Applied Text Analytics &
Natural Language Processing

with Dr. Mahdi Roozbahani
& Wafa Louhichi

Perceptron

GT

Learning Objectives

In this lesson, you will learn another linear text classifier

- Binary classification

- Linear separability
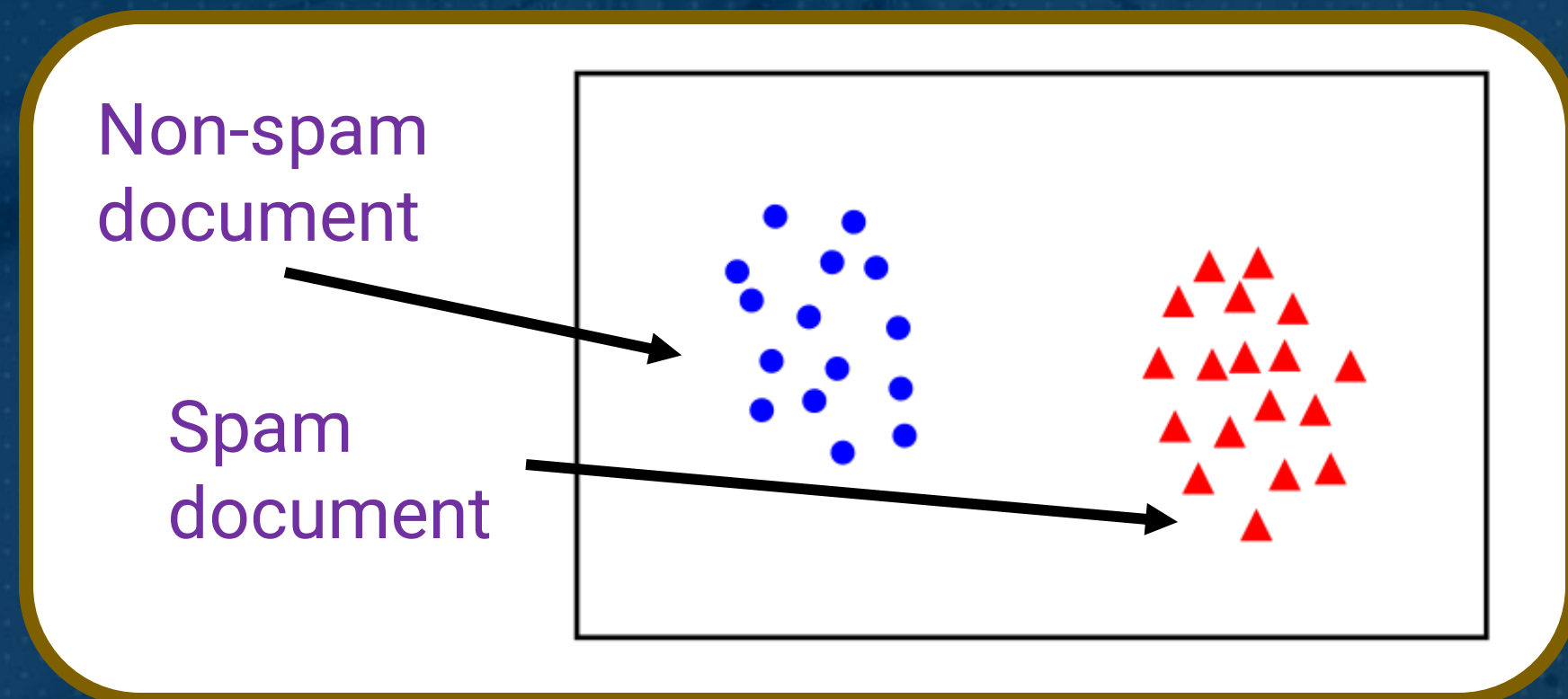
- Perceptron algorithm

Binary Classification

Given training data (x_i, y_i) for $i = 1 \ldots N$, with $x_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\}$, learn a classifier $f(x)$ such that
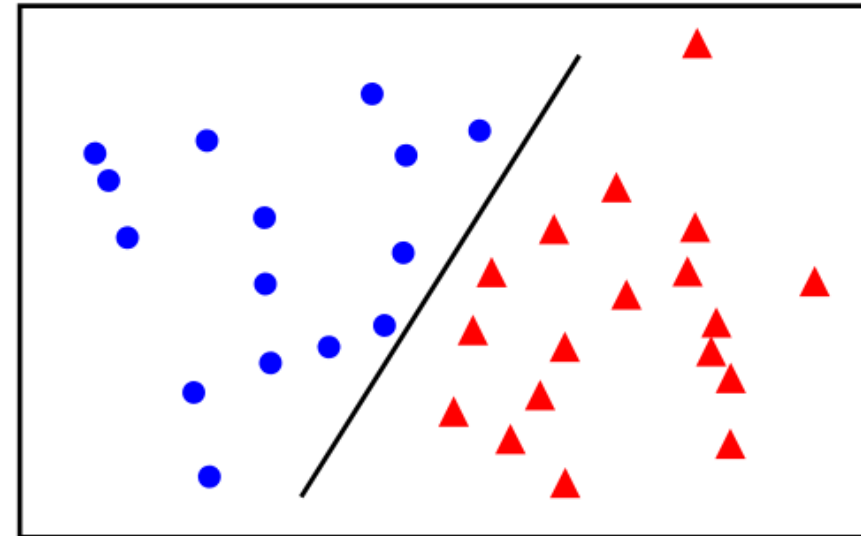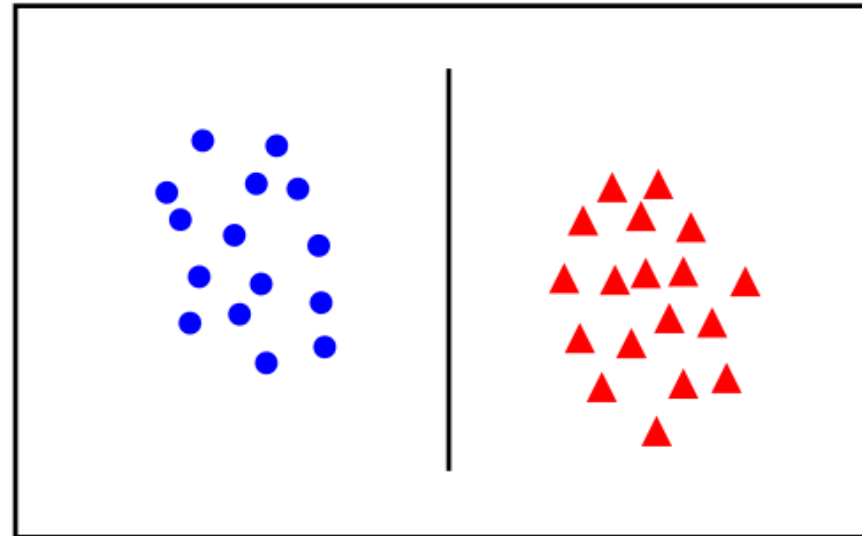
$$f(x_i) \begin{cases} \geq 0 & +1 \quad \text{Non-spam document} \\ < 0 & -1 \quad \text{Spam document} \end{cases}$$
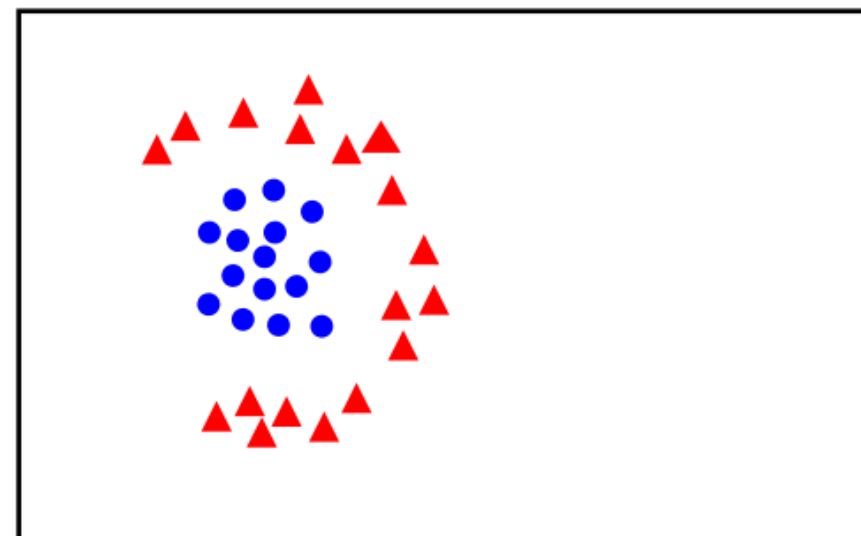
i.e. $y_i f(x_i) > 0$ for a correct classification

Non-spam document

Spam document

Linear Separability
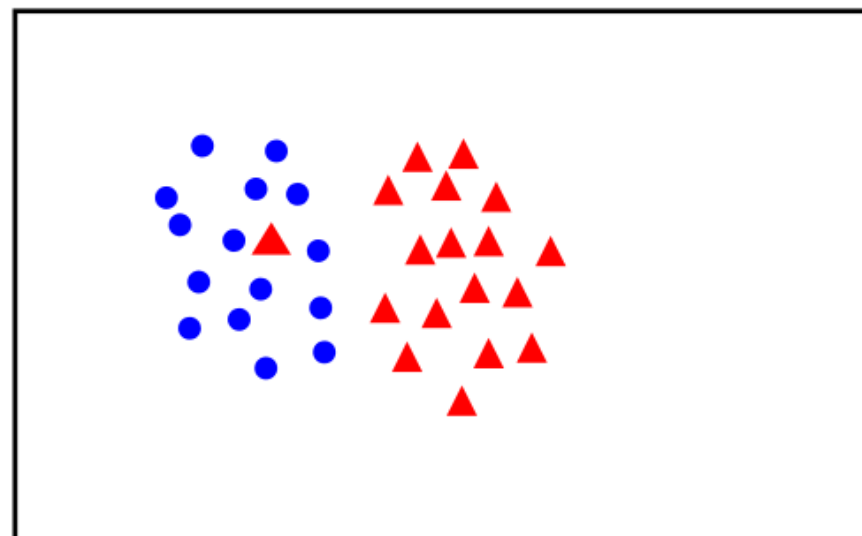
Linear Classifier

A linear classifier has the form
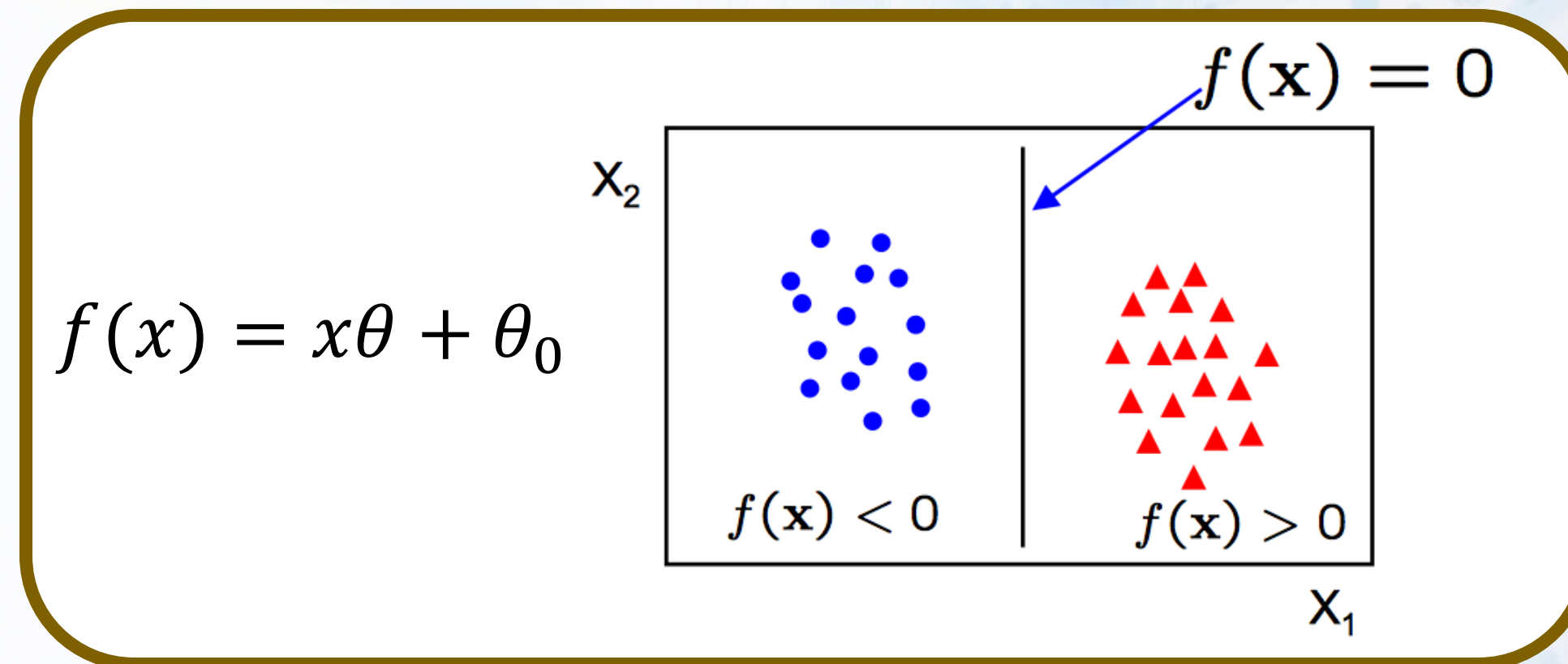
$$f(x) = x\theta + \theta_0$$

- In 2D, the discriminant is a line
- θ is the **normal** to the line, θ_0 and is the **bias term**
- θ is known as the model **parameters** or the **weight vector**

Linear Classifier (higher dimension)

A linear classifier has the form

$$f(x) = x\theta + \theta_0$$
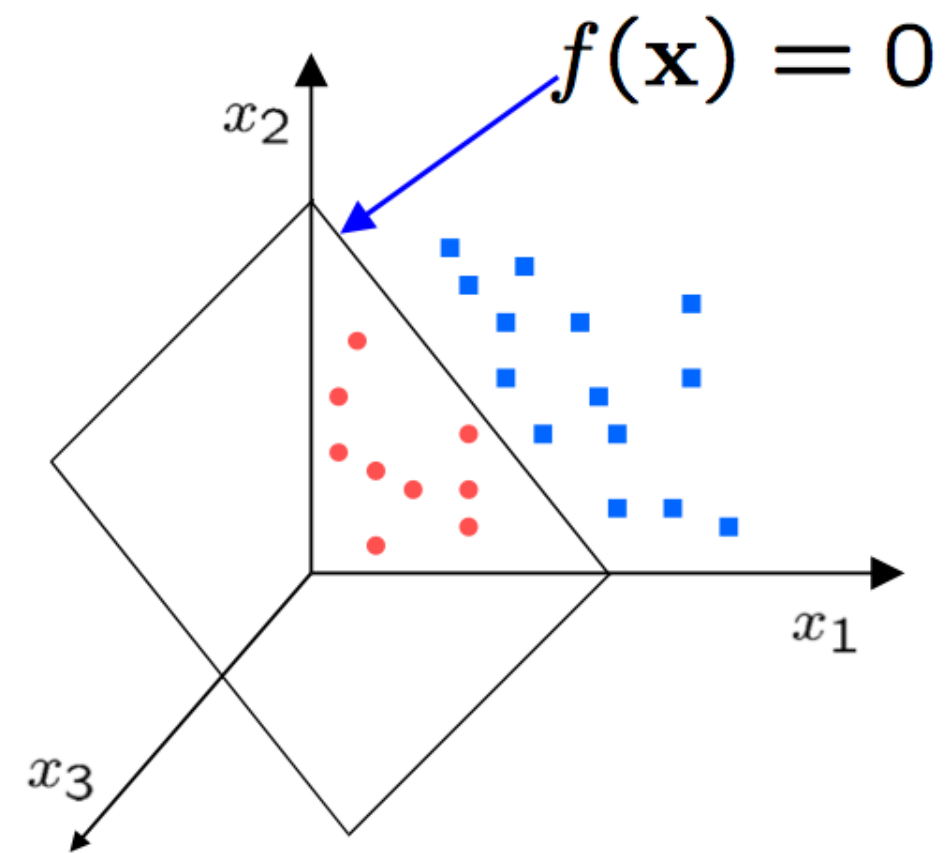
- In 3D the discriminant is a plane, and in nD it is a hyperplane

The Perceptron Classifier

Considering x is **linearly separable** and y has two labels of $\{-1,1\}$

$$f(x_i) = x_i\theta \qquad \text{Bias is inside } \theta \text{ now}$$

How can we separate datapoints with **label 1** from datapoints with **label** -1 using a **line**?

Perceptron Algorithm:

- Initialize $\theta = 0$
- Go through each datapoint $\{x_i, y_i\}$

  - If x_i is misclassified, then $\theta^{t+1} \leftarrow \theta^t + \alpha\, y_i x_i$

- Until all datapoints are correctly classified

Misclassified

Ex. $y_i f(x_i) < 0$

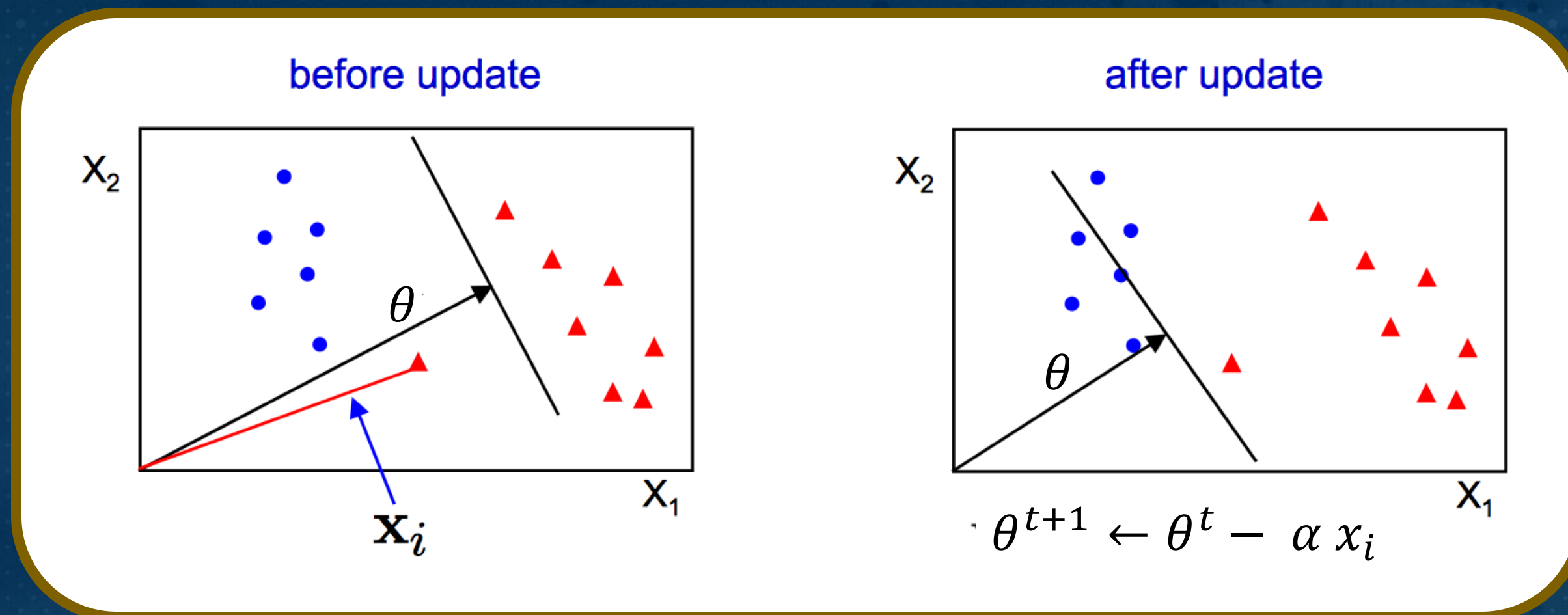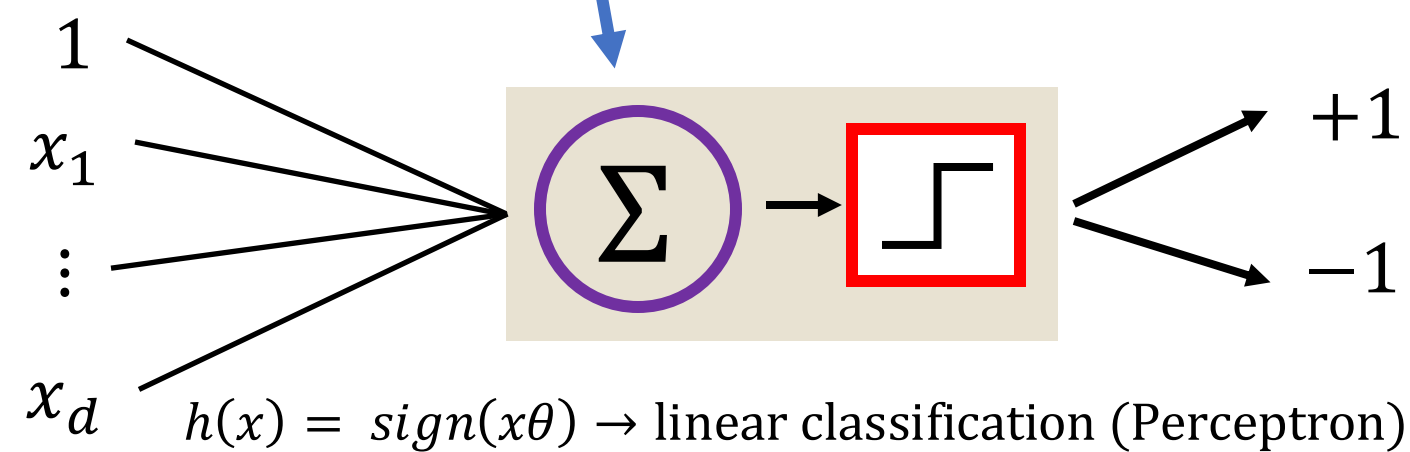actual   predicted
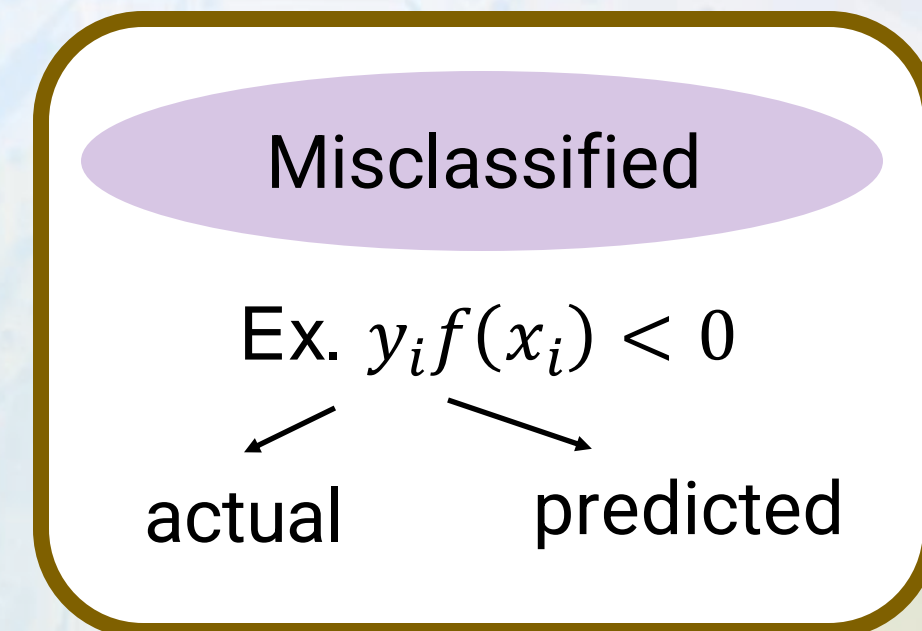
Perception Algorithm

- Initialize $\theta = 0$
- Go through each datapoint $\{x_i, y_i\}$
    - If x_i is misclassified, then $\theta^{t+1} \leftarrow \theta^t + \alpha y_i x_i$

- Until all datapoints are correctly classified

$$f(x) = x\theta + \theta_0$$

$$= \sum_{i=0}^{d} x_i \theta_i = \theta_0 + \theta_1 x_1 + \cdots + \theta_d x_d$$

Hard classification (perceptron)

$h(x) = sign(x\theta) \rightarrow$ linear classification (Perceptron)

Advantages and Disadvantages of Perceptron

- Advantages:
    - Very simple algorithm
    - It is fast and does not require many parameters
    - Quick training to optimize parameters

- Disadvantages:
    - It just works for linearly separable data
    - It will not provide a unique decision boundary

Summary

- We know what it means by a data to be linearly separable

- We learned about perceptron