

KB per acquisire la conoscenza nascosta nei dati

di Roberto Bello (aprile 2008)

LA STATISTICA (di Trilussa)

Sai ched'è la statistica? E' 'na cosa
che serve pe' fa' un conto in generale
de la gente che nasce, che sta male,
che more, che va in carcere e che sposa.

Ma pe' me la statistica curiosa
è dove c'entra la percentuale,
pe' via che, lì, la media è sempre eguale
puro co' la persona bisognosa.

Me spiego, da li conti che se fanno
seconno le statistiche d'adesso
risurta che te tocca un pollo all'anno:
e, se nun entra ne le spese tue,
t'entra ne la statistica lo stesso
perché c'è un antro che se ne magna due.

La Business Intelligence e il pollo di Trilussa

La statistica molto spesso sbaglia o, per meglio dire, sbagliano i suoi utilizzatori.

Sbagliano quando applicano gli strumenti statistici di aggregazione su frammenti informativi provenienti da oggetti o situazioni fra di loro del tutto differenti.

Prima frammentano, poi mescolano ed infine aggregano.

Per finire pretendono di sentenziare.

Così i ricercatori sulle tendenze politiche spezzettano le opinioni degli intervistati, mescolano le singole risposte, aggregano, incrociano ed infine sentenziano certezze che sono attribuibili solo agli intervistati virtuali che loro hanno creato, non esistenti nella realtà e sicuramente non riconducibili ai singoli individui o a gruppi omogenei di intervistati.

In modo analogo la Business Intelligence rende disponibili degli strumenti di analisi dei dati in grado di tagliare i dati e poi di ricomporli in strutture multidimensionali nelle quali le peculiarità informative delle situazioni di partenza sono state distrutte.

Così con la Business Intelligence si mescolano aziende di diversi settori, fatturati, organici, mercati, abitudini di pagamento, cambiando di volta in volta le variabili di incrocio dei dati.

Le decisioni che poi si prendono, a quali soggetti (o situazioni) potrebbero essere applicate avendo distrutto il patrimonio informativo globale dei soggetti (o situazioni) di partenza?

Per fare un esempio, se avessi un archivio di animali mammiferi nel quale fossero compresi anche uomini e primati, potrei ottenere come risultato che i mammiferi hanno mediamente circa tre zampe.

Dove trovo un mammifero che abbia mediamente tre zampe?

Per fare della vera statistica occorre conservare il più possibile intatto il patrimonio informativo dei dati di partenza del soggetto o della situazione sotto esame.

Le tecniche derivate dalle reti neurali usano un approccio all'analisi dei dati del tutto differente.

Non richiedono all'utente di definire le variabili da incrociare, impedendogli così di formulare incroci assurdi.

Richiedono unicamente di inserire il numero massimo dei gruppi che l'algoritmo dovrà creare.

Non distruggono il patrimonio informativo dei dati di partenza, ma elaborano sempre i dati del soggetto (o situazione) in rapporto ai dati degli altri soggetti (o situazioni).

Conservano tutte le informazioni attribuibili al soggetto o alla situazione in esame e classificano i soggetti (o situazioni) in gruppi nei quali i soggetti (o situazioni) sono fra di loro simili.

Tecniche più sofisticate sono in grado di segnalare quali siano le variabili significative di aggregazione e quali siano i valori di aggregazione per ogni gruppo creato.

Segnalano anche quali siano le variabili non influenti nella classificazione.

Tecniche ancora più sofisticate possono elaborare qualsiasi tipo di insieme di dati sentenziando se nell'archivio sono presenti delle informazioni oppure sono presenti solo numeri o caratteri fra di loro non legati da relazioni interne.

Il modello deve seguire i dati e non viceversa (J. B. Benzecri)

L'apprendimento per induzione e le reti neurali

L'induzione è una modalità molto importante di apprendimento per le creature viventi.

Uno dei primi filosofi ad essere ricorso a questo concetto fu Aristotele, il quale, attribuendo a Socrate il merito di averla scoperta, sosteneva che l'induzione fosse, appunto, "*il procedimento che dai particolari porta all'universale*" (*Top.*, I, 12, 105 a 11).

Sempre secondo Aristotele non sono né i sensi per via induttiva, né la razionalità per via deduttiva, a dare di per sé garanzia di verità, bensì soltanto *l'intuizione intellettuale*: essa consente di cogliere *l'essenza* della realtà fornendo dei principi validi e universali, da cui il ragionamento sillogistico trarrà delle conclusioni coerenti con le premesse.

L'apprendimento vita ed evoluzione sono fra di loro correlati.

Infatti la vita e' evoluzione ed evoluzione e' apprendimento di ciò che e' necessario alla sopravvivenza.

Apprendimento e' la capacita' di elaborare le informazioni con intelligenza critica.

Quindi l'elaborazione critica delle informazioni e' vita.

Un semplice esempio può illustrare come si apprende per induzione.

Ipotizziamo di aver di fronte una persona che non abbia mai visto dei contenitori di uso comune come bicchieri, bottiglie, barattoli, tazze, vasi, scatole, fiaschi, boccali, calici, tetrapack e via dicendo.

Senza alcun commento mostro in successione esempi reali di oggetti appartenenti alle categorie sopra descritte.

La persona può guardare, odorare, toccare, soppesare e far suonare gli oggetti mostrati.

Dopo aver esaminato un sufficiente numero di oggetti, la persona facilmente sarà in grado di raggruppare gli oggetti in categorie contenenti gli oggetti fra di loro *globalmente* simili, privilegiando, in questo caso, la forma dell'oggetto alle altre caratteristiche di peso, materiale e colore.

Ad apprendimento avvenuto, io potrei presentare un altro oggetto a forma di bicchiere di altro colore, di altro materiale, di altro peso e di altro suono ottenendo comunque la collocazione dell'oggetto nella categoria dei bicchieri.

L'apprendimento ha consentito alla persona di riconoscere gli aspetti dell'oggetto utili per *passare dal particolare all'universale* trascurandone gli aspetti non influenti.

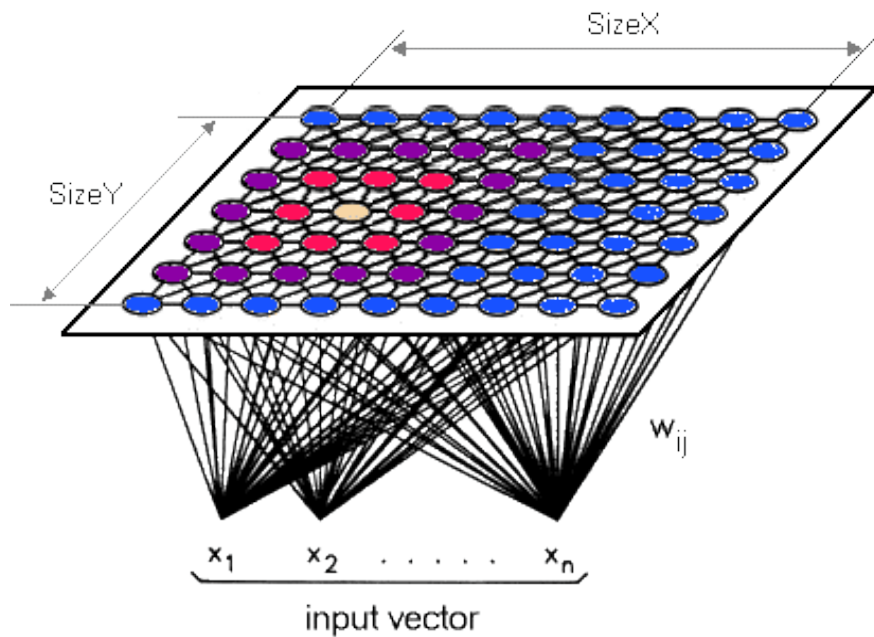
Gli algoritmi basati sulle reti neurali, con particolare riferimento alle mappe di Kohonen (SOM Self Organizing Map), si basano sui principi appena illustrati nell'esempio.

Tale modello di Rete Neurale riflette in modo significativo i meccanismi biologici del sistema nervoso centrale; molti studi hanno infatti dimostrato che sulla superficie della corteccia cerebrale esistono delle zone ben definite, ciascuna delle quali risponde ad una precisa funzione sensoria o motoria.

Ogni neurone si specializza a rispondere a determinati stimoli attraverso un'interazione continua con i neuroni confinanti. Avremo quindi zone riservate all'udito, alla visione, all'attività motoria etc., e la demarcazione spaziale tra i diversi gruppi è tanto netta che si parla di formazione di "bolle di attività".

Il modello di Rete Neurale presentato da Kohonen imita il comportamento sopra descritto. L'architettura è abbastanza semplice; la rete è formata da una griglia rettangolare, detta anche "strato di Kohonen", composta dai neuroni del livello di output, ciascuno dei quali occupa una precisa posizione ed è collegato a tutte le unità di ingresso.

I pesi delle connessioni tra il livello di input e quello di output vengono aggiornati grazie al processo di apprendimento, mentre le connessioni tra i neuroni del livello di output presentano pesi che producono eccitazione tra i neuroni limitrofi e inibizioni tra neuroni lontani.



Architettura di una mappa di Kohonen

Il fenomeno della “bolla” viene schematizzato attraverso l’introduzione del concetto di “vicinanza tra neuroni”; all’inizio della fase di apprendimento il vicinato contiene tutti i neuroni e si riduce progressivamente al crescere del numero di iterazioni del processo di addestramento.

Le reti SOM vengono applicate a molti problemi pratici; hanno la capacità di scoprire in modo autonomo proprietà importanti tra i dati di input e quindi sono particolarmente utili nei processi di Data Mining, soprattutto per problemi di classificazione.

L’algoritmo di apprendimento per le reti di Kohonen parte dalla fase di inizializzazione dei pesi sinaptici, i quali devono assumere valori casuali nell’intervallo (0,1) ed essere differenti per ogni neurone. Successivamente vengono presentati alla rete dei valori di input e l’algoritmo adottato permetterà alla rete di auto-organizzarsi e correggere i pesi dopo ogni presentazione degli ingressi, fino al raggiungimento di uno stato di equilibrio.

Le reti di Kohonen vengono definite anche reti competitive poiché si basano sul principio di competizione tra i neuroni per vincere e rimanere attivi; solo i pesi delle unità attive vengono modificati. L’unità vincente i^* è quella che possiede il potenziale di attivazione maggiore; quanto più un’unità è attiva per un certo pattern di ingresso tanto più il vettore dei pesi sinaptici è simile a tale pattern. In base a questo assunto è possibile trovare l’unità vincente calcolando la distanza euclidea tra il vettore di input ed il relativo vettore di pesi sinaptici:

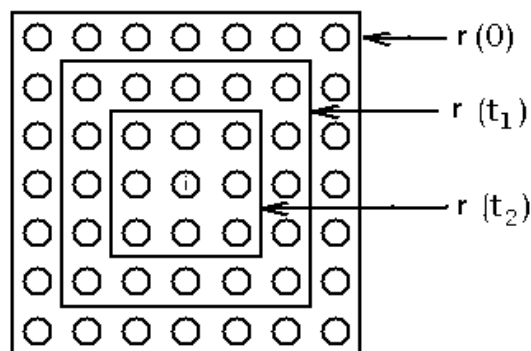
A questo punto si seleziona il neurone i^* a cui corrisponde la distanza minima.

Una volta determinato il neurone vincitore viene effettuato un aggiornamento automatico dei pesi del neurone stesso e di quelli che fanno parte del suo vicinato, in base ad una

regola di tipo hebbiano.

Risulta evidente che questo tipo di apprendimento competitivo spinge i pesi verso i vettori di ingresso a cui l'unità risponde maggiormente. Si crea in questo modo una mappa relativa alle caratteristiche degli input dove unità limitrofe rispondono a determinati stimoli di ingresso grazie alla similarità dei pesi sinaptici.

Minore è la dimensione del vicinato, minore è il numero di unità dello strato di Kohonen i cui pesi vengono modificati significativamente, quindi è maggiore la capacità dei neuroni di differenziarsi e di acquisire dettagli ma anche la complessità del sistema di apprendimento. Secondo Kohonen l'ampiezza della funzione di vicinanza deve essere fatta variare, scegliendola inizialmente in modo da coprire tutte le unità dello strato e facendola decrescere gradualmente. In questo modo si passerà dall'apprendimento delle caratteristiche di ingresso di base all'apprendimento di dettagli e specializzazione delle aree nel rispondere a determinati stimoli.



Rappresentazione della diminuzione graduale del vicinato

Conclusa la fase di addestramento la rete è in grado di fornire risposte in corrispondenza di nuovi input presentati. La proprietà di generalizzazione deriva dal fatto che anche i neuroni vicini a quello selezionato vengono modificati, la rete dovrà quindi auto-organizzarsi in aree costituite da un ampio insieme di valori attorno all'input da cui apprende, e questo farà sì che se viene presentato un input mai visto prima ma con caratteristiche simili, la rete sarà in grado di classificarlo nel modo appropriato. Rispetto agli algoritmi di tipo supervisionato, inoltre, il processo di apprendimento auto-organizzato risulta efficiente anche se vengono utilizzati dati di input incompleti o contenenti errori, caratteristica che rende queste reti particolarmente adatte ad essere applicate nel processo di Data Mining.

KB - Generalità

L'applicativo KB è un sistema di estrazione della conoscenza contenuta nei dati

basato su di un algoritmo delle mappe di Kohonen rivisto e modificato da Roberto Bello.

KB può elaborare qualsiasi tabella di dati sia numerici, sia testuali, tabella nella quale la prima riga della colonna sia destinata alla descrizione delle colonne / variabili e la prima colonna sia destinata ai codici (arbitrari) di identificazione delle registrazioni / casi.

In KB sono comprese delle funzionalità con l'obiettivo di:

- normalizzare i dati di tipo numerico rapportandoli allo scarto quadratico medio o al valore massimo a scelta dell'utilizzatore
- trasformare i dati alfanumerici in dati numerici opportunamente resi fra di loro equidistanti
- inserire funzioni statistiche in grado di valutare la bontà dei risultati di classificazione per ogni singolo gruppo e globalmente
- scrivere diversi archivi in output contenenti:
 1. le registrazioni / casi ordinati per codice di gruppo di classificazione
 2. le informazioni statistiche sintetiche sulle caratteristiche dei diversi gruppi in rapporto anche con gli indici statistici riferibili all'intera popolazione delle registrazioni / casi
 3. la matrice di addestramento finale avente l'errore minimo

Le reti neurali hanno il riconosciuto difetto di essere delle *scatole nere* nel senso che sono in grado di classificare ma non *spiegano*:

- quali siano le caratteristiche di ogni gruppo
- quali siano le colonne / variabili che, per ogni gruppo, sono importanti nella classificazione
- quali siano i gruppi maggiormente omogenei al loro interno
- se, nella sua globalità, la tabella di input contiene informazioni nel senso di relazioni fra le variabili oppure se la tabella è soltanto un insieme di numeri e di caratteri senza alcuna valenza di tipo induttivo.

KB risolve i problemi appena descritti.

KB è stato scritto da Roberto Bello nel linguaggio Ruby e, in virtù di tale linguaggio, è funzionante su tutti i sistemi operativi che supportano Ruby, ad esempio:

- Unix
- Linux

- DOS
- Microsoft Windows 95/98/2000/2003/NT/XP/Vista
- Mac OS X
- BeOS
- Amiga
- Acorn RISC OS
- OS/2
- Syllable

KB è stato tradotto da Guido Antonelli e da Luigi Bisone in linguaggio C compilabile senza modifiche in ambiente Windows e in ambiente Linux.

KB può essere ceduto in versione compilata C secondo modalità da concordare.

KB è reso disponibile come web service in www.xxxx.com.

KB - Modalità di utilizzo

Mandando in esecuzione KB, sono richieste le seguenti informazioni (esempio sull'archivio conten.txt):

Archivio di Input

-> conten.txt

conten.txt è l'archivio in formato txt contenente la tabella delle registrazioni / casi da classificare.

Contenuto di conten.txt (Contenitori)

descrizione	forma	materia	altezza	colore	peso	manico
bicch_vino	troncocono	vetro	15	trasparente	15	no
bicch_birra	troncocono	porcellana	18	diversi	25	no
tazza	troncocono	ceramica	10	bianco	30	si
tazza_latte	troncocono	terracotta	15	blu	35	si
tazza3	troncocono	vetro	20	trasparente	35	si
boccale	cilindro	terracotta	25	bianco	40	si
bottiglia	cilindro_cono	vetro	40	verde	120	no
bottiglione	cilindro_cono	vetro	50	verde	170	no
tazzina	troncocono	ceramica	6	bianco	20	si
tazza_tea	troncocono	terracotta	7	bianco	30	si

fiasco	sfera_cono	vetro	40	verde	150	no
bicch_dessert	troncocono	vetro	17	trasparente	17	no
flute	troncocono	cristallo	17	trasparente	17	no
damigiana	sfera_cono	vetro	80	verde	15000	no
bicch_1	troncocono	peltro	10	peltro	20	no
bicch_2	troncocono	plastica	9	bianco	4	no
bicch_3	troncocono	metallo	8	grigio	20	no
bottiglia1	cilindro_cono	vetro	40	trasparente	125	no
bottiglia2	cilindro_cono	vetro	45	opaca	125	no
bottiglia3	cilindro_cono	vetro	35	verde	125	no
tetra1	parallelepipedo	accoppiato	40	diversi	20	no
tetra2	parallelepipedo	plastica	40	diversi	21	no
tetra3	parallelepipedo	cartone	40	diversi	22	no

Numero Gruppi (3 - 20)

-> 4

il valore 4 è la radice quadrata del numero massimo dei gruppi di classificazione: in questo caso 16; poiché la matrice di addestramento è di forma quadrata, il numero massimo dei gruppi di addestramento può solo essere il quadrato del valore digitato.

Normalizzazione (Max, Std, Nessuna)

-> m

il valore m (M) indica che si richiede che i dati numerici siano normalizzati rapportandoli tutti al valore massimo della colonna / variabile.

E' sconsigliato inserire il valore N (nessuna) soprattutto in presenza di variabili fra di loro molto dissimili come range fra valore minimo e valore massimo.

Epoche addizionali (n x min err da 1 a 99) -> 50

KB, come tutti gli algoritmi delle reti neurali, esegue dei cicli di calcolo facendo delle iterazioni di apprendimento su tutte le registrazioni / casi in input.

Proseguendo nelle iterazioni, l'errore di addestramento normalmente diminuisce e KB memorizza il nuovo minimo e duplica la matrice di addestramento della matrice minima.

Avanzando il calcolo l'errore minimo smette di scendere e invece risale.

Quasi sempre però, continuando l'elaborazione, il minimo riprende a scendere fino al verificarsi della condizione di un minimo non ulteriormente migliorabile.

Il parametro delle *epoche addizionali* ha proprio lo scopo di non bloccare l'elaborazione al

verificarsi del primo non miglioramento dell'errore minimo di iterazione.

Il valore da assegnare al parametro delle epoche addizionale dipende dalla natura e dalle numerosità delle registrazioni / casi; le informazioni di log che appaiono durante il calcolo sono di aiuto nella definizione del valore del parametro considerando la vicinanza / lontananza dell'ultimo errore minimo dalla fine del calcolo.

KB produce i seguenti output:

Archivio di Output/Classif.origin (conten_M_g4_e50_out.txt)

E' identico all'archivio in input con l'aggiunta della colonna di attribuzione del codice del gruppo di appartenenza.

Gruppo	descriz.	forma	materia	altezza	colore	peso	manico
G_03_02	bicch_vino	troncocono	vetro	15,0	trasparente	15,0	no
G_04_03	bicch_birra	troncocono	porcellana	18,0	diversi	25,0	no
G_04_01	tazza	troncocono	ceramica	10,0	bianco	30,0	si
G_04_01	tazza_latte	troncocono	terracotta	15,0	blu	35,0	si
G_04_01	tazza3	troncocono	vetro	20,0	trasparente	35,0	si
G_04_01	boccale	cilindro	terracotta	25,0	bianco	40,0	si
G_02_01	bottiglia	cilindro_cono	vetro	40,0	verde	120,0	no
G_02_01	bottiglione	cilindro_cono	vetro	50,0	verde	170,0	no
G_04_01	tazzina	troncocono	ceramica	6,0	bianco	20,0	si
G_04_01	tazza_tea	troncocono	terracotta	7,0	bianco	30,0	si
G_01_02	fiasco	sfera_cono	vetro	40,0	verde	150,0	no
G_03_03	bicch_dessert	troncocono	vetro	17,0	trasparente	17,0	no
G_03_04	flute	troncocono	cristallo	17,0	trasparente	17,0	no
G_01_01	damigiana	sfera_cono	vetro	80,0	verde	15000,0	no
G_04_04	bicch_1	troncocono	peltro	10,0	peltro	20,0	no
G_04_04	bicch_2	troncocono	plastica	9,0	bianco	4,0	no
G_04_04	bicch_3	troncocono	metallo	8,0	grigio	20,0	no
G_02_03	bottiglia1	cilindro_cono	vetro	40,0	trasparente	125,0	no
G_01_01	bottiglia2	cilindro_cono	vetro	45,0	opaca	125,0	no
G_01_01	bottiglia3	cilindro_cono	vetro	35,0	verde	125,0	no
G_01_04	tetra1	parallelepipedo	accoppiato	40,0	diversi	20,0	no

G_01_04	tetra2	parallelepipedo	plastica	40,0	diversi	21,0	no
G_01_04	tetra3	parallelepipedo	cartone	40,0	diversi	22,0	no

Archivio di Output/Classif.sort (conten_M_g4_e50_outsrt.txt)

E' identico all'archivio precedente ma le registrazioni / casi sono ordinati in sequenza di codice del gruppo di appartenenza.

Gruppo	descriz.	forma	materia	altezza	colore	peso	manico
G_01_01	damigiana	sfera_cono	vetro	80,0	verde	15000,0	no
G_01_01	bottiglia2	cilindro_cono	vetro	45,0	opaca	125,0	no
G_01_01	bottiglia3	cilindro_cono	vetro	35,0	verde	125,0	no
G_01_02	fiasco	sfera_cono	vetro	40,0	verde	150,0	no
G_01_04	tetra1	parallelepipedo	accoppiato	40,0	diversi	20,0	no
G_01_04	tetra2	parallelepipedo	plastica	40,0	diversi	21,0	no
G_01_04	tetra3	parallelepipedo	cartone	40,0	diversi	22,0	no
G_02_01	bottiglia	cilindro_cono	vetro	40,0	verde	120,0	no
G_02_01	bottiglione	cilindro_cono	vetro	50,0	verde	170,0	no
G_02_03	bottiglia1	cilindro_cono	vetro	40,0	trasparente	125,0	no
G_03_02	bicch_vino	troncocono	vetro	15,0	trasparente	15,0	no
G_03_03	bicch_dessert	troncocono	vetro	17,0	trasparente	17,0	no
G_03_04	flute	troncocono	cristallo	17,0	trasparente	17,0	no
G_04_01	tazza	troncocono	ceramica	10,0	bianco	30,0	si
G_04_01	tazza3	troncocono	vetro	20,0	trasparente	35,0	si
G_04_01	boccale	cilindro	terracotta	25,0	bianco	40,0	si
G_04_01	tazzina	troncocono	ceramica	6,0	bianco	20,0	si

G_04_01	tazza_tea	troncocono	terracotta	7,0	bianco	30,0	si
G_04_01	tazza_latte	troncocono	terracotta	15,0	blu	35,0	si
G_04_03	bicch_birra	troncocono	porcellana	18,0	diversi	25,0	no
G_04_04	bicch_1	troncocono	peltro	10,0	peltro	20,0	no
G_04_04	bicch_2	troncocono	plastica	9,0	bianco	4,0	no
G_04_04	bicch_3	troncocono	metallo	8,0	grigio	20,0	no

Archivio di Output/Medie, SQM, CV (conten_M_g4_e50_medsd.txt)

L'archivio contiene le Medie, i Massimi, gli Sqm e i CV dei valori normalizzati di tutta la popolazione.

Colonna	forma	materia	altezza	colore	peso	manico
Media	0,31521	0,28502	0,43788	0,34161	0,50931	0,26086
Max	1,0	1,0	1,0	1,0	1,0	1,0
Sqm	0,37020	0,32073	0,28283	0,28490	0,28826	0,43910
CV	1,17444	1,12530	0,64589	0,83398	0,56599	1,68325

Archivio di Output/CV dei gruppi (conten_M_g4_e50_cv.txt)

L'archivio contiene le informazioni atte a misurare la qualità della classificazione.

Il valore contenuto in **ogni cella** rappresenta l'importanza dei valori della colonna / variabile nel gruppo della riga considerata: più il valore è prossimo allo zero e maggiormente la colonna variabile è significativa nella classificazione.

Nel caso il valore sia uguale a zero, la variabile / colonna in quel gruppo avrà valori identici: esempio gruppo G04_04 per la colonna / variabile "forma"

Il valore contenuto nelle **celle dell'ultima colonna** (media CV) indicano se il gruppo è omogeneo al suo interno: più il valore è prossimo allo zero e maggiore sarà la somiglianza delle registrazioni / casi fra di loro all'interno del gruppo in esame.

Sono omogenei al loro interno i gruppi G_01_01 e G_04_04, mentre non lo è il gruppo G_04_01.

Da ultimo è opportuno confrontare i valori contenuti in ogni colonna con il valore corrispondente contenuto nell'ultima riga CV Pop., oppure in alternativa i valori della penultima riga CV Gruppi con i corrispondenti dell'ultima riga CV Pop.

Dove la differenza è significativa la colonna variabile è stata globalmente importante nella classificazione e soprattutto dove nella penultima riga compare il valore zero; dove invece

la differenza è minima oppure addirittura il valore della penultima riga è superiore a quello dell'ultima, la colonna / variabile non ha globalmente contribuito alla classificazione.

In base alle considerazioni appena riportate le colonne / variabili “forma”, “materia”, “peso” e “manico” hanno globalmente contribuito alla classificazione, mentre per le colonne / variabili “altezza” e “colore” ciò non è avvenuto.

Gruppo	forma	materia	altezza	colore	peso	manico
G_01_01	0,60207	0,00000	1,09701	0,93769	0,14697	0,00000
G_01_02	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
G_01_04	0,60207	0,66500	1,09476	0,84786	0,48190	0,00000
G_02_01	0,85146	0,00000	1,54823	1,19905	0,16061	0,00000
G_02_03	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
G_03_02	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
G_03_03	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
G_03_04	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
G_04_01	1,90394	0,66500	1,40891	0,89372	0,86211	0,00000
G_04_03	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
G_04_04	0,00000	0,63392	1,09786	1,12161	0,39978	0,00000
CV Gruppi	0,30104	0,31696	1,9744	1,02965	0,27338	0,00000
CV Pop.	1,17444	1,12530	0,64589	0,83398	0,56599	1,68325

Archivio di Output/Grid addestram (conten_M_g4_e50_grid.txt)

L'archivio contiene i valori della matrice di addestramento con *errore minimo*.

Questa matrice può essere utilizzata dalla funzione di *riconoscimento* di nuove registrazioni / casi che possono essere classificati sulla base di quanto in precedenza appreso.

La matrice ha 3 dimensioni: prima parte del codice di gruppo, seconda parte del codice di gruppo e numero identificativo della colonna / variabile.

Segue il valore ricavato in fase di addestramento.

Gruppo1	Gruppo2	Colonna / Variabile	Valore
000001	000001	000001	0.579170
000001	000001	000002	0.000021
000001	000001	000003	0.622006

000001	000001	000004	0.612741
000001	000001	000005	0.645304
000001	000001	000006	0.000000
000001	000002	000001	0.599792
000001	000002	000002	0.046732
000001	000002	000003	0.580049
.....

DA COMPLETARE.....