1

Capitolo J52
intelligenza artificiale e matematica

Contenuti delle sezioni

23 pagine

J52 a. matematica utilizzata nei sistemi AI

J52 a.01 L'area AI ha fatto costantemente riferimento alla matematica, così come l'industria elettronica ha continuamente dovuto fare riferimento alla fisica atomica, alla scienza dei materiali e alla chimica molecolare.

La cosiddetta Intelligenza artificiale simbolica, come le scienze fisiche, ingegneristiche, economiche e naturali, si è servita sistematicamente dell'analisi infinitesimale, dell'algebra lineare, della logica booleana, del calcolo logico del primo ordine, della statistica e del calcolo delle probabilità.

Inoltre la definizione degli algoritmi che sono stati necessari in tutte le fasi dello sviluppo dei sistemi applicativi si è servita delle teorie combinatorie, in particolare della teoria dei grafi e degli ipergrafi, della teoria della complessità computazionale, della teoria della ottimizzazione discreta e della teoria dei codici.

In tutti questi studi si possono riconoscere fasi nelle quali l'area AI attinge semplicemente alla matematica e fasi, in genere più avanzate, nelle quali nuove esigenze portano a sviluppi e risultati innovativi nei settori della matematica che vengono coinvolti.

J52 a.02 Con gli sviluppi del machine learning (ML) sono emerse nuove esigenze che hanno condotto a innovazioni più radicali, in buona parte argomenti di indagini tuttora in pieno sviluppo e dalle quali ci si possono aspettare interessanti aperture di orizzonti.

Un elemento che sembra una novità sostanziale deriva dal fatto che un sistema di machine learning, di apprendimento automatico, dipende strettamente dall'insieme D_1 dei dati con i quali è stato addestrato e dall'insieme dei dati D_2 che gli vengono affidati pe essere esaminati adattandoli a un modello parametrico assunto per spiegare processi che hanno generano gli elementi di D_1.

Un sistema ML per riuscire ad apprendere, ossia per riuscire a migliorare i parametri del modello che adotta, deve proporre meccanismi in grado di fornire generalizzazioni dei dati D_1 che si accordino con i dati D_2.

Le reti neurali forniscono funzioni I/O dipendenti dai dati molto più pesantemente di quanto è accaduto alle funzioni di interesse applicativo della tradizione matematica e scientifica.
Attualmente si pone il problema di capire a fondo perché e in quali tipi di circostanze abbiano successo le suddette funzioni I/O delle reti dotate di molti strati (anche milioni) e di moltissimi nodi (molti miliardi).

Resta quindi aperta l'esigenza di comprendere i principi secondo i quali vanno costruiti i sistemi di ML, esigenza concettuale le cui conseguenze evidentemente possono avere importanti ricadute pratiche.

J52 a.03 Presentiamo ora una rapida rassegna degli argomenti matematici che si incontrano negli sviluppi dell'Intelligenza artificiale e segnaliamo che per raccoglierli ci siamo riferiti ad alcuni dei più importanti testi del settore, i due volumi di Russell e Norvig intitolato "Intelligenza artificiale. Un approccio moderno" nella sua IV edizione e nel testo "Mathematics for machine Learning" di Deisenroth, Faisal e Ong.

Diamo per scontata la utilizzazione dei primi elementi della Matematica e dell'analisi infinitesimale come numeri interi, razionali e reali, insiemi, relazioni, funzioni, successioni, serie, passaggio al limite, derivate e integrali, argomenti già nelle prime 15 sezioni.

Grande importanza è dedicata all'algebra lineare, lo studio degli spazi costituiti da vettori, entità che si possono sommare e moltiplicare per scalari, ovvero che possono essere combinate linearmente.

Nella AI, oltre ai vettori talora chiamati geometrici tridimensionali, a componenti reali, caratterizzabili con intensità e direzione, utili per esprimere varie grandezze fisiche, si utilizzano vettori s costituiti da sequenze numeriche con un definito numero n di componenti e funzioni a valori reali o complessi (in particolare polinomi). Questo modo di fare, derivato dalla cosiddetta analisi funzionale, apre la strada al trattamento di molte trasformazioni lineari di funzioni con strumenti che generalizzano le manovre, realizzabili molto concretamente, sopra le matrici, le strutture descrivibili come scacchiere contenenti numeri.

J52 a.04 Nelle elaborazioni che consentono di risolvere problemi riguardanti vettori e matrici giocano ruoli importanti le rappresentazioni dei vettori come combinazioni lineari di particolariinsiemi di vettori che costituiscono le cosiddette basi per gli spazi nei quali i vettori si collocano; la scelta di basi opportune serve anche a tenere sotto controllo le trasformazioni tra vettori espresse da matrici e accade che molti problemi si risolvono trovando trasformazioni inverse (ovvero matrici inverse).

Il calcolo vettoriale e il calcolo matriciale, all'interno dell'algebra lineare, costituiscono una strumentazione importante ed efficace per molti capitoli della matematica, per molteplici sviluppi di carattere geometrico e di loro numerosissime applicazioni nella statistica, nell'ingegneria, nell'economia e non solo; di conseguenza anche in molte attività che si avvalgono dell'Intelligenza artificiale.

Le idee sviluppate nell'algebra lineare sono alla base dei metodi per la soluzione dei sistemi di equazioni lineari, strumenti matematici che consentono di esprimere (talora in modo approssimato) il comportamento di molti fenomeni che riguardano le evoluzioni di corpi solidi, di masse fluide, di popolazioni da studiare statisticamente, di processi studiati nell'economia e nella sociologia.
Molti di questi processi sono decisamente complessi ma si riescono a studiare cominciando con approssimarli mediante trasformazioni lineari.

Risultano quindi di grande importanza la possibilità, garantita dall'algebra lineare, di affrontare con chiarezza lo studio di vettori e matrici e, la possibilità di manipolare vettori e matrici con grande efficienza mediante apparecchiature elettroniche sempre più spinte e che spesso sono progettati per implementare con elevata efficienza algoritmi estremamente accurati e ingegnosi.

Tra i sistemi elettronici capaci di raggiungere le maggiori prestazioni ricordiamo i supercomputers in grado di effettuare elaborazioni massicciamente parallele, i sistemi multicore e le recenti graphical processing units (GPU), le tensor processing units (TePU, i tensori sono entità che generalizzano le matrici), i modelli GPT prodotti da Cerebras e i modelli LLaMA prodotti da MetaAI in grado di controllare 65 miliardi di parametri.

J52 a.05 Per giungere a soluzioni stringenti di problemi numerici si devono trattare lunghezze ed angoli e queste grandezze vengono trattate sistematicamente nell'ambito della geometria analitica euclidea.

Le costruzioni della geometria analitica si estendono proficuamente ai calcoli sulle funzioni e alle collezioni di dati quantitativi che si possono considerare campionature di funzioni con le quali si intendono esprimere leggi fisiche o empiriche in senso lato (negli ambiti delle scienze naturali, dell'ingegneria, dell'economia, della sociologia, ...).

Evidentemente la geometria analitica è essenziale per la robotica e per la costruzione di automatismi semoventi, compresi quelli che si muovono in gruppi di membri interagenti e addirittura in sciami interconnessi, finalizzati al controllo della interconnettività di veicoli spaziali, in particolare per le costellazioni di satelliti.

Per molte delle relative problematiche servono procedimenti geometrici per la ricostruzione, a partire da più immagini bidimensionali, spesso riprese da sensori in movimento, di mappe tridimensionali delle porzioni di spazio che si intendono monitorare e sulle quali si deve agire.

Queste ricostruzioni nella frequente presenza di ostacoli per la visuale dei sensori devono necessariamente limitarsi a fornire mappe probabilistiche.

J52 a.06 Un ruolo molto importante nell'algebra lineare dal punto di vista delle necessità del machine learning viene svolto dalla decomposizione SVD, singular value decomposition, delle matrici, anche non quadrate, e quindi più generale della classica decomposizione di matrici quadrate mediante autovettori ortogonali.

Si dimostra che ogni matrice A di profilo $m \times n$ si può decomporre nel prodotto$U\Sigma V^\top$ con U di profilo $m \times m$, V di profilo $n \times n$ e Σ con entrate diverse da 0 solo sulla diagonale principale; va segnalato che i vettori costituenti la U e quelli costituenti la V in genere non sono ortogonali.

La decomposizione SVD trova applicazione in operazioni effettuate nella AI che vanno dal metodo dei minimi quadrati per il curve fitting di dati da osservazioni, alla soluzione dei sistemi lineari, alla capacità di approssimare matrici di un dato rango con matrici di rango inferiore.
Segnaliamo che la sostituzione di una matrice con la sua decomposizione SVD spesso consente di effettuare calcoli numerici più robusti nei confronti dei possibili errori di arrotondamento e che l'approssimazione di un matrice con matrici più ridottr apre la possibilità di efficaci applicazioni di machine learning come la riduzione della dimensionalità e la modellizzazione tipica per la compressione e il clustering dei dati.

J52 a.07 Ricordiamo anche i procedimenti di eliminazione da immagini di rumore dovuto a difetti nella ricezione, le operazioni di attenuazione degli annebbiamenti e la riparazione nelle immagini di particolari danneggiati.

Molti procedimenti per la precisazione di modelli, nella AI ma non solo, consistono nella ricerca di valori dei parametri che controllano le modalità con le quali i modelli stessi giustificano i dati.

In molti campi, in particolare nel machine learning, si rende necessario esaminare modelli dipendenti da un gran numero di parametri e in questi casi si tratta di individuare i parametri che massimizzano i valori di una funzione obiettivo dipendente da molte variabili.

Questi procedimenti possono essere classificati come metodi per la soluzione di problemi di ottimizzazione. A questo genere appartengono: la regressione lineare, elaborazione con la quale si cercano parametri che esprimono linearmente la verosimiglianza del modello; i codificatori delle reti neurali con i parametri (anche molti miliardi) consistenti nei pesi dei nodi e i bias degli strati con gli scopi della riduzione della dimensionalità dei parametri e della compressione dei dati; i modelli di misture gaussiane per la modellizzazione delle distribuzioni per i quali si tratta di ottimizzare i parametri di localizzazione e di dispersione delle distribuzioni componenti le misture stesse.

Questi procedimenti si concentrano sulle funzioni a valori reali di più variabili reali e si inquadrano nel calcolo vettoriale; essi inoltre presentano molti collegamenti con nozioni probabilistiche.

Gli strumenti matematici partono dalle nozioni di derivata, integrale e sviluppo in serie di Taylor delle funzioni-RtR e toccano le derivate parziali delle funzioni-RntR, e i gradienti e gli Jacobiani delle funzioni-RntRm, fino ai gradienti delle matrici.

In molte applicazioni di machine learning per la ricerca di punti di minimo di funzioni obiettivo in spazi di molte dimensioni si effettuano ricerche delle traiettorie di massima discesa del gradiente della funzione obiettivo e per evitare calcoli troppo onerosi si adottano metodi di retropropagazione.

J52 a.08 Più in generale si adotta la tecnica della differenziazione numerica, risultata decisamente preferibile alla differenziazione simbolica.

Tecniche anche più elaborate vengono sviluppate per le derivate parziali di ordine superiore, in particolare tecniche di linearizzazione per serie di Taylor multivariate.

Un agente retto dalla AI in genere deve operare in condizioni di incertezza dovute alla complessità degli ambienti nei quali agisce e/o motivate dal fatto che le sue osservazioni sono parziali.
Si deve quindi ricorrere spesso al calcolo delle probabilità, dove questa grandezza esprime l'incapacità dell'agente di giungere a decisioni definitive e riassume le sue credenze rispetto alle sue percezioni.

Serve per questo la teoria delle decisioni che tiene conto degli obiettivi che l'agente si pone e delle sue credenze per scegliere come migliore azione da intraprendere quella che massimizza la funzione di utilità che si è assunta.

J52 a.09 Per elaborare i valori di probabilità vanno definite le probabilità a priori o non condizionate e le probbilità a posteriori o condizionate; esse riguardano proposizioni semplici e proposizioni complesse che vanno analizzate con le regole derivate dagli assiomi del calcolo delle probabilità.

Se si riuscisse a trattare la distribuzione di probabilità congiunta completa si potrebbe rispondere a ogni interrogazione, ma spesso questa è troppo complessa. Per ridurre questa complessità occorre individuare sottoinsiemi indipendenti di variabili casuali che consentono di fattorizzare la probabilità congiunta.

Spesso di devono calcolare probabiità sconosciute partendo da probabilità condizionate note e per quasto si ricorre alla regola di Bayes che per due eventi a e b afferma:

$$\mathcal{P}(b|a) = \frac{\mathcal{P}(a|b)\mathcal{P}(b)}{\mathcal{P}(a)}$$

Questa regola risulta utile quando si conosce l'"effetto" di una "causa" ignota o decisamente incerta leggendo la regola nella forma che segue, di evidente utilità in attivitàdi natura diagnostica.

$$\mathcal{P}(causa|effetto) = \frac{\mathcal{P}(effetto|causa)\mathcal{P}(causa)}{\mathcal{P}(effetto)}$$

In generale si devono affrontare variabili probabilistiche a più valori, e la regola di Bayes porta a un insieme di equazioni e quindi richiede esami complessi e calcoli impegnativi.
In effetti nel passato la regola di Bayes è stata molto trascurata e ha potuto essere ampiamente adottata solo quando sono stati disponibili strumenti di calcolo automatico molto efficienti.

J52 a.10 Una applicazione semplice riguarda i cosiddetti modelli di Bayes ingenui, nei quali si ha una causa C che influenza direttamente vari effetti E_1, E_2,...,E_n accertati e che, quando la causa venisse accertata, risulterebbero condizionalmente mutuamente indipendenti.
Abbiamo allora la regola

$$\mathcal{P}(C|E_1, E_2, ..., E_n) = \mathcal{P}(C) \cdot \prod_{i=1}^{n} \mathcal{P}(E_i|C)$$

Per trattare le dipendenze tra le variabili che si incontrano nei problemi reali piuttosto articolati sono utili le reti bayesiane, digrafi aciclici i cui nodi rappresentano le varibili casuali, discrete o continue, e i cui archi esprimono la dipendenza del nodo (variabile) di ingresso dal no (variabile) di uscita; ogni nodo q è etichettato dai valori attribuiti alle probabilità che i vari nodi predecessori p_i influenzino q.

Sulle reti bayesiane si effettuano inferenze consistenti nel valutare le distribuzioni di probabilità per insiemi variabili di query, oggetti dell'indagine, a partire da insiemi di variabili esprimenti evidenze.
Sono ora disponibili algoritmi di inferenza esatta che valutano somme di prodotti di probabilità nel modo più efficiente possibile.

Nel caso privilegiato di rete bayesiana ridotta a un poliabero, ossia a un digrafo unilateralmente connesso, l'inferenza è soltanto lineare, ma per le reti bayesiane generali l'inferenza può risultare intrattabile.

Per evitare questo genere di impasse si adottano semplificazioni delle reti bayesiane mediante procedimendi di clustering che permettono di operare sui polialberi, mediante tecniche di campionamento casuale come la pesatura per verosimiglianza e gli algoritmi Monte Carlo per catene di Markov, MCMC.
Le reti di Markov sono modelli costituiti da grafi non orientati che approssimano le reti bayesiane.
Viene adottata anche la tecnica della eliminazione delle variabili che si dimostra sostanzialmante equivalente alla programmazione dinamica non seriale.
Con queste tecniche si riescono a fornire stime ragionevoli anche per grandi reti bayesiane.

J52 a.11 Un altro approccio adottato per affrontare le difficoltà che si riscontrano nel fornire input esatti ai sistemi intelligenti si serve dei fuzzy sets, gli insiemi sfumati, introdotti da Lofty Zadeh nel 1965; si tratta di funzioni aventi come dominio uno spazio S e una di tali entità, che denotiamo con E, viene espressa da una funzione di distribuzione su S che fornisce per ogni punto del suo dominio la probabilità della sua appartenenza ad E.
Questi insiemi sono utilizzati per costruire sistemi di controllo di prodotti industriali basati su regole, dette fuzzy rules, concernenti la corrispondenza fra input a valori reali e parametri di output.
I controlli fuzzy si utilizzano congiuntamente a metodi probabilistici quando oltre a trattare l'incertezza di proposizioni ben definite, occorre affrontare la vaghezza della corrispondenza fra i termini di un modello formale e gli oggetti del mondo reale che essi rappresentano.

J52 a.12 Un trattamento probabilistico, evidentemente, si rende necessario anche per modellare gli ambienti mutevoli nel tempo nei quali operano gli agenti AI.
Per questo si ricorre ai modelli di Markov, i quali assumono che lo stato X_t di un sistema in evoluzione su tempi discreti dipenda solo da un numero finito di stati precedenti. I più semplici tra questi modelli assumono che vi sia dipendenza dal solo stato X_{t-1} immediatamente precedente l'attuale.

Le attività inferenziali che li riguardano si servono di operazioni di base quali:
filtraggio o stima dello stato credenza alla quale si può dare una forma del tipo $\mathcal{P}(X_t|e_1, ..., e_t)$ che si basa su tutte le evidenze e_i disponibili;
predizione, calcolo di una espressione della forma $\mathcal{P}(X_{t+k}|e_1, ..., e_t)$ per qualche k intero positivo;
smoothing o regolarizzazione, calcolo della distribuzione a posteriori di uno stato passato utilizzando tutte le evidenze tramite espressioni della forma $\mathcal{P}(X_k|e_1, ..., e_t)$ con $k = 0, 1, ..., t-1$;
spiegazione più probabile, valutazione a partire da una sequenza σ di osservazioni della sequenza più probabile degli stati che hanno portato alla σ mediate espressioni della forma $argmax_{x_{1:t}}[\mathcal{P}(X_{1:t}|e_1, ..., e_t)]$;

A questo punto ci limitiamo a ricordare i nomi dei modelli di inferenza temporale, dei modelli di Markov nascosti, dei filtri di Kalman e delle reti byesiane dinamiche (DBN) delle quali i primi due modelli citati sono casi particolari.

Per descrivere e programmare con versatilità flessibilità i comportamenti probabilistici, sulle orme di personalità quali Leibniz, Jacob Bernoulli,De Morgan, Boole, Charles Peirce, Carnap e Keynes, si è cercato un linguaggio formale.
Questo importante problema si è portato a soluzione verso la fine del 1900 individuando alcuni linguaggi Turing-completi, funzionali ed espressivi che consentono di definire modelli probabilistici di domini anche molto complessi.

J52 a.13 Presentiamo ora qualche considerazione generale sui metodi di ottimizzazione, metodi che si presentano in numerose forme e che sono ampiamente utilizzati dagli agenti che si incontrano nell'area dell'Intelligenza artificiale a causa delle numerosissime occasioni nelle quali gli agenti stessi devono fare scelte accurate di insiemi di parametri che rappresentano situazioni complesse e che devono risultare il più possibile convenienti.

Con il termine ottimizzazione matematica, o con il suo sinonimo programmazione matematica, in generale si intende la selezione, entro un dato insieme di alternative di un elemento considerato il migliore secondo un ben definito criterio.

Si pongono problemi di ottimizzazione in tutte le discipline quantitative, in tutti i rami dell'ingegneria, nell'informatica, nell'economia e in tanto altro.
In modo più tecnico diciamo che un problema di ottimizzazione consiste nel trovare un massimo o un minimo di una funzione di più variabili.

La principale distinzione della ottimizzazione prevede i due sottocampi della ottimizzazione discreta e della ottimizzazione continua.
Nella ottimizzazione discreta si tratta di trovare un oggetto discreto che potrebbe essere un numero intero, un grafo, una permutazione o un'altra struttura discreta da scegliere in un insieme contabile, ossia numerabile o finito.
Nella ottimizzazioe continua si richiede di trovare un valore ottimo per l'argomento di una funzione continua (un minimo o un massimo per una funzione a valori reali).

Spesso la funzione da ottimizzare è sottoposta a vincoli. A questa funzione ni diversi campi applictivi si danno nomi quali funzione obiettivo, funzione utilità, funzione costo e funzione perdita.

Ai problemi di ottimizzazione si sono dedicati personalità come Fermat, Newton, Lagrange e Gauss, ma questa disciplina ha assunto una ben determinata fisionomia soprattutto in seguito ai contributi di Leonid Kantorovich intorno al 1939 e di George Dantzig intorno al 1947.

J52 a.14 Sono molti e differenziati i sottocampi della ottimizzazione matematica; segnaliamo i principali.

Programmazione convessa: studia i casi nei quali la funzione obiettivo è concava o convessa e l'eventuale insieme vincolante è convesso.

Programmazione lineare (LP): programmazione convessa con la funzione obiettivo lineare e l'insieme vincolante costituito da un poligono, un poliedro o un politopo, in ogni caso una figura limitata.

Pogrammazione quadratica: la funziome obiettivo presenta termini quadratici, e i vincoli sono dati da uguaglianze o disuguaglianze lineari; se i vincoli determinano un insieme ammissibile convesso si ha un caso particolare di programmazione convessa.

Programmazione geometrica: funzione obiettivo e vincoli dati da polinomi, ossia polinomi della forma

$$f(x_1, ..., x_n) \; = \; \sum_{h=1}^{k} c_h \, {x_1}^{a_{1h}} \cdots {x_k}^{a_{1k}}$$

con le variabili X_j e i coefficienti c_h che assumono valori reali positivi; si tratta di un caso particolare di programmazione convessa.

Programmazione intera: programmazione lineare con una parte delle variabili e dei vincoli ch possono assumere solo valori interi; non èimpegnativa quanto la programmazione convessa, ma è più difficoltosa della programmazione lieare.

Programmazione frazionale: studia l'ottimizzazione del rapporto di due funzioni non lineari.

Programmazione stocastica: studia casi nei quali alcuni vincoli dipendono da variabili casuali.

Ottimizzazione combinatria: studia problemi nei quali l'insieme delle soluzioni accettabili è discreto.

Ottimizzazione infinito-dimensionale: studia i casi nei quali l'insieme delle soluzioni accettabili è un sottoinsieme di uno spazio infinito-dimensionale, in genere uno spazio di funzioni di generi particolari.

J52 a.15 Si distinguono inoltre i casi di ottimizzazione in contesti dinamici, cioè nei quali si devono trovare soluzioni che variano nel tempo; ci limitiamo a segnalare i sottocampi che seguono.
Calcolo delle variazioni
Teoria del controllo ottimo.
Programmazione dinamica
Programmazione matematica con vincoli di equilibrio.

Vicino alle tecniche di ottimizzazione si trova il problema della determinazione di strategie afferenti alla teoria dei giochi.

J52 a.16 Segnaliamo infine sette problemi matematici generali la cui soluzione si ritiene capace di fornire solidi fondamenti alla impostazione generale del machine learning.
Questi problemi sono stati segnalati come cruciali dalle discussioni a riguardo svolte nel corso del Congresso Internaziomnale dei Matematici del 2022.

Qual'è il ruolo della profondità delle reti neurali?

Quale caratteristica dell'architettura delle reti neurali influisce sulle prestazioni del deep learning?

Perchè la discesa del gradiente stocastico converge a minimi locali buoni nonostante la non convessità del problema?

Perché le grandi reti neurali non vanno in overfit?

Perché le reti neurali funzionano meglio in ambienti descrivibili con un numero elevato di dimensioni?

Quali generi di features, di caratteristiche qualificanti, dei dati vengono apprese dalle architertture profonde?

Le reti neurali sono in grado di sostituire gli algoritmi numerici altamente specializzati nelle scienze naturali?

J52 b. influenza della Intelligenza Artificiale sulla matematica

J52 b.01 Anche una disciplina con una tradizione plurimillenaria come la Matematica ha cominciato a servirsi degli strumenti provenienti dall'area della Intelligenza artificiale.

Forse più precisamente si dovrebbe dire che fra le due discipline c'è stato un complesso di reciproci scambi di apporti e di stimoli, ma fino a pochi anni fa l'influenza della AI sulla matematica è stato scarsamente riconosciuto.

Tra i primi successi dell'AI ci sono le prime procedure per il calcolo simbolico automatizzato (in particolare per lavalutazione simbolica degli integrali) e per la dimostrazione meccanica di teoremi (in particolare per alcuni teoremi della Mathematica di Russel e Whitehead).
Queste da un lato hanno contribuito a far crescere le aspettative verso l'AI nel suo primo periodo e dall'altro hanno dato impulso alla matematica sperimentale su argomenti diversi da quelli concernenti il calcolo numerico approssimato.

J52 b.02 Questo già si era imposto e maturato nelle attività applicative della matematica, a cominciare dall'astronomia e dalle attività più quantitative dell'antichità e che erano cresciute con le esigenze della navigazione, delle rivoluzioni della macchina a vapore e dell'elettricità fino alla crescita trionfale in seguito alla rivoluzione digitale dell'ingegneria e della scienza dei computers.

La prima crisi dell'AI dovuta alla sottovalutazione dell'esplosione combinatoria ha stimolato gli studi sulla complessità computazionale e in particolare la definizione dei problemi NP-completi.
Questi contributi sono andati progressivamente crescendo insieme allo sviluppo di nuovi algoritmi, fino alla recente disponibilità di interactive theorem provers e proof checkers di grande efficienza; attualmente si sviluppano dimostrazioni automatiche che producono files di dimensioni sui 50 TB.

Agli studi sulla dimostrazione automatica di teoremi si sono successivamente affiancati gli studi per la verifica della validità delle dimostrazioni, importante attività di controllo per i teoremi più ardui ed elaborati, e le attività di miglioramento delle dimostrazioni attraverso la ricerca di catene dimostrative ottimali.

J52 b.03 La recente vistosa crescita della potenza computazionale consentita dalle computer farms e dai chips e wafers della nanoelettronica hanno portato allo sviluppo del cosiddetto "brute reasoning", una sorta di successore del "brute computing" o "number crunching" dei supercomputers della fina del secolo scorso. Le procedure di brute reasoning, oltre a permettere catene inferenziali di lunghezza umanamente insostenibili, consentono di stabilire se una entità caratterizzata da proprietà si trova entro un dominio di possibilità costruttive estremamente vasto.

J52 b.04 Un altro settore della matematica che ha adottato abbastanza prontamente metodi cresciuti nell'area AI è lo studio dei problemi inversi, in particolare nel settore del trattamento delle immagini nel quale i metodi di apprendimento automatico sono stati usati per risolvere problemi fortemente mal posti come la riduzione del rumore (denoising), l'eliminazione dell'appannamento (deblurring), la ricostruzione di parti di figure danneggiate (inpainting) e la tomografia computerizzata per angoli limitati.

In effetti, stante la mancanza di modelli matematici precisi per le immagini, il settore del trattamento delle immagini è particolarmente adatto alle applicazioni dei metodi di apprendimento da parte di automatismi.

Il settore delle equazioni differenziali invece si è accostato lentamente ai metodi dell'area AI a causa della iniziale poca evidenza dei vantaggi che questi avrebbero potuto portare allo studio delle equazioni alle derivate parziali, equazioni che si possono considerare modelli matematici molto ben definiti al livello simbolico e che non invogliano ad adottare metodi di apprendimento statistico.

Recentemente tuttavia il fatto che le reti neurali profonde riescono a battere le difficoltà dovute alla crescita della dimensionalità, quando si affrontano problemi da ambientare in spazi di elevate dimensioni, ha condotto questo settore a pensare a un paradigma alternativo.
In effetti dal 2017 i campi nei quali di incontrano l'analisi numerica delle derivate parziali e l'Intelligenza artificiale si sono andate estendendo.

J52 b.05 Un genere di vantaggio che la Matematica ha ricavato dalle ricerche nell'area AI, soprattutto dalle recenti ricerche sul machine learning, è la cresciuta interdisciplinarieteà fra teorie cresciute sulla spinta di esigenze relativamente settoriali. Questo riguarda in particolare la teoria dei giochi, molta della ottimizzazione matematica, la teoria della decisione, la teoria del controllo.

J52 b.06 Molti argomenti per i quali la Matematica ha ricevuto risultati, spunti e stimoli dalla tecnologia ICT nei quali si possono vedere idee che vengono dibattute nell'area AI hanno riguardato l'area della matematica sperimentale.
I compiti e i ruoli che sono stati attribuiti a questa sottodisciplina si possono riassumere con l'elenco che segue.

Ottenere intuizione e visione
Scoprire nuovi schemi e relazioni.
Utilizzo di visualizzazioni grafiche per suggerire principi matematici sottostanti.
Testare e soprattutto falsificare congetture
Esplorare un possibile risultato per vedere se è degno di una prova formale.
Suggerire approcci per una prova forma
Sostituire lunghe catene deduttive ottenute manualmente con derivazioni computerizzate
Confermare risultati derivati analiticamente.

J52 b.07 Tra i risultati matematici conseguiti con attività sperimentali ricordiamo i seguenti.
Sono state ottenute più dimostrazioni del teorema dei quattro colori sia servendosi di una dimostrazione limitativa seguita da una verifica al computer esaustiva sulle configurazioni rimaste incerte, sia servendosi di un software di portata generale per dimostrazioni automatiche di teoremi.
Sono stati trovati controesempi alla congettura di Eulero sulla somma di potenze intere di numeri interi.
È stata ottenuta la dimostrazione della congettura di Keplero sul più ridotto impacchettamento di sfere uguali attraverso un lungo complesso di deduzioni e verifiche sperimentali .
È sytata dimostrata la non esistenza di piani proiettivi di ordine 10
È stato dimostrato che una configurazione minima di Sudoku univocamente risolvibile richiede 17 entrate esplicitate.
Sono stati valutati numericamente con precisione molto elevata i valori forniti da serie, prodotti infiniti e integrali definiti per trovare espressioni lineari con coefficienti interi e costanti matematiche per i suddetti valori
Sono state prodotte con la computer grafica molte immagini di strutture geometriche (ed esempio della trasformazione di Möbius e del gruppo di Shottky) per ottenere evidenti convinzioni sulla validità di congetture collegate e spingere ad indagini più approfondite.

J52 b.08
Questo genere di sostegni sono stati sviluppati per vari problemi fisico-matematici.
Si è giunti a trovare una soluzione analitica mediante una generalizzazione della funzione W di Lambert per la struttura dello ione $H_2{}^+$) (caso di problema dei tre corpi quantistico) attraverso soluzioni numeriche di situazioni speciali. In conseguenza si è scoperto legame prima ignorato fra teoria della gravità e meccanica quantistica.
Nell'ambito della meccanica relativistica dei molti corpi si è dimostrata con una tolleranza dell'inverso del quadrato della velocità della luce una equivalenza fra potenziali fra particelle prima di giungere alla dimostrazione matematica di tale equivalenza e alla conseguente rivalutazione della teoria della nonlocalità quantistica di Wheeler-Feynman.
Nell'ambito dell'ottica lineare la verifica numerica dello sviluppo in serie dell'inviluppo del campo elettrico prodotto da impulsi ultrabrevi attraverso mezzi non isotropi ha fatto scoprire la mancanza di un termine nella versione precedentemente adottata per il detto sviluppo.
Vari patterns di configurazioni matematiche sono stati trovati accidentalmente, per serendipità, attraverso indagini numeriche non guidate da argomentazioni.
Sono stati individuati gli attrattori di Lorentz nei sistemi dinamici analizzati numericamente per chiarire anomalie di un modello per il tempo atmosferico.
La spirale di Ulam è stata osservata inattesamente; lo stesso è accaduto per i numeri di Ulam.
la costante di Feigenbaum è stata individuata da osservazioni numeriche prima di essere individuata con una dimostrazione rigorosa.

Un'altro genere di supporti che la Intelligenza artificiale sta cominciando ad offrire alla ricerca matematica consiste nell'indirizzamento euristico per la scoperta di nuove proprietà in campi circoscritti portato avanti avvalendosi di procedure connessionistiche addestrate con basi di conoscenze matematiche che toccano i suddetti settori.

J52 b.09 Altri temi interessanti riguardano innovazioni collegate a propostecome il Mathematical knowledge management, il QED manifesto e le tecnologie per la comunicazione delle conoscenze matematiche come il Mathematical Markup Language (MathML) e altre proposte provenienti dal consorzio W3C.

In vari momenti si sono discussi i problemi riguardanti i futuri sviluppi prevedibili o auspicabili della ricerca matematica che hanno portato a previsioni spesso ampiamente disattese, anche a causa di interessi nei confronti dei finanziamenti che hanno ridotto l'obiettività delle analisi.
Queste discussioni recentemente si sono intrecciate con le analoghe riguardanti il futuro della tecnologia e il futuro della educazione matematica (per il quale in particolare si à impegnato Laszlo Lovasz).
Probabilmente questi dibattiti in futuro dovranno tenere conto soprattutto degli sviluppi della AI, degli strumenti che renderà disponibili e delle metodologie, che saprà sviluppare, fpossibilmente numerose e innovative fino alla rivoluzionarietà.

J52 b.11 Si pone anche il problema dei sistemi AI e in particolare dei chatbots nell'insegnamento della matematica, con il timore che questi possano costituire irresistibili spinte ad imbrogliare, ma anche con il riconoscimento della loro efficienza.
Alcuni matematici praticano il cosiddetto insegnamento bilingue che alterna insegnamento tradizionale e uso degli assistenti delle dimostrazioni, i proof assistants, per la verifica degli esercizi svolti. Si osserva anche che questi assistenti possono fornire all'umano una comprensione più completa delle questioni affrontate.

Recentemente sono stati sviluppati modelli LLM addestrati con i cosiddetti "math-rich data set"' che aiutano a risolvere problemi matematici seri.
Mentre fino a qualche tempo fa sembrava utopia, si pone la prospettiva del cosiddetto "automated mathematician" e si pone il problema che questi sistemi superino l'uomo anche nel fare matematica innovativa.
Contrariamente a quanto pensato finora, i matematici cominciano a porsi il problema di perdere posti di lavoro perché sostituiti da macchine. A questo proposito va segnalato che taluni si lamentano del fatto che finora su questi temi c'è stata poca discussione. Si ritiene anche che il problema della comprensione del successo dei metodi di deep learning sia un problema per la matematica stessa e insieme a questo il problema della armonizzazione della capacità logica con la intuizione.
Taluni ritengono che per queste tematiche la matematica sia la cartina di tornasole.

J52 c. Il libro bianco sulla AI della AMS

J52 c.01 Nel luglio del 2024 l'American Mathematical Society, un organismo che raccoglie un gran numero di matematici professionisti da tutte le parti del mondo, ha pubblicato un libro biancosulla AI considerando cruciale la potenzile influenza di questa disciplina sull'intera matematica.

Artificial intelligence (AI), broadly construed, has already changed the working lives of mathematicians, and has the potential to radically alter the profession.
While the new developments come with challenges, they also bring numerous opportunities for research in mathematics, in support of teaching and mentoring roles, and for activities of the AMS. As a general principle, the AMS should be open to the potential benefits of developments in AI. This should be carried out with careful and thoughtful experimentation with new technologies; at the same time, disclosure of how and when AI is being used should be encouraged until such time as the consequences are more clearly understood.

J52 c.02 The AI Advisory Group has developed white papers summarizing current issues of importance for the following policy committees:

- Committee on the Profession: "Questions artificial intelligence raises for the mathematics profession" a broad discussion of how AI might affect our profession. - Committee on Education: "Artificial intelligence: Challenges and opportunities in postsecondary mathematics education" discussing what we should think about as educators - ethical concerns, new opportunities, and curriculum development. - Committee on Publications: "Artificial intelligence: Publishing in mathematics)" which considers implications of AI for peer review, research integrity, copyright, and publications. - Committee on Equity, Diversity, and Inclusion: "Equity, diversity, inclusion, and artificial intelligence: Issues for mathematicians to consider" discusses research opportunities around fairness and equity, as well as ethical questions around training data and large language models. - Committee on Science Policy: "Harnessing the power of science policy with mathematics" - how AI and mathematics can enrich one another, and how to enable interested mathematicians to get involved.

J52 c.03 These are intended as living documents to be updated regularly with resources and guidance. Mathematics does not stand alone in these discussions. Many other fields are being transformed, or have already been transformed, through new technologies, and the AMS should remain in discourse with other professional societies and partners to learn from their experiences and, where appropriate, collectively represent issues of broad importance.
The AMS AI Advisory Group strongly encourages all mathematicians to seriously consider the potential impact of AI on their own work, and to participate in a broad community discussion about this impact.

J52 c.04 Domande sollevate dall'intelligenza artificiale per la professione matematica
Below we list some questions that AI raises for the mathematics profession. We discuss the role of mathematicians in conversations about AI in the public sphere, the effects AI may have on the economics of the profession, and the role that AI may play in our nonresearch and research work lives. What role can mathematicians play in the public conversation around AI? We argue that mathematicians should play a role in this conversation because: – We have a domain to offer in which AI success can be measured objectively: Mathematical text and particularly computer formalized mathematical text can be checked for correctness. – We are more immune than other fields to being intimidated by mathematical or technical language. – We have a mode of understanding that could be used to create clear standards for messy questions (fairness, explainability, attribution, appropriate use of data). To

illustrate the last point, we point to the role mathematicians have played in quantifying questions of gerrymandering.1 We think mathematicians should participate in conversations concerning: – The suitability of AI tools for a particular application in the public sphere – The assessment of biases in AI tools under development, and methods by such tools could be regulated – Questions on where training data comes from and quantifying attribution for AI generated work How might AI affect the economics of the mathematics profession? The mathematics profession today benefits from the fact that mathematical skills are valued in numerous well-paid jobs from industry to academia. These jobs are the destinations for current and former undergraduate mathematics students, undergraduate mathematics majors, math grad students, postdocs, mathematics educators, and professional mathematicians.2 The decades-long boom in the tech industry has buoyed the mathematics profession as public investment has lessened. If tech work requires less of a mathematical mindset, how does this affect the economic opportunities of mathematics students? And if student enrollments fluctuate, does this affect the size of mathematics departments?3 As AI tools for teaching and grading mathematics develop, will this affect the staffing levels required to teach mathematics courses? This could have far-reaching effects on educators at all levels, including the availability of mathematics jobs and graduate teaching stipends. How can AI assist mathematicians with the nonresearch aspects of our work and should it?

– In the drafting of papers: typesetting handwritten mathematics, drawing diagrams in LaTeX, fixing LaTeX bugs, drafting prose for introduction or background sections, transcribing dictated notes, helping with grammar and spelling (particularly for non-native speakers)4 – In improving the accessibility of the mathematics literature: by translating mathematical texts between languages, by transcribing spoken mathematics into text, and by narrating written mathematics (particularly in LaTeX or involving diagrams) – In the review of papers: summarizing the technical contents for editors in a different field, verifying correctness of mathematical arguments (if proofs are accompanied by computer formalizations), identifying related work5 – In hiring and admissions: filtered large application pools according to some committee-designed attributes – In many aspects of teaching: assistance in grading, in lecture note preparation, in creating homework and exams (perhaps modifying for individual students or sections) – In administrative work: creating documents for internal use, writing emails6 To be clear, we do not necessarily endorse all of these potential applications. It's quite likely that mathematicians will want to be directly involved in the development of some of these tools, in the way that LaTeX has always been partially developed by hobbyists. We hope for a vibrant ecosystem that is not dependent on a small number of big tech providers or expensive proprietary software. As a community, we must have a proactive conversation about the ethics of the development of and uses of AI. – All of the hypothetical software mentioned above will have been trained on the output of countless hours of careful work by mathematics professionals. Does the AMS want to defend the collective output of mathematicians over the decades?7 – What uses of AI software in decisions with human impacts (grading, hiring, admissions, peer review) are fair and appropriate? – How do we compensate mathematicians (and tech companies) who develop useful tools for our community while ensuring their broad access? – Should we develop norms concerning the disclosure of AI in writing that reflect nuances of different hypothetical use cases?8 – How do we ensure proper attribution of mathematical ideas that are disseminated by AI? What roles might AI play in mathematics research? We can envision a few potential uses of AI in mathematics research. Others will undoubtedly emerge. A computer proof assistant (such as Agda, Coq, HOL Light, Isabelle, Lean, -) is a software program that checks the correctness of mathematical arguments in these logical languages and may provide automation to help construct such arguments.9 Independent of developments in AI, researchers have been working in the last decades to make such computer proof assistants more powerful and user-friendly and mathematicians are increasingly adopting them to formally verify work in their own areas. Large language models

(LLMs) can develop text, written in a conversational language but also in programming languages, and notably in the logical languages of computer proof assistants. One plausible pipeline is that LLMs could iteratively develop such proofs by incorporating the feedback on line-by-line correctness provided by proof assistants.

J52 h. altro

J52 h.01 Nel titolo la scienza con la tradizione più lunga e la tecnologia più recente e alla moda
Della AI si parla tanto con discorsi piuttosto frammentari e scollegati dalla Matematica con toni spesso drammatici e dedicando molto spazio a lamentarsi della complessità degli argomenti.
Qui cerco di collegarli sostenendo una visione unitaria che deve coinvolgere una bella fila di soggetti.
Problemi Soluzioni Algoritmi Matematica Fisica Informatica Intelligenza artificiale
Accenno soltanto alle complessità che riguardano anche le applicazioni (che già ora toccano quasi tutto).
I problemi universalmente presenti: toccano tutti i viventi, sono noti dalla preistoria, soprattutto dal periodo della nascita dell'agricoltura.
Si sono prospettate molte soluzioni: prima particolari e tendenzialmente approssimate, urgenti e tuttofare; poi più accurate, specialistiche e sistematiche. Le soluzioni si accompagnano con lo sviluppo di strumenti (ruota, aratro, imbarcazioni, armi, abaci, ...).

J52 h.02 Attraverso le soluzioni dei problemi e i relativi strumenti si riescono a spiegare molti cambiamenti storici e culturali.
Dalla ricerca di soluzioni di alto livello discendono gli algoritmi (e i metodi). Gli algoritmi sono indicaziono operative accurate e dettagliate dovendo essere prive di ambiguità rivolte ad esecutori umani ma oggi anche artificiali che devono essere molto accurati e diligenti ma non necessariamente onniscenti sul problema e innovativi.
Alla soluzione dei problemi impegnativi devono contribuire molti agenti e questi devono saper interoperare e quindi devono saper intercomunicare.
È necessario precisare linguaggi specialistici e artificiali (come i linguaggi delle espressioni matematiche, delle formule chimiche, della biologia). a cominciare dagli insiemi di caratteri (come ASCII e Unicode).
La precisione porta ai linguaggi formali da collocare in un A^* con il ruolo dell'ambiente e quindi alla notazione $w \in \mathsf{A}^*$; si prospettano le trasmissioni di stringhe fra agenti capaci di distinguere i caratteri e quindi alla giustapposizione, a prefissi, suffissi, infissi, riflessione di stringhe.
In particolare $|^*$ porta alla nozione di numero naturale e quindi alla somma e allo zero e alla notazione $n \in \mathbb{N}$.
Per esprimere significati si devono introdurre stringhe strutturate, coppie, liste, riflessione, sostituzioni,
La nozione
Le coppie di caratteri o interi o stringhe conducono al prodotto cartesiano e alle potenze cartesiane. In particolare si considerano il prodotto di naturali, le potenze naturali di naturali e le notazioni binarie, decimali, posizionali degli interi naturali.
Si cercano operazioni inverse capaci di contribuire a risolvere problemi: si comincia dalla differenza l'inversa della somma e quindi si arriva ai numeri negativi e a $\mathbb{Z}$.
Si osserva che si sta portando avanti una costruzione basata su semplificazione di problemi e soluzioni che consiste in una progressiva modellizzazione analitica che riguarda il dualismo tra situazioni reali osservabili e loro formalizzazione e che si giustifica con la ragionevolezza e l'accuratezza del procedere e che propone una fiducia condivisibile ovvero una condivisibilità affidabile.
Si cerca operazione inversa del prodotto fra interi e si giunge a quoziente e resto se ci si limita agli interi o all'estensione degli interi ai razionali e alla notazione $\mathbb{Q}$ per il loro insieme.
$\mathbb{Q}$ va visto come parte di $\mathbb{N} \times \mathbb{N} = \mathbb{N} \times \mathbb{N}$ e in grado di eaprimere pendenze e direzioni nel piano.
In $\mathbb{N} \times \mathbb{N}$ si introducono varie nozioni geometriche: vettori, quadranze, spread, traslazioni, parallelismo, riflessioni, simmetria, invarianza, parità.

Queste si estendono a $\mathbb{Q} \times \mathbb{Q}$, il piano razionale nel quale si aggiungono le nozioni di rotazione, angolo dotato di seno e coseno, circonferenza, omotetia; per gestire queste nozioni quantitativamente sono disponibili le terne pitagoriche di numeri razionali, estensioni delle terne pitagoriche classiche con componenti interi.

Si introducono le relazioni entro $\mathbb{Z} \times \mathbb{Z}$ ed entro $\mathbb{Q} \times \mathbb{Q}$ e le più particolari funzioni-ZtZ e funzioni-QtQ; inoltre si estendono queste nozioni a 3 e più dimensioni e con 3D la possibilità di studi approssimati dello spazio della fisica classica.

L'insieme dei razionali si rivela insufficiente a trovare $\sqrt{2}$ e anche $\sqrt{n}$ per ogni intero positivo non quadrato, cioè a risolvere un problema inverso. Occorre ampliare $\mathbb{Q}$ all'insieme dei numeri algebrici i quali corrispondono alla possibilità di effettuare costruzioni geometriche piane con riga e compasso, in particolare trovare il punto intersezione fra circonferenza $Circ(\mathbf{0}, 1)$ e funzione $y = x$ e trovare zeri di polinomi a coefficienti interi o razionali.

Anche questo insieme si rivela insufficiente a trovare zeri di funzioni $\sin(x)$ e $\cos(x)$, ovvero a ottenere la quadratura del cerchio. Si impone la introduzione di un insieme numerico più ampio.

Prima di questo vanno introdotte le macchine di Turing strumento formale dotato di nastro illimitatamente estendibile e di semplici istruzioni che riesce ad effettuare tutte le elaborazioni formali viste e tutti gli algoritmi incontrati.

Una macchina di Turing con un set di dati può arrestarsi o proseguire illimitatamente generando sempre nuovi dati; in questo caso si dice illimitatamente generatrice.

Ogni A^* e può essere generato ordinatamente da una tale macchina; in particolare può esserlo NMb o anche $\mathbb{N} \times \mathbb{N}$, $\mathbb{Z} \times \mathbb{Z}$, $\mathbb{Q} \times \mathbb{Q}$, $\mathbb{N}^{\times 3}$ e insiemi simili.

Diciamo insieme numerabile un tale insieme.

Ogni insieme ottenibile da una macchina di Turing si dice ricorsivo.

Sono insiemi ricorsivi gli insiemi ottenuti da macchine che generano un sottoinsieme di un ambiente numerabile ottenuto con un algoritmo selettivo che sicuramente si arresta.

Data una macchina di Turing generatrice qualsiasi con un set di dati qualsiasi può darsi che non si arresti. Essa genera un insieme ricorsivo che può non essere numerabile.

Si dimostra che non esiste macchina di Turing in grado di decidere in un numero finito di passi se una arbitraria macchina di Turing con dati arbitrari si arresterà o procederà illimitatamente.

Questo costituisce un limite della matematica strettamente collegato con i teoremi di incompletezza di di Gödel; con questo si accorda la scelta di questo discorso di dare importanza alla affidabilità condivisa senza pretendere una verità della matematica sostenuta dalla logica, disciplina molto impegnativa.

Si precisa facilmente una macchina generatrice con ordine in grado di procedere a generare A^*, il momoide libero sull'alfabeto ordinato A, secondo l'ordine lunghezza-lessicografico.

In particolare si può procedere a generare l'insieme delle macchine di Turing e l'insieme delle macchine di Turing che procedono a generare successioni convergenti di numeri razionali, o equivalenti intervalli di $\mathbb{Q}$ telescopici e convergenti. Queste successioni comprendono le scritture numeriche binarie o decimali illimitate costruibii e si possono definire numeri reali costruibili; il loro insieme si denota con $\mathbb{R}_{\mathsf{C}}$ ed evidentemente è sovrainsieme dell'insieme dei numeri algebrici che a sua volta è sovrainsieme di $\mathbb{Q}$.

Con questi insiemi costruibili si viene a disporre della nozione di costruzione che in linea di principio è illimitata e della nozione di infinito potenziale.

Si congettura che tutti i problemi ben definiti si possano trattare nell'ambito di questi insiemi, se vogliamo servendosi di macchine di Turing o di altri meccanismi formali equivalenti (i cosiddetti automatismi Turing completi). Questa affermazione viene detta congettura di Church-Turing e viene accettata dalla quasi totalità degli studiosi della computabilità.

Con queste nozioni si potrebbero portare avanti tutte le attività finalizzate alle soluzioni affidabili di problemi ben definiti.
Accade però che le descrizioni finalizzate alle visioni generali dei procedimenti, degli algoritmi e dei metodi in termini di approssimazioni razionali dei numeri costruibili risulta troppo prolissa e pesante da formulare, da comunicare e da padroneggiare servendosi anche dell'intuizione.
Risulta invece conveniente fare riferimento alla logica introducendo assiomaticamente gli insiemi e i numeri reali. Questo richiede di formulare la teoria assiomatica degli insiemi, la teoria assiomatica dei numeri reali e la teoria della dimostrazione formale. Si tratta di teorie molto impegnative molto consolidate e accettate dalla massima parte dei matematici e degli studiosi delle discipline quantitativo-strutturali; qui le possiamo considerare attendibilmente valide e limitarci a utilizzare i loro risultati.
Abbiamo quindi l'insieme dei numeri reali $\mathbb{R}$ che possiamo pensare distesi sulla retta reale, retta dotata di un'origine corrispondente al numero 0 e a una lunghezza uguale a 1 riguardante l'intervallo tra 0 e 1. a questa retta appartengono tutti gli interi, i razionali e i costruibili, numeri costituenti un insieme numerabile.
Si dimostra invece, per assurdo, che l'insieme dei numeri forniti da una sequenza infinita di cifre binarie ha un cardinale superiore ad $\aleph_0$ che chiamiamo cardinale del continuo e denotiamo con $\aleph_1$, ossia poniamo $\aleph_1 := |\mathbb{R}|$.
Abbiamo quindi una situazione paradossale con l'insieme dei reali costruibili numerabile e l'insieme dei reali non costruibili più che numerabile; nessuno dei reali noncostruibili si può utilizzare per un problema concreto perché se lo fosse sarebbe approssimabile, ossia costruibile. Però l'insieme dei reali consente presentazioni utili alla comunità matematico-scientifica e quindi gli sviluppi per scopi generali della matematica e delle sue applicazioni se ne servono abbondantemente.

Dalla seconda guerra mondiale a oggi stiamo assistendo alla vistosa crescita dell'elettronica e dell'informatica in termini di conoscenze disponibili, di strumenti realizzati, di diffusione e di influenza sociale e culturale.

Confrontando il computer che ha guidato Apollo 11 nel 1969 e un odierno comune smart phone, si constata che i dispositivi di memoria sono cresciuti di un fattore 1-7 milioni, che le velocità di calcolo sono aumentate di circa 60 000 volte, che il costo delle componenti è diminuito di un fattore 100 milioni e che la diffusione è almeno un milione di volte maggiore.

Questo ha consentito una grande crescita della usabilità delle tecnologie della informazione e della comunicazione (ICT) e ha consentito di affrontare una ampia varietà di problemi, in particolare di problemi afferenti a discipline che solo da pochi anni si stanno dotando di procedimenti, strumenti e atteggiamenti quantitativi.

Oggi gli uomini possono risolvere molti più problemi, con molti più dati disponibili, con maggiore profondità e con molto minore fatica.

Vediamo a grandi linee le tappe del mondo dei computers.

Dopo i calcolatori elettromeccanici ed i calcolatori elettronici, tra il 1945 e gli ultimi anni 1950 sono stati alcuni computers in pochi esemplari con circuiti a valvole termoioniche e memorie a tamburo.
Dal 1955 si sono adottati circuiti a transistors e memorie a nuclei di ferrite, molto più piccoli, più veloci, meno energivori e di maggiore durata. Con questi elaboratori attribuiti a una seconda generazione si sono avuti i primi modelli con molti esempleri e la possibilità di usarli con linguaggi di programmazione di alto livello (Fortran e COBOL) abbastanza indipendenti dall'hardware.

Sono andate aumentando le unità periferiche anch'esse controllate da transistors: nastri e dischi magnetici, telescriventi che hanno consentito l'utilizzo a distanza, anche tramite le prime reti.

All'inizio degli anni 1960 si sono adottati i primi circuiti integrati su germanio e soprattutto su silicio e sono stati prodotti i primi supercomputers per affrontare calcoli tecnico scientifici sempre più impegnativi; inoltre sono stati utilizzati come memorie i transistor MOSFET, metal-oxide semiconductor field-effect transistor. I computers di questo periodo si sono attribuiti a una terza generazione.

Alla fine degli anni 1960 sono comparsi i primi computers attribuiti alla quarta generazione basati sui microprocessori, resi possibili dalla rapida crescita della densità di transistors sui chips in accordo con la legge di Moore; questa legge dice che ogni paio d'anni raddoppia la densità dei circuiti su un chip di circuito integrato e costituisce un riferimento per la continua veloce crescita delle prestazioni complessive dei computers.
Negli anni 1970 si sono imposti i microprocessori che includevano CPU, memoria RAM e ROM e logica per le operazioni di I/O.
I microprocessori hanno consentito la produzione di piccoli minicomputers e poco dopo dei personal computers (il primo Olivetti P6060, dotato di floppy discs, stampante, display per 32 caratteri e linguaggio BASIC). Questo ha consentito una enorme diffusione delle attività informatiche e della relativa cultura; inoltre ha spinto verso lo sviluppo di Internet da parte dell'agenzia DARPA della Difesa USA. Questo ha ridotto l'importanza di IBM, il preedenza la compagnia nettamente prevalente
Negli anni 1980 si sono imposti computers nei piccoli ambienti di lavoro e per videogiochi, all'inizio molto diversificati (con la popolarità di Apple II e Commodore), in seguito unificati dallo standard de facto del computer IBM con sistema operativo di Microsoft.
Negli anni 1990 la continua crescita dell'elettronica digitali ha consentito di produrre le workstations, piccole macchine molto potenti per calcoli tecnico scientifici molto impegnativi in piccoli laboratori dotate in particolare di grandi prestazioni grafiche; questo ha fatto crescere il settore dei video giochi.
Lo sviluppo più ricco di conseguenze di quegli anni ha riguardato il WWW, il World Wide Web, iniziato nel 1989 da Tim Bernrs-Lee presso il CERN, reso liberamente disponibile nel 1993 e accolto dalle maggiori aziende informatiche e da un gran numero di istituzioni intorno al 1997.
Il Web ha poi visto una successiva continua crescita sia sul piano tecnico, che su quello della adozione di massa.
Con WWW è cresciuta la concreta interoperabilità dei sistemi con la conseguenza della crescita di collaborazioni a distanza in molti settori scientifici, tecnologici, medici e finanziari e con l'esplosione della globalizzazione dei mercati.
La crescita della interattività e della interoperabilità per persone e organismi è ulteriormente cresciuta a partire dal 2007 con la disponibilità dei telefoni cellulari e satellitari e con la messa in orbite a bassa quota di intere costellazioni di satelliti artificiali. Si è quindi arrivati ad una attuale disponibilità di circa 5 miliardi di accessi al WWW e di telefoni smart con capacità elaborative molto elevate. La disponibilità di sistemi per la comunicazione di grande efficienza e versatilità porta al grande vantaggio di abbreviare drasticamente i tempi richiesti dalle innovazioni.

Sul piano dell'hardware si sono avute continue rilevanti innovazioni. vengono resi disponibili sistemi multiprocessori capaci di grandi quantità di calcoli paralleli; vengono costruite server farms, e processor clusters costituiti da milioni di unità di calcolo con potenze di calcolo ben superiori alle precedenti le quali vengono erogate a utenti remoti collegati tramite linee ad alta velocità di trasmissione che stanno facendo nascere sistemi applicativi di portata globale. Si è sviluppato il cloud computing con server farms in grado di fornire a utenti remoti un ampio spettro di servizi comprendente oltre alla potenza di calcolo, alte prestazioni interattive, gestione di grandi quantità di dati (big data), assistenza software, aggiornamento e potenziamento della strumentazione e consulenza alla progettazione.

Vengono sviluppati sistemi digitali con alte prestazioni specialistiche in particolare per la grafica pro-

fessionale e per video games (GPU di Nvidia e TPU di Google) e per implementare reti neurali per il deep learning (TPU, chip WSE-2 di Cerebras con $2.6\,10^{12}$ transistors).

Si nutrono grandi aspettative per i sistemi per il quantum computing, ossia di unità computazionali che sfruttano fenomeni quantistici molto diverse dei vari tipi di CPU che operano sui bits, sui possibili valori 0 e 1: un computer quantistico si serve dei cosiddetti qbits, ciascuno dei quali esprime una sovrapposizione di 0 e 1.
I principi generali e gli algoritmi di base del quantum computing sono ben definiti, ma la loro costruzione effettiva e sicura presenta varie difficoltà dovute al comportamento molto delicato dei qbits. Solo nel 2019 si sono avute macchine in grado di gestire 54 qbits.

L'idea dell'Intelligenza artificiale nasce assieme ai primi computers negli anni 1940 per opera di personalità lungimiranti (Wiener, Turing, McCulloch con Pits, Shannon, Minsky, McCarthy, Chomsky) che si interrogano sui limiti delle possibilità dei computers, oltre alle riconosciute loro capacità di eseguire velocemente calcoli numerici e di manipolare grandi quantità di dati.

Nel periodo iniziale incontriamo:
le lucide considerazioni di Turing che sposta il problema dell'essenza della intelligenza sulla problematicità del distinguere fra le risposte di un umano e quelle di una macchina (test di Turing);
la costruzione di Minsky di SNARC, la prima macchina con reti neurali;
il programma di Strachey per giocare a Dama;
la costruzione da parte di Newell e Simon di "Logic Theorist", programma in grado di dimostrare alcuni dei teoremi enunciati nel testo "Principia Mathematica" di Russell e Whitehead.

Nel 1956 presso il Darmouth College si è svolto un seminario con la partecipazione di Shannon, Minsky, McCarthy e altri pionieri dal quale sono usciti il nome Artificial Intelligence e la definizione del suo obiettivo, la produzione di strumenti e procedimenti in grado di svolgere compiti che in precedenza avevano richiesto l'intelligenza umana.
Negli anni successivi sono stati ottenuti sensibili successi nei campi dei calcoli matematici simbolici, delle dimostrazioni di teoremi, nei giochi come dama e scacchi, nel controllo di frasi della lingua inglese e nella adozione di tecniche formali derivate dalla logica per l'esecuzione di procedimenti deduttivi.
Si è allora diffuso un clima di ottimismo e l'area AI ha promesso rapidi progressi che in USA e UK hanno ottenuto rilevanti finanziamenti.
Molte promesse richiedevano l'esecuzione di scelte che avrebbero richiesti tempi di calcolo enormi (esplosione combinatoria dei casi da analizzare) pesantemente sottovalutati.
Alla fine degli anni 1960 Minsky e Papert hanno sostenuto l'opportunità di concentrarsi su situazioni semplici e tendenzialmente artificiali a imitazione dei modelli semplificati sui quali si è basata la fisica.
Questo atteggiamento ha portato ai primi successi nella visione artificiale, ai primi bracci robotici, e al sistema robotico SHRDLU di Winograd in grado di eseguire ordini impartiti in inglese.
In quegli anni sono stati delineati un programma 'AI debole", più realistico, volto al conseguimento di risultati che rispondono ad esigenze concrete circoscritte e un programma "AI forte" più ambizioso e ottimista che sostiene l'opportunità di concentrare gli sforzi su obiettivi il cui conseguimento apra la possibilità di successi di ampia portata.
Nel 1958 lo psicologo Frank Rosenblatt ha ripreso il metodo delle reti neurali e del connessionismo costruendo Perceptron macchina per l'analisi di fotografie di persone e altre immagini. Questa macchina dava risultati solo in parte attendibili ma ha dimostrato di essere in grado di migliorarsi tenendo conto dei risultati parziali.
Tuttavia il libro "Percepron" di Minsky e Papert ha mostrato le limitazioni di questo genere di macchine provocando un lungo abbandono delle ricerche sul commessionismo.

Le previsioni esagerate fatte intorno al 1960 non hanno portato ai risultati promessi, soprattutto quelli sulla traduzione automatica e sul controllo di macchine militari e questo ha portato al taglio dei finanziamenti in UK e USA intrno al 1973.
Questo ha portato al cosiddetto primo "AI winter" con un sensibile discredito dell'intero settore.

Si sono però avuti tre eventi che hanno preparato un terzo periodo di ripresa dell'AI.
Si è sviluppato nel 1972-73 lo studio della complessità computazionale (Cook e Karp) che ha portato a una visione più matura dello studio degli algoritmi che da è prograssivamente molto cresciuto.

Nei primi anni 1970 gli ambienti industriali e ministeriali giapponesi hanno deciso di investire sulla robotica e su altre prospettive della AI per sostenere l'industria manufatturiera ottenendo alcune interessanti realizzazioni.

Negli anni 1970 è cresciuta notevolmente la aspettativa di crescita dell'elettronica digitale (legge di Moore) con la disponibilità di processori più potenti, meno costosi e più maneggevoli con la disponibilità di minicomputers e di quantità mai viste di macchine versatili come i PCs e la conseguente maggiore fiducia nelle iniziative basate sull'informatica.
Questa maggiore disponibilità verso iniziative innovative di ambienti materiali e culturali è stata ulteriormente rafforzata dalla prima diffusione dell'uso delle reti di computers (Internet).

La prospettiva d1 una forte continua crescita della ICT ha attratto interessi economici, industriali e scientifici verso i progetti basati su tecnologie innovative e in particolare verso l'area AI.
In particolare gli ambienti medici che nei primi anni 1970 avevano accolto scarsamente i primi sistemi esperti per la cura della salute, hanno cominciato ad accettarli e ad adottare il loro uso nella diagnostica, per migliorare le attività cliniche e per formare una visione globale dei problemi sanitari.
Agli ambienti industriali e della ricerca si sono resi disponibili processori, workstation e mainframes sensibilmente più potenti, dotati anche di prestazioni simboliche e parallele, e si sono sviluppate ricerche metodologiche più accurate.
Questo ha portato a crescenti successi, in particolare nella robotica industriale e ha consentito di lasciare alle spalle il primo AI winter.

In quegli anni si sono effermati i sistemi esperti, sistemi software in grado di risolvere problemi in domini di conoscenze circoscritti servendosi di regole di scelta suggerite da persone competenti.
Hanno avuto un notevole successo con chiari vantaggi economici i sistemi per la produzione e l'assemblaggio di prodotti industriali elaborati, molte grandi compagnie hanno costituito al loro interno reparti per l'AI e si è andato formando un nuovo cpm[arto industriale.
In questo periodo il mondo della ricerca in AI ha dato sempre maggiore importanza alla gestione delle conoscenze riconosciute come fattori essenziali per il successo di molte procedure avanzate; è quindi nata una ingegneria della conoscenza.
Nel 1981 il governo giapponese ha varato il progetto Fifth Generation con obiettivi ambiziosi come sistemi per sostenere conversazioni e traduzioni automatiche e interpretare immagini. Questo ha indotto i governi USA e UK ad aumentare gli investimenti in AI
Sono cresciuti gli studi sulle procedure per la deduzione automatica diventati molto più versatili grazie alla introduzione del linguaggio Prolog e ad altri strumenti di programmazione logica.
Negli anni 1980 sono ripresi gli studi sul connessionismo quasi contrapposti alla AI simbolica della inferenza automatica per opera di Hinton e Rumelhart con la proposta di reti neurali a molti strati dimostratesi in grado di apprendere come migliorare il proprio comportamento e risultate utili per l'OCR, il riconoscimento dei caratteri scritti e per il riconoscimento del parlato.
Queste applicazioni richiedevano massicce elaborazioni e avrebbero dovuto aspettere la disponibilità di strumenti elettronici molto più efficienti per imporsi largamente.

I successi ottenuti hanno portato a entusiasmi che poco dopo, insieme alla constatazione che molti obiettivi della Fifth Generation giapponese erano mancati, si sono ritorti in una bolla specultiva e a un secondo AI winter durato all'incirca dal 1987 al 1993, ma con una più prolungata cattiva fama della AI considerata una attività di sognatori inconcludenti.

In questo periodo si sono sviluppate serie analisi critiche dei metodi dell'AI che hanno condotto a concezioni nuove per l'imitazione dei comportamenti umani.
Si è imposto l'atteggiamento volto alla costruzione di agenti AI dotati di un corpo in grado di percepire l'esterno e di muoversi e agire materialmente con la convinzione che le capacità sensomotorie e le prestazioni intuitive degli agenti fossero da curare non meno dei ragionamenti deduttivi. Inoltre negli anni 1990 si è iniziato a immettere negli studi di AI idee dell'economia e della teoria delle decisioni (Giudea Pearl) e lo sviluppo sistematico di metodi probabilistici.
D'altra parte in quegli anni si erano consolidati i sistemi per giocare giochi impegnativi fino ad arrivare nel 1997 alla vittoria del sistema Deep Blue della IBM sul campione internazionale degli scacchi Kasparov, fatto che ha avuto ampia risonanza. A questo proposito va anche segnalato che la velocità di Deep Blue era 10 milione di volte superiore a quella della prima macchina giocatrice di scacchi disponibile nel 1951.

Negli anni 1990 un evento della massima importanza è stato lo sviluppo del WWW (Tim Berners Lee), la sua libera disponibilità dal 1993, la sua adozione da parte delle maggiori industrie e delle istituzioni nel 1997 e l'inizio della sua diffusione inarrestabile nei paesi dotati di tecnologie digitali intorno al 1999. Questo ha portato alla crescita delle collaborazioni aziendali, commerciali e culturali, ha supportato la globalizzazione dei commerci e delle produzioni ha fatto crescere la circolazione delle conoscenze e delle idee e ha portato ad ambienti più favorevoli alle innovazioni, particolarmente a quelle digitali e della AI.
Questo genere di crescita è stato ulteriormente rafforzato dalla disponibilità dal 2007 degli smart phones e dalla loro rapida diffusione e integrazione con internet. Oggi sono disponibili circa 5 miliardi di connessioni a Internet e di telefoni cellulari, sono ampie le possibilità di comunicazione a distanza, anche grazie ai cavi sottomarini e alle costellazioni di molte migliaia di satelliti‘ e sono molteplici le attività basate sulla interoperabilità.
Un importante fenomeno è stata la crescita di imprese che utilizzano le tecnologie ICT e AI per attività fortemente innovative e investono in continua innovazione elevate percentuali dei loro ricavi (enormi). Si tratta in particolare di MicroSoft, Apple, Facebook (Meta), Google (Alphabet) e Amazon dagli USA; di Sony dal Giappone; di Samsung dalla Corea del Sud; di Baidoo, Tancent e Alibaba dalla Cina.
Queste compagnie supportano iniziative di portata tendenzialmente globale e hanno una grande capacità di influire anche sulla cultura, ad esempio tramite canali di streaming e produzione di intrattenimento.
In questo nuovo quadro globale, al quale vanno aggiunti i continui progressi dell'elettronica giunta allo stadio della nanoelettronica dei dispositivi che riguardano le distanze interatomiche, la AI ha trovato nuovi appoggi, nuovi stimoli e nuovi sbocchi ed ha continuato a crescere.

L'enorme traffico delle informazioni consentito da Internt e dalla telefonia mobile multifunzionale, oltre all'inizio della IoT della rete Internet delle cose, ha portato a una disponibilità di dati utilizzabili per molteplici scopi mai prima neppure pensabile.
È quindi cresciuta la cosiddetta scienza dei dati che si avvale soprattutto di dati ricavati dalle attività digitali.

Per la gestione dei grandi volumi di dati si sono sviluppate tecniche specifiche che costituiscono il mondo dei cosiddetti Big Data.
Attualmente i dati costituiscono fattori determinanti per molte decisioni e per nuove iniziative.
Tra le loro applicazioni vanno citate: registrazioni governative per scopi civili e statistici; applicazioni finanziarie per operazioni commerciali, per decidere investimenti e per operazioni assicurative; azioni per la gestione dei rischi; attività per la cura della salute attraverso la gestione delle cartelle cliniche digitali, la diagnosi assistita da computer, la ricerca esplorativa in biomedicina.
La tecnologia dei big data è essenziale per la conduzione delle attività scientifiche più impegnative.

A questo proposito ricordiamo:
il contenimento dell'impatto della pandemia del COVID-19;
la raccolta di dati astronomici della Sloan Digital Sky Survey;
la raccolta di dati per la simukazione del clima da parte della NASA;
sequenziamento del genoma umano e raccolte di dati biologici;
gli esperimenti condotti presso il Large Hadron Collider del CERN
le attività delle grandi aziende tecnologiche come Amazon, eBay, Facebook e Google e delle grandi organizzazioni commerciali e per la gesrione delle carte di credito.
le molteplici attività di alcuni stati, soprattutto Cina, India, USA, UK e Israele vanno dal controllo di opinioni e comportamenti, alla gestione della cura della salute, dalle competizioni elettorali, al controllo della fruizione di farmaci e droghe da parte della popolazione.