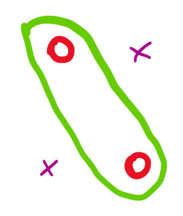
XOR PROBLEM AND MULTI-LAYER PERCEPTRON (MLP)

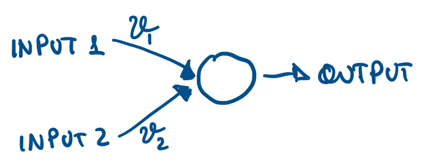
(PAG 62, CAP 3, SECTION 3.2.2 OF THE BOOK NEURAL NETWORK A SYSTEMATIC INTRODUCTION)

MINSKY AND PAPERT (1969) DEMONSTRATED THAT
THE XOR PROBLEM CANNOT BE SOLVED BY A PERCEPTRON

INPUT 1	INPUT 2	OUTPUT
1	1	0
1	0	1
0	1	1
0	0	0



PERCEPTRON MODEL



$0v_1 + 0v_2 < t \Rightarrow t > 0$ IF THE FIRST IS VALID

$0v_1 + 1v_2 > t \Rightarrow v_2 > t > 0$

$1v_1 + 0v_2 > t \Rightarrow v_1 > t > 0$

$\Rightarrow v_1 > 0,\ v_2 > 0 \quad v_1 + v_2 > 2t$

$1v_1 + 1v_2 < t \Rightarrow v_1 + v_2 < t$

$t > 2t$

INCONSISTENCE

⇒ THERE IS NO CONFIGURATION FOR A PERCEPTRON WHICH SOLVE XOR PROBLEM

PERCEPTRON CAN COMPUTE ONLY LINEAR DECISION BOUNDARIES!

TO SOLVE ⇒ ADD HIDDEN LAYER

PROBLEM ⇒ HOW TO TRAIN?
HOW TO DETERMINE HIDDEN PARAMETERS?

NON LINEAR SEPARABILITY

WHEN POINTS WITH LABEL $y=1$ CANNOT BE SEPARED WITH A LINE (HYPERPLANE) FROM THOSE WITH LABEL $y=0$ THEN THE PROBLEM IS NOT PERCEPTRON COMPUTABLE

⇒ PERCEPTRON CAN COMPUTE ONLY FUNCTIONS ON SET OF SEPARABLE POINTS

DEFINITION LINEAR SEPARABILITY 3.3 N-N A SYSTEMATIC APPROACH

TWO SETS OF POINTS A AND B IN A n-DIMENSIONAL SPACE ARE CALLED LINEAR SEPARABLE IF $n+1$ REAL NUMBER $w_1 \dots w_{n+1}$ EXIST, SUCH THAT EVERY POINT $(x_1, \dots x_n) \in A$ SATISFIES $\sum_{i=1}^{n} x_i w_i \geq w_{n+1}$ AND EVERY POINT $(x_1, \dots x_n) \in B$ SATISFIES $\sum_{i=1}^{n} w_i x_i < w_{n+1}$

HOW MANY LINEAR SEPARABLE FUNCTIONS OF n BINARY ARGUMENTS EXISTS?

$n=2 \Rightarrow$ 14 OUT OF 16 POSSIBLE BOOLEAN FUNCTIONS ARE SEPARABLE (LINEARLY)

$n=3 \Rightarrow$ 104 OF 256 ARE LINEAR SEPARABLE

$n=4 \Rightarrow$ 1882 OUT OF 65536 FUNCTION ARE LINEARLY SEPARABLE

THERE IS NO FORMULA WHICH EXPRESS THIS AS FUNCTION OF n

WE NEED PROOF

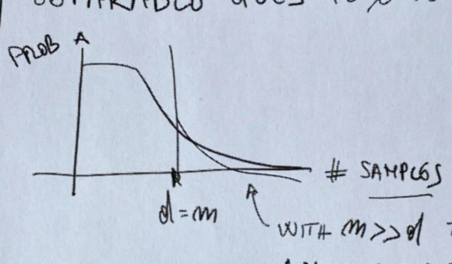
WE CAN GET THAT INCREASING n THE PROBABILITY OF A FUNCTION TO BE LINEAR SEPARABLE DECREASE DRASTICALLY

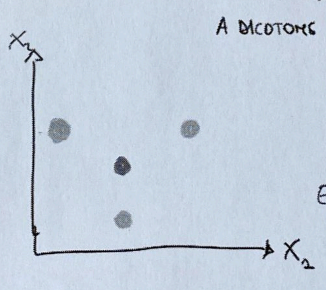
NOTE: USUALLY OUR INPUT IS HUGE WRT n (AND CONTINUOUS)

NON SEPARABILITY & LINEAR MACHINE

$X = \{x^{(1)}, x^{(2)}, \dots, x^{(m)}\} \in \mathbb{R}^d$

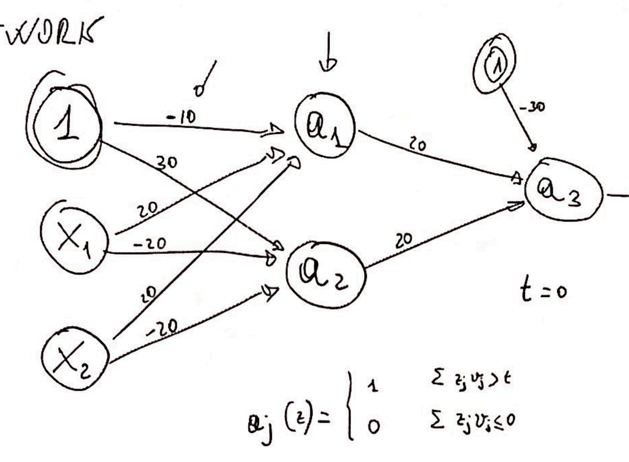
- GIVEN m points THERE ARE 2^m dicotomies (POSSIBILITY OF SEPARATION IN 2 GROUPS)
  FOR 2 POINTS ⇒ 2^2 DICOTOMIES
- ONLY ABOUT $\sqrt{d}$ dicotomies ARE LINEAR SEPARABLE
- WITH $m > d$ THE PROBABILITY THAT X IS LINEARY SEPARABLE GOES TO 0 VERY FAST





XOR NETWORK

CAPTER 6.1.2, 6.1.3 NN, A SYSTEMATIC INTRODUCTION



$a_j(z) = \begin{cases} 1 & \sum \geq 0 \\ 0 & \sum < 0 \end{cases}$

INPUT (0,0)

a_1: $0 \cdot 20 + 0 \cdot 20 - 10 = -10 < 0 \Rightarrow 0$

a_2: $0 \cdot -20 + 0 \cdot -20 + 30 = 30 > 0 \Rightarrow 1$

OUTPUT PURE

a_3: $0 \cdot 20 + 1 \cdot 20 - 30 = -10 < 0 \Rightarrow 0$

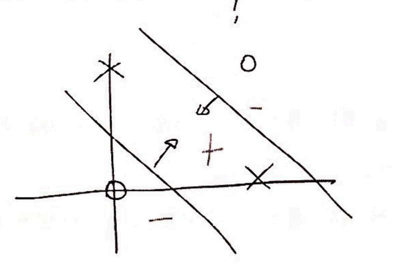
(1,1)

a_1: $1 \cdot 20 + 1 \cdot 20 - 10 = 30 > 0 \Rightarrow 1$

a_2: $1 \cdot -20 + 1 \cdot -20 + 30 = -10 < 0 \Rightarrow 0$

a_3: $1 \cdot 20 + 0 \cdot 20 - 30 = -10 < 0 \Rightarrow 0$

? COMPUTE FOR (0,1) AND (1,0) → EXERCISE



+ POSITIVE
− NEGATIVE

MULTIPLE HYPERPLANE

IN SUM

ESISTONO PROBLEMI CHE NON SONO RISOLVIBILI UTILIZZANDO FUNZIONI LINEARI. TALI PROBLEMI POSSONO RISOLVERSI UTILIZZANDO ARCHITETTURE PIU' COMPLESSE CHE PERMETTONO DI TRASFORMARE L'INPUT ATTRAVERSO FUNZIONI NON LINEARI.

(A) • RECAP PERCEPTRON MODEL (SUPERVISED LEARNING ALGORITHM)

1) MODEL: PARAMETRIZED FUNCTION, ONE WEIGHT FOR INPUT. FIRE IF $\geq t$
2) TRAINING SET $T: x^{(i)} \in \mathbb{R}^n$, $y^{(i)} \in \{0,1\}$: SET OF COUPLE $(x^{(i)}, y^{(i)})$
3) LEARNING ALGORITHM – ROSENBLATT 1958 | INTUITION MATH
4) PERFORMANCE MEASUREMENT (COST FUNCTION) $J(\theta) = \sum_{i=1}^{m} 1\{\hat{y}^{(i)} \neq y^{(i)}\}$

GOAL: CLASSIFICATION – BINARY $y^{(i)} \in \{0,1\}$

$\min_{\theta} J(\theta) = 0$

(B) • TRAINING SET, VALIDATION SET, TEST SET
- RANDOM SPLITTING
- CROSS VALIDATION

(E) • EVALUATION ON TEST (SAME AS COST FUNCTION)

(F) • CAN WE USE FOR REGRESSION? ⇒ NO SIGN FUNCTION
- CAN WE INTERPRET IN A PROB FASHION? ⇒ SIGMOID (NON LINEAR)
- KERNEL & BASIS FUNCTION REPRESENTATION

1969
- LINEAR SEPARABILITY DEFINITION RISPETTO AL MODELLO PERCEPTRONE, HOW MANY SEPARABLE FUNCTION
- XOR PROBLEM ⇒ POSSIBLE SOLUTION ⇒ PROBLEM WITH LEARNING
- NON SEPARABILITY & LINEAR MACHINE: DICOTOMIES

• TYPE OF LEARNING (1)
- SUPERVISED
- UNSUPERVISED
- REINFORCEMENT
- RECOMMENDER SYSTEMS

(C) • LEARNING THE THRESHOLD
⇒ CHANGE DATA $[1, x^{(i)}]^T$
⇒ ADD PARAMETERS TO THE MODEL θ_0

(D) DATA NORMALIZATION
• EACH COMPONENT
$\frac{x_j - \bar{x}_j}{\sigma_j}$, $\frac{x_j - x_j^{min}}{x_j^{max} - x_j^{min}}$
⇒ $x_j \in [0,1]$

⇒ DATA DEPENDENCE

$\sigma_j = \sqrt{\frac{\sum_{i=1}^{m} (x_j^{(i)} - \mu_j)^2}{m}}$

$\mu_j = \sum_{i=1}^{m} \frac{x_j^{(i)}}{m}$