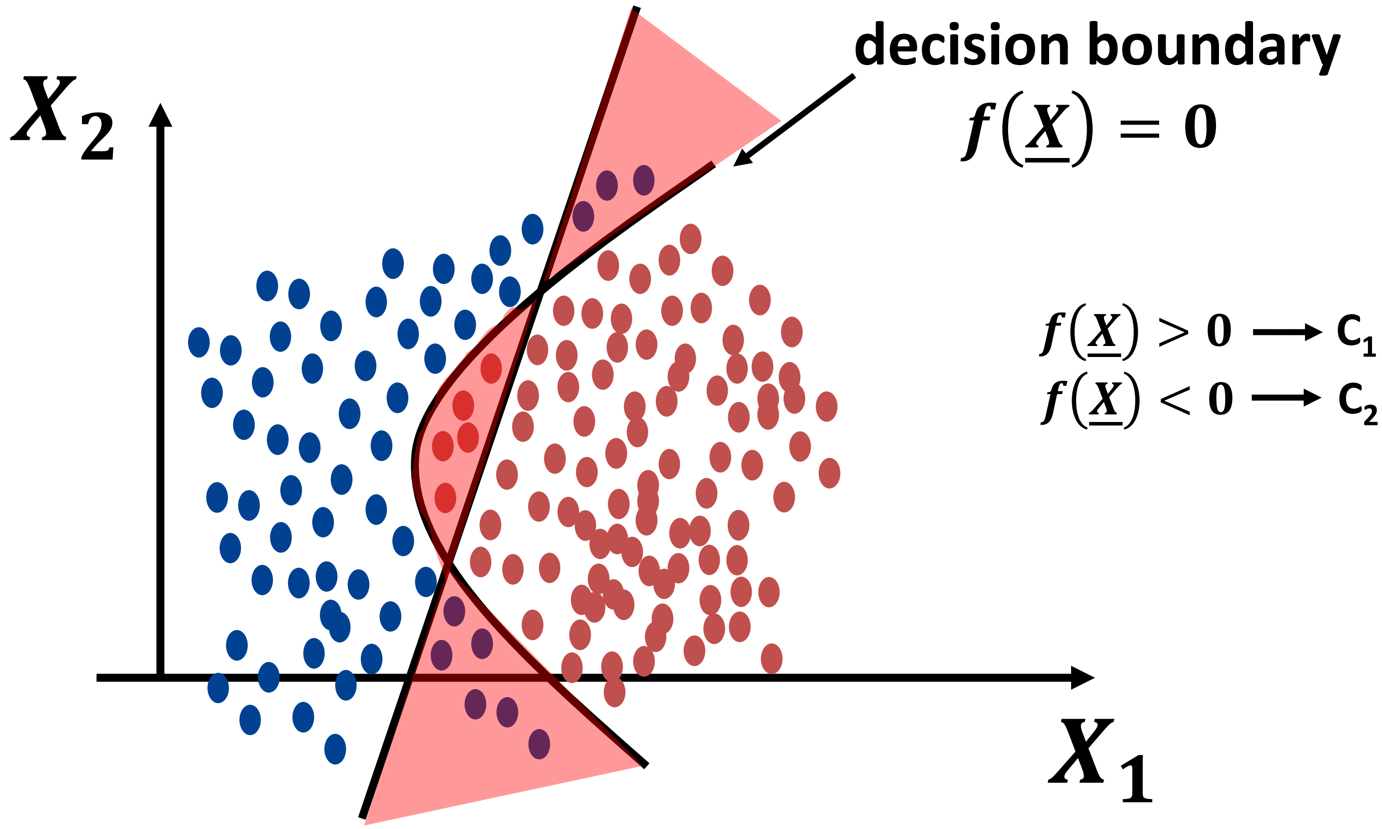
Linearly separable

Types of Problems



- For some problems like this above, a non-linear decision boundary would be more appropriate (for a non-linear classifier)

Types of Problems



- The linear classifier gives 14 errors, while the non-linear classifier gives zero error → choose the non-linear classifier

Acknowledgment

- These slides have been created relying on lecture notes of Prof. Dr. Amir Atiya