

SOCIOLOGIA E RICERCA SOCIALE

Rivista fondata da Gianni Statera

anno XXIV, n. 70, 2003 nuova serie

Comitato scientifico:

A. Abruzzese (Roma I), S.S. Acquaviva (Padova), G.D. Amendola (Bari), A. Baratta (Saarbrücken), F. Barbano (Torino), G. Barbero (Roma I), G. Bechelloni (Firenze), L. Benadusi (Roma I), E. Campelli (Roma I), A. Cavazzani (Cosenza), M. Corsale (Roma I), R. De Giorgi (Lecce), A. de Lillo (Milano), P. De Nardis (Roma I), M. Fedele (Roma I), G. Genaro (Catania), F. Leonardi (Catania), M. Livolsi (Milano Iulm), P. Marconi (Roma I), C. Marletti (Torino), A. Marradi (Firenze), G. Marselli (Napoli), F. Martinelli (Roma I), U. Melotti (Roma I), M. Morcellini (Roma I), M. Negrotti (Urbino), G. Pagliano (Roma III), A. Palazzo (Pisa), G.P. Prandstraller (Bologna), M. Santuccio (Padova), G. Sertorio (Torino), G. Sivini (Cosenza), M. Toscano (Pisa).

Direttore: Enzo Campelli

Redazione:

M.S. Agnoli, L. Ciampi, A. Fasanella, A. Marinelli, C. Lombardo, S. Mauceri e S. Tusini (segreteria di redazione), S. Vergati.

Direzione e redazione: Via Savoia, 80 - 00198 Roma

Amministrazione e distribuzione: v.le Monza, 106 - 20127 Milano - tel. (02)2837141 - Casella postale 17175 - 20100 Milano

Abbonamento 2003: Italia € 46,50, Estero € 73,00; da versare sul ccp 17562208 intestato a FrancoAngeli s.r.l., Milano

Per i soci Ais - Associazione italiana di sociologia in regola con le quote 2002: € 44,00

SOCIOLOGIA E RICERCA SOCIALE USUFRUISCE DI UN CONTRIBUTO CNR

Testata iscritta nel registro della stampa ex art. 11 L. 416/81, al n. 2048 del 12.1.1987 (PCM, Div. X Editoria)

Autorizzazione del Tribunale di Milano n. 762 del 3.12.1999 - Direttore responsabile: Alda D'Eusanio - Quadrimestrale - Sped. in abbonamento postale - 45% - art. 2, comma 26/b, Legge 662/96 - Filiale di Milano - Copyright 2002 © by FrancoAngeli s.r.l. - Stampa: Tipomozza, Milano

I quadrimestre 2003 - finito di stampare nel settembre 2003

Sommario

Saggi

- M.A. Fabiano, *La sociologia di Alfredo Niceforo. Le trasformazioni del positivismo italiano* pag. 5

Strumenti e ricerche

- C. Corposanto, G. Manzo, *Disuguaglianze educative e loro trasformazioni nel tempo: problemi tecnici e metodologici di un'analisi compiuta mediante reti neurali artificiali* » 43
- S. Calandi, *Il campionamento: analisi del concetto di rappresentatività* » 70
- A.N. Feldman, *L'uso dell'Analisi Statistica dei Dati Testuali in una ricerca sistematica* » 96
- G. Franchi, *«Vergini si diventa e non si nasce»: l'infibulazione, tra differenza culturale e violazione dei diritti umani* » 112

Note, rassegne e dibattiti

- P. Cavara, *Destra e sinistra: sulla riproposizione critica delle categorie politiche moderne* » 149
- F. Pezzulli, *Una nuova emigrazione?* » 160

Strumenti e ricerche

*Disuguaglianze educative e loro trasformazioni nel tempo: problemi tecnici e metodologici di un'analisi compiuta mediante reti neurali artificiali**

di Cleto Corposanto** e Gianluca Manzo***

1. Oggetto e contesto problematico dell'analisi

Le reti neurali artificiali possono essere proficuamente utilizzate per studiare l'evoluzione temporale delle disuguaglianze di opportunità educative? Qual è il grado di affidabilità e la portata esplicativa dei risultati cui esse conducono? Quali i problemi tecnici che le caratterizzano se applicate a questo specifico aspetto del più vasto oggetto della stratificazione sociale?

Il presente articolo si pone l'obiettivo di rispondere a questo tipo di interrogativi osservando il comportamento di uno specifico modello neurale artificiale - il «perceptrone multistrato o multilivello» - all'opera su alcuni dati disponibili per il nostro paese.

Un tale interesse conoscitivo s'inserisce in un dibattito tecnico-metodologico che ha attraversato la sociologia della stratificazione sociale da oltre un ventennio: lo studio dei cambiamenti temporali delle disuguaglianze educative, infatti, ha costantemente posto specifici e complessi problemi di analisi statistica.

Gli anni '60 e, soprattutto, '70 furono monopolizzati dall'utilizzo della regressione lineare nella forma articolata dei modelli path-analitici. Nel quadro della *status attainment research*, a seguito dello studio capitale di Blau e Dun-

* Il presente articolo rappresenta il risultato di una riflessione comunemente condotta dai due autori che ha preso spunto dalle ricerche da tempo portate avanti dal primo sui temi della classificazione nelle scienze sociali - si veda, in particolare, Corposanto, 2001 - e dal lavoro di tesi (*Disuguaglianza sociale, opportunità educative e reti neurali artificiali. Teoria e metodo a fondamento di nuovi modelli di analisi dei dati*) discusso dal secondo nel settembre 2001 presso la Facoltà di sociologia dell'Università di Trento. Sebbene quanto segue derivi dal nostro sforzo comune, per motivi organizzativi i paragrafi 1, 2, 5 e l'appendice statistica si devono a Gianluca Manzo mentre i paragrafi 3 e 4 a Cleto Corposanto.

** Professore associato presso il Dipartimento di scienze umane e sociali della Facoltà di sociologia dell'Università di Trento; cleto.corposanto@soc.unitn.it.

*** Dottorando in Epistemologie e metodologia delle scienze sociali presso le Università Paris IV-Sorbonne e Trento; mgianluca@wanadoo.fr.

can (1967), analizzare l'evoluzione temporale della struttura della stratificazione sociale rinviava a comparare i parametri di un modello di regressione stimato separatamente per gruppi di età. Alcuni studi ormai classici - Duncan (1967), Treiman (1970), Duncan, Featherman e Duncan (1972), Jencks e collaboratori (1972), Featherman e Hauser (1976), Halsey (1977) - forniscono un chiaro esempio di quest'impostazione tecnica.

Il bilancio che col trascorrere degli anni e delle ricerche si poté trarre sull'andamento temporale delle asimmetrie educative non fu particolarmente incoraggiante: alla fine degli anni '70, il solo elemento sul quale si sarebbe potuto ragionevolmente convenire consisteva nel riconoscere il carattere estremamente contraddittorio dei risultati empirici.

Si deve a Robert Mare il merito di avere spiegato formalmente la natura di queste contraddizioni. Lo studioso americano, in due articoli successivi pubblicati agli inizi degli anni '80 - «Social background and school continuation» (1980); «Change and stability in educational stratification» (1981) -, mostra che i modelli di regressione lineare sono uno strumento statisticamente inadeguato per studiare l'andamento temporale delle disuguaglianze educative. Due dimensioni analitiche fondamentali devono essere distinte a questo proposito: da un lato, la distribuzione dei titoli di studio e la loro struttura in termini di dispersione, dall'altro, l'allocazione dei singoli in questa struttura ed i principi o i meccanismi che la regolano (1981, pp. 73-4). Mare mostra matematicamente che i parametri dei modelli di regressione confondono sistematicamente queste due fonti di cambiamento sottostanti l'associazione tra origine sociale ed esito educativo (*ibid.*, pp. 75-76, 78). L'instabilità dei risultati deriva dunque dall'incapacità tecnica della regressione di scorporare dal legame tra le due variabili gli effetti derivanti dalla loro struttura o dispersione. Lo studioso americano dimostra formalmente che il principale merito della regressione logistica è esattamente quello di controllare gli effetti delle distribuzioni marginali e di restituire il valore dell'associazione tra origine e titolo di studio al netto di questi effetti (1980, p. 297; 1981, p. 74).

Come osservato da Blossfeld e Shavit (1993, p. 4), il contributo di Robert Mare ha costituito una specie di rivoluzione concettuale e metodologica nello studio degli andamenti temporali delle disuguaglianze di opportunità educative.

Attualmente, infatti, la maggior parte delle ricerche empiriche condotte su questo fenomeno si avvalgono di una strumentazione statistica basata sulla regressione logistica e sulla modellistica log-lineare (Benedusi, 1993; Blossfeld, 1993; Blossfeld e Shavit, 1993; Cobalti, 1990; Cobalti, 1992; Cobalti e Schizzerotto, 1993; Cobalti e Schizzerotto, 1994; Garnier e Raffalovich, 1984; Jonsson, 1993; Jonsson e Mills, 1993a e 1993b; Jonsson e Erikson, 2000; Kerckhoff e Trott, 1993; Muller e Karle, 1993; Schizzerotto, 1997; Shavit e Westerbeek, 1997; Vallet e Thélot, 2000).

Ciononostante è ugualmente evidente che le acquisizioni tecnico-metodologiche derivanti dai lavori di Mare non sono state ancora in grado di convincere l'intera comunità scientifica tanto da omogeneizzarla attorno ad una pratica di ricerca standardizzata.

Ne sono dimostrazione, in primo luogo, importanti ricerche sugli andamen-

ti temporali della stratificazione sociale e educativa - per es., Vrooman e Dronkers (1986); Treiman e Yip (1989) - che continuano ad operare attraverso versioni più o meno articolate e complesse della regressione lineare multipla. Qualche autore ha più generalmente osservato che vi sono alcuni segni di una ripresa di vigore degli studi condotti in termini di «path models of stratification» (Grusky, 2001, p. 14449).

Particolarmente significativi, inoltre, ci sembrano alcuni elementi della sociologia francese della stratificazione. Essa, senza dubbio profondamente influenzata dalle analisi realizzate da Raymond Boudon, in particolare, nel lavoro oramai classico *L'inégalité des chances* (1973), sembra integrare con difficoltà la svolta realizzatasi nella comunità scientifica internazionale a seguito dei lavori di Mare.

Prova ne è il dibattito tecnico-metodologico - Barbut (1984 e 1985); Combesse (1984 e 1985); Florens (1984); Gremy (1984); Merlié (1985); Prévot (1985) - svoltosi sulle pagine della *Revue française de sociologie* negli anni 1984 e 1985. I numerosi articoli che, partendo dai problemi tecnici sollevati dal lavoro di Boudon citato poco sopra, cercarono di fare chiarezza sulla possibilità di studiare i cambiamenti temporali delle asimmetrie educative attraverso misure basate su semplici proporzioni, infatti, sembrano completamente ignorare che gli articoli di Robert Mare, in particolare quello del 1981¹, avevano già dato una risposta definitiva alla questione. Inoltre, è interessante notare che la più importante ricerca comparativa organizzata a livello internazionale sul fenomeno delle disuguaglianze educative e su suoi cambiamenti nel tempo (Blossfeld e Shavit, 1993), la quale si dà esplicitamente come quadro tecnico-metodologico comune quello ispirato dai lavori di Mare, non contempla la Francia. Infine, vi sono giovani ricercatori (per es., Bulle, 2002) che dedicano parte delle loro energie intellettuali a denunciare le insufficienze per gli studi di stratificazione della misura di associazione - gli *odds-ratios* - su cui si fondano matematicamente tanto la regressione logistica che la modellistica log-lineare².

La complessità tecnico-metodologica della questione è dunque evidente.

Dal nostro punto di vista, non si tratta tuttavia di rimettere in discussione le acquisizioni derivanti dai lavori di Robert Mare: è, infatti, nostra convinzione che quella parte della tecnica statistica fondata sugli *odds-ratios* sia la scelta adeguata allorché si voglia indagare l'evoluzione temporale delle distanze relative tra i gruppi sociali negli esiti scolastici³.

Il nostro obiettivo dunque non è quello di cercare di stabilire l'adeguatezza

1. Si tenga presente che in questo articolo Mare propone una critica esplicita e formale all'impostazione tecnica di Boudon (1981, pp. 74-6), mostrando come essa sia del tutto assimilabile ad un particolare tipo di modello di regressione lineare (modello lineare di probabilità). In questo modo, Mare spiega efficacemente talune contraddizioni innegabili esistenti nei risultati empirici del sociologo francese.

2. Vi sono ovviamente alcune eccezioni a quello che potremmo definire il «caso francese»: i lavori recenti di Vallet (1999 e 2000) ne rappresentano l'esempio più evidente.

3. Teniamo a precisare che questo non significa minimamente che non vi siano alcuni aspetti delle proposizioni dello studioso americano suscettibili di essere migliorate: da questo punto di vista, ci sentiamo molto vicini ad alcuni recenti lavori come quello di Breen e Jonsson, 2000.

dei modelli neurali per lo studio di questo specifico aspetto della stratificazione sociale al fine di sostituirli all'analisi log-lineare: più semplicemente, cercheremo di vedere se essi sono ugualmente utilizzabili. Ciò implicherà principalmente tentare di capire se essi sono idonei ad indagare il legame tra origine e titolo di studio al netto degli effetti che la struttura delle variabili indipendenti e dipendenti può avere sull'associazione in quanto tale.

2. Reti neurali artificiali e perceptrone multilivello⁴

Le reti neurali artificiali possono essere definite come una famiglia di modelli computazionali non lineari la cui struttura matematica è nata dall'idea di riprodurre la configurazione ed il funzionamento del sistema nervoso umano. Tre elementi essenziali sono presenti in ogni modello neurale: un certo numero di «nodi o neuronodi»; un certo numero di «connessioni o pesi»; un processo iterativo di modificazione del valore di quest'ultimi, detto «processo di apprendimento». Da un punto di vista matematico, i «nodi» non sono nient'altro che una funzione composta: i segnali di ingresso sono sintetizzati in essi sotto forma additiva e ponderata; a questa somma pesata, essi applicano una funzione matematica, detta «funzione di attivazione», che produce un valore responsabile dell'attivazione o della non attivazione del nodo. Le «connessioni» sono concretamente semplici parametri numerici che possono assumere valori positivi o negativi: la conoscenza di una rete - l'informazione che essa estrae dai dati che analizza - è incorporata nella struttura valoriale dei pesi. Il «processo di addestramento» è un algoritmo iterativo che modifica progressivamente tale struttura: l'obiettivo ultimo di questo processo è quello di dare alla rete la capacità di generalizzare a dati sconosciuti le informazioni estratte dai dati esaminati.

Secondo la configurazione assunta dai nodi, la struttura delle connessioni e, soprattutto, il tipo d'algoritmo d'apprendimento implementato, si possono avere innumerevoli topologie neurali. In particolare, con riferimento a quest'ultimo elemento, si è soliti operare una generale distinzione tra «reti ad ap-

4. Per una discussione a diversi livelli di approfondimento e di specificità delle reti neurali artificiali, si possono utilmente consultare i seguenti lavori: Bellacicco, 1997; Bellacicco e Lauro, 1997; Breda, 1999; Breda e Buscema, 1999; Buscema, 1999a, voll. 1-2 e 1999b; Carrella, 1995; Cheng e Titterington, 1994; Corposanto, 2001; Davino, Gerghi e Vistocco, 1997; Davino, Mola e Siciliano, 1997; Davino e Vistocco, 1997; Ferrari, 1996; Floreano, 1996; Guizzardi, 1997; Maarit e Hakkinen, 2000; Massimi, 1999; Matera, 1999; Paik, 2000a e 2000b; Parisi, 1989 e 1999; Patarnello, 1992; Pizzuti, 1999; Ripley, 1993; Rumelhart *et al.*, 1986; Schiavo e Lisi, 1997; Schumacher, Rossener e Vach, 1996; Stone, 1986; Vach *et al.*, 1996; Vicino, 1999; Zeng, 1999; Woelfel, 1993. Rispetto all'estensione della letteratura tecnica sulla modellistica neurale, gli studi volti ad applicarle in campo sociologico sono estremamente rari: nel più generale ambito delle scienze sociali, eccettuato un promettente filone in campo economico-finanziario, le applicazioni più frequenti sembrano essere quelle di natura politologia (Zeng, 1999 e bibliografia ivi citata: nota 3, p. 521; Woelfel, 1993). Per quanto riguarda l'argomento specifico affrontato in questo articolo, a nostra conoscenza, si tratta del primo tentativo in tal senso; più in generale, siamo riusciti a trovare solo due articoli volti ad utilizzare le reti neurali artificiali per studiare le disegualianze sociali (Meraviglia, 1996 e 1999; in senso molto più largo, Carbone e Piras, 1999).

prendimento supervisionato» e «reti ad apprendimento non supervisionato o per auto-organizzazione». Con il primo termine si indicano quei modelli in cui la risposta corretta viene fornita esternamente alla rete ed il processo di mutamento dei valori sinaptici si realizza tramite un progressivo confronto tra questa risposta corretta e la risposta fornita dalla rete con l'obiettivo ultimo di minimizzare una qualche misura di errore. I «modelli auto-organizzati», invece, si caratterizzano per la totale assenza di risposte target fornite esternamente e realizzano l'apprendimento unicamente in conformità a qualche regola di plasticità sinaptica osservando autonomamente le regolarità presenti nei dati.

Lo specifico tipo di rete che utilizzeremo nel presente articolo fa parte della prima classe di modelli, quelli ad apprendimento supervisionato: segnatamente, si tratta della topologia neurale conosciuta sotto il nome di «perceptrone multi-strato o multi-livello».

Morfologicamente, essa si caratterizza per la presenza di tre classi funzionali di nodi: uno strato di nodi di *inputs*, i quali ricevono l'informazione dall'esterno e la trasmettono alla rete; uno strato di unità di *outputs*, il quale fornisce il risultato della rete; uno (o più) strato di nodi nascosti - detti «unità hidden» - che riceve segnali dalle unità d'ingresso e li trasmette a quelle d'uscita. Questo strato intermedio è particolarmente importante: è in esso, infatti, che viene immagazzinata la conoscenza della rete.

Per quanto riguarda la struttura delle connessioni, essa è di tipo monodirezionato: ogni unità dello strato precedente è connessa con tutte le unità dello strato successivo senza che vi siano né legami tra unità del medesimo strato né connessioni tra strati non immediatamente successivi. L'informazione passa dunque dagli strati di *input* a quelli di *output* senza alcuna forma di circolarità o di ricorrenza: in questo senso, si parla spesso di «rete feed-forward».

Trattandosi dell'algoritmo d'apprendimento, il perceptrone multistrato utilizza una procedura - messa a punto verso la meta degli anni '80 (Rumelhart *et al.*, 1986, pp. 324-7, 352-61) - nota come «regola delta generalizzata» o «algoritmo di back-propagation». Essa consiste nella minimizzazione di una funzione di errore che calcola lo scarto tra la risposta data dalla rete e quella corretta. I pesi vengono aggiornati proporzionalmente alle derivate parziali di tale funzione secondo un processo «a ritroso» che muove cioè dagli strati inferiori a quelli superiori passando per gli strati intermedi. Questo meccanismo di propagazione all'indietro dell'errore costituisce uno degli elementi di maggiore forza del perceptrone multi-strato: fino alla sua apparizione, infatti, i modelli neurali non erano in grado di mutare il valore delle proprie connessioni se non per uno strato.

Sono principalmente due gli elementi matematici e computazionali che rendono questa topologia neurale estremamente attraente tanto in generale quanto più in particolare per l'analisi empirica che abbiamo intenzione di realizzare nel presente lavoro.

In primo luogo, si è dimostrato che un perceptrone multistrato con almeno uno strato di unità *hidden* possiede la proprietà di un «approssimatore universale» ovvero sia d'un operatore computazionale in grado di stimare qualunque tipo di funzione (Cheng *et al.*, 1994, p.20; Rumelhart *et al.*, 1986, p. 319; Vach *et al.*, 1996, p. 684; Paik, 2000a, pp. 430-1; Zeng, 1999, p. 501; Woelfel, 1993). In particolare, se le unità nascoste contengono una funzione di attiva-

zione non lineare, questo modello neurale si è rivelato estremamente sensibile alla scoperta di relazioni non lineari tra le variabili in input (Buscema, 1999a, p. 61; Floreano, 1996, p. 79; Paik, 2000a, pp. 430, 432; Woelfel, 1993). Alcuni autori al fine di sottolineare questa caratteristica parlano delle reti multistrato come «modelli non-lineari input-output» o, più semplicemente, «modelli o metodi non lineari» (Maarit e Hakkinen, 2000, p. 242; Paik, 2000b, p. 5; Zeng, 1999, pp. 500-1, 512).

Secondariamente, il perceptrone multilivello a *back-propagation* intrattiene una relazione di analogia particolarmente forte ed evidente con una parte importante della statistica tradizionale. Se, in generale, l'avviso secondo cui le reti neurali artificiali restano dei metodi statistici multivariati sembra diffondersi progressivamente (Maarit e Hakkinen, 2000, p. 242), le reti multistrato sono sempre più insistentemente associate e paragonate ai modelli di regressione - lineare o non, secondo il tipo di funzione di attivazione presente nei neuroni (Cheng *et al.*, 1994, p. 16; Stone, 1986, pp. 453 e ss.; Ripley, 1993, p. 59; Woelfel, 1993, p. 67). Più in particolare, poi, si sostiene che, laddove la funzione di attivazione delle unità nascoste assume una forma logistica, il perceptrone multistrato è completamente equivalente ad un modello di regressione logistica (Davino, Mola e Siciliano, 1997, p. 25; Paik, 2000a, p. 430 e 2000b, p. 5; Schumacher *et al.*, 1996; Zeng, 1999).

Si tratta di un elemento particolarmente apprezzabile per i nostri fini, nella misura in cui la maggior parte delle analisi contemporanee condotte sugli andamenti temporali delle disuguaglianze educative si basano su questa tecnica di analisi o sulla modellistica log-lineare. L'equivalenza tra reti multistrato ed analisi logistica, quindi, sembrerebbe attribuire alle prime le stesse «virtù statistiche» delle seconde quanto al problema - richiamato nel paragrafo introduttivo - degli effetti delle distribuzioni marginali sulla consistenza delle relazioni tra le variabili.

Queste potenzialità del perceptrone multistrato a *back-propagation* non devono tuttavia far dimenticare alcuni aspetti problematici che lo caratterizzano e che pongono specifiche difficoltà di natura operativa: nel quadro degli obiettivi del presente articolo, ci limitiamo ad esplicitarne due.

Primariamente, si tratta della definizione dell'architettura della rete: nella misura in cui i nodi di *input* e di *output* sono automaticamente definiti dalla struttura rispettivamente delle variabili indipendenti e dipendente, il problema si riduce alla costruzione di uno strato adeguato di nodi hidden. A questo riguardo, si incontra ciò che è stato talvolta definito il «dilemma della BP» (Buscema, 1999b, pp. 189-90): da un lato, un numero ridotto di unite intermedie impedisce alla rete di cogliere in profondità le relazioni esistenti nei dati, dall'altro, uno strato di nodi nascosti troppo articolato potrebbe appiattire la rete sui dati di addestramento riducendone la capacità di generalizzazione - problema detto di «overfitting» (Floreano, 1996, p. 92). Rispetto all'assenza in letteratura di una risposta definitiva e teoricamente fondata⁵, nella nostra ana-

5. Sebbene in generale si è portati a considerare il processo di selezione dell'architettura ottimale della rete come un processo per tentativi ed errori (Schiavo e Lisi, 1997, p. 113;

lisi adotteremo un'ottica esplorativa ed incrementale che ci porterà a costruire per ogni relazione indagata una pluralità di architetture neurali caratterizzate da una complessità crescente tanto per numero di strati intermedi che per numero di nodi all'interno di questi ultimi.

In seconda istanza, vi è una difficoltà strettamente legata a questo primo aspetto: di fronte ad una pluralità di strutture multistrato, come selezionare quella migliore? Altrimenti detto, quali criteri utilizzare per valutare la bontà della performance dei nostri modelli?

In ordine di importanza decrescente, adotteremo sistematicamente e congiuntamente i parametri decisionali seguenti. Prima di tutto si osserverà il comportamento classificatorio della rete nei termini della «correttezza della classificazione» che essa è in grado di fornire ovverosia della capacità di assegnare i casi ai gruppi cui questi appartengono realmente (Buscema, 1999a, vol. 1, pp. 82-3). Con il termine «classification rate» si indicherà una semplice misura che calcola il numero di casi correttamente classificati sul totale dei casi analizzati (Meraviglia, 1999, p. 206)⁶. Secondariamente, compareremo costantemente il classification rate ottenuto sui dati sconosciuti con quello esibito dalla rete sul set di addestramento: nel caso in cui sia presente un problema di «overfitting», quest'ultimo dovrebbe essere più elevato del secondo. A parità di comportamento classificatorio, pertanto, privilegeremo sempre la rete caratterizzata dal classification rate più basso in fase di addestramento. In terzo luogo, prenderemo in considerazione una misura di errore - Rmse, «root mean square error» (calcolata come la radice quadrata dell'errore quadratico medio) - volta a fornire un'informazione sulla distanza media intercorrente tra l'output della rete e quello campionario⁷. Infine, un ultimo criterio valutativo consisterà nel grado di complessità del modello: come orientamento generale, per ragioni di parsimonia e d'interpretabilità, a parità di comportamento classificatorio, si sceglierà sempre la rete caratterizzata dalla struttura più semplice.

Armata di questi criteri strettamente operativi, non rimane che osservare come le nostre reti multistrato a *back-propagation* con funzione di attivazione logistica nei nodi intermedi si comportano allorché calate nel contesto dei dati empirici.

Zeng, 1999, p. 522, nota 13), bisogna riconoscere che sono disponibili in letteratura un certo numero di indicazioni operative (Carrella, 1995, p. 208; Ferrari, 1996, p. 51; Floreano, 1996, pp. 93-4, 96; Ripley, 1993, p. 49; Zeng, 1999, p. 522, nota 13): noi utilizzeremo tali accorgimenti a mo' di guida generale. Si osservi, d'altro canto, che questa fase dell'analisi neurale è ragionevolmente assimilabile a ciò che nella statistica tradizionale - si pensi all'analisi log-lineare - è la fase di «selezione del modello» (Cheng *et al.*, 1994, pp. 4-5; Schumacher *et al.*, 1996, p. 670).

6. Nella misura in cui essa sarà calcolata sui casi che la rete non ha analizzato in fase di addestramento, questa misura rappresenta un'indicazione della sua capacità generalizzatrice. Per questa ragione, la si può ritenere una buona misura della performance della rete: «(...) un buon modello è quello che generalizza bene su dati nuovi», sostiene Zeng (1999, p. 508); «la previsione diventa il test principale, conclusivo nel rifiutare o meno una determinata specificazione», nota Guizzardi (1999, p. 54).

7. Consci del carattere approssimativo e problematico di questa misura (Buscema, 1999a, vol. 1, pp. 86-7), tuttavia, la utilizzeremo nel proseguo dell'analisi come criterio valutativo secondario e sussidiario.

3. Dati e variabili

Il contesto spazio-temporale dell'analisi empirica oggetto del presente articolo è rappresentato dal nostro paese durante i primi ottant'anni del ventesimo secolo.

I dati su cui si baseranno le elaborazioni seguenti sono quelli raccolti nel 1985 dall'indagine di mobilità italiana condotta da Barbagli, Capecchi, Cobalti, de Lillo e Schizzerotto, contenenti informazioni su 5016 soggetti di entrambi i generi (occupati e non) e di età compresa tra i 18 ed i 65 anni⁸. Questo campione, nazionalmente rappresentativo, è stato costruito mediante un'articolata procedura di campionamento a stadi⁹.

Noi utilizzeremo un sotto-campione di 4107 casi. Rispetto al campione originario, sono stati eliminati tutti coloro che al momento dell'intervista erano studenti o che avevano meno di 24 anni, il che permette di lavorare con soggetti che hanno tutti terminato la propria carriera educativa (la stessa operazione è realizzata da Cobalti e Schizzerotto, 1993). A causa delle specifiche caratteristiche tecniche della modellistica neurale, tale campione sarà sistematicamente articolato in tre insiemi di dati rigorosamente distinti: il *training set* (insieme di dati sul quale la rete compie il proprio addestramento) la cui numerosità ammonta sempre all'80% del campione originario; il *testing set* (insieme di dati per i quali la rete non conosce la risposta classificatoria e sui quali viene testata la capacità predittiva e generalizzativa della rete medesima), la cui numerosità ammonta sempre al 10% del campione originario; il *validation set* (insieme di dati sul quale viene monitorato il comportamento della rete mentre il processo di addestramento è ancora in corso¹⁰), la cui numerosità ammonta sempre al 10% del campione originario. Nell'assegnazione dei casi ai tre gruppi, si seguirà sistematicamente una procedura rando-

8. A parziale integrazione di questi dati e per un'esigenza specifica verranno utilizzati alcuni dati sul contesto svedese raccolti nel 1974: il campione che ci interesserà ammonta a 5828 casi. Tali dati sono stati tratti dal più ricco ed articolato *dataset* costituito nell'ambito del progetto Casmin diretto da Muller e Goldthorpe, progetto che ha dato origine ad importanti pubblicazioni come quelle di Erikson e Goldthorpe, 1992 e di Muller e Karle, 1993.

9. Siamo consci di esporci all'obiezione di utilizzare dei dati «obsoleti». Vi è ciononostante una buona ragione che giustifica questa scelta: l'obiettivo stesso della nostra analisi - valutazione della bontà di uno strumento di analisi dei dati eccentrico rispetto alla statistica tradizionale per lo studio di un fenomeno altamente complesso come l'andamento temporale delle disuguaglianze educative - ci ha consigliato di utilizzare dei dati la cui qualità fosse indubbia e sui quali fossero già state fatte numerose analisi. Il primo elemento consente di eliminare quante più possibili fonti di variazioni esterne che potrebbero influenzare il comportamento della modellistica neurale; il secondo fornisce un insieme coerente di risultati cui confrontare quelli ottenibili mediante le reti. Si tenga presente, in ogni caso, che alcune analisi condotte sui dati più recenti raccolti nel 1997 nell'ambito della prima ondata dell'«Indagine longitudinale sulle famiglie italiane» condotta dall'Università di Trento, forniscono un'immagine di sostanziale stabilità piuttosto che di cambiamento della struttura delle disuguaglianze sociali del nostro paese (Pisati e Schizzerotto, 1999a e 1999b).

10. Questo *dataset* è utilizzato dai quei software che utilizzano come procedura per limitare il problema dell'*overfitting* quella di interrompere la fase di addestramento allorché l'errore sul *validation set* incomincia ad aumentare.

mica automaticamente generata dal *software* così come la lettura dei casi operata dalla rete assumerà anch'essa una sequenza casuale¹¹.

Per quanto riguarda la misura delle variabili, la posizione sociale d'origine sarà operativizzata tanto in termini di classe sociale che di raggruppamenti distributivi. Dal primo punto di vista, si adopererà lo schema di classe ad otto posizioni - «imprenditori», «liberi professionisti», «dirigenti», «classe media impiegatizia», «piccola borghesia urbana», «piccola borghesia agricola», «classe operaia urbana», «classe operaia agricola» - lungamente discusso da Schizzerotto (1988 e 1993, pp. 63-9) e testato empiricamente in modo approfondito da Cobalti e Schizzerotto (1994). Quanto all'aspetto distributivo della stratificazione, utilizzeremo la scala di desiderabilità sociale delle occupazioni - articolata in un continuum unilineare che prende valori da 9.97 a 90.20 - costruita per il nostro paese da De Lillo e Schizzerotto (1985).

La posizione educativa dei rispondenti, invece, sarà operativizzata come il più alto titolo di studio formalmente conseguito, secondo le modalità «senza titolo o licenza elementare», «licenza di scuola media inferiore», «licenza di scuola media superiore», «laurea». Adopereremo tale schema di misurazione anche per il livello d'istruzione del padre e della madre del rispondente allorché quest'informazione sarà inserita tra le variabili indipendenti al fine di operativizzare le risorse culturali della famiglia di origine. A tal fine si utilizzerà spesso anche una misura dell'istruzione dei genitori in termini di numero di anni di scuola massimamente ottenuti, secondo una scala di valori che va da un minimo di 1 anno ad un massimo di 19.

Un'ultima considerazione prima di passare all'analisi empirica: nelle nostre intenzioni, questo schema di operativizzazione delle variabili si vorrebbe espressivo di un certo numero di convinzioni teoriche che, per limiti di spazio, ci limitiamo ad elencare.

In primo luogo, l'idea che la stratificazione sociale debba essere concettualizzata tanto nei suoi aspetti gerarchico-relazionali quanto nella sua articolazione distributiva: da qui la considerazione congiunta di uno schema di classe e di una scala di prestigio. Secondariamente, l'idea che la posizione sociale di

11. Il *software* di elaborazione dati utilizzato in quest'analisi è «Neural connection 2.0», prodotto e commercializzato dalla Spss Inc. (Spss Inc., 1997). Questo programma di implementazione delle reti neurali, parallelamente alla procedura del *validation set* per ridurre il problema dell'*overfitting*, consente anche di adottare un'ulteriore strategia nota come «save best net», in base alla quale ad ogni ciclo di addestramento viene utilizzata la matrice delle connessioni che ha fornito i risultati migliori. Tale accorgimento sarà adottato nella nostra analisi, seguendo una linea operativa diffusa in letteratura (Carboni e Piras, 1999; Meraviglia, 1999). Si consideri inoltre che: in merito al tempo di addestramento, tutte le reti saranno addestrate per 400 cicli; relativamente all'inizializzazione dei pesi, si seguirà pressoché sempre la seguente procedura: estrazione casuale dei valori da una distribuzione uniforme multivariata all'interno di un *range* pari a $-0.1/+0.1$; quanto ai valori del tasso di apprendimento e del *momentum* essi assumeranno, per tutte le reti, valori diversi a seconda degli *step* in cui si articola il processo di addestramento: il tasso di apprendimento assume una sequenza di valori pari a 0.9/0.6/0.4/0.2 mentre il *momentum* assumerà valori pari a 0.1/0.3/0.5/0.6. Questo tipo di combinazione risponde all'esigenza di costruire un processo di apprendimento progressivamente più sofisticato che eviti, al tempo stesso, i minimi locali della funzione di errore.

origine debba essere costantemente concettualizzata anche in relazione alle risorse di natura simbolica e culturale: da qui la presenza sistematica nei modelli che si costruiranno del livello educativo parentale. Da ultimo, l'idea secondo cui anche l'istruzione si compone tanto di un aspetto discreto, gerarchico e relazionale - aspetto credenzialistico del titolo di studio - che di una dimensione continua, lineare e distributiva - istruzione come investimento «capitalistico»: da qui l'inserimento di una variabile ordinale e di una variabile metrica per operativizzare la posizione educativa dei singoli.

4. L'analisi empirica mediante reti neurali artificiali

Qual è l'evidenza empirica attualmente disponibile per il nostro paese quanto alle evoluzioni temporali delle disuguaglianze di opportunità educative?

In estrema sintesi, considerando unicamente quegli studi che partendo dalle acquisizioni tecnico-metodologiche degli inizi degli anni ottanta si avvalgono dell'analisi logistica e log-lineare, si ha a che fare principalmente con due classi di risultati.

Da un lato, un certo numero di ricerche (Cobalti, 1990; Cobalti e Schizzerotto, 1993 e 1994; Schizzerotto, 1990) evidenziano una situazione di sostanziale stabilità delle asimmetrie educative tranne una riduzione localizzata alle classi agricole. Questo risultato sarebbe d'altro canto completamente coerente con quello messo in evidenza a livello internazionale dalla più importante ricerca comparativa realizzata agli inizi degli anni '90 (Blossfeld e Shavit, 1993, fatta eccezione per la Svezia e l'Olanda). Dall'altro lato, tuttavia, alcune ricerche - per ora ancora minoritarie - hanno individuato una tendenza perequativa consistente e generalizzata all'opera ai livelli inferiori del nostro sistema scolastico (Benedusi, 1993; Shavit e Westerbeek, 1997). Il che sarebbe consistente con un certo numero di lavori empirici condotti separatamente su diverse realtà nazionali - Svezia, Olanda, Inghilterra, Francia - volti a rivedere ed ad approfondire i risultati emersi dalla ricerca comparativa diretta da Shavit e Blossfeld (per es., Jonsson, 1993; Jonsson e Mills, 1993b, p. 421; Jonsson e Erikson, 2000, p. 370; Blossfeld e Shavit, 1993, pp. 16-17; Jonsson e Mills, 1993b; Vallet e Thélot, 2000, pp. 3, 19, 21, 22, 26-29, 31).

In che modo, dunque, le reti neurali artificiali si inseriscono in questo quadro di ricerca? Per rispondere a questo interrogativo, abbiamo costruito le seguenti tre coorti di età, seguendo la ripartizione adottata da alcune analisi condotte sugli stessi dati (Cobalti, 1990, p. 202; Cobalti e Schizzerotto, 1993, p. 161): nati tra il 1920 ed il 1933; tra il 1934 ed il 1947; tra il 1948-1961. Secondariamente, si è proceduto a stimare una serie di reti multistrato all'interno di ogni coorte: il comportamento classificatorio delle reti dovrebbe restare stabile tra le coorti nell'ipotesi di un'invarianza temporale degli effetti dell'origine sul destino educativo dei singoli; mutare nel senso di un peggioramento della percentuale di corrette classificazioni nel caso di un indebolimento del legame tra origine e titolo di studio.

Le tab. 1 riporta il miglior modello neurale stimato all'interno di ogni coor-

te¹²: l'ipotesi di una stabilità generalizzata delle disuguaglianze educative non sembra corroborata dalla reti neurali artificiali.

Tab. 1 - Var. ind.: classe sociale paterna e materna, prestigio occupazionale paterno, titolo studio paterno e materno, anni scuola paterni e materni; Var. dip.: titolo studio rispondente

Coorte	Struttura rete	Funz. Attiv.	% training	% testing	Rmse
20-33 (N: 1072)	27*2*4	sigmoide	75,73	77,57	0,278
34-47 (N: 1368)	27*10*8*4	sigmoide	58,80	59,52	0,363
48-61 (N: 608)	27*6*4	sigmoide	47,21	51,82	0,393
<i>Modello su intero campione</i>					
(N: 3695)	27*7*4	sigmoide	56,39	57,84	0,363

Il comportamento delle reti muta sistematicamente in connessione con il gruppo di età considerato, evidenziando un effetto d'interazione in senso equitativo tra origine e coorte. Nella coorte più anziana la rete permette di classificare correttamente il destino educativo del 77,57% - con un incremento dunque di circa venti punti percentuali rispetto all'intero campione non strutturato per età. Questo valore scende poi a 59,52% nella coorte intermedia per abbassarsi ancora fino a 51,82% tra i soggetti più giovani: una differenza di circa 26 punti percentuali che, unitamente all'assenza di qualunque problema di *overfitting*, non sembra lasciare dubbi sul fatto che secondo le reti il potere esplicativo delle variabili di origine perde di consistenza passando da una coorte all'altra.

Nella misura in cui le ricerche realizzate nel nostro paese che sostengono una stabilità delle asimmetrie educative, hanno ciononostante riscontrato una attenuazione dello svantaggio educativo tradizionalmente proprio delle classi agricole (Cobalti, 1990, Cobalti e Schizzerotto, 1993, pp. 171, 175, 178 e 1994, pp. 173-4), si potrebbe ipotizzare che il peggioramento della capacità predittiva mostrato dalle reti sia dovuto al solo miglioramento delle chances educative di queste classi: eliminandole dall'analisi, dunque, i modelli dovrebbero presentare analoghe percentuali di corrette classificazioni tra le coorti.

La tab. 2 riporta i risultati dei modelli stimati in assenza della piccola borghesia agricola e della classe operaia agricola: sebbene il comportamento classificatorio delle reti tenda ad essere più omogeneo rispetto a quello ottenuto sull'intero spettro delle classi sociali, non sembra scomparso un sistematico effetto d'interazione tra origine e coorte. La percentuale di corrette classifica-

12. Per limiti di spazio abbiamo deciso di riportare per ogni analisi unicamente il modello selezionato e non l'intera serie di architetture costruite: quest'ultima è ovviamente disponibile presso gli autori. Un'opzione puramente stilistica, invece, è alla base della scelta di presentare tutte le tabelle nell'*appendice statistica*.

zioni si riduce da 57,38% nella coorte più anziana a 48,60% in quella più giovane, passando per un valore di circa 63% nella coorte intermedia, valore che sembra ben attribuibile a qualche fattore distorsivo presente nei dati complice la ridotta numerosità campionaria sui cui si lavora (il leggero grado di *overfitting* presente nei modelli può ricondursi alla stessa fonte). Non vi è dubbio dunque che il minor potere classificatorio delle reti così come un campo di variazione più limitato del *classification rate* indicano chiaramente che le classi agricole sono state interessate da un consistente processo perequativo e che esso corrisponde ad una larga parte delle trasformazioni della struttura dei vantaggi e degli svantaggi educativi intercorrenti tra le classi sociali. Ciononostante, nella misura in cui un sistematico peggioramento del comportamento classificatorio delle reti persiste, non sembra del tutto ragionevole imputare tale tendenza perequativa interamente alla situazione di queste classi.

Tab. 2 - Var. ind.: classe sociale paterna e materna, prestigio occupazionale paterno, titolo studio paterno e materno, anni scuola paterni e materni: Var. dip.: titolo studio rispondente (Analisi senza piccola borghesia agricola e classe operaia agricola)

Coorte	Struttura rete	Funz. Attiv.	% training	% testing	Rmse
20-33 (N: 608)	27*1*4	sigmoide	65,63	57,38	0,353
34-47 (N: 840)	27*9*4	sigmoide	52,30	63,10	0,353
48-61 (N: 1069)	27*12*4	sigmoide	49,12	48,60	0,395

Al fine di incrementare il grado di comparabilità tra la nostra analisi e quelle condotte sugli stessi dati agli inizi degli anni '90, abbiamo pensato di misurare la posizione sociale d'origine nel loro stesso modo ovvero sia attraverso la sola posizione di classe familiare costruita secondo il «principio di dominanza» (Cobalti, 1990; Cobalti e Schizzerotto, 1993 e 1994).

Tab. 3 - Var. ind.: classe sociale familiare: Var. dip.: titolo studio rispondente

Coorte	Struttura rete	Funz. Attiv.	% training	% testing	Rmse
20-33 (N:1072)	8*15*4	sigmoide	72,81	75,70	0,300
34-47 (N: 1255)	8*1*4	sigmoide	57,17	57,94	0,368
48-61 (N:1368)	8*6*4	sigmoide	41,86	43,07	0,410

Le tabb. 3 e 4 mostrano i risultati dei modelli stimati rispettivamente con e senza le classi agricole: come si vede chiaramente, anche in questo caso il comportamento classificatorio delle reti indica nettamente una perdita di rilevanza dell'origine sociale nel processo d'ottenimento di un titolo di studio. Senza ripeterci nel commento delle tabelle, ci preme tuttavia porre l'accento su una differenza rispetto ai risultati evidenziati nella fase precedente dell'analisi: il valore costantemente più basso della percentuale di corrette classificazioni esibito dai sei modelli neurali scelti. Si tratta di un risultato importante

perché indica che una misurazione quanto più adeguata dell'origine sociale necessita di uno spettro di indicatori più ampio della semplice classe sociale familiare.

Tab. 4 - Var. ind.: classe sociale familiare; Var. dip.: titolo studio rispondente (Analisi senza piccola borghesia agricola e classe operaia agricola)

Coorte	Struttura rete	Funz. Attiv.	% training	% testing	Rmse
20-33 (N: 607)	8*15*4	sigmoide	58,55	50,82	0,375
34-47 (N: 829)	8*3*4	sigmoide	48,86	45,78	0,400
48-61 (N: 1060)	8*6*4	sigmoide	43,04	38,68	0,410

Rispetto al quadro dell'evidenza empirica delineato poco sopra, dunque, la modellistica neurale nella versione qui adottata sembra confermare l'ipotesi minoritaria di un declino delle disuguaglianze di opportunità educative nel nostro paese piuttosto che il risultato maggiormente consolidato di una persistenza e resistenza al cambiamento del complesso di vantaggi e svantaggi educativi costruito attorno all'origine sociale.

Trattandosi, come detto, di un risultato ancora minoritario, ottenuto per di più con una tecnica di analisi pressoché inutilizzata per lo studio di questo fenomeno, ci sembra ragionevole approfondire l'analisi in due direzioni.

In primo luogo, abbiamo pensato di testare la bontà delle reti analizzando un contesto - quello svedese - per il quale vi è completo accordo in letteratura sull'esistenza di una forte e consistente tendenza perequativa in materia di disuguaglianze educative (Blossfeld e Shavit, 1993, pp. 15-7; Jonsson, 1993; Jonsson e Erikson, 2000, p. 370; Jonsson e Mills, 1993b, p. 421; Shavit e Westerbeek, 1997, p. 91; Goldthorpe, 1996, pp. 498-9; Goldthorpe e Breen, 1997, pp. 182-3). Nel caso in cui le reti fornissero un risultato in linea con quelli ottenuti tramite l'analisi logistica e log-lineare, disporremmo di una conferma esterna, per così dire, della bontà e dell'affidabilità di questo strumento di analisi dei dati.

Secondariamente, potrebbe essere utile osservare se le reti sono in grado di confermare alcune proposizioni derivabili dagli studi, in particolare quello di Shavit e Westerbeek (1997), che hanno sostenuto un indebolimento nel nostro paese del legame tra origine e destino educativo: nel caso di un risultato positivo, si potrebbe parlare di una «conferma dall'interno» dei risultati ottenuti mediante reti neurali artificiali.

Quanto al primo aspetto, quindi, basandosi sui dati lungamente analizzati all'interno del progetto Casmin (Erikson e Goldthorpe, 1992; Muller e Karle, 1993), si è provveduto in primo luogo a ricodificare tanto lo schema di classe quanto la struttura dei titoli di studio tradizionalmente impiegati per questi dati, così da rendere entrambe le variabili analoghe a quelle fin qui utilizzate nelle nostre analisi. In un secondo tempo, abbiamo costruito tre serie di reti volte a stimare la relazione tra classe sociale paterna e titolo di studio del rispondente nelle tre differenti coorti - nati tra il 1910 ed 1924; tra il 1925 ed il 1939; tra