



UNIVERSITÀ DEGLI STUDI DI FIRENZE  
SCUOLA DI INGEGNERIA - DIPARTIMENTO DI INGEGNERIA  
DELL'INFORMAZIONE

---

Tesi di Laurea Triennale in Ingegneria Informatica

**APPRENDIMENTO DEMOCRATIZZATO  
AUTO-ORGANIZZATO**

*Candidato*  
Leonardo Pampaloni

*Relatore*  
Romano Fantacci

*Correlatori*  
Benedetta Picano

---

Anno Accademico 2022/2023

# Indice

|   |    |
|---|----|
| Introduzione                                  | i  |
| 1 Algoritmo di apprendimento federato         | 1  |
| 2 Apprendimento democratizzato: concetti      | 3  |
| 3 Apprendimento democratizzato: Progettazione | 7  |
| 4 Conclusioni                                 | 16 |
| Bibliografia                                  | 17 |

# Introduzione

Le intelligenze artificiali sono riuscite al giorno d'oggi a risolvere innumerevoli problemi della vita reale come ad esempio il controllo avanzato di alcuni sistemi di automazione, le telecomunicazioni e la robotica. Anche in campo mobile vi sono molti servizi che incorporano IA che sfruttano i dati degli utenti al fine di proporre sistemi altamente personalizzati per le specifiche dell'utente come ad esempio la stessa tastiera qwerty dei nostri dispositivi che è in grado di memorizzare come scriviamo e 'impara' a proporre le prossime parole. Sfruttando le caratteristiche personali dei vari utenti non solo viene migliorata l'esperienza personale ma aiutano anche a controllare meglio i loro dispositivi.

La crescente preoccupazione per la privacy dei dati nei vari framework di machine learning esistenti ha alimentato un crescente interesse nello sviluppo di paradigmi di machine learning distribuiti come i framework di apprendimento federato che illustreremo in seguito.

# Capitolo 1

## Algoritmo di apprendimento federato

Il Federated Learning è una tecnologia di apprendimento automatico distribuito che permette di addestrare un modello di intelligenza artificiale su un gran numero di dispositivi, mantenendo i dati su ciascun dispositivo. FL (federated learning) è strutturato in modo che gli agenti di apprendimento si coordinino tramite un server centrale per addestrare un modello di apprendimento globale in modo distribuito. Questi agenti ricevono il modello di apprendimento locale in base ai set di dati disponibili. Quindi restituiscono i modelli di apprendimento aggiornati al server per l'aggiornamento del modello globale tramite un'operazione di aggregazione senza rivelare i dati di addestramento privato agli altri. In pratica, il set di dati privato raccolto presso ciascun agente è sbilanciato, altamente personalizzato per alcune applicazioni come la scrittura a mano e il riconoscimento vocale e presenta caratteristiche non indipendenti e non identicamente distribuite. Pertanto il processo iterativo di aggiornamento del modello globale migliora la generalizzazione del modello ma danneggia anche le prestazioni personalizzate degli

agenti. Quindi gli algoritmi FL esistenti non possono gestire in modo efficiente la relazione coesiva tra la capacità di generalizzazione e personalizzazione del modello di apprendimento addestrato. Vi sono stati alcuni tentativi di studiare e migliorare le prestazioni personalizzate di FL utilizzando un algoritmo di media federata personalizzata (Per-FedAvg) basato su un framework di meta-apprendimento. Si è anche provato ad avere un framework personalizzato adattivo in cui è stata adottata una combinazione di modelli locali e globali per ridurre l'errore di generalizzazione. Tuttavia in entrambi i casi la relazione coesiva tra generalizzazione e personalizzazione non è stata adeguatamente analizzata.

In questo elaborato svilupperemo un nuovo framework di apprendimento distribuito che può estendere direttamente lo schema FL convenzionale per risolvere collettivamente un compito di apprendimento comune ai vari agenti. Diversamente dagli algoritmi FL esistenti per la costruzione di un singolo modello generalizzato, manteniamo modelli di gruppo gerarchici autoorganizzanti. Di conseguenza adottiamo il clustering gerarchico agglomerativo e aggiorniamo periodicamente la struttura gerarchica in base alla somiglianza nelle caratteristiche di apprendimento degli utenti. In particolare proponiamo la generalizzazione gerarchica in forma ricorsiva. Per risolvere il complesso problema dato dalla forma ricorsiva, consideriamo l'algoritmo di apprendimento distribuito: DemLearn.

## Capitolo 2

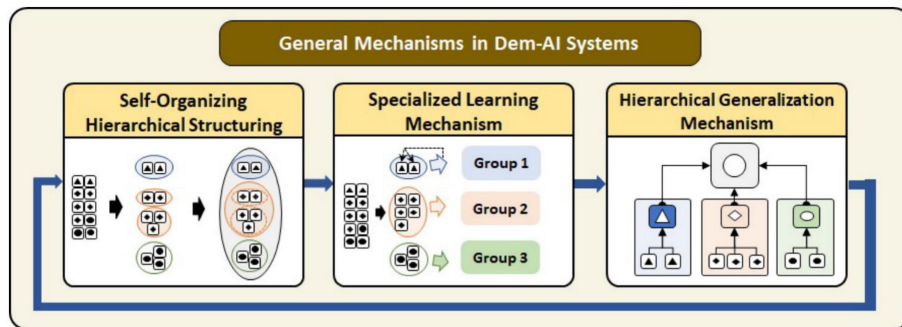
# Apprendimento democratizzato: concetti

L'algoritmo proposto utilizza lo schema bottom-up per eseguire in modo iterativo l'apprendimento locale risolvendo problemi di apprendimento personalizzati e aggiornando gerarchicamente i modelli generalizzati per gruppi a livelli superiori.

Diversamente dall'algoritmo di apprendimento federato, il Dem-AI introduce una struttura gerarchica auto-organizzante per risolvere compiti di apprendimento complessi singoli/multipli mediando i contributi di un gran numero di agenti di apprendimento in un apprendimento collaborativo. Secondo le differenze nelle loro caratteristiche, gli agenti di apprendimento formano gruppi appropriati che possono essere specializzati affinché agenti simili affrontino i compiti di apprendimento. Questi gruppi specializzati sono auto-organizzati in una struttura gerarchica e costruiscono collettivamente la conoscenza di apprendimento generalizzata condivisa per migliorare le loro prestazioni di apprendimento riducendo i problemi individuali dovuti ai dati locali sbilanciati e altamente personalizzati. In particolare, il sistema di apprendimento

consente ai nuovi membri del gruppo di accelerare il loro processo di apprendimento con la conoscenza del gruppo esistente e incorporare le loro nuove conoscenze di apprendimento espandendo la capacità di generalizzazione dell'intero gruppo. In questi sistemi ogni agente è libero di unirsi ad uno dei qualsiasi gruppi e possiedono lo stesso potere inteso come numero dei componenti che varierà poi nel tempo.

Consideriamo quindi i punti principali del Dem-AI.



- Definizione ed obiettivo

L'apprendimento democratizzato studia processi duali ,accoppiati e operanti insieme, specializzato-generalizzato in una struttura gerarchica auto-organizzante di sistemi di apprendimento distribuiti su larga scala. I processi specializzati e generalizzati devono operare congiuntamente verso un obiettivo di apprendimento finale identificato come l'esecuzione di un apprendimento collettivo da agenti di apprendimento prevenuti, che si impegnano ad apprendere dai propri dati utilizzando le loro limitate capacità di apprendimento. In quanto tale, l'obiettivo finale di apprendimento del sistema Dem-AI è stabilire un meccanismo per risolvere collettivamente compiti di apprendimento complessi comuni (singoli o multipli) da un gran numero di agenti di apprendimento

- Processo specializzato

Questo processo viene utilizzato per sfruttare le capacità di apprendimento specializzate degli agenti di apprendimento e dei gruppi specializzati sfruttando i dati raccolti. Incorporando la conoscenza generalizzata dei gruppi di livello superiore creati dal meccanismo di generalizzazione, gli agenti di apprendimento possono aggiornare i parametri del loro modello per ridurre il bias nel loro apprendimento personalizzato. Pertanto, l'obiettivo di apprendimento personalizzato ha come obiettivi: eseguire un apprendimento specializzato e riutilizzare la conoscenza gerarchica generalizzata disponibile.

- **Processo generalizzato**

Il meccanismo di generalizzazione incoraggia i membri del gruppo a condividere la conoscenza quando svolgono compiti di apprendimento con caratteristiche simili e costruiscono livelli gerarchici di conoscenza generalizzata. La conoscenza gerarchica generalizzata aiuta il sistema Dem-AI a mantenere la capacità di generalizzazione per ridurre il bias degli agenti di apprendimento e affrontare in modo efficiente i cambiamenti ambientali o eseguire nuovi compiti di apprendimento.

- **Struttura gerarchica auto-organizzante**

La struttura gerarchica dei gruppi specializzati e le relative conoscenze generalizzate sono costruite e regolate secondo un principio di autoorganizzazione basato sulla somiglianza degli agenti di apprendimento. In particolare, questo principio governa l'unione di piccoli gruppi per formare un gruppo più grande che alla fine migliora le capacità di generalizzazione di tutti i membri. Pertanto, i gruppi specializzati ai livelli più alti nella struttura gerarchica hanno più membri e possono costruire una conoscenza più generalizzata e meno distorta adattandosi più



rapidamente ai nuovi ambienti.

- Transizione nel duplice processo specializzato-generalizzato

Il processo specializzato diventa sempre più importante rispetto al processo generalizzato durante il periodo di formazione. Di conseguenza, il sistema di apprendimento non solo si evolve per acquisire capacità di specializzazione dai compiti appresi, ma perde anche la capacità di affrontare i cambiamenti ambientali come nuovi agenti di apprendimento e nuovi compiti di apprendimento. Nel frattempo, la struttura gerarchica del sistema Dem-AI è auto-organizzata e si è evoluta da un alto livello di plasticità a un alto livello di stabilità, cioè da gruppi specializzati instabili a gruppi specializzati ben organizzati. La transizione del sistema di apprendimento Dem-AI è illustrata in figura con tre sub-meccanismi iterativi come la generalizzazione, l'apprendimento specializzato e il meccanismo di strutturazione gerarchica. Di conseguenza, la transizione del processo duale specializzato-generalizzato rappresenta le fasi di un tipico quadro di apprendimento democratizzato. In quella transizione, gli agenti di apprendimento sono raggruppati in base alle somiglianze dei loro compiti di apprendimento nella fase iniziale. Quindi, il processo generalizzato aiuta nella costruzione di una conoscenza gerarchica generalizzata per i gruppi specializzati dal basso verso l'alto e incoraggia i membri del gruppo a essere vicini. Nel frattempo, i processi di apprendimento specializzati sfruttano l'apprendimento personalizzato per sfruttare i loro set di dati distorti incorporando la conoscenza di gruppo generalizzata di livello superiore dai gruppi di livello superiore a quelli di livello inferiore.

## Capitolo 3

# Apprendimento democratizzato: Progettazione

In questo elaborato, studiamo un nuovo algoritmo di apprendimento distribuito che consiste in: raggruppamento gerarchico, generalizzazione gerarchica e meccanismi di apprendimento con un compito di apprendimento comune per tutti gli agenti di apprendimento.

- Meccanismo di raggruppamento gerarchico

Per costruire la struttura gerarchica del sistema Dem-AI con gruppi di apprendimento specializzati pertinenti, adottiamo l'algoritmo di clustering gerarchico agglomerativo ovvero l'implementazione del dendrogramma da scikit-learn, basato sulla somiglianza o non somiglianza di tutti gli agenti di apprendimento. Il metodo del dendrogramma viene utilizzato per esaminare le relazioni di somiglianza tra gli individui ed è spesso utilizzato per l'analisi dei cluster in molti campi di ricerca. Durante l'implementazione, la topologia ad albero del dendrogramma

viene costruita unendo le coppie di agenti o cluster che hanno la distanza minore tra loro, seguendo lo schema bottom-up. Di conseguenza, la distanza misurata è considerata come le differenze nelle caratteristiche degli agenti di apprendimento (ad esempio, parametri del modello locale o gradienti della funzione dell'obiettivo di apprendimento). Poiché otteniamo prestazioni simili implementando il clustering basato su parametri o gradienti del modello, in quanto segue presentiamo solo un meccanismo di clustering utilizzando i parametri del modello locale.

Dati i parametri del modello locale  $\omega_n = (\omega_{n,1}, \dots, \omega_{n,M})$  dell'agente di apprendimento  $n$ , dove  $M$  è il numero di parametri di apprendimento, la misura della distanza tra due agenti  $\phi_{n,l}$  è derivata in base al modello euclideo distanza come  $\phi_{n,l} = \|\omega_n - \omega_l\|$ . Inoltre, consideriamo il metodo del collegamento medio per il calcolo della distanza tra un agente e un cluster utilizzando la distanza euclidea tra i parametri del modello dell'agente e i parametri del modello medio dei membri del cluster. Di conseguenza, la struttura gerarchica ad albero è sotto forma di un albero binario con molti livelli. Richiederà quindi, costi computazionali e di archiviazione inutilmente elevati per mantenere i dati ed è un modo inefficiente per mantenere un gran numero di modelli generalizzati di basso livello per piccoli gruppi. Di conseguenza, manteniamo solo i  $K$  livelli superiori nella struttura ad albero e scartiamo la struttura di livello inferiore. Pertanto, al livello  $K$  più alto, il sistema potrebbe avere due grandi gruppi che hanno un gran numero di agenti di apprendimento.

- Generalizzazione gerarchica e meccanismo di apprendimento

La struttura gerarchica di livello  $K$  emerge attraverso il clustering ag-

glomerante. Di conseguenza, il sistema costruisce  $K$  livelli della generalizzazione, come in Figura. Pertanto, proponiamo problemi di apprendimento generalizzato gerarchico (HGLP) per costruire questi modelli generalizzati per gruppi specializzati in una forma ricorsiva, a partire dalla costruzione del modello globale  $w^K$  al livello superiore  $K$  come segue.

Problema HGLP al livello  $K$ .

$$\begin{aligned} \min_{W^{(K)}} L^{(K)} &= \sum_{i \in S_k} \frac{N_{g,i}^{(K-1)}}{N_g^{(K)}} L_i^{(K-1)}(w_i^{(K-1)} | D_i^{(K-1)}) \\ \text{s.t. } w^{(K)} &= w_i^{(K-1)}, \forall i \in S_k \end{aligned} \quad (1)$$

Dove  $W^{(K)} = (w^{(K)}, w_1^{(K-1)}, \dots, w_{|S_k|}^{(K-1)})$ ,  $S_k$  è l'insieme dei sottogruppi del gruppo di livello superiore e  $L_i^{(K-1)}$  è la funzione di perdita del sottogruppo  $i$  dato il suo set di dati collettivo  $D_i$ . La funzione obbiettivo è pesata su una frazione del numero di agenti totali  $N_g^{(K-1)}$  del sottogruppo  $i$  e  $N_g^{(K)}$  agenti totali. Quindi i sottogruppi che hanno più agenti di apprendimento hanno un impatto superiore sul modello globale a livello  $K$ . I rigidi vincoli in (2) incoraggiano questi sottogruppi a condividere un modello di apprendimento comune. Per preservare le capacità di specializzazione di ciascun sottogruppo, questi vincoli (2) potrebbero essere allentati utilizzando termini prossimali aggiuntivi nell'obiettivo. In questo modo, il problema incoraggia i modelli di apprendimento del sottogruppo ad avvicinarsi al modello globale ma non essendo necessariamente uguali. Pertanto, il problema rilassato HGLP' è definito come segue. Problema HGLP' al livello  $K$ .

$$\min_{W^{(K)}} \sum_{i \in S_k} \frac{N_{g,i}^{(K-1)}}{N_g^{(K)}} (L_i^{(K-1)}(w_i^{(K-1)} | D_i^{(K-1)}) + \frac{\mu_K}{2} \|w^{(K)} - w_i^{(K-1)}\|^2)$$

(3)

Dove  $\mu_K$  denota il compromesso tra la perdita di apprendimento e il vincolo di generalizzazione che impone ai modelli di apprendimento di gruppo di essere vicini al modello globale  $w^{(K)}$ . Poiché il set di dati è distribuito e disponibile solo presso gli agenti di apprendimento, il problema (3) al livello superiore  $\mathbf{K}$  può essere risolto partendo prima dal problema dei suoi membri. Di conseguenza, la struttura gerarchica generalizzata è emersa naturalmente seguendo lo schema bottom-up in cui i modelli di apprendimento ai livelli inferiori vengono aggiornati prima di risolvere i problemi generalizzati di livello superiore del gruppo superiore. Nello specifico, il problema (3) può essere decentralizzato e risolto dal seguente problema di ciascun sottogruppo  $i$  al livello  $K - 1$ . Problema HGLP per ogni gruppo  $i$  al livello  $K - 1$ .

$$\begin{aligned} \min_{W^{(K-1)}} \sum_{j \in S_{i,K-1}} \frac{N_{g,j}^{(K-2)}}{N_g^{(K)}} (L_j^{(K-2)}(w_j^{(K-2)} | D_j^{(K-2)})) \\ + \frac{\mu_{K-1}}{2} \|w_j^{(K-2)} - w_i^{(K-1)}\|^2 + \frac{\mu_K N_{g,i}^{(K-1)}}{2N_g^{(K)}} \|w^{(K)} - w_i^{(K-1)}\|^2 \end{aligned}$$

Dove  $W^{(K-1)} = (w_i^{(K-1)}, w_1^{(K-2)}, \dots, w_{|S_{i,K-1}|}^{(K-2)})$  e facciamo una forma di approssimazione generale del problema generalizzato di apprendimento per il gruppo  $i$  a livello  $\mathbf{K}$  dato il precedente modello generalizzato  $w^{(k+1)}$  nel seguente modo.

Problema HGLP per ogni gruppo  $i$  al livello  $K$ .

$$\begin{aligned} \min_{W^{(K)}} \sum_{j \in S_{i,K}} \frac{N_{g,j}^{(K-1)}}{N_g^{(K)}} (L_j^{(K-1)}(w_j^{(K-1)} | D_j^{(K-1)})) \\ + \frac{\mu_K}{2} \|w_j^{(K-1)} - w_i^{(K)}\|^2 + \frac{\mu_{K+1} N_{g,i}^{(K-1)}}{2N_g^{(K)}} \|w^{(K+1)} - w_i^{(K)}\|^2 \end{aligned} \quad (4)$$

Dove  $W^{(K)} = (w_i^{(k)}, w_1^{(k-1)}, \dots, w_{|s_{i,k}|}^{(k-1)})$ ,  $N_{g,i}^{(k)}$  è il numero di agenti di apprendimento del gruppo  $i$ , e  $w^{(k+1)}$  è il modello di apprendimento del gruppo superiore al livello  $k + 1$  a cui appartiene il gruppo  $i$ . Poiché esiste accoppiamento tra i livelli superiore e inferiore, e il set di dati di addestramento è decentralizzato, il problema di apprendimento (4) del gruppo  $i$  al livello  $k$  non può essere risolto direttamente. Quindi, analogamente a FL, la perdita di apprendimento del gruppo può essere distribuita tra i membri del gruppo [2]. Di conseguenza, l'obiettivo del problema di gruppo ha i restanti termini prossimali che costringono i modelli di apprendimento nei diversi livelli ad essere vicini l'uno all'altro. Pertanto, il modello di apprendimento è costruito con il modello del gruppo superiore a livello  $k + 1$  e i membri del gruppo a livello  $k - 1$  modellano risolvendo il seguente problema:

$$\min_W \sum_{j \in S_{i,K}} \frac{\mu_k N_{g,j}^{(K-1)}}{N_g^{(K)}} \|w_j^{(k-1)} - w\|^2 + \frac{\mu_{K+1} N_{g,i}^{(k)}}{N_g^{(K)}} \|w^{(K+1)} - w\|^2 \quad (5)$$

La forma chiusa della soluzione ottima del problema (5) può essere facilmente derivata impostando il gradiente a zero come segue:

$$\sum_{j \in S_{i,K}} \mu_k N_{g,j}^{(k-1)} (w_j^{(k-1)} - w^*) = \mu_{k+1} N_{g,i}^{(k)} (w^* - w^{(k+1)})$$

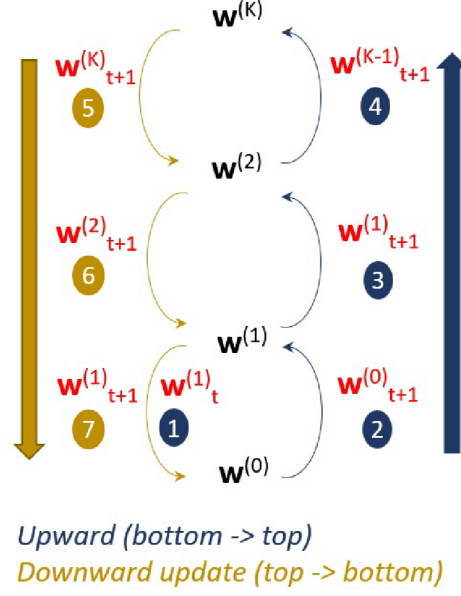
Pertanto, il modello di apprendimento del gruppo  $i$  può essere aggiornato come

$$w_i^{(k)} = \alpha w^{(k+1)} + (1 - \alpha) \sum_{j \in S_{(i,k)}} \frac{N_{g,j}^{(k)}}{N_{g,i}^{(k)}} w_j^{(k-1)} \quad (6)$$

Dove  $\alpha = \mu_{(k+1)} / (\mu_k + \mu_{(k+1)})$ . Il parametro di compromesso  $\alpha$  può essere regolato successivamente negli esperimenti per controllare il contributo dei modelli di apprendimento dei membri del gruppo superiore e del gruppo. Data la soluzione in forma chiusa (6), l'accoppiamento tra diversi livelli può essere approssimato suddividendo gli aggiornamenti del modello a ciascun livello tramite lo schema bottom-top e poi top bottom. In Fig. 3, mostriamo l'illustrazione degli aggiornamenti gerarchici proposti. In particolare, i gruppi e i membri inferiori vengono aggiornati prima dei gruppi superiori. E poi, i gruppi superiori aggiornati trasmettono i parametri ai loro membri del gruppo per completare un ciclo di aggiornamento. Al livello più basso, ciascun agente di apprendimento  $n$  può effettivamente eseguire il processo di addestramento locale per adattare i propri dati privati al problema di apprendimento personalizzato utilizzando gli ultimi modelli gerarchici generalizzati come segue.

PLP al livello 0

$$w_n^{(0)} = \arg_{w \in W} \min L_n^{(0)}(w | D_n^{(0)}) + \frac{\mu}{2} \|w - w_n^{(1)}\|^2 \quad (7)$$



Dove  $L_n^{(0)}$  è la funzione di perdita di apprendimento personalizzata per il compito di apprendimento dato il suo set di dati personalizzato  $D_n^{(0)}$ ,  $N_{(n,g)}^{(k)}$  è il numero di agenti di apprendimento del gruppo di livello- $k$  a cui appartiene l'agente  $n$ . Risolvendo il problema PLP, l'agente di apprendimento può aggiornare il proprio modello personalizzato  $w_n^{(0)}$  appartenente al set di modelli di deep learning parametrizzato a  $W$ . In questo livello personalizzato, il numero di membri del gruppo è 1.

- Algoritmo di apprendimento democratizzato.

Ispirati dagli algoritmi FedAvg e FedProx, adottiamo la suddetta analisi ricorsiva e il meccanismo di clustering gerarchico per sviluppare un nuovo algoritmo di apprendimento democratizzato, chiamato Dem-Learn. I dettagli sono presentati di seguito. Ogni agente  $n$  usa il modello del gruppo superiore al livello 1 (cioè,  $w_{(n,t)}^{(1)}$ ) come modello di apprendimento iniziale. Successivamente, l'agente risolve iterativamente il problema PLP nell'equazione (8) in base al metodo del gradiente.



Il modello client aggiornato verrà inviato al server centrale per eseguire il clustering gerarchico e l'aggiornamento dal livello 1 generalizzato al livello  $K$ . Dopo ogni giro globale  $\tau$ , la struttura gerarchica viene ricostruita in base ai cambiamenti nel modello di apprendimento personalizzato degli agenti. I modelli di apprendimento generalizzato dei gruppi vengono aggiornati, rispettivamente, in modalità bottom-top e top-bottom, seguendo (9) e (10) come approssimazione della soluzione in forma chiusa (6). Ciò consente ai sottogruppi di livello inferiore di contribuire con le proprie conoscenze all'aggiornamento del modello di gruppo. In cambio, ricevono (e incorporano) la migliore conoscenza generalizzata dai gruppi superiori che accresce la capacità di generalizzazione dei loro modelli di apprendimento locali. Inoltre, introduciamo un trucco di amplificazione nell'aggiornamento dal basso verso l'alto per i primi cinque round per accelerare la fase iniziale del processo di apprendimento. Di conseguenza, l'aggiornamento dai membri del gruppo [cioè,  $w_{t+1}^{(k)}$  in (9)] viene moltiplicato con una costante di 1,15 nei primi cinque round.

**Data:** this text

**Result:** how to write algorithm with L<sup>A</sup>T<sub>E</sub>X2e

```

1 initialization;
2 while not at end of this document do
3   | read current;
4   | if understand then
5   |   | go to next section;
6   |   | current section becomes this one;
7   | else
8   |   | go back to the beginning of current section;
```

**Algorithm 1:** How to write algorithms

```

1  $n \geq 0$   $y = x^n$   $y \Leftarrow 1$   $X \Leftarrow x$   $N \Leftarrow n$  while  $N \neq 0$  do
2  $N$  is even  $X \Leftarrow X \times X$   $N \Leftarrow \frac{N}{2}$  This is a comment  $N$  is odd
    $y \Leftarrow y \times X$   $N \Leftarrow N - 1$ 
```

# Capitolo 4

## Conclusioni

dfghdghdgh ...

## Bibliografia