Grokking: transizione di fase e generalizzazione nel Deep Learning

Autore: Francesco Zinghinì | **Data:** 13 Febbraio 2026

Immaginate di osservare uno studente che si prepara per un esame complesso. Per settimane, sembra non capire nulla: ripete a memoria le risposte senza logica, sbaglia ogni volta che la domanda viene posta in modo leggermente diverso e i suoi risultati nei test di prova sono disastrosi. Poi, improvvisamente, in un singolo pomeriggio, tutto cambia. Non solo smette di sbagliare, ma dimostra di aver intuito i principi fondamentali della materia, riuscendo a risolvere problemi mai visti prima. In ambito umano lo chiamiamo “momento eureka”. Nel mondo dell’**Intelligenza Artificiale**, e più specificamente nel **deep learning**, questo fenomeno esiste, è documentato e porta il nome curioso di **Grokking**.

Per anni, i ricercatori hanno creduto che l’apprendimento delle macchine fosse un processo graduale e lineare: più dati forniamo, più l’errore scende. Tuttavia, recenti studi sulle dinamiche di addestramento degli **LLM** (Large Language Models) e delle reti neurali hanno svelato che l’apprendimento può avvenire attraverso transizioni di fase improvvise. Il **Grokking** rappresenta uno dei misteri più affascinanti e controintuitivi del **progresso tecnologico** attuale: è il punto di svolta in cui un modello smette di “barare” memorizzando i dati e inizia realmente a generalizzare le regole sottostanti.

Memorizzazione vs Generalizzazione: Il dilemma dell’apprendimento

Per comprendere la stranezza del Grokking, dobbiamo prima analizzare come impara tradizionalmente un algoritmo di **machine learning**. L’obiettivo di qualsiasi rete neurale è minimizzare l’errore tra la sua previsione e il risultato reale. Durante l’addestramento, l’AI vede milioni di esempi. Inizialmente, la rete cerca di adattare i suoi parametri (i “pesi”) per far combaciare gli input con gli output.

Esistono due modi per farlo:

- **Memorizzazione (Overfitting):** L’AI impara a memoria ogni singolo esempio del dataset di training. È come uno studente che impara a memoria le risposte del libro di testo. Se gli chiedi esattamente quella domanda, risponde bene. Se cambi una virgola, fallisce.
- **Generalizzazione:** L’AI comprende la regola logica che governa i dati. Non ha bisogno di ricordare l’esempio specifico, perché possiede l’algoritmo mentale per derivare la soluzione.

Fino a poco tempo fa, si pensava che se una rete neurale veniva addestrata troppo a lungo sui dati di training (overfitting), le sue prestazioni sui dati nuovi (test set) sarebbero peggiorate drasticamente. Il Grokking ha ribaltato questa convinzione: in certi casi, se si continua a spingere l’addestramento *molto oltre* il punto in cui l’AI sembra aver memorizzato tutto, accade il miracolo. La curva di errore sui dati nuovi crolla improvvisamente verso lo zero. L’AI ha “capito”.

Copyright © 2026 TuttoSemplice.com - Tutti i diritti riservati

La fisica delle reti neurali: Transizioni di fase

Perché accade questo? Gli scienziati oggi, in questo 2026 sempre più dominato dall'**automazione** cognitiva, utilizzano metafore prese in prestito dalla fisica per spiegarlo. Il passaggio dalla memorizzazione alla generalizzazione è simile a una **transizione di fase**, come l'acqua che si trasforma in ghiaccio. L'acqua si raffredda gradualmente rimanendo liquida, finché, raggiunta una soglia critica, la sua struttura molecolare si riorganizza completamente in cristalli solidi.

All'interno di una **architettura neurale**, accade qualcosa di simile. Durante la lunga fase di "plateau" (dove sembra non succedere nulla di buono), l'algoritmo sta segretamente riorganizzando i suoi circuiti interni. La memorizzazione è una soluzione "facile" ma inefficiente: richiede pesi grandi e complessi per codificare ogni singola eccezione. La generalizzazione, al contrario, è una soluzione "elegante" e semplice, ma difficile da trovare nello spazio matematico delle possibilità.

Sotto la pressione costante dell'ottimizzazione (spesso guidata da una tecnica chiamata "weight decay", che punisce la complessità inutile), la rete neurale abbandona improvvisamente la strategia faticosa della memoria per scivolare nella valle stabile della comprensione logica. È qui che modelli come **ChatGPT** o i suoi successori evoluti trovano la loro vera potenza.

Il paradosso dei Benchmark e la "Doppia Discesa"

Il fenomeno del Grokking è strettamente legato a un altro concetto che ha scosso le fondamenta della teoria dell'apprendimento: la "Double Descent" (Doppia Discesa). Tradizionalmente, i grafici di performance mostravano una forma a "U": l'errore scende e poi risale se ci si allena troppo. Nel regime del

Copyright © 2026 TuttoSemplice.com - Tutti i diritti riservati

deep learning moderno, vediamo invece che l'errore scende, risale leggermente, e poi scende di nuovo, ancora più in basso, se si ha la pazienza di aspettare.

Questo ha implicazioni enormi per i **benchmark** utilizzati per valutare le AI. Spesso, un modello viene scartato perché le sue prestazioni iniziali sono mediocri. La scoperta del Grokking ci suggerisce che forse non stavamo guardando abbastanza a lungo. Alcuni **algoritmi** che sembravano fallimentari potrebbero essere stati semplicemente interrotti un attimo prima della loro transizione di fase verso l'intelligenza.

Immaginate quante architetture promettenti sono state gettate nel cestino negli ultimi dieci anni semplicemente perché i ricercatori hanno spento i computer troppo presto. Oggi, la consapevolezza di questo fenomeno sta cambiando il modo in cui progettiamo e addestriamo i modelli, spingendoci verso tempi di calcolo più lunghi ma con la promessa di reti molto più robuste e capaci di ragionamento astratto.

Cosa significa per il futuro dell'AI?

Capire il Grokking non è solo una curiosità accademica; è la chiave per rendere l'**Intelligenza Artificiale** più efficiente e sicura. Se riusciamo a prevedere quando avverrà il Grokking, o meglio ancora, a indurlo artificialmente prima del tempo, possiamo ridurre drasticamente l'energia e i dati necessari per addestrare i futuri LLM.

Inoltre, questo fenomeno getta luce sulla "Black Box" (la scatola nera). Quando una rete "grokka", i suoi circuiti interni tendono a diventare più ordinati e interpretabili. Questo apre la strada alla *Mechanistic Interpretability*, la scienza

Copyright © 2026 TuttoSemplice.com - Tutti i diritti riservati

che cerca di leggere nel pensiero delle macchine. Invece di avere un groviglio incomprensibile di numeri, potremmo arrivare a vedere chiaramente i “circuiti” logici che l’AI ha costruito per risolvere un problema, garantendo che il **progresso tecnologico** rimanga sotto il nostro controllo e comprensione.

Conclusioni

Il fenomeno del Grokking ci ricorda che l’intelligenza, sia essa biologica o artificiale, non è sempre un processo di accumulo lineare. A volte, è un salto nel buio che porta a una nuova forma di chiarezza. Mentre continuiamo a sviluppare sistemi sempre più complessi, da assistenti virtuali avanzati a sistemi di guida autonoma, riconoscere e sfruttare questi momenti di “illuminazione algoritmica” sarà fondamentale.

Non stiamo solo costruendo macchine che calcolano più velocemente; stiamo costruendo architetture capaci di distillare l’ordine dal caos, passando dalla bruta memorizzazione alla raffinata comprensione. E forse, studiando come le macchine imparano all’improvviso, potremmo scoprire qualcosa di nuovo anche su come avvengono le nostre intuizioni umane, chiudendo il cerchio tra creatore e creazione.

Domande frequenti

Che cos’è il fenomeno del Grokking nell’Intelligenza Artificiale?

Il Grokking è un comportamento controintuitivo osservato nel deep learning in cui una rete neurale, dopo un lungo periodo di stagnazione o apparente memorizzazione meccanica dei dati, subisce un’improvvisa transizione di fase. In questo momento preciso, le prestazioni del modello migliorano drasticamente sui nuovi dati, poiché l’algoritmo smette di basarsi sulla

Copyright © 2026 TuttoSemplice.com - Tutti i diritti riservati

semplice memorizzazione e inizia a generalizzare, comprendendo le regole logiche sottostanti al problema.

Qual è la differenza tra memorizzazione e generalizzazione nel machine learning?

La memorizzazione, spesso associata all'overfitting, si verifica quando l'AI impara a memoria ogni singolo esempio del training set ma fallisce se i dati vengono leggermente modificati. La generalizzazione, invece, avviene quando la rete assimila i principi fondamentali e l'algoritmo logico che governano i dati, permettendole di risolvere correttamente problemi mai visti prima. Il Grokking rappresenta proprio il salto qualitativo dalla prima alla seconda modalità.

Perché il Grokking viene paragonato a una transizione di fase della fisica?

Gli ricercatori utilizzano la metafora della transizione di fase, simile al passaggio dell'acqua allo stato solido, per spiegare la riorganizzazione interna della rete neurale. Durante l'addestramento prolungato, l'algoritmo abbandona improvvisamente le configurazioni complesse e inefficienti della memorizzazione per assestarsi su circuiti interni più semplici ed eleganti. Questo cambiamento strutturale permette di raggiungere una comprensione stabile e profonda, riducendo l'errore quasi a zero.

Cosa indica la teoria della Double Descent in relazione al Grokking?

La Double Descent, o Doppia Discesa, descrive un andamento dell'errore che contraddice le teorie classiche: l'errore scende, risale temporaneamente peggiorando le prestazioni, e poi crolla nuovamente se si ha la pazienza di continuare l'addestramento. Il Grokking spiega questa seconda discesa finale, suggerendo che molti modelli scartati in passato perché sembravano fallimentari erano semplicemente stati interrotti un attimo prima di

Copyright © 2026 TuttoSemplice.com - Tutti i diritti riservati

raggiungere la loro vera capacità di ragionamento astratto.

In che modo il Grokking aiuta a interpretare il funzionamento delle reti neurali?

Il fenomeno del Grokking ha implicazioni dirette sulla Mechanistic Interpretability, la scienza che tenta di decifrare la cosiddetta scatola nera dell'AI. Quando una rete effettua il grokking, i suoi circuiti interni tendono a diventare più ordinati e meno caotici rispetto alla fase di memorizzazione. Questo permette ai ricercatori di identificare con maggiore chiarezza i percorsi logici sviluppati dalla macchina, rendendo l'intelligenza artificiale più trasparente, sicura e controllabile.

Copyright © 2026 TuttoSemplice.com - Tutti i diritti riservati