Unit 1 - lecture 2: Linear Classifier and Perceptron

Objectives:

- Understand the concepts of Feature vectors and labels, Training Set and Test Set classifier, Training Error, Test Error and the Set of Classifiers.
- Derive the mathematical presentation of linear classifiers.
- Understand the intuitive and formal definition of linear separation.
- Use the perceptron algorithm with and without offset.

Basic Concepts:

Feature vector x: Provide the context for classifier to make predictions. $X \in \mathbb{R}^d$

labels Y: Targets or outputs. $Y \in \{-1; +1\}$

Supervised learning: Task where you are given the input and the corresponding output that you want. And you're supposed to learn irregularity between the two, in order to make predictions for future examples.

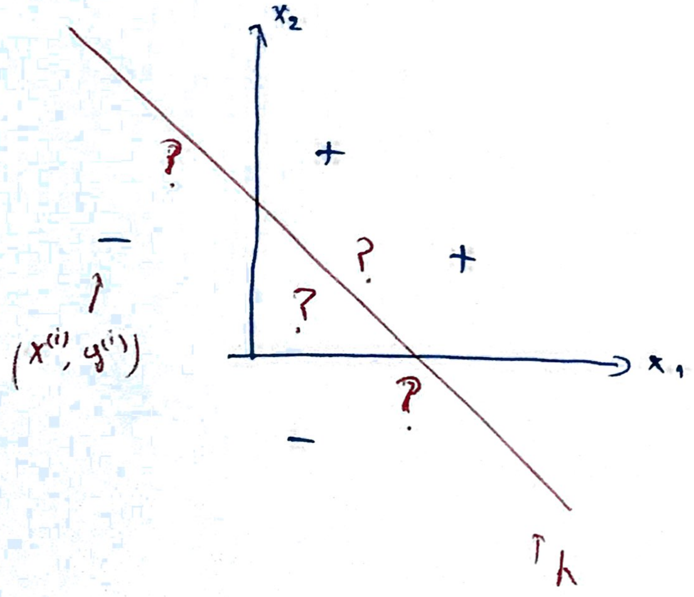
Training Set: $S_m = \{ (x^{(i)}, y^{(i)}), i = 1, \dots m \}$

Classifier: $h : \mathbb{R}^d \rightarrow \{-1; 1\}$, $h(x) = +$ or -1 Divide the space in two halfs.

Training Error: Evaluate how good that classifier is $\mathcal{E}_m(h) = \frac{1}{m} \sum_{i=1}^{m} [[\underbrace{h(x^{(i)}) \neq y^{(i)}}_{1 \text{ if error} \quad 0 \text{ otherwise}}]]$

Test Error: $\mathcal{E}(h)$

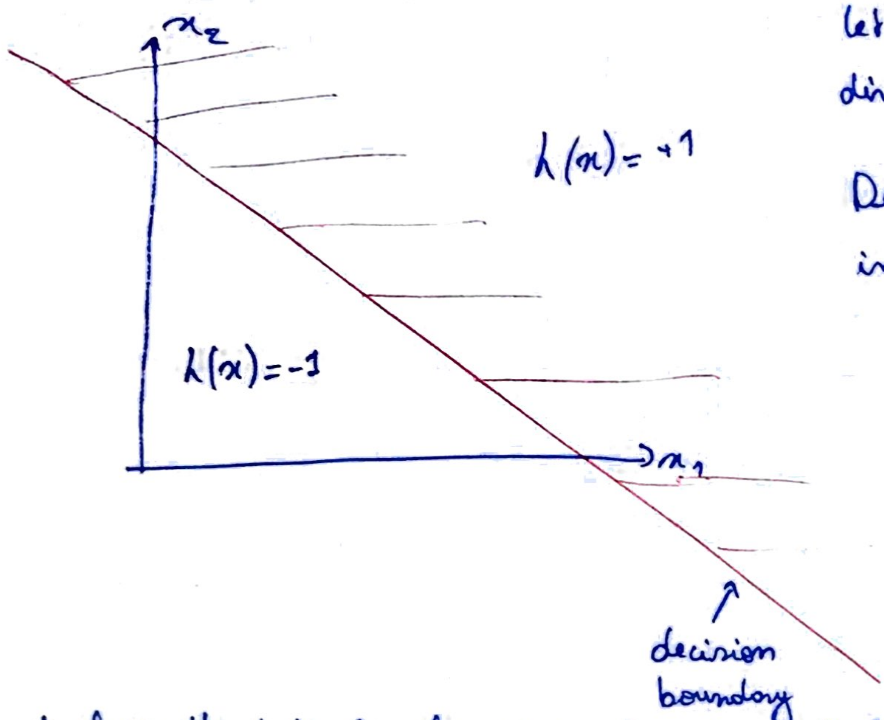
Set of classifiers: $h \in \mathcal{H}$



$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$\mathcal{E}_m(h) = 0$ no error of classification for the training set.

Linear Classifiers Mathematically Revisited:



Let's take a specific classifier that divides the space into 2 parts.

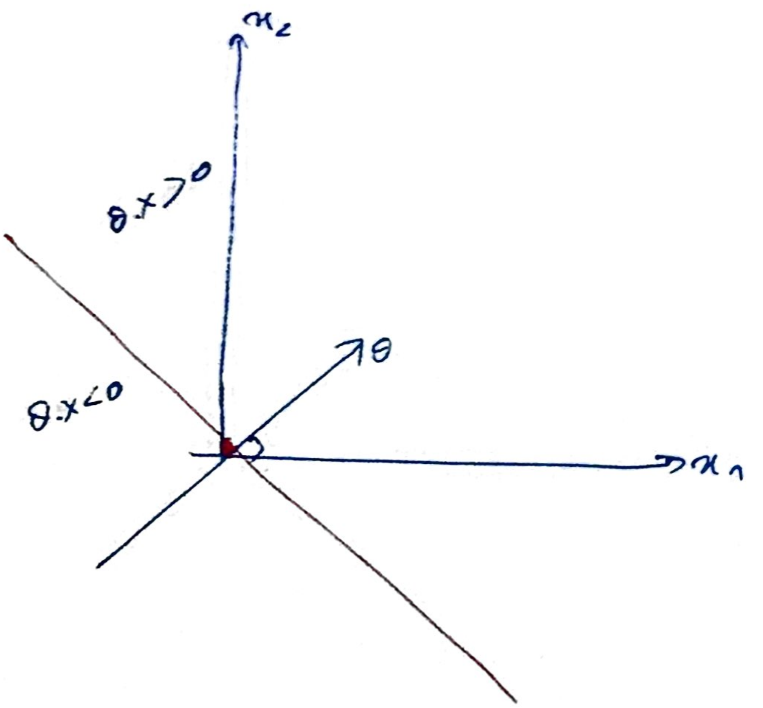
Decision boundary:

in 1-D : point

2-D : line

3-D : plane

We must choose the best classifier given a training set: 1. linear classifier through origin:



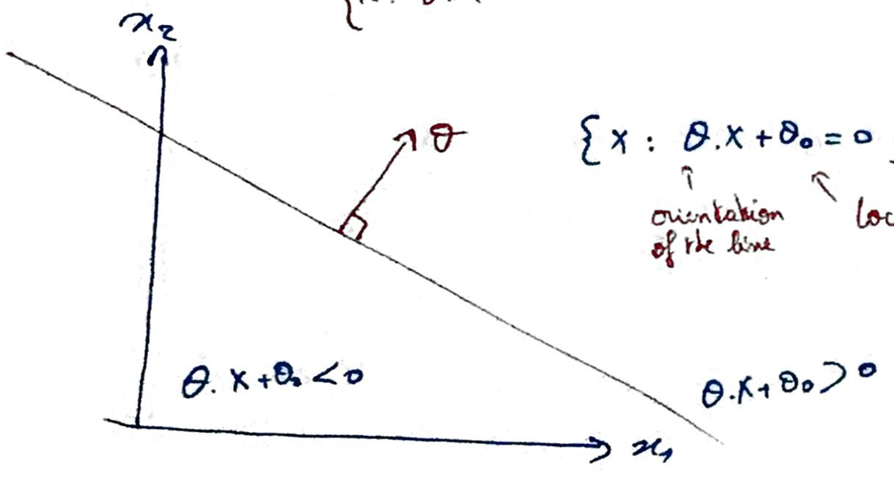
$\{X : \theta_1 x_1 + \theta_2 x_2 = 0\}$ dec. boundary

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$\{X : \theta \cdot X = 0\}$

dot product

$\{h(x;\theta) = \text{sign}(\theta . x)$
$\theta \in \mathbb{R}^d\}$ Set of linear classifiers through origin



$\{X : \theta . X + \theta_0 = 0\}$ dec. boundary.

orientation of the line

location

$\{h(X;\theta,\theta_0) = \text{sign}(\theta . X + \theta_0)$
$\theta \in \mathbb{R}^d, \theta_0 \in \mathbb{R}\}$

Set of linear classifiers.

Linear Seperation:

Given θ and θ_0, a linear classifier $h: X \rightarrow \{-1, 0, +1\}$ is a function that

Outputs → $+1$ if $\theta . X + \theta_0 > 0$
→ 0 if $\theta . X + \theta_0 = 0$
→ -1 if $\theta . X + \theta_0 < 0$

$\Longrightarrow h(x) = \text{sign}(\theta . X + \theta_0)$

$x^{(i)}$: feature vector

$y^{(i)}$: label

$\text{sign}(\theta . X + \theta_0)$: Output of the classifier h

For the i-th training data $(x^{(i)}, y^{(i)})$, $\underline{y^{(i)}}$ take conventionally $\begin{cases} -1 \\ +1 \end{cases}$

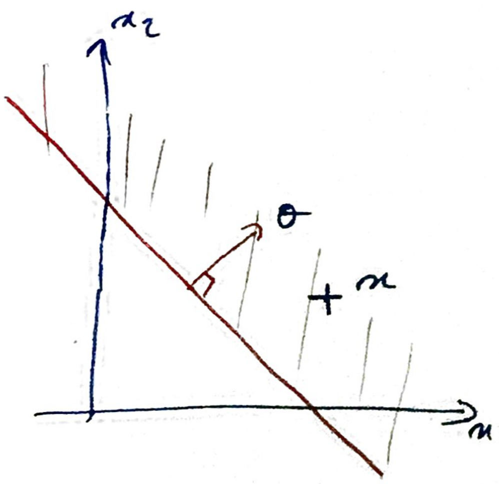
For the i-th training data $(x^{(i)}, y^{(i)})$, $\underline{\text{sign}(\theta . x^{(i)} + \theta_0)}$ take $\begin{cases} -1 \\ 0 \\ +1 \end{cases}$

Basics:

$y^{(i)}(\theta . x^{(i)}) > 0$ When → $y^{(i)} > 0$ and $\theta . x^{(i)} > 0$
→ $y^{(i)} < 0$ and $\theta . x^{(i)} < 0$

$y^{(i)}(\theta . x^{(i)}) > 0 \Longrightarrow$ $x^{(i)}$ label and classified result match

$y^{(i)}(\theta . x^{(i)}) < 0 \Longrightarrow$ $x^{(i)}$ label and classified result do not match.



Perceptron Algorithm:

Training error for a linear classifier (through origine):

$$\mathcal{E}_m(h) = \frac{1}{m} \sum_{i=1}^{m} [[\ h(x^{(i)}) \neq y^{(i)}\ ]]$$

We can write this down for linear classifiers through origine slightly differently:

$$\mathcal{E}_m(\theta) = \frac{1}{m} \sum_{i=1}^{m} [[\ \underbrace{y^{(i)}(\theta \cdot x^{(i)})} \leq 0\ ]]$$

if the sign of y^i and $\theta \cdot x^i$ is the same $\Rightarrow$ no error

if $y^i(\theta \cdot x^i) = 0$, the example lies on the decision boundary, we count it as an error.

Training Error for the general linear classifier:

$$\mathcal{E}_m(\theta, \theta_0) = \frac{1}{m} \sum_{i=1}^{m} [[\ y^{(i)}(\theta \cdot x^{(i)} + \theta_0) \leq 0\ ]]$$

Learning algorithm: perceptron

$\theta = 0$

if $y^{(i)}(\theta \cdot x^{(i)}) \leq 0$ then $\rightarrow$ • if error we update θ to correct the error.

$\theta = \theta + y^{(i)} x^{(i)}$ • the first example is always an error

---

Perceptron $(\{(x^{(i)}, y^{(i)}), i = 1, \dots, m\}, T)$

T we go through to training set many times $\rightarrow$ T time

$\theta = 0$ (vector), $\theta_0 = 0$ (scalar)

for $t = 1, \dots T$ do

for $i = 1, \dots m$ do

if $y^{(i)}(\theta \cdot x^{(i)} + \theta_0) \leq 0$ then

$$\begin{cases} \theta = \theta + y^{(i)} x^{(i)} \\ \theta_0 = \theta_0 + y^{(i)} \end{cases} \Bigg| \Rightarrow \begin{bmatrix} \theta \\ \theta_0 \end{bmatrix} = \begin{bmatrix} \theta \\ \theta_0 \end{bmatrix} + y^{(i)} \begin{bmatrix} x^{(i)} \\ 1 \end{bmatrix}$$

return θ, θ_0