

APPENDIMENTO SUPERVISATO

COSA SIGNIFICA APPRENDERE?

Supponiamo di avere un dataset:

$$D = \langle X, Y \rangle = \{x_i, y_i\}_{i \in \mathbb{N}}$$

x_i : DATI

y_i : LABEL

$$x_i \in \mathbb{R}^n, y_i \in \mathbb{R}^m$$

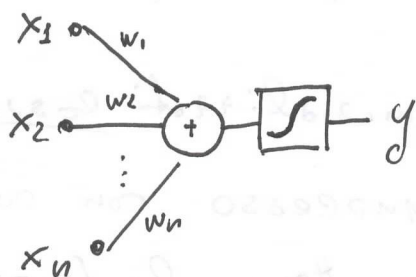
Apprendere significa trovare una funzione che ad ogni x_i associa il suo y_i .

$$\text{APPRENDERE} \iff \text{TROVARE } f: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

$x_i \mapsto y_i$

NEURONI ARTIFICIALI

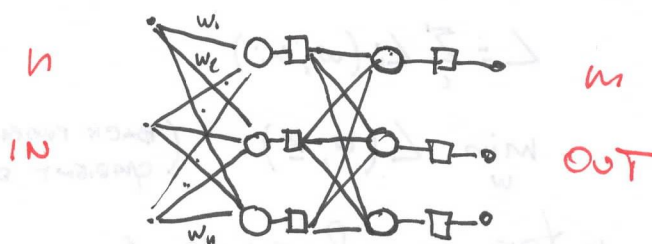
$$f: \mathbb{R}^n \rightarrow \mathbb{R}$$



$$y = x_1 w_1 + x_2 w_2 + \dots + x_n w_n$$

w_1, w_2, \dots, w_n PESI O PARAMETRI

Usando più neuroni possiamo realizzare una rete complessa:



$$f: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

Il teorema UFA (Universal Function Approximation) afferma che è sempre possibile trovare una rete neurale che approssimi qualunque funzione continua, su un dominio compatto.

Questo significa che, dato un numero sufficiente (potenzialmente INFINITO) di neuroni ~~usando~~ una rete neurale può apprendere qualsiasi cosa!!!

Le reti neurali vengono solitamente usate per:

previsioni \longleftrightarrow "REGRESSIONI"
classificazioni

INFERENCE significa TRARRE UNA CONCLUSIONE -

INFERENCE vs. TRAINING

Dato una rete neurale con pesi inizializzati casualmente definiamo ADDESTRAMENTO quel processo con cui modifichiamo i pesi DELLA RETE. per ottenere la funzione che meglio approssima la $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$

- è
P
O
C
A
1. Costruiamo $L_i(w, x_i) = y_i - \hat{y}_i \quad \forall i \in N$
 2. $L = \sum_i L_i(w, x_i)$
 3. $\min_w L(w, x)$ (BACK PROPAGATION GRADIENT DESCENT)
 4. ritorniamo al punto 1

Questo procedimento avviene, di solito, su un unico elaboratore.

Dato un'insieme di macchine è poss. h. le distribuire
il processo di apprendimento su più macchine?

Sì!

FEDERATED LEARNING.

1. Il dataset viene splittato
su più macchine

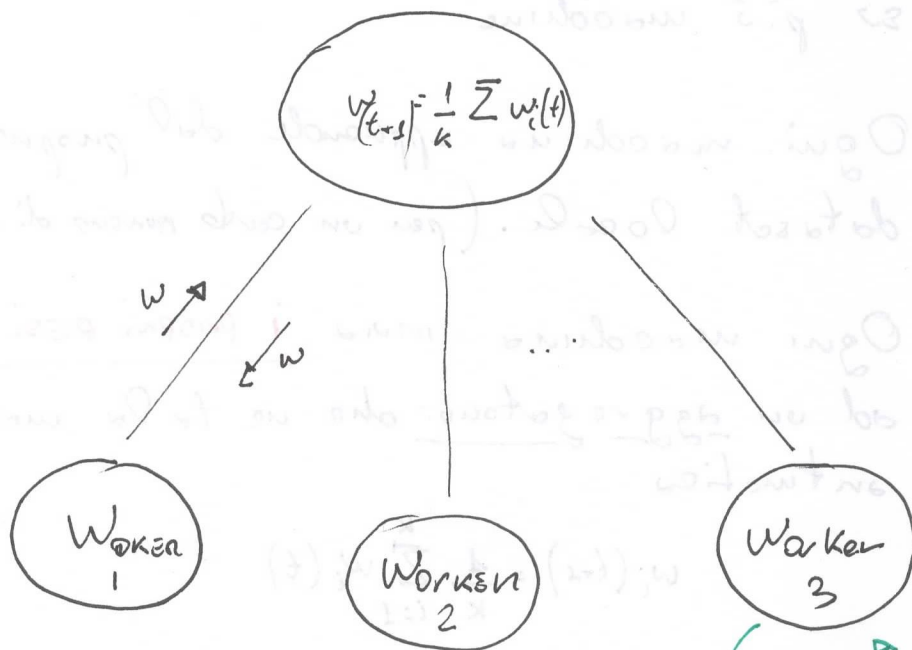
2. Ogni macchina apprende dal proprio
dataset locale. (per un certo numero di epoche) NON
PATI !!

3. Ogni macchina invia i propri pesi
ad un aggregatore che ne fa la media
aritmetica ROUND

$$w_i(t_{r+1}) = \frac{1}{K} \sum_{i=1}^K w_i(t)$$

4. L'aggregatore manda ~~all'indietro~~ i pesi
ottenuti dal contributo di tutti i worker
ad ogni worker...

Ognuno di essi aggiorna il proprio
modello interno.



↑
D
O
O
O
↑

epoch

