L'elaborazione d'informazione nelle reti neurali

If a machine is expected to be infallible, it cannot also be intelligent.

A. Turing

Elena Agliari *Dipartimento di Matematica "Guido Castelnuovo" – Sapienza Università di Roma.*

Adriano Barra *Dipartimento di Matematica & Fisica "Ennio De Giorgi" – Università del Salento.*

L'intento di queste note è mostrare come due *modi operandi* tipici, rispettivamente, della Matematica (l'inferenza statistica) e della Fisica (la meccanica statistica) possano fungere da pilastri concettuali sui quali erigere una *teoria delle reti neurali*, a dire, un telaio logico-deduttivo nel quale rappresentare le reti di neuroni e dal quale evincere, come proprietà emergenti delle stesse, le capacità intellettive superiori di cui queste fruiscono. Le abbiamo volutamente chiamate emergenti, poiché, come mostreremo nella prima sezione, il singolo neurone può essere relegato ad un semplice interruttore, un sommatore a soglia rumoroso, lontano dal manifestare le suddette capacità che quindi scaturiscono dalla rete in quanto tale e non dai suoi singoli costituenti elementali. Nel resto delle scritto discuteremo un *paradigma teorico minimale*: divideremo il processo della *cognizione* in due momenti, quello dell'*apprendimento* ("learning") e quello dell'*impiego* di ciò che si è appreso ("retrieval") e mostreremo come l'inferenza statistica sia il linguaggio naturale per il primo momento mentre la meccanica statistica lo sia per il secondo (ed il confine tra le due alquanto sfumato). Per descrivere questi processi sfrutteremo due modelli paradigmatici per l'apprendimento automatico artificiale e per le reti neurali biologiche, ovvero, rispettivamente, la macchina di Boltzmann e la rete di Hopfield. Infine, nella chiusura dello scritto, mostreremo come, dal punto di vista astratto della processazione d'informazione, questi modelli siano due facce di un'unica medaglia, rendendo apprendimento ed impiego (d'informazione) un tutt'uno, i.e. il fenomeno cognitivo.

L'intersezione tra le Neuroscienze e l'Intelligenza Artificiale

Il modo in cui il cervello rappresenta ed elabora le informazioni sul mondo, l'emergenza della coscienza e la potenziale ricaduta applicativa delle macchine intelligenti sono oggigiorno tra i temi più caldi della Scienza. Recenti progressi nella comprensione del funzionamento del cervello (e.g., attraverso registrazioni multielettrodo per sondare l'attività cerebrale) e nel miglioramento delle prestazioni della sua controparte sintetica (e.g., attraverso nuove tecnologie come le GPU, la creazione di enormi database e lo sviluppo di algoritmi per l'elaborazione di questi *big data*) hanno suscitato profondo interesse e clamore. Tra gli scopi di questo articolo divulgativo è anche l'apprezzare come la modellistica matematica[1] – che è sempre stata soggiaciente tanto alle investigazioni nelle neuroscienze quanto in intelligenza artificiale (IA) – abbia eretto paradigmi validi in entrambe le declinazioni della processazione d'informazione e sia stata di fatto il loro *trait d'union* sin dalla genesi di queste discipline. In effetti, le neuroscienze e l'IA sono finalmente abbastanza mature da interagire ed autosostenersi reciprocamente, promuovendo ulteriormente il loro sviluppo: iniziative come lo Human Brain Project in Europa, il progetto BRAIN negli Stati Uniti, il programma Brain-MINDS in Giappone ed il China Brain Project hanno dimostrato l'entusiasmo e l'impegno dell'intera comunità scientifica su scala globale, tuttavia, l'interesse nel decifrare il *codice neurale* non è nato con quest'ultima ondata di entusiasmo, alimentata, di fatto, da progressi tecnologici più che concettuali. In effetti, come l'era moderna delle neuroscienze ha le sue radici nel metodo rivoluzionario di Golgi per colorare i neuriti e nelle indagini pioneristiche di Cajal perpetrate più di un secolo fa, alla stessa stregua l'IA pone le sue fondamenta nella macchina di Babbage (primordiale rotativa da calcolo derivata dal telaio tessile), in ultima istanza frutto della rivoluzione industriale inglese avvenuta oltre due secoli orsono.

In estrema sintesi, e restringendo il nostro discorso dalla metà del Novecento ai nostri giorni (dove si concentrano i principali risultati), una volta effettuati i primi esperimenti sulla comunicazione elettrica tra neuroni, immediatamente modelli matematici che mimassero l'emissione di impulsi elettrici analoghi a quelli sperimentalmente rivelati iniziarono a proliferare (dai neuroni di Hodgkin e Huxley, a quelli di Stein, a quelli di Nakubo, etc.) fino a culminare nel *perceptrone* di Rosenblatt del 1958 dove si offriva una prima trattazione logica dei neuroni astratti e delle loro capacità (suggerendone un impiego artificiale). Questa prima ondata di entusiasmo frenò bruscamente nel 1969 con l'uscita del libro *Perceptron* di Minsky e Papert, i quali mostrarono come questi modelli matematici di neuroni per la computazione spontanea fossero inadatti a risolvere perfino banali operazioni di logica elementare (e.g., tecnicamente si fermavano allo *XOR* perché erano intrinsecamente classificatori linari): questa doccia fredda, che fece sprofondare quest'intersezione tra le due Scienze in quello che è chiamato *the winter time* nella Comunità, fu in realtà una felice cornucopia poiché (tenendo a mente che nel mentre una meccanica statistica complessa per i modelli di campo medio era ormai a buon punto [1]) spostò l'attenzione dal singolo neurone alle reti di neuroni. Prendendo spunto dalla geniale idea di Hebb sull'apprendimento sinaptico, nel 1982 Hopfield [2] – ed indipendentemente Little e Amari – sviluppò un modello minimale di rete neurale che aveva capacità emergenti di gran lunga superiori a quelle che i singoli neuroni, per quanto *complessificati*, riuscissero ad avere. La rete di Hopfield è un grafo completamente connesso sui cui nodi vivono dei neuroni binari (on/off) ed i cui archi mimano le connesioni sinaptiche tra gli stessi e possono essere associati a pesi (efficacie sinaptiche) sia positivi che negativi: dal punto di vista della meccanica statistica questo sistema complesso è un vetro di spin (si veda, per una definizione di vetro di spin, la *lezione mancata*). Come mostreremo, questo sistema presenta comportamenti non banali che naturalmente elessero la meccanica statistica a disciplina cardine per lo studio teorico di questi modelli di reti neurali. In particolare,

[1] Nella *lezione mancata* di questo numero di Ithaca dedicato all'Intelligenza Artificiale approfondiamo alcune basi teoriche necessarie per una migliore comprensione della modellistica affrontata nel presente lavoro divulgativo (ed in molti altri di questo volume): in particolare, tutte le Hamiltoniane usate in questo articolo sono forme quadratiche.

tra il 1979 ed il 1980, Parisi [3] mise in luce nei vetri di spin proprietà ultrametriche che, da un punto di vista cognitivo, si è tentati associare alla categorizzazione genere-specie che spontaneamente tendiamo a fare. Qualche anno più tardi, nel 1985 Amit, Gutfreund e Sompolinsky [4] studiarono il modello di Hopfield sfruttando idee e tecniche meccanico-statstiche originariamente sviluppate per l'indagine dei vetri di spin ottenendo la prima trattazione sistemica interamente meccanico statistica di una rete neurale[2].
In queste note informali ripercorreremo in primis la via di Hopfield, ispirato dalle reti biologiche, per poi approdare a modelli di impiego nella controparte artificiale per infine mostrare come le due vie, la biologica e l'artificiale, siano coincidenti dal punto di vista astratto della processazione di informazione mediante generiche reti frustrate[3].

Dinamica di singolo neurone

In questa sezione prenderemo confidenza con gli attori principali del nostro cervello, i *neuroni* (i nodi della rete neurale) e le loro connessioni, dette *sinapsi* (gli archi della rete)[4]. Queste componenti evolvono su scale così distanti che, come sovente accade in Fisica (e.g., nell'approssimazione di Born-Oppeneimer in struttura della materia o in tutta la termodinamica adiabatica), possiamo trattare separatamente le relative dinamiche: quando ci preoccuperemo dei neuroni

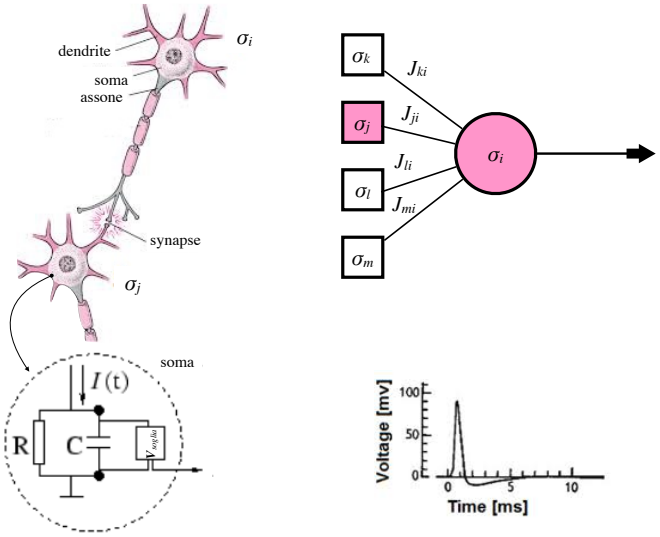

Figura 1: *Rappresentazione schematica di un neurone biologico (a sinistra) e di un neuron artificiale/logico (a destra). Nel primo sono mostrati alcuni dettagli biologici (assone, dendriti, soma ed un contatto sinaptico con un neurone afferente dal basso) ed il loro* equivalente circuitale *come integratore RC, mentre nel secondo si vede come questo sommi gli stimoli provenienti dai neuroni adiacenti, pesandoli mediante il filtro sinaptico J, ed eventualmente producendo un segnale a sua volta (in basso a destra mostriamo inoltre l'andamento temporale tipico di un singolo spike).*

considereremo le sinapsi *congelate* (come gli accoppiamenti nei vetri di spin), viceversa quando ci interesseremo alla dinamica sinaptica potremo tralasciare l'influenza di quella neurale (poiché lo stato dei neuroni potrá essere *mediato via*).
Un'altra similitudine con la Fisica, anticipata nella sezione precedente, risiede nel fatto che tanto i neuroni possono essere (dal punto di vista della processazione di informazione) fondamentalmente in due stati ("on", emettono un segnale elettrico ed "off", rimangono quiescienti) alla stregua degli spin di Ising, tanto le sinapsi possono essere sia eccitatorie (mimando gli accoppiamenti positivi) che inibitorie (mimando quelli negativi), in ultima istanza permettendoci così di sancire che, da una prospettiva meccanico statistica, la rete neurale è un vetro di spin.
A seguire presenteremo due modelli stilizzati di neurone, il primo (il modello integrate-and-fire di Stein) ambisce a fornire un supporto ai fisiologi alle prese con le reti neurali biologiche, mentre il secondo (il modello logico di McCulloch e Pitts) costituisce un tassello fondamentale

[2] L'impiego della meccanica statistica offriva un ulteriore vantaggio pratico in quanto non richiede una descrizione particolarmente dettagliata dei componenti del sistema oggetto di studio (e.g., non è necessaria una conoscenza minuziosa della posizione e della velocità di ciascuna particella per sviluppare un modello di gas per determinarne pressione e temperatura) ed effettivamente negli anni '80 le informazioni disponibili sulla struttura microscopica delle reti neurali biologiche erano ancora piuttosto limitate.

[3] Per la comprensione, cruciale, dell'aggettivo *frustrato* si veda di nuovo la *lezione mancata*.

[4] Più precisamente, un neurone è costituito da un soma, i.e., il corpo cellulare, da un "cavo di uscita dove eventualmente propagare il segnale elettrico" chiamato *assone* e da molti "cavi di entrata", che formano l'*albero dendritico* al quale neuroni afferenti mandano i loro stimoli: l'assone di un neurone afferente ed il dendrite del neurone ricevente sono connessi mediante le sinapsi: dal punto di vista della processazione d'informazione, i costituenti passivi (i.e., assoni e dendriti) non giocano un ruolo saliente e verranno perciò trascurati per semplicità.

nella controparte artificiale.

- Neurone (biologico) di Stein
  Nel neurone di Stein gli elementi cardine sono la differenza di potenziale della membrana esterna del neurone (a dire il voltaggio che si registra mettendo un elettrodo all'interno del soma ed uno all'esterno), la sua resistenza R, la sua capacità C, le correnti afferenti da neuroni esterni e la possibilità di generare a sua volta una corrente (un segnale elettrico impulsato che viene chiamato *spike*[5]), si veda la figura 1. Il tutto si amalgama nell'equazione differenziale per l'evoluzione del potenziale di membrana V_i per il neurone i-esimo che si legge

$$\frac{dV_i}{dt} = -\frac{V_i}{\tau} + \sum_{j\neq i}^{N} J_{ij} \sum_{k}^{T} \delta(t - t_{kj} - d_{ij}), \quad (1)$$

  dove $\tau := RC$ funge da costante di tempo di questo "circuito integratore", mentre la corrente afferente al neurone in esame è costituita dalla somma (lineare) dei contributi provenienti dai vari neuroni ad esso connesso, ognuno pesato mediante l'efficacia sinaptica J_{ij} e ricevuto stocasticamente a tempi diversi t_{kj} e ritardati dalla propagazione stessa mediante i d_{ij}. Questa dinamica neurale è dissipativa, in virtù del termine $-V_i$, ma continuamente rinvigorita da stimoli esterni: poiché le sinapsi possono sia aumentare la differenza di potenziale del neurone afferente (sinapsi eccitatorie) sia diminuirla (sinapsi inibitoria), l'evoluzione del potenziale sinaptico compie un moto erratico e se questo raggiunge una soglia critica per la stabilità della membrana, semplificando oltremodo, questa "si rompe temporaneamente", dando luogo ad un impulso elettrico, i.e. lo spike, rivolto ai neuroni riceventi (ognuno dei quali lo peserà, in concerto con altri afferenti da neuroni terzi, positivamente o negativamente in ragione della sinapsi che congiunge l'assone di output con i dendriti di input)[6]. Un esempio della dinamica neurale fino all'emissione di uno spike è mostrato in figura 2.

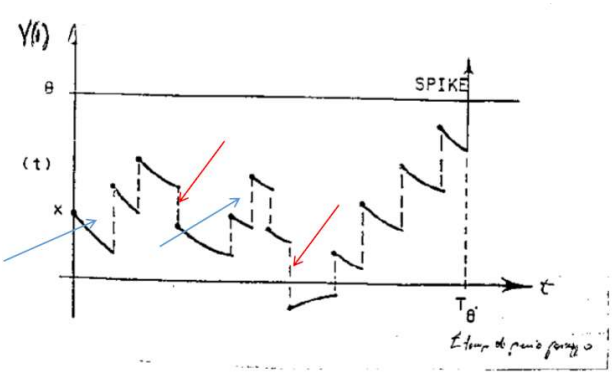

Figura 2: *Esempio dell'evoluzione temporale erratica del potenziale di membrana di un neurone fino alla sua generazione dello spike. I salti discontinui sono dovuti a spikes provenienti da neuroni afferenti (in blu filtrati da sinapsi eccitatorie ed in rosso da sinapsi inibitorie).*
Nota: in basso si legge in corsivo "tempo di primo passaggio": è la scrittura di Daniel Amit, pioniere delle reti neurali (l'immagine e' presa dal suo corso di Reti Neurali, tenuto in Sapienza dagli anni novanta fino al 2005).

- Neurone (artificiale) di McCulloch&Pitts
  In questo neurone logico, si veda la figura 1, si trascurano i dettagli fisici (quali le dissipazioni ed i ritardi di propagazione) e si mette il fuoco solamente sulle capacità di calcolo dello stesso. Usando gli stessi simboli del neurone di Stein possiamo scrivere

$$V_i(t + \Delta t) = \Theta\left(\sum_{j\neq i}^{N} J_{ij} I_j - V_{\text{soglia}}\right), \quad (2)$$

  dove Θ è la funzione di Heaviside: lo stato di uscita del neurone (i.e., il potenziale della sua membrana) è fisso a zero a meno che la somma dei contributi elettrici afferenti non superi una soglia V_{soglia}, in qual caso il neurone emette il *potenziale d'azione*, cosa che si avverte notando che lo stato di uscita del neurone diventa uno. Nel neurone artifi-

[5] La genesi dello spike è dovuta ad un brusco crollo della stabilità della membrana cellulare, la quale, se destabilizzata dai continui spikes a sua volta ricevuti, si apre per dar luogo a sua volta al prosieguo della comunicazione nervosa lungo l'assone della cellula in questione, alla volta degli alberi dendritici di altri neuroni con cui questo è connesso.

[6] Per un approfondimento sulle reti neurali biologiche si veda il contributo di Paolo Del Giudice & Maurizio Mattia in questo volume.

ciale si possono quindi rappresentare i due stati logici di Boole[7].

Immaginando di scrivere ora l'evoluzione del potenziale di Stein, l'eq. 1, o di quello di McCulloch&Pitts, l'eq. 2, non più per il singolo neurone i-simo, ma per tutti gli N neuroni che compongono la rete (risultando quindi in un sistema di N equazioni differenziali accoppiate), la presenza sottostante di una rete di neuroni interconnessi appare nitida, alla pari delle piuttosto limitate capacità di processare l'informazione da parte del singolo neurone: l'atto di cognizione deve emergere come un fenomeno collettivo della rete (motivo per cui, a seguire, chiamiamo in causa la Meccanica Statistica come modus operandi per investigarlo).

La cognizione nelle reti neurali

Come abbiamo visto nella precedente sezione, ogni singolo neurone può vivere (in prima approssimazione) in due stati: quiesciente o emettitore di segnale. Prendiamo a prestito dalla Fisica lo spin di Ising $\sigma_i \in \{-1, +1\}$ per caratterizzarlo in maniera tale che, finché V_i è minore della soglia, si ha $\sigma_i = -1$, mentre quando V_i raggiunge la soglia per l'emissione dello spike $\sigma_i \to +1$. Consideriamo ora una rete costituita da N neuroni $\{\sigma_i\}_{i=1,...,N}$ e scriviamo la legge evolutiva per il generico neurone i-esimo come

$$\sigma_i(t+1) = \text{sign}[\tanh(\beta h_i(t)) + \eta_i(t)], \quad (3)$$

$$h_i(t) = \sum_{j \neq i}^{N} J_{ij}\sigma_j(t) + h_i^{ext}, \quad (4)$$

dove h_i è il campo afferente sul neurone σ_i in esame (ed è costituito dalla somma lineare di tutti gli stimoli prodotti dai neuroni afferenti $\{\sigma_j\}_{j\neq i}$, pesati con le rispettive efficacie sinaptiche J_{ij} e da un eventuale stimolo esterno h^{ext}) mentre la tangente iperbolica rilassa l'assunto di assenza di rumore tacito nella funzione a gradino di McCullch&Pitts: η_i è una variabile casuale uniformemente distribuita in $[-1, +1]$ e $\beta \in \mathbb{R}^+$ è un parametro che modula la stocasticità del processo in maniera tale che quando $\beta \to \infty$ la dinamica è deterministica e lo spin si allinea alla direzione del campo h_i (e si riottiene il comportamento logico di McCulloch&Pitts); quando $\beta \to 0$ i campi diventano impercettibili e la serie temporale degli stati neurali diventa una sequenza di Bernoulli di questionabile interesse per gli scopi di questo scritto. In un contesto meccanico-statistico classico β gioca il ruolo dell'inverso della temperatura (in unità opportune)[8].
Questa dinamica si può scrivere, per l'intera rete, in termini probabilistici, introducendo la probabilita' $P_t(\sigma)$ di trovare, al passo di aggiornamento t, la rete in un generico stato $\sigma := \{\sigma_1, ..., \sigma_N\}$, tra i 2^N possibili, come

$$P_{t+1}(\sigma) = \prod_{i=1}^{N} \frac{e^{\beta\sigma_i h_i(\sigma(t))}}{2\cosh[\beta\sigma_i h_i(\sigma(t))]},$$

alla volta di un processo di Markov

$$P_{t+1}(\sigma) = \sum_{\sigma'} W[\sigma;\sigma']P_t(\sigma'),$$

con W opportuna matrice di transizione. È possibile dimostrare che questo processo è ergodico[9] e, per $t \to \infty$, $P_t(\sigma)$ converge ad un'unica distribuzione stazionaria. Inoltre, se ci restringiamo a considerare efficacie sinaptiche simmetriche (i.e., $J_{ij} = J_{ji}$), la dinamica soddisfa il *bilancio dettagliato*, il quale garantisce che lo stato stazionario a cui il sistema rilassa sia uno stato di equilibrio e la relativa distribuzione abbia la forma funzionale della distribuzione di Gibbs

$$\lim_{t\to\infty} P_t(\sigma) =: P(\sigma) \propto \exp[-\beta H(\sigma|J)] \quad (5)$$

per qualche opportuna funzione costo $H(\sigma|J)$ (o Hamiltoniana se si vuole preservare il gergo fisico). Questa informazione, come vedremo nelle due prossime sezioni, è cruciale tanto per l'apprendimento quanto per l'impiego di ciò che si è appreso. Nel seguito, per chiarezza espositiva, tratteremo prima il richiamo alla memoria di informazione precedentemente appresa e dopo ci concentreremo sul processo di apprendimento[10].

[7] Per un approfondimento sulle reti neurali artificiali si veda il contributo di Giorgio Buttazzo in questo volume.

[8] Si veda a questo proposito il riquadro "La temperatura ubriaca" nella *lezione mancata*.

[9] L'ergodicità vale quasi ovunque in β, ovvero ad eccezione del limite $\beta \to \infty$; per maggiori dettagli rimandiamo a [5, 6].

[10] Si veda anche il contributo di Daniele Tantari in questo volume per un simile approccio.

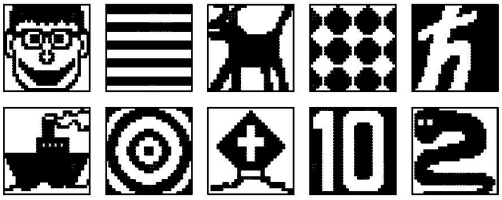

Figura 3: *Esempi di dieci immagini in bianco e nero formate da 30×30 pixel, incamerate da una rete di Hopfield di 900 neuroni binari.*

Il paradigma minimale del retrieval: meccanica statistica

Alla volta di un'Hamiltoniana efficace da inserire nell'esponenziale di Gibbs nella distribuzione(5), introdotti gli N neuroni di Ising $\{\sigma_i\}_{i=1,...,N}$, dobbiamo specificare meglio la matrice sinaptica J_{ij}. Per fare questo introduciamo il generico concetto di *pattern*, ξ che non è altro che un'informazione codificata in un linguaggio binario: un pattern può essere un concetto, una parola, un'immagine, etc.. Qui assumeremo che un pattern rappresenti un'immagine in bianco e nero, codificata attraverso una stringa di lunghezza fissa, costituita da N bit (si veda figura 3).
Per semplicità (ma non solo[11]) lavoreremo solo con patterns random: un pattern $\xi = (\xi_1, \xi_2, ..., \xi_N) \in \{-1, +1, \}^N$ si genera estraendo l'elemento ξ_i secondo la probabilità $P(\xi_i = +1) = P(\xi_i = -1) = 1/2$, per ogni $i \in (1, ..., N)$. Inoltre, poiché non vogliamo usare un'intera rete neurale (cioè N neuroni) per gestire un unico pattern ξ, ma per gestirne $P \sim O(N)$, distingueremo i vari pattern attraverso un'etichetta: $\xi \to \xi^\mu$, $\mu \in (1, ..., P)$. Siccome sappiamo che una qualunque (ragionevole) dinamica neurale stocastica converge necessariamente verso le configurazioni corrispondenti ai minimi della funzione costo $H(\sigma|J)$, il punto fondamentale ora è *incastonare* questi pattern nei minimi della funzione costo in maniera tale che la termalizzazione del sistema (garantita dal bilancio dettagliato) faccia evolvere lo stato neuronale da una configurazione iniziale $\sigma(0)$ ad una configurazione $\sigma(t) = \xi^\mu$ stabile per tutti i tempi successivi (almeno entro un certo grado di errore); questo fenomeno viene interpretato come la ricostruzione del pattern μ-esimo[12]. Il particolare pattern ξ^μ a cui il sistema spontaneamente rilassa dipenderà dallo stato iniziale $\sigma(0)$, interpretato come input della rete. Detto in altre parole, vogliamo definire la matrice delle interazioni J in modo che le configurazioni neurali corrispondenti a ciascuno dei P pattern costituiscano degli attrattori, si vedano le figure 4 e 5. A questo scopo, la scelta più intuitiva è $H(\sigma|J) \sim -N^{-1} \sum_{\mu=1}^{P} (\xi^\mu \cdot \sigma)^2$, infatti, se la rete

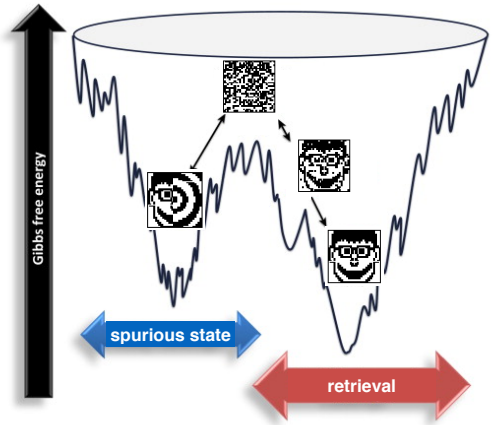

Figura 4: *Rappresentazione grafica dell'energia (libera) del modello di Hopfield. Nel regime in cui la rete lavora al meglio, gli stati corrispondenti ai pattern sono minimi globali e, poiché le dinamiche stocastiche a cui ogni neurone obbedisce devono convergere ad essi, la termalizzazione nella meccanica statistica coincide con il riconoscimento dei pattern mediante memoria associativa nella teoria delle reti neurali. Il profilo energetico presenta anche minimi locali, ovvero stati mestastabili, che tipicamente corrispondono a combinazioni degli stessi pattern memorizzati, anche detti stati spuri; il rilassamento ad uno di questi stati viene considerato come un errore da parte della memoria.*

[11] Si potrebbe obiettare – a ragione [7] – che una teoria random non abbia granché senso come *oscillatore armonico* di una teoria per le reti neurali. Osserviamo però che per il teorema di compressione di Shannon se la rete in esame è in grado di gestire P patterns casuali sarà certamente in grado di gestirne almeno lo stesso numero se correlati o con struttura al loro interno: la teoria random può fornire utili limiti ed offre un classico quadro di riferimento (dove tutto fattorizza asintoticamente) [8]. Di contro è parimenti d'obbligo, convenire che molti dei problemi interessanti in IA sono proprio legati alla presenza di struttura nei dataset (si veda a tal proposito il contributo di Matteo Marsili in questo volume e si approcciano coerentemente con telai inferenziali *profondi*, i.e a molti strati [9], per i quali si vedano i contributi di Guido Sanguinetti e Carlo Lucibello nel presente volume.)

[12] Si pensi per esempio come immagine ad un volto a noi ben noto: se ci viene presentato solo un dettaglio, per esempio gli occhi, subito noi sappiamo riconoscere, ovvero riportare alla memoria (fare *pattern recognition* mediante memoria associativa) l'intero volto.

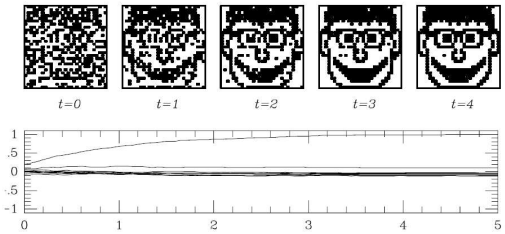

Figura 5: *Richiamo di un pattern. Al crescere del suo parametro d'ordine da zero verso uno (cioè al termalizzare della rete), si vede il ricostruirsi della* faccia *che via via appare alla memoria. Parimenti sotto viene mostrata l'evoluzione delle* 10 *magnetizzazioni di Mattis (una per pattern incamerato) e si osserva come una -quella relativa al pattern richiamato- converga monotonamente al suo massimo mentre le altre nove oscillino (con intensità $\sim N^{-1/2}$) intorno allo zero.*

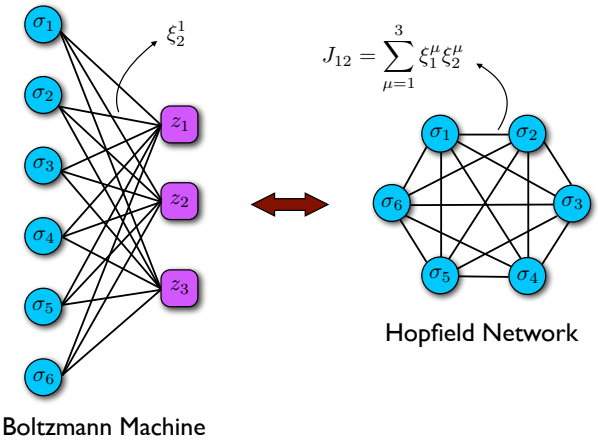

Figura 6: *Rappresenzazioni schematiche di una macchina di Boltzmann (sinistra), archetipo della macchina che apprende in machine learning, e di una rete di Hopfield (destra), oscillatore armonico delle reti biologiche, i.e. memorie associative che eseguono pattern recognition.*

non riconosce alcun pattern, l'energia è circa nulla, mentre se riconosce un certo pattern la rete acquista un'energia $-\mathcal{O}(N)$, di gran lunga più conveniente[13]. Il peso sinaptico tra il neurone i ed il neurone j risulta pertanto $J_{ij} = \sum_\mu \xi_i^\mu \xi_j^\mu$: si implementa in maniera naturale l'idea di apprendimento di Donald Hebb (erede di Ivan Pavlov), per cui *neuroni che emettono congiuntamente e reciprocamente segnali elettrici rinforzano i relativi canali di comunicazione (qui schematizzati nella matrice sinaptica)*[14] [10]. Possiamo quindi scrivere l'Hamiltoniana di Hopfield come

$$H_H(\sigma|J) = -\frac{1}{2N}\sum_{i,j}^{N,N}\sum_{\mu=1}^{P}(\xi_i^\mu \xi_j^\mu)\sigma_i\sigma_j \quad (6)$$

$$= -\frac{N}{2}\sum_{\mu=1}^{P} m_\mu^2, \quad (7)$$

dove il pedice H nell'Hamiltoniana sta per Hopfield e nella seconda riga abbiamo introdotto i P parametri d'ordine *magnetizzazioni di Mattis*, definiti come $m_\mu := N^{-1}\sum_{i=1}^{N}\xi_i^\mu\sigma_i$ (per avere una rappresentazione grafica della rete si veda Figura 6, grafo di destra). A questo punto è elementare notare che, al fine di minimizzare l'energia H_H alla rete convenga avere una[15] magnetizzazione di Mattis pari ad uno: ai neuroni della rete, per vivere comodi, conviene organizzarsi e questa loro organizzazione spontanea produce proprio il richiamo alla memoria di patterns precedentemente appresi.
Lo schema di apprendimento Hebbiano genera nel profilo energetico non solo i P minimi globali corrispondenti ai pattern memorizzati, ma anche un grandissimo numero (esponenzialmente crescente in N) di minimi locali tipicamente corrispondenti a "misture" di pattern, anche detti stati spuri (e.g., lo stato $\sigma_i = \mathrm{sgn}(\xi_i^1+\xi_i^2+\xi_i^3)$, per $i = 1, ..., N$). Quando il rapporto tra il numero di pattern ed il numero di neuroni è troppo alto (i.e., $\alpha = P/N > \alpha_c \approx 0.14$), questi minimi locali dominano il panoramo energetico ed il sistema non è più in grado di richiamare correttamente. Da queste considerazioni emerge anche che, a nostro avviso, questo modello debba rientrare nei cosiddetti *sistemi complessi*: questo non stupisce poichè,

[13] Osserviamo che, per costruzione, i pattern $\{\xi^\mu\}_{\mu=1,...,P}$ sono tra loro ortogonali (almeno nel limite di N grande), ovvero $\xi^\mu \cdot \xi^\nu = 0$, per ogni $\mu \neq \nu$. Di conseguenza, la ricostruzione di un pattern, diciamo ξ^ν, comporta che lo stato neuronale sia (circa) ortogonale a tutti gli altri pattern e di conseguenza l'unico contributo non nullo all'energia $H(\sigma|J)$ proviene dal termine ν-esimo della sommatoria.

[14] Questa idea è molto semplice ed è di fatto *economia idraulica*: se il neurone i sta emettendo incessantemente segnali al neurone j e viceversa, mentre non ne sta mandando al neurone k, conviene ampliare il canale di comunicazione tra i e j e diminuire quello tra i e k per preservare l'omeostasi della rete e minimizzare la congestione di segnali, proprio come si cerca di minimizzare le impedenze in una rete idraulica.

[15] Come ricordato prima, poiché i pattern sono ortogonali, la rete può richiamare correttamente solo una memoria per volta.

essendo le sinapsi (cioè gli accoppiamenti tra i neuroni) tanto eccitatorie (i.e. positive) quanto inibitorie (i.e. negative), il modello di Hopfield è, dal punto di vista della Meccanica Statistica, un particolare esempio (molto interessante) di spin glass [16]. Dal punto di vista statistico questo modello cattura e riproduce fedelmente funzioni di correlazione a due punti, che forniscono una sufficiente caratterizzazione per pattern randomici appartenenti allo scenario classico descritto dai Teoremi del Limite Centrale. Se volessimo cimentarci in problemi molto più complessi (e.g., la comprensione del linguaggio naturale, la comprensione dei dati generati al CERN per lo studio della fisica delle alte energie o il tracciamento automatico per il contenimento di una pandemia[17]) probabilmente dovremmo optare per funzioni costo in grado di inferire bene funzioni di correlazione a molti punti: questo risulterebbe in un telaio inferenziale lontano da una distribuzione di Gibbs di un'Hamiltoniana quadratica: molti dei progressi di cui l'IA si fa artefice in questi anni avvengono mediante architetture *deep* [9] (che non affrontiamo in questo articolo).

Il paradigma minimale del learning: inferenza statistica

In questa sezione dedicata all'apprendimento inizieremo descrivendo un classico modello di rete neurale artificiale, per poi mostrare come, in ultima istanza, questa via non sia altro che una rilettura del paradigma Hebbiano, opportunamente declinato per le macchine piuttosto che per la materia organica.
Sempre restringendoci a reti per le quali valga il bilancio dettagliato[18], uno dei mattoni fondamentali sui quali si erigono oggi architetture da calcolo per il moderno *deep learning* è certamente la Macchina di Boltzmann (*Boltzmann machine* [14, 15]): questa è una rete bipartita, con uno strato di neuroni visibile, a cui viene fornito input dall'esterno, ed uno strato di neuroni nascosto, atto a scovarne le correlazioni (si veda figura 6, grafo di sinistra).
Nello specifico assumiamo che lo strato di neuroni visibile sia composto da N spin di Ising $\sigma_i = \pm 1, i \in (1, ..., N)$, mentre lo strato nascosto sia composto da P spin di Ising[19] $\tau_\mu = \pm 1, \mu \in (1, ..., P)$, assumiamo inoltre che esistano delle connessioni sinaptiche unicamente tra neuroni di strati diversi, quindi la matrice sinaptica sia una matrice $N \times P$ il cui ingresso generico chiamiamo $\xi_i^\mu \in \mathbb{R}$ con $i \in (1, ..., N)$ e $\mu \in (1, ..., P)$. Dal punto di vista della meccanica statistica, questa macchina null'altro è che uno spin glass bipartito, per cui la funzione costo (o Hamiltoniana) H_B, dove il pedice ricorda il nome Boltzmann, e relativa misura di Gibbs, si scrivono rispettivamente

$$H_B(\sigma, \tau|\xi) = \frac{-1}{\sqrt{N}} \sum_{i,\mu}^{N,P} \xi_i^\mu \sigma_i \tau_\mu - \sum_i^N h_i^{ext} \sigma_i \quad (8)$$

$$P_\xi(\sigma, \tau) = \frac{e^{-\beta H_B(\sigma,\tau|\xi)}}{Z_B}. \quad (9)$$

dove $\mathbf{h}^{ext}$ è un campo esterno al fine di far interagire la rete con il mondo esterno e $Z_B := \sum_\sigma^{2^N} \sum_\tau^{2^P} e^{-\beta H_B(\sigma,\tau|\xi)}$ è un fattore di normalizzazione, chiamato *funzione di partizione* in Meccanica Statistica[20].
Mentre il processo di richiamo coinvolge la dinamica neurale e fonda le sue radici teoriche nella Meccanica Statistica (in ultima istanza nell'estremizzazione dell'entropia vincolata da opportune funzioni costo, per dirla à la Jaynes), il processo di apprendimento coinvolge la dinamica sinaptica e fonda le sue radici teoriche nell'Inferenza

[16] In relazione alla sezione sul *Riduzionismo Statistico* nella Lezione Mancata, la rete di Hopfield ha una funzione costo quadratica (inevitabilmente, poichè abbiamo richiesto un minimo parabolico nel caso di correlazione tra lo stato della rete ed il pattern, i.e., $H_H \sim -(\sigma \cdot \xi)^2$) ed abbiamo visto come questo risulti in sostanziale accordo con le osservazioni empiriche di Pavlov ed Hebb sulle caratteristiche che la matrice sinaptica debba presenzare (la regola d'apprendimento di Hebb).

[17] Questi tre temi sono trattati, rispettivamente, nei contributi di Valerio Basile, Konstantinos Bachas & Alfredo Braunstein, Luca Dall'Asta ed Alessandro Ingrosso nel presente volume.

[18] La richiesta della simmetria degli accoppiamenti ci permette di parlare di una fetta, comunque cospicua, di modelli ma esclude importanti macchine e relativi algoritmi di apprendimento, probabilmente prime tra tutte la *back-propagation* nelle reti *feed-forward*, il *reinforcement learning* e le *convolutional networks*.

[19] In questo articolo divulgativo non entreremo in tecnicismi, ma la scelta del tipo di neurone da usare influenza in maniera cruciale il funzionamento della macchina [12], si vedano a tal proposito i contributi di Aurèlienne Decelle e Carlo Lucibello nel presente volume.

[20] Ed è importante notare che nel tentare di calcolarla con forza bruta, tipicamente fallendo, ci si imbatte in un conto NP.

Statistica. In questo parallelo, facciamo una prima distinzione sostituendo all'entropia di Shannon della Meccanica Statistica, la mutua entropia $D(P|Q)$ di Kullback-Leibler che, introdotte due distribuzioni $P(\sigma)$ e $Q(\sigma)$ definite sullo stesso spazio di probabilità, ne misura la similarità e si scrive

$$D(P|Q) = \sum_{\sigma} P(\sigma) \ln \left(\frac{P(\sigma)}{Q(\sigma)} \right),$$

in maniera tale che quest'osservabile sia sempre non negativa ed uguale a zero se e soltanto se $P = Q$ quasi ovunque. Vogliamo usare questo strumento matematico, per esempio estremizzandolo opportunamente, per fare machine learning: la macchina di Boltzmann, assunto che abbia matrici sinaptiche simmetriche, ammette una rappresentazione probabilistica $P(\sigma, \tau)$ in termini di una distribuzione di Gibbs di un'opportuna funzione costo $H(\sigma, \tau|\xi)$. Sottilineiamo nuovamente che, in questo contesto, sono le ξ -le connessioni sinaptiche- i parametri liberi da poter variare a nostro vantaggio per far apprendere la rete (ovvero per l'estremizzazione della mutua entropia).
Sempre con fare pratico, per prendere confidenza, immaginiamo di avere M realizzazioni eventualmente rumorose (i.e., un campione casuale semplice) di uno dei previi pattern (cioè M immagini dello stesso soggetto), tutte di dimensione N (formate cioè da N pixel binari) che possiamo indicare come $\{\tilde{\sigma}_i^{(\alpha)}\}_{i=1,\dots,N}^{\alpha=1,\dots,M}$. Assumiamo che le M realizzazioni siano state generate identicamente ed indipendentemente dalla stessa distribuzione $Q(\tilde{\sigma})$. Ricordiamo che $Q(\tilde{\sigma})$ è incognita, ma possiamo stimarne i momenti a partire dagli esempi a disposizione, mentre $P_\xi(\sigma)$ ha solo la forma funzionale fissata (ed è una comoda famiglia di esponenziali peraltro). Scriviamo come punto di partenza la distribuzione congiunta $R(\tilde{\sigma}|\xi)$ dei dati e di $P_\xi(\sigma, \tau)$

$$\begin{aligned} R(\tilde{\sigma}|\xi) &= \prod_{\alpha=1}^{M} P_\xi(\tilde{\sigma}^{(\alpha)}, \tau) = e^{\sum_{\alpha=1}^{M} \ln P_\xi(\tilde{\sigma}^{(\alpha)}, \tau)} \\ &=: e^{\mathcal{L}(\xi|\tilde{\sigma}, \tau)}, \end{aligned} \tag{10}$$

quindi, invece di estremizzare $R(\tilde{\sigma}, \tau|\xi)$, possiamo usare $\mathcal{L}(\xi|\tilde{\sigma}, \tau)$, e parimenti non cambia nulla se alla stessa aggiungiamo e sottraiamo l'entropia empirica $S_M(Q) = -M^{-1} \sum_{\alpha=1}^{M} \ln Q(\tilde{\sigma}^{(\alpha)})$ ottenendo

$$\mathcal{L}(\xi|\tilde{\sigma}, \tau) = -\frac{1}{M} \sum_{\alpha=1}^{M} \ln \left(\frac{Q(\tilde{\sigma}^{(\alpha)})}{P_\xi(\sigma, \tau)} \right) - S_M(Q),$$

cioe' $\mathcal{L}(\xi|\tilde{\sigma}) = D_M(P|Q) - S(Q)$, dove la mutua entropia empirica si legge $D_M(P|Q) = \frac{1}{M} \sum_{\alpha=1}^{M} \ln \left(\frac{Q(\tilde{\sigma}^{(\alpha)})}{P_\xi(\sigma)} \right)$. La minimizzazione di $D_M(P|Q)$[21] ci porta a regole di apprendimento di sicura convergenza (nonostante nulla ci dica sui tempi per raggiungerla [13]).
Una volta applicata alla macchina di Boltzmann, nello scenario di apprendimento supervisionato (che ora introduciamo mediante il concetto di *media clamped*), questa procedura inferenziale risulta in una "ricetta" per l'apprendimento particolarmente intuitiva: introduciamo delle medie *clamped*, cioè dove lo strato visibile σ è costretto di volta in volta a *vedere* le M immagini, cioè tale che $P(\sigma) = \sum_{\tilde{\sigma}} P(\tilde{\sigma}) \delta(\sigma - \tilde{\sigma})$ (per esempio supplendo alla rete le varie immagini mediante il campo $\mathbf{h}$) e chiamiamo la media di Boltzmann di una generica osservabile O $\langle O \rangle =: \sum_{\sigma,\tau} O(\sigma, \tau) P_\xi(\sigma, \tau) / \sum_{\sigma,\tau} P_\xi(\sigma, \tau)$ mentre chiamiamo *clamped* la media con lo strato visibile forzato sull'immagine $\langle . \rangle_{clamped}$. La regola di apprendimento che otteniamo estremizzando l'entropia di Kullback-Leibler in questa maniera si legge

$$\Delta \xi_i^\mu = \epsilon \beta \left(\langle \sigma_i \tau_\mu \rangle - \langle \sigma_i \tau_\mu \rangle_{clamped} \right). \tag{11}$$

Quello che operativamente cerca di fare la macchina è quindi di riprodurre, operando in modalità autonoma, le correlazioni statistiche che la macchina vede, forzata a guardare il dataset.

Questo schema, sommariamente descritto, è chiamato *contrastive divergence* [14] ed è alla base di numerosi algoritmi di *machine learning*: a prima vista, poco sembra avere a che fare con le reti neurali di Hopfield ed il loro apprendimento Hebbiano di ispirazione biologica, ma non è così. Se infatti calcoliamo esplicitamente la funzione di partizione della macchina di Boltzmann

[21] e.g., mediante la regola del gradiente, cioè, chiamato ϵ un -piccolo- parametro di learning, $\xi(t + \epsilon) = \xi(t) - \epsilon \nabla_\xi D_M(P|Q)$.

marginalizzando sui neuroni τ scriviamo

$$
\begin{aligned}
Z_B &= \sum_{\sigma}^{2^N} \sum_{\tau}^{2^P} e^{\frac{\beta}{\sqrt{N}} \sum_{i,\mu}^{N,P} \xi_i^\mu \sigma_i \tau_\mu} \\
&= \sum_{\sigma}^{2^N} \prod_{\mu=1}^{P} 2\, e^{\ln\cosh(\frac{\beta}{\sqrt{N}} \sum_i^N \xi_i^\mu \sigma_i)} \\
&\sim \sum_{\sigma}^{2^N} e^{\frac{\beta}{2N} \sum_{i,j}^{N,N} \sum_\mu^P \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j} = Z_H . \quad (12)
\end{aligned}
$$

Come le due funzioni di partizione, quella della macchina di Boltzmann e quella della rete di Hopfield coincidono[22], così, con un semplice argomento Bayesiano si può infatti mostrare che per ogni dataset generato da un pattern, la macchina di Boltzmann fa evolvere i suoi pesi in maniera tale che, una volta espressa in termini duali di rete Hebbiana, la matrice sinaptica inferita abbia come ingresso generico J_{ij} proprio $J_{ij} = \hat{\xi}_i^\mu \hat{\xi}_j^\mu$, dove i cappucci rappresentano le medie campionarie (stimatori eccellenti) dei pixel dell'immagine contenuta nel dataset relativo al pattern μ: almeno all'interno di uno scenario elementare, alle prese con datasets e patterns randomici, c'è completa armonia nell'investigazione teorica tra reti neurali ispirate dalla biologia e reti neurali artificiali di impiego nel machine learning [16, 17].

Alcune considerazioni conclusive

Prima del cosiddetto *winter time*, venivano prodotti modelli matematici per i neuroni sia d'ausilio ai neurofisiologi per la comprensione dell'intelligenza biologica, sia come processori sintetici di informazione, alla volta dell'IA. Il modello di Hopfield per il retrieval (e la duale macchina di Boltzmann per il learning) discusso in questo articolo è nato con la fine del *winter time*: le critiche al perceptrone di Rosenblatt erano in realtà critiche all'analisi di singolo neurone ed hanno saggiamente spostato l'attenzione degli Scienziati dai neuroni alle reti di neuroni.

[22] E quindi anche la macchina di Boltzmann gode del diagramma di fase di Hopfield (si veda la figura 8), dove in questo caso $\alpha = P/N$ rappresenta il rapporto tra la grandezza dello strato nascosto e quella dello strato libero).

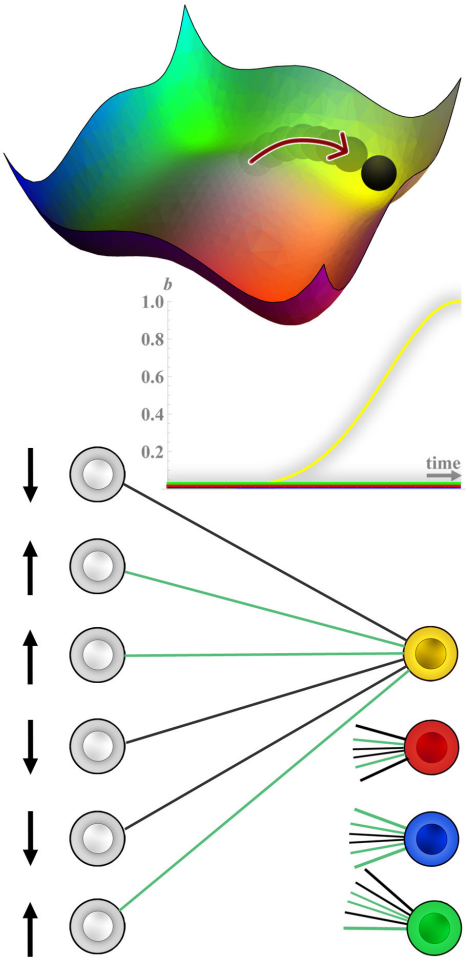

Figura 7: *Rappresentazione schematica di una macchina per il riconoscimento del colore: questa macchina di Boltzmann giocattolo è stata addestrata con la* contrastive divergence *a discriminare tra quattro colori. Di conseguenza, il panorama di energia (libera) della sua corrispettiva rete duale di Hopfield presenta quattro minimi (uno per ogni colore imparato). Se un colore viene in seguito presentato allo strato visibile della macchina (per esempio il* giallo*, rappresentato in un alfabeto binario fatto di spin sulla sinistra), i pesi che connettono questo strato a quello nascosto (se la macchina è stata addestrata con successo) sono tali da supplire al neurone nascosto la cui attivazione è responsabile dell'identificazione del giallo il campo massimo che lo strato visibile può produrre, lasciando i campi volti agli altri tre neuroni nascosti a valori minuscoli rispetto al primo. Intuitivamente si capisce anche perché ci sia un'α massimo in queste reti associative (si vedano le linee critiche nel diagramma di fase di Figura 8): chiaramente, tenendo fissato N, all'aumentare di P il numero di minimi in questo paesaggio deve continuare ad aumentare, ma in ultima istanza questi inizieranno a mischiarsi e fondersi uno nell'altro, rendendo il riconoscimento vacuo.*

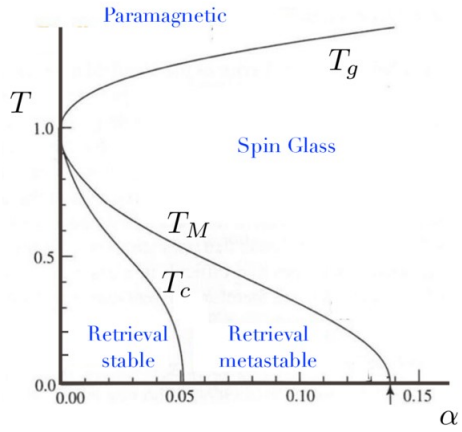

Figura 8: *Diagramma di fase della rete di Hopfield. Nel piano T, α – dove T rappresenta il rumore nella rete mentre $\alpha = P/N$ il suo carico (cioè il rapporto tra il numero di patterns che la rete deve gestire ed il numero di neuroni di cui essa e' composta) – troviamo diverse regioni, cioè diversi comportamenti della rete. Ad alte temperature, dove i neuroni non possono percepirsi reciprocamente, il sistema si comporta come un gas di spin, cioè un paramagnete ergodico. Abbassando la temperatura, partendo da carico nullo, la rete funziona bene come memoria associativa ed è in grado di riconoscere e distinguere tutti i P patterns fino alla prima linea critica, che sancisce il confine dello* stable retrieval*, dove si ha una transizione di fase e la rete entra in un nuovo regime dove i patterns sono ancora richiamabili, ma minimi spuri iniziano a dominare il paesaggio. Questo avviene in tutta la regione* retrieval metastable*, oltre la quale, per α ancora maggiori, la rete perde le sue capacità di pattern recognition e rimane uno spin glass senza proprietà di memoria.*

In questa era *post winter* nella quale viviamo circondati sempre più dall'IA, anche grazie alla dualità tra reti di Hopfield (archetipi dell'apprendimento biologico) e macchine di Boltzmann (oscillatori armonici dell'apprendimento artificiale) discussa in questo scritto, queste due branche della Scienza, che oggi si potrebbero chiamare Neurobiologia ed IA, si reincontrano (lavorando in connubio proprio come avvenne il secolo scorso alle prese con la dinamica di singolo neurone), ma a livello piu alto: questa volta sono le reti di neuroni ad interessare, le interazioni sopra i soggetti. Infatti, uno degli aspetti più significativi del modello di Hopfield è stato quello di spostare il focus dal singolo neurone alla rete [18]: la silente rivoluzione di pensiero per la quale si sposta l'attenzione dal singolo, quale che esso sia[23], alle interazioni tra i singoli, le reti che questi formano, sta introducendo nell'intera Comunità Scientifica una nuova prospettiva con cui vedere la Scienza a tutto tondo[24].

Parimenti significativo, a nostro avviso, sempre nella formulazione di Hopfield delle reti neurali, è il concetto cardine di *diagramma di fase*, importato dalla Meccanica Statistica grazie agli sforzi di Amit, Gutfreund e Sompolinsky[25]: uno sguardo alla Figura 8 ci mostra che esiste una ben definita regione, nel piano rumore $T := 1/\beta$, carico della rete $\alpha = P/N$ (definito come il rapporto tra il numero di patterns che la rete deve gestire ed il numero di neuroni a disposizione), dove la rete funziona (cioè esegue spontaneamente e correttamente pattern recognition) chiamata *Retrieval Stable*, mentre fuori da quella regione la rete si comporta in maniera diversa: nella regione *Retrieval metastable* funziona con un certo margine di errore che dipendende significativamente dall'inizializzazione della rete; per livelli di rumore proibitivi (limite di alta T) la rete diventa un inutile paramagnete ergodico e per carichi troppo grandi (limite di alto α) diventa uno spin glass senza proprietà di richiamo. La conoscenza del diagramma di fase è di un'importanza cruciale per progettare una rete: per esempio, in questa formulazione con i patterns casuali, è inutile farle immagazzinare un numero di patterns uguale alla metà del numero di neuroni (cioè farla lavorare ad $\alpha = 0.5$) poiché, per quel valore di carico P/N, la rete non può riuscire ne ad imparare ne quindi a fare riconoscimento: è nostro credo

[23] si pensi nell'evoluzione della nostra cultura a quanto il *soggetto* sia stato centrale (e.g., il passaggio dall'antropocentrismo all'eliocentrismo).

[24] Si studiano ad oggi reti ovunque: reti sociali, reti di proteine, reti immunitarie, reti geniche, reti di telecomunicazioni e trasporti, etc. [19]

[25] Il lettore potrebbe questionare sull'impiego della misura di Gibbs (che sappiamo valere per la meccanica statistica canonica all'equilibrio) anche in un regime di stato stazionario fuori dall'equilibrio (si osservi che una configurazione di minimo del modello di Hopfield implichi, nel caso di pattern random in esame, che metà dei neuroni sia "on" e metà sia "off" e pertanto si verifichi un circolo di corrente stazionaria nella rete (un *treno di spikes*) fintanto che una $m^\mu \sim O(1)$). La prospettiva con cui Jaynes legge la meccanica statistica dovrebbe essere la chiave di lettura con cui vedere anche l'analisi meccanico statistica del modello di Hopfield (si veda a tal proposito la *lezione mancata* ed il contributo di Michele Castellana dedicato all'uso dell'inferenza di massima entropia).

che una conoscenza opportuna della meccanica statistica dei sistemi complessi possa offrire una chiave di lettura per governare il dilagare dell'IA nelle prossime decadi[26].

Volendo forzare la mano, ci sono due facetamente provocatorie riflessioni (che ben si prestano ad interpretare alcuni comportamenti bizzarri della nostra società): se si crede che queste reti Hebbiane possano effettivamente essere dei ragionevoli modelli di memoria associativa per i moduli corticali del cervello [11], il fatto che questi sistemi possano imparare qualunque cosa, compresi patterns composti unicamente da rumore bianco, fornisce uno spunto di riflessione. L'altro è che, se si aumenta l'esposizione di informazione alla rete (ci si sposta in α significativamente verso destra nel digramma di fase di Figura 8) la rete smette di funzionare opportunamente e confonde qualunque cosa, cosa che potrebbe fornire un'intuitiva spiegazione sull'apparente correlazione temporale tra l'avvento di internet e della globalizzazione (che ci hanno sovraesposti all'informazione) e la genesi dei complottisti...

Per chiudere, osserviamo che le reti neurali, in quanto particolari *reti elettriche*, sono state le prime reti biologiche ad essere state studiate a nostro avviso perchè nella prima metà del secolo scorso si aveva una conoscenza soddisfacente della Fisica classica (in particolare le equazioni di Maxwell ne sugellavano il trionfo e supplivano a perfezione le necessità della Fisiologia) mentre Biochimica e Genetica hanno avuto i loro primi trionfi nella seconda metà del secolo (e probabilmente molti ne avranno nel presente e nei prossimi): ad oggi una fetta della comunità scientifica (alla quale anche gli autori del presente scritto divulgativo appartengono [20]) è dedita allo scovare meccanismi Hebbiani, in ultima istanza forme di cognizione, anche in reti biochimiche (ed a diverse le scale): quando il sistema immunitario intraprende una specifica risposta non riconosce forse l'antigene di qualche patogeno che vuole usarci? E quando, dopo averlo eliminato, evita che ci re-infettiamo (proteggendoci con dei linfociti ad uopo) non sviluppa forse memoria?

Mentre le risposte *si* sono ovvie in entrambi i casi, rendersi conto che anche questi lo faccia in maniera Hebbiana [21] è sorprendente e suggerisce una sorta di universalità di processazione d'informazione, sulla quale c'è ancora molto lavoro ancora da fare e che potrebbe portare ad una nuova auspicabile intersezione tra l'Intelligenza Artificiale e la Biocomplessità: nella prossima medicina di precisione si sta tentando per certi aspetti una prima sintesi di questo connubio.

Infine, come ulteriore approfondimento rimandiamo con piacere al volume speciale che il *Journal of Physics A: mathematical & theoretical* ha dedicato quest'anno al tema *statistical physics & machine learning* [22].

[26] Si veda, a tal proposito, il contributo di Jean Barbier nel presente volume.

[1] M. Mezard, G. Parisi, M.A. Virasoro: *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, World Scientific Publishing Company, Singapore (1987).

[2] J.J. Hopfield: *Neural networks and physical systems with emergent collective computational abilities*, Proceedings of the National Academy of Sciences, 79 (1982) 2554.

[3] G. Parisi: *Infinite number of order parameters for spin-glasses*, Physical Review Letters 43 (1979) 1754.

[4] D.J. Amit, H. Gutfreund, H. Sompolinsky: *Storing infinite numbers of patterns in a spin-glass model of neural networks*, Physical Review Letters 55 (1985) 1530.

[5] D.J. Amit: *Modeling brain function: the world of attractor neural networks*, Cambridge University press, Cambridge (UK) (1992).

[6] A.C.C. Coolen, R. Kuhen, P. Sollich: *Theory of neural information processing systems*, Oxford University Press, Oxford (2005).

[7] L. Zdeborova: *Understanding deep learning is also a job for physicists*, 16 Nature Physics (2020) 1.

[8] H.S. Seung, H. Sompolinsky, N. Tishby: *Statistical mechanics of learning from examples*, Physical review A 45 (1992) 6056.

[9] Y. LeCun, Y. Bengio, G.E. Hinton: *Deep learning*, Nature 521 (2015) 436.

[10] D. Hebb: *The organization of behavior: a neuropsychological theory*, Psychology Press, Road Hove (UK) (2005).

[11] E. Schneidman, et al.: *Weak pairwise correlations imply strongly correlated network states in a neural population*, Nature 440 (2006) 1007.

[12] A. Barra, G. Genovese, P. Sollich, D. Tantari: *Phase transitions in Restricted Boltzmann Machines with generic priors*, Physical Review E 96 (2017) 042156.

[13] S. Kirkpatrick, M. Vecchi: *Optimization by simulated annealing*, Science 220 (1983) 671.

[14] D.H. Ackley, G.E. Hinton, T.J. Sejnowski: *A learning algorithm for Boltzmann machines*, Cognitive Science 9 (1985) 147.

[15] R. Salakhutdinov, G.E. Hinton, *Deep Boltzmann machines*, Proc. International Conference on Artificial Intelligence and Statistics (AISTATS) Clearwater Beach, Florida, USA. Volume 5 of JMLR:W&CP 5.(2009).

[16] A. Barra, A. Bernacchia, E. Santucci, P. Contucci: *On the equivalence of Hopfield networks and Boltzmann machines*, Neural Networks 34 (2012) 1.

[17] E. Agliari, A. Annibale, A. Barra, A.C.C. Coolen, T. Tantari: *Multitasking associative networks*, Physical Review Letters 109 (2012) 268101.

[18] M.E.J. Newman: *The structure and function of complex networks*, SIAM review 45 (2003) 167.

[19] G. Caldarelli: *Scale-free networks: complex webs in nature and technology*, Oxford University Press, Oxford (2007).

[20] E. Agliari, A. Barra, L. Dello Schiavo, A. Moro: *Complete integrability of information processing by biochemical reactions*, Scientific Reports 6 (2016) 1.

[21] E. Agliari, A. Annibale, A. Barra, A.C.C. Coolen, D. Tantari: *Retrieving infinite numbers of patterns in a spin-glass model of immune networks*, Europhysics Letters 117 (2017) 28003.

[22] E. Agliari, A. Barra, P. Sollich, L. Zdeborova (Editors): *Machine Learning and Statistical Physics: Theory, Inspiration, Application*, Journal of Physics A: Special Issue (2020).

Elena Agliari: è ricercatrice in Fisica Matematica presso Sapienza Università di Roma, dove insegna -tra i vari- *Modelli di Reti Neurali*. Si occupa principalmente di Meccanica Statistica dei Sistemi Complessi, Teoria dei Grafi e Processi Stocastici, con particolare attenzione alle loro applicazioni nella Biologia e nell'Intelligenza Artificiale.

Adriano Barra: è professore associato in Fisica Matematica presso l'Università del Salento, dove insegna -tra gli altri- *Metodi Matematici per l'Intelligenza Artificiale*. Si occupa di principalmente di Meccanica Statistica dei Sistemi Complessi, Teoria dei Grafi e Processi Stocastici, con particolare attenzione alle loro applicazioni nella Biologia e nell'Intelligenza Artificiale.