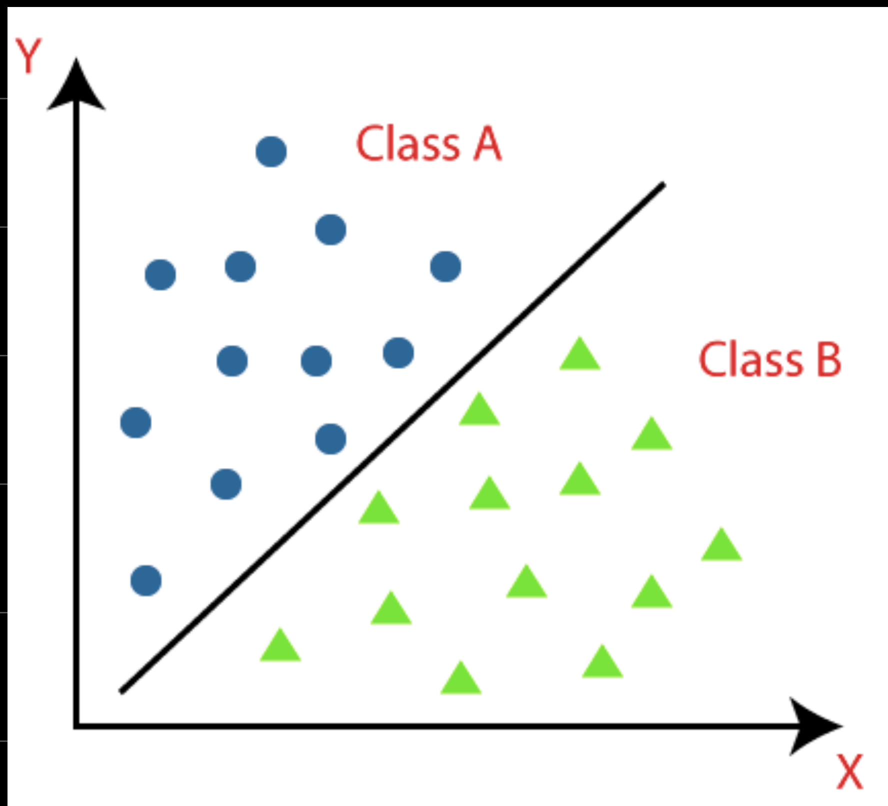
Chapter 4. Classification

Example:



Task of Classification: given sample pts with labels $(x^{(1)}, y_1), \cdots, (x^{(n)}, y_n)$ with $y_i = \pm 1$, find a hyperplane/hypersurface to classify $x^{(1)}, \cdots, x^{(n)}$ based on the sign of $y_1, \cdots, y_n$.

4.1 Perceptron

4.1.1. Motivation

Credit Approval Example

n: customers
m: features

Applicant information

	age	gender	salary	Yrs of residence	Yrs in job	...	...	Current debt	y	
X_1	x_{11}	x_{21}	x_{31}	...	...	...	...	x_{m1}	good	y_1
X_2	x_{12}	x_{22}	x_{32}	...	...	...	...	x_{m2}	bad	y_2
... X_i										... y_i
...										...
X_n	x_{1n}	x_{2n}	x_{3n}	...	...	...	...	x_{mn}	good	y_n

$\sum_{i=1}^{m} w_i\, x_i > threshold$ Approve credit

$\sum_{i=1}^{m} w_i x_i < threshold$ Deny credit

$$h(x) = \text{sign}\left\{\underbrace{\left[\sum_{i=1}^{m} w_i x_i\right]}_{\text{Credit score}} - threshold\right\}$$

For convenience: $threshold \longrightarrow w_0 x_0$
where $x_0 = 1$ Artificial feature (coordinate)

$$h(x) = \text{sign}\left(\sum_{i=0}^{m} w_i x_i\right)$$

4.1.2 Theory

Def: A perceptron is an algorithm for learning a binary classifier of the form:

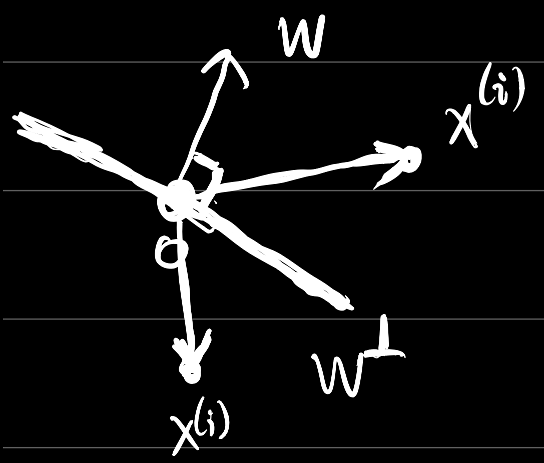
$$h(x) \overset{\text{def}}{=} \text{sign}\left(w_1 x_1 + \cdots + w_p x_p + b\right)$$

where $w_1, \cdots, w_p$ are weights and b is the bias.

Remark: (1) Without loss of generality, we may assume $b=0$. Otherwise replace $x = (x_1, \cdots, x_p)^T$ by $\tilde{x} = (1, x_1, \cdots, x_p)^T$ and $w = (w_1, \cdots, w_p)^T$ by $\tilde{w} = (-b, w_1, \cdots, w_p)^T$.

(2) Given a unit vector $w \in \mathbb{R}^p$ with $\|w\|=1$, the set $w^{\perp} \overset{\text{def}}{=} \{x \in \mathbb{R}^p : x \cdot w = 0\}$

is the hyperplane orthogonal to w.



A sample pt $x^{(i)}$ is "above" $w^{\perp}$ if and only if

$$x^{(i)} \cdot w > 0$$

A sample pt $x^{(i)}$ is "below" $w^{\perp}$ if and only if

$$x^{(i)} \cdot w < 0$$

(3) We say the perceptron <u>correctly classifies</u> $x^{(i)}$ if $h(x^{(i)}) = \text{sign}(w \cdot x^{(i)}) = y_i$.

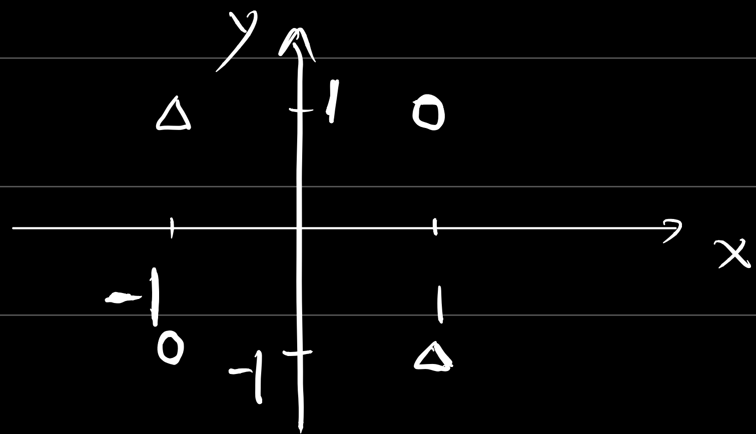
or equivalently. $y_i \cdot (w \cdot x^{(i)}) > 0$

We say the perceptron <u>misclassifies</u> $x^{(i)}$ if

$$h(x^{(i)}) = \text{sign}(w \cdot x^{(i)}) \neq y_i ,$$

or equivalently, $\quad y_i\,(w\cdot x^{(i)}) < 0.$

(4) Not all sample pts can be classified using hyperplanes, e.g.



We say $(x^{(1)}, y_1), \cdots, (x^{(n)}, y_n)$ are <u>linearly separable with const $\gamma > 0$</u> if $\exists$ a unit vector $w^* \in \mathbb{R}^p$, $\|w^*\| = 1$ such that

$$y_i \cdot (w^* \cdot x^{(i)}) > \gamma > 0 \quad i = 1, \cdots, n.$$