

MICHELE LAURELLI

LE LEGGI FONDAMENTALI DELL'INTELLIGENZA ARTIFICIALE

VERSO UNA NUOVA SCIENZA

tecniche nuove

Sommario

Prefazione *(di Giovanni Landi)* ..... **VII**

Introduzione ..... **IX**
La macchina di Anticitera IX

Guida alla lettura ..... **XV**

LEZIONE I – **La Legge della rappresentazione** **1**
I. L'intuizione 1
II. La fondazione 5
Il vettore: l'atomo della rappresentazione 5
Lo spazio: dove la conoscenza abita 7
La matrice: trasformare le rappresentazioni 7
Il tensore: la rappresentazione delle strutture 9
Dimensionalità: il vuoto ad alte dimensioni 10
Lo spazio latente: la rappresentazione nascosta 11
L'embedding: dare coordinate al significato 13
III. L'incarnazione 13
AlexNet e le CNN: la rivoluzione delle immagini 14
Word2Vec: la rivoluzione silenziosa 16
I Transformer: la rappresentazione contestuale 20
IV. I limiti 21
Il primo limite: ogni scelta esclude qualcosa 22

Il secondo limite: le rappresentazioni ereditano i pregiudizi dei dati 23
Il terzo limite: non esiste la mappa universale 25
V. La mappa 26

LEZIONE II – La Legge dell'incertezza 29
I. L'intuizione 29
II. La fondazione 31
La probabilità come linguaggio 31
Bayes: imparare dall'evidenza 34
Maximum likelihood: la domanda capovolta 36
Il compromesso bias-varianza 39
III. L'incarnazione 42
Naive Bayes: l'incertezza al lavoro 42
Le reti bayesiane: ragionare con le cause 44
I modelli latenti e l'algoritmo EM 48
Le VAE: l'incertezza nel deep learning 51
Generare come atto probabilistico 52
IV. I limiti 55
L'illusione della certezza 55
Il sovradattamento come malattia della sicurezza 56
Il no free lunch: non esiste il modello universale 60
V. La bussola 62

LEZIONE III – La Legge dell'apprendimento 65
I. L'intuizione 65
II. La fondazione 68
L'entropia: misurare l'imprevedibilità 68
La divergenza KL: misurare l'ignoranza del modello 71
L'informazione mutua: cosa i dati dicono l'uno sull'altro 73
Minimum Description Length (MDL): il rasoio formalizzato 76
III. L'incarnazione 80
L'overfitting come fallimento della compressione 80
La cross-entropy: comprimere come obiettivo di addestramento 81
La predizione del token successivo come compressione 84
IV. I limiti 88
Il confine tra struttura e rumore 88

Cosa non si può apprendere 90
La compressione perfetta come miraggio 92
V. La distillazione 97

LEZIONE IV – **La Legge dell'ottimizzazione 99**
I. L'intuizione 99
II. La fondazione 102
La discesa del gradiente 102
Il learning rate: l'arte del passo giusto 105
Il momentum: la memoria della discesa 107
La backpropagation: distribuire la colpa 109
SGD e il rumore come alleato 112
Adam: l'ottimizzatore adattivo 115
III. L'incarnazione 118
Addestrare una rete passo per passo 118
Il fine-tuning: scendere più in profondità 122
RLHF: quando la funzione obiettivo diventa umana 124
IV. I limiti 126
I punti sella e i minimi locali 126
Il gradiente che svanisce 131
La Legge di Goodhart: quando l'obiettivo tradisce 135
I plateau: il silenzio del paesaggio 137
V. La discesa 138

LEZIONE V – **La Legge dei limiti 141**
I. L'intuizione 141
II. La fondazione 144
I confini del calcolabile 144
Il Teorema di Rice e i limiti strutturali 148
La complessità: calcolabile ma impraticabile 150
Il No Free Lunch rivisitato 155
I limiti della generalizzazione 156
III. L'incarnazione 160
Quando i modelli fondazionali incontrano i propri confini 160
Il distribution shift: un limite statistico ed epistemico 161
I limiti del ragionamento: un limite computazionale e architetturale 165
Il test della scala: l'emergenza e i suoi dubbi 169

IV. I limiti 173
Una tassonomia dei limiti, e cosa farne 173
V. La Quinta Legge 179

Epilogo 181
La macchina ritrovata 181

Ringraziamenti 187

Bibliografia 189

Prefazione

di Giovanni Landi

Heidegger diceva che la Filosofia è come un enorme impero che con la nascita di ogni nuova scienza va lentamente disgregandosi. Appena una disciplina (la Fisica, la Biologia) acquista coscienza di sé, ed è capace di delimitare con precisione i propri confini, dichiara la propria indipendenza e da provincia dell'impero diventa uno Stato autonomo con leggi e costituzione proprie.

Questa opera, almeno nelle intenzioni dell'autore, è la dichiarazione di indipendenza dell'Intelligenza Artificiale in quanto scienza: indipendenza non solo dalla Filosofia madre di tutto il sapere, ma anche dalle province circostanti, che per troppo tempo l'hanno considerata un'appendice incapace di reggersi sulle proprie gambe, in primis l'Informatica che per la prima volta vede messo in discussione il proprio monopolio sulla "digitalizzazione" del mondo.

Come tutte le dichiarazioni di indipendenza, anche questa è un sapiente *mélange* di certezza grandiloquente e di trasparente umiltà. Ed è giusto che sia così.

Ci sono le Leggi, i Principi, le Definizioni, i Postulati, tutti ben ordinati e legati tra loro in intima coordinazione e armonia (stilistica e concettuale); ogni capitolo descrive con chiarezza gli elementi e le tecniche di lavoro dell'IA, e chiunque ne sia genuinamente interessato farà bene a prepararsi a notti di studio come ai tempi dell'Università. Questo che avete sotto gli occhi è solo l'inizio del lavoro che l'autore promette. L'IA è una scienza

giovane, piena di vigore, che ha tanto da dire e vuole farsi sentire; un'opera come questa riflette esattamente questo stato d'animo, e il plauso per chi ha concepito e realizzato un tale lavoro è giustamente dovuto.
In questa grandiosa costruzione annunciata come da una cattedra ormai irraggiungibile da ogni umana obiezione, però, fin dalle prime pagine affiorano parole come "limite", "incertezza", "perdita necessaria", "miraggio". Questo in un certo senso è rassicurante, la troppa baldanza o certezza di sé non è mai un buon segno, ed è un ulteriore merito dell'autore quello di essere riuscito a mantenere sobrietà di linguaggio e di concetto in questo nostro periodo di annunci sull'IA uno più roboante dell'altro (e usiamo l'aggettivo "roboante" con moderazione).
Un lavoro senza pecche, dunque, che si può leggere e rileggere con intensità diverse a seconda del personale impegno professionale, culturale o altro di ognuno di noi. Ora che credo di averlo situato sia concettualmente che storicamente non penso di poter aggiungere nulla se non l'augurio di buona lettura.

P.S. Io e Michele condividiamo la passione per i romanzi di Isaac Asimov. Approfitto di questa comune debolezza per un'ultima osservazione o avvertimento che dir si voglia: l'impero disgregantesi di cui parla Martin Heidegger ha (come l'Impero Galattico del ciclo della Fondazione) un unico inevitabile destino, e cioè la nascita di un Secondo Impero, non la frammentazione continua in entità sempre più piccole. La Filosofia non potrà mai essere eliminata dalla nascita di nuove scienze, poiché le scienze riguardano le conoscenze ma la Filosofia riguarda il Sapere. Sono sicuro che questa mia ultima e un po' criptica affermazione condannerà me e Michele ad ancora lunghe conversazioni negli anni a venire.

LEZIONE I

La Legge della rappresentazione

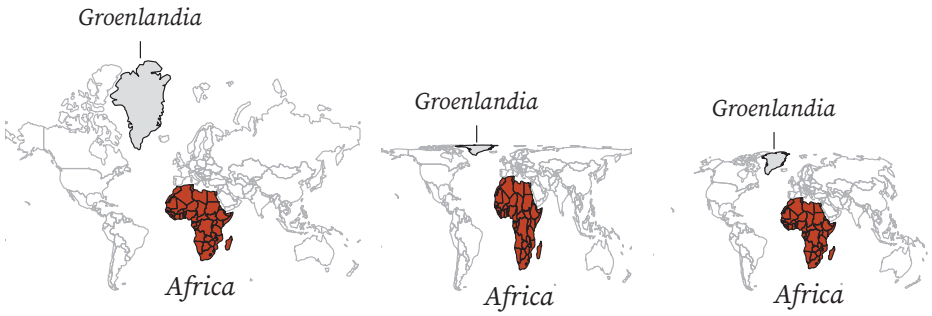
PRIMA LEGGE

Ogni forma di intelligenza artificiale opera allo stesso modo: trasformando le rappresentazioni del mondo in uno spazio matematico.

La scelta della rappresentazione determina il limite di ciò che il sistema può apprendere e produrre. Nessun algoritmo, per quanto sofisticato, può compensare una rappresentazione inadeguata.

I. L'intuizione

Nel 1569, Gerardus Mercatore pubblicò una mappa del mondo che avrebbe cambiato la navigazione per i secoli a venire. Non perché avesse scoperto nuove terre (le terre erano le stesse che già si conoscevano), ma perché aveva cambiato il modo di rappresentarle. La proiezione di Mercatore trasformava la superficie curva della Terra in un piano rettangolare, con una proprietà specifica: preservava gli angoli. Dunque, una linea retta segnata sulla mappa corrispondeva a una rotta percorribile a bussola costante. Per la prima volta, un marinaio poteva tracciare una linea tra due punti e seguirla. Il mare e il pianeta non erano cambiati; era cambiata la loro rappresentazione.



Mercatore
(angoli)

Gall-Peters
(aree)
aree vere: 13,7 a 1

Winkel-Tripel
(compromesso)

Figura 1.1 - La stessa Terra in tre proiezioni. (*Fonte*: Natural Earth, Admin 0 – Countries 110m; rapporto d'area su geometrie con area sferica, R = 6 371 008,8 m.)

Eppure, la mappa di Mercatore mente. La Groenlandia appare grande quanto l'Africa, quando in realtà è quattordici volte più piccola. Le regioni polari si gonfiano a dismisura. Infatti, ogni proiezione cartografica (non solo quella di Mercatore) distorce qualcosa: si possono preservare gli angoli o le aree, le distanze o le forme, ma mai tutto contemporaneamente. Non esiste una mappa perfetta del mondo, esiste solo la mappa giusta per la domanda che si sta ponendo.

Questa è, in forma di metafora, la Prima Legge dell'Intelligenza Artificiale. Ogni sistema di AI prende un frammento del mondo (un'immagine, una frase, il comportamento di un mercato finanziario, la struttura di una proteina) e lo traduce in numeri. Questa traduzione è ciò che chiamiamo *rappresentazione*: ed è possibile pensarla come una mappa matematica della realtà (Marr, 1982; Goodfellow, Bengio e Courville, 2016). Allo stesso modo delle mappe cartografiche, la rappresentazione non è la realtà. È una proiezione, che preserva alcune proprietà e ne sacrifica altre.

E la scelta di cosa preservare e cosa sacrificare viene prima di tutto il resto.

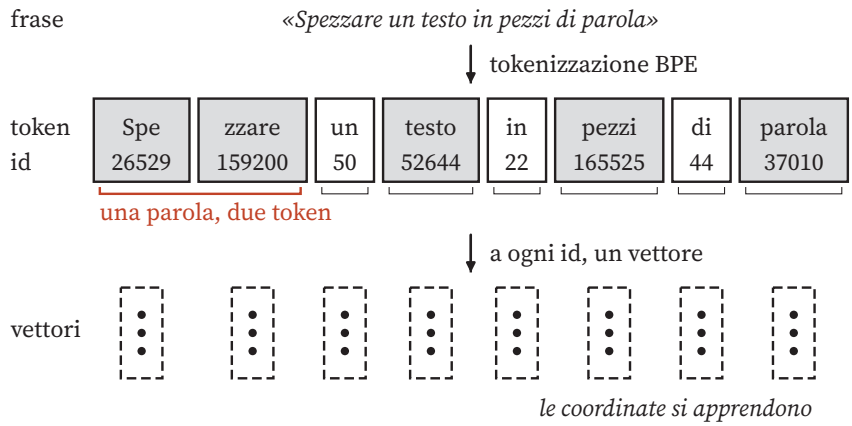


Figura 1.2 - "Spezzare un testo in pezzi di parola" tokenizzato con il BPE di XLM-RoBERTa. (*Fonte*: SentencePiece [BPE] di XLM-RoBERTa base, FacebookAI [Conneau *et al.*, 2019], licenza MIT.) La tokenizzazione a sotto-parole spezza anche l'italiano: "Spezzare" diventa due token. Le parentesi e i fondini alternati sotto i token mostrano le parole di partenza.

Pensiamo a qualcosa di concreto. Vogliamo insegnare a un computer a distinguere le foto dei gatti da quelle dei cani (Figura 1.3). Prima ancora di scegliere quale metodo di apprendimento usare, c'è una domanda a cui rispondere: *come descriviamo una foto attraverso il linguaggio dei numeri?*

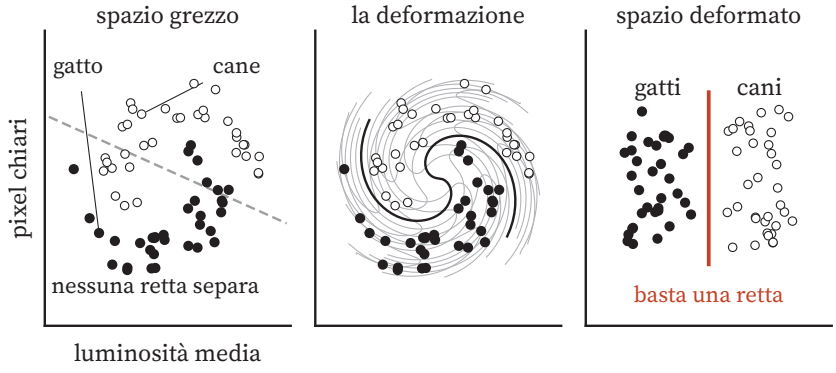


Figura 1.3 - Gli stessi punti in tre pannelli.

Un'opzione è prendere ogni foto e ridurla a due soli numeri: un numero segnerà quanto è luminosa in media, e un altro invece quanti punti chiari contiene. Un'altra ipotesi è quella di usare come riferimento tutti i milioni di pixel che costituiscono l'immagine, uno per uno. Si potrebbe anche provare

LEZIONE II

La Legge dell'incertezza

SECONDA LEGGE

Ogni sistema di AI opera in condizioni di informazione incompleta. L'incertezza non è un difetto da eliminare: è una proprietà fondamentale del mondo da modellare.

I. L'intuizione

Nella Lezione I abbiamo stabilito un principio: tutto comincia dalla mappa, e la qualità della mappa determina il limite superiore di ciò che un sistema può imparare. Ma c'è un fatto che quella lezione ha deliberatamente lasciato in ombra: il mondo che stiamo cercando di mappare non sta fermo, non è nitido e non si lascia descrivere completamente.

Un medico guarda una radiografia e deve decidere se quella macchia è un tumore o un'ombra. Non ha tutte le informazioni: l'immagine è bidimensionale, il corpo è tridimensionale, e la macchia potrebbe essere mille cose diverse (Figura 2.1). Eppure deve decidere. Oppure: un filtro antispam riceve un'email con oggetto "Congratulazioni, hai vinto!" e deve classificarla. È spam? Probabilmente sì. Ma "probabilmente" non è "certamente", e da qualche parte nel mondo c'è sicuramente un concorso vero che manda email con quell'oggetto. O ancora: un sistema di guida autonoma vede una

sagoma sul bordo della strada. È un pedone che sta per attraversare, o un cartello pubblicitario? Ha pochi millisecondi per decidere, e le conseguenze di una decisione sbagliata sono radicalmente diverse nei due casi.

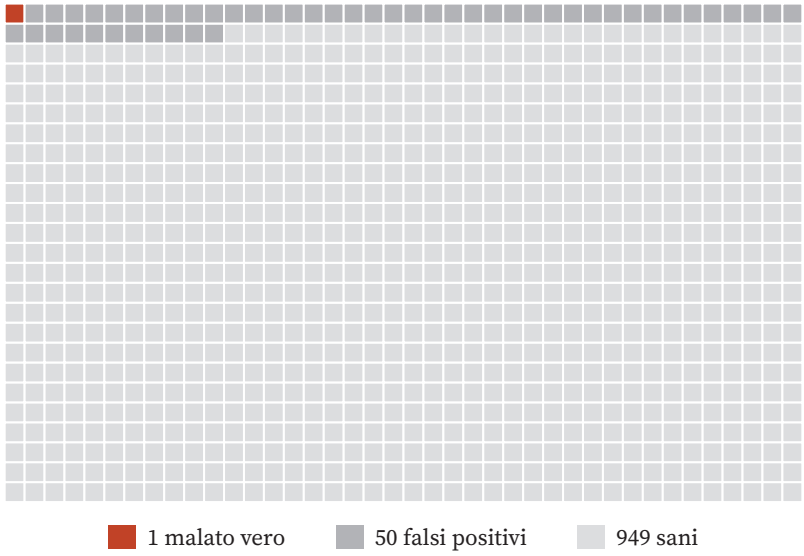


Figura 2.1 - Mille persone sottoposte a un test accurato al 95% per una malattia che colpisce una persona su mille.

In ognuna di queste situazioni, il problema non è la mancanza di un buon algoritmo. Il problema è che l'informazione disponibile non basta a eliminare il dubbio. Il mondo non collabora: è incompleto dove vorresti che fosse chiaro, e rumoroso dove vorresti che fosse nitido. Qualsiasi sistema che debba agire in un mondo così deve trovare il modo di prendere decisioni ragionevoli *nonostante* il dubbio.

Questa è la Seconda Legge. Non dice che l'incertezza sia un ostacolo da superare con più dati o più potenza di calcolo. Dice qualcosa di più radicale: l'incertezza è una proprietà del mondo, non un difetto del nostro sguardo su di esso. Un sistema che la ignora non è più efficiente; è più fragile. Un sistema che impara a conviverci, a quantificarla e a gestirla, è un sistema che può operare nel mondo così com'è, non nel mondo come vorremmo che fosse.

Nella Prima Legge, la domanda fondamentale era: *cosa stai dando da guardare al sistema?* Nella Seconda, la domanda è diversa e complementare: *quanto puoi fidarti di ciò che stai guardando?*

LEZIONE III

La Legge dell'apprendimento

TERZA LEGGE

Apprendere è comprimere. Un sistema apprende quando estrae da un insieme di osservazioni un modello più semplice delle osservazioni stesse. Un sistema che si limita a memorizzare non ha appreso.

I. L'intuizione

Alla fine della Lezione II ci siamo lasciati con un'osservazione: l'incertezza è legata a filo doppio con la compressione. È il momento di sciogliere quel nodo.

Partiamo da una domanda semplice. Qual è la differenza fra uno studente che ha capito la fisica newtoniana e uno che ha memorizzato tutte le risposte previste dall'esame? Durante l'esame i due sono indistinguibili, perché entrambi rispondono correttamente. La differenza emerge dopo, davanti a una domanda nuova, di un tipo che non figurava nel materiale: il primo estrae dal proprio repertorio le regole sottostanti, le applica al caso inedito e trova la risposta; il secondo resta smarrito, perché non possiede una formula già pronta. Il vero apprendimento ha avuto luogo là dove sono state interiorizzate le regole capaci di generare le risposte, non le risposte stesse.

Lo studente che ha capito possiede un modello compresso della realtà. Le tre leggi di Newton occupano poche righe, eppure permettono di prevedere il moto di qualsiasi corpo, da una mela alle sonde interplanetarie. Le singole risposte dell'esame sono potenzialmente migliaia. Le leggi che determinano tutte quelle risposte sono soltanto tre. Passare dalle migliaia di casi alle tre leggi è un atto di compressione, ed è esattamente ciò che intendiamo con "imparare".

L'idea ha radici profonde, e per coglierne la portata conviene tornare al laboratorio in cui è nata. Alla fine degli anni Quaranta, nei Bell Telephone Laboratories del New Jersey, lavorava un giovane matematico e ingegnere di nome Claude Shannon. I Bell Labs erano il braccio di ricerca di AT&T, e il problema che assillava i loro ingegneri era concretissimo: trasmettere voce e dati attraverso cavi e onde radio nel modo più affidabile ed economico possibile, sapendo che ogni canale è disturbato dal rumore e che ogni bit costa. Shannon, durante la guerra, si era occupato di crittografia militare, un altro contesto in cui la domanda "quanta informazione c'è davvero in un messaggio, e quanta se ne può togliere senza perderne il senso?" diventa drammaticamente concreta. Da quell'incrocio fra telefonia, crittografia e matematica pura nacque, nel 1948, A Mathematical Theory of Communication, che avrebbe ridefinito non solo le telecomunicazioni ma il modo stesso in cui pensiamo alla conoscenza (Figura 3.1).

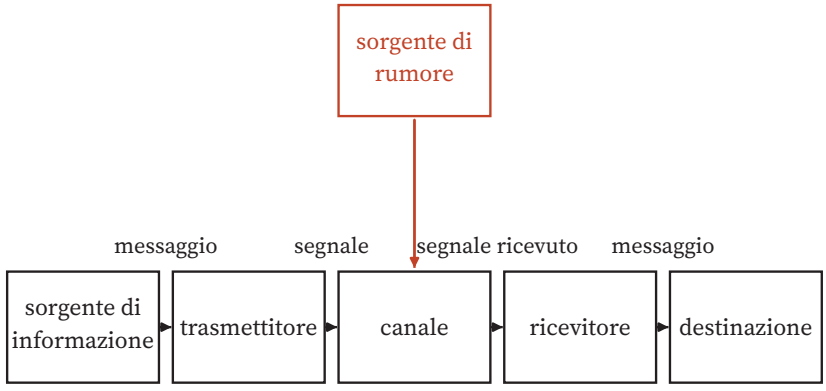


Figura 3.1 - Lo schema della comunicazione di Shannon (1948) con la nomenclatura originale.

La domanda di partenza sembrava squisitamente ingegneristica: quanti bit servono, al minimo, per trasmettere un messaggio attraverso un canale, e

LEZIONE IV

La Legge dell'ottimizzazione

QUARTA LEGGE

Ogni processo di apprendimento artificiale è una ricerca di parametri nello spazio. La geometria della superficie di errore determina ciò che il sistema può imparare e come vi convergerà.

I. L'intuizione

Alla fine della Lezione III abbiamo lasciato un'immagine: un alambicco che distilla l'essenza dai dati grezzi. Ma non abbiamo detto come avvenga la distillazione. Sappiamo che apprendere è comprimere; con quale meccanismo, però? Come fa una rete neurale con milioni di parametri a trovare la configurazione giusta?

La risposta è una metafora che vale la pena prendere quasi alla lettera: proviamo a immaginare il nostro modello che scende da una montagna.

Immaginiamo un paesaggio sterminato, pieno di colline, valli, altipiani e crepacci (Figura 4.1). Ogni punto del paesaggio corrisponde a una configurazione dei parametri del modello, e l'altitudine di quel punto è l'errore: ovvero quanto il modello sbaglia con quei parametri. Le cime sono configurazioni terribili (errore alto). Le valli profonde sono configurazioni eccellenti

(errore basso). Addestrare il modello significa partire da un punto qualsiasi del paesaggio e cercare di raggiungere la valle più profonda.

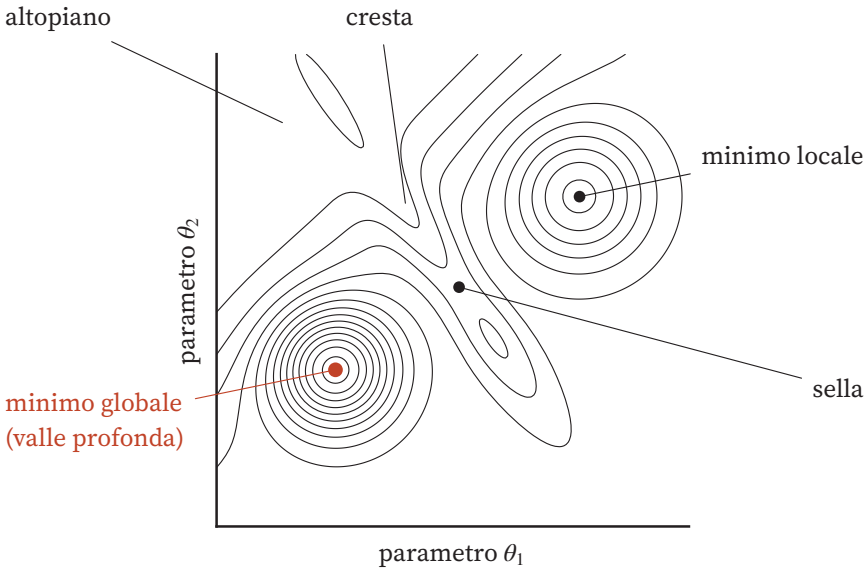


Figura 4.1 - Il paesaggio dell'errore usato in tutta la lezione, visto dall'alto come mappa di curve di livello.

Il problema è che il paesaggio è immenso. Un modello con un milione di parametri vive in uno spazio a un milione di dimensioni. Non possiamo "guardare" il paesaggio dall'alto per trovare la valle migliore. Non possiamo esplorarlo sistematicamente: i punti possibili sono più numerosi degli atomi nell'universo. L'unica cosa che possiamo fare, in ogni istante, procedendo, è guardare il terreno *sotto i nostri piedi* e capire in quale direzione scende.

È come trovarsi sulla cima di una montagna avvolti nella nebbia, senza mappa (Figura 4.2). Non vediamo la valle, e non sappiamo nemmeno se la valle più vicina è la più profonda. Possiamo solo sentire la pendenza del terreno sotto le suole, e fare un passo nella direzione che scende. Poi un altro, e poi un altro ancora, sperando che, passo dopo passo, la discesa ci porti in un buon posto. Il fatto che, contro ogni ragionevole aspettativa, ci finisca quasi sempre, è uno dei piccoli miracoli del nostro mestiere.

LEZIONE V

La Legge dei limiti

QUINTA LEGGE

I limiti computazionali, statistici ed epistemici dell'AI non sono difetti in attesa di correzione. Sono proprietà strutturali. Vanno studiati, non rimossi.

I. L'intuizione

Le prime quattro lezioni hanno costruito una sorta di edificio concettuale, le cui pareti reggono e i cui pilastri si appoggiano gli uni agli altri secondo una logica precisa. Si è cominciato con una **mappa**, la rappresentazione, vale a dire il modo in cui un sistema di intelligenza artificiale traduce in numeri un frammento del mondo che vuole comprendere; poi si è introdotta la **nebbia** che inevitabilmente avvolge ogni mappa, l'incertezza, che è il linguaggio in cui si esprime la conoscenza parziale del cartografo; poi è venuto il *distillato*, ovvero la compressione, l'operazione attraverso cui un modello estrae dai dati la struttura essenziale e scarta il rumore accidentale; infine si è descritto il *processo* attraverso cui questa estrazione avviene concretamente, l'ottimizzazione, che è la discesa lungo il paesaggio dell'errore alla ricerca dei suoi punti più bassi. Insieme, questi quattro elementi rispondono in modo articolato alla domanda "come funziona un sistema di intelligenza

Michele Laurelli

LE LEGGI FONDAMENTALI DELL'INTELLIGENZA ARTIFICIALE

Verso una nuova scienza

Nel 1900 un relitto romano restituisce ottantadue frammenti di bronzo, scambiati per zavorra. Cinquant'anni per non archiviarli come inutili, altri cinquanta per capire cosa fossero: un calcolatore astronomico del secondo secolo avanti Cristo. Il meccanismo funzionava da sempre. Mancava la cornice per leggerlo.

L'intelligenza artificiale, oggi, è quello stesso grumo di bronzo. Funziona magnificamente e se ne parla ovunque, ma senza un vocabolario condiviso per descriverne davvero natura e limiti.

Questo libro apre i Fondamenti di Intelligenza Artificiale enunciando le sue cinque leggi su cui si regge l'intero impianto: come rappresenta il mondo, convive con l'incertezza, apprende, si ottimizza, incontra i propri confini. Non un manuale di programmazione, non un saggio sull'impatto sociale: il primo tentativo sistematico di costruire l'AI come disciplina scientifica autonoma.



Michele Laurelli, nato nel 1991, vive a Savona e lavora a Milano. Si occupa di intelligenza artificiale da due prospettive che raramente coincidono: quella di chi la costruisce e quella di chi la studia. Come imprenditore ha fondato Algoretico, società di riferimento in Italia per l'intelligenza artificiale privata, e Scriba AI, dove dirige lo sviluppo AI; come docente a contratto insegna Intelligenza Artificiale presso l'Università Vita-Salute San Raffaele e altri atenei. Autore di quattro libri precedenti e di articoli scientifici.

tecniche nuove
www.tecnichenuove.com

ISBN 978-88-481-8203-4

€ 24,90