

APPENDICE 3 (capitolo 10)

LA STRATEGIA ANALITICA: SELF ORGANIZING MAP

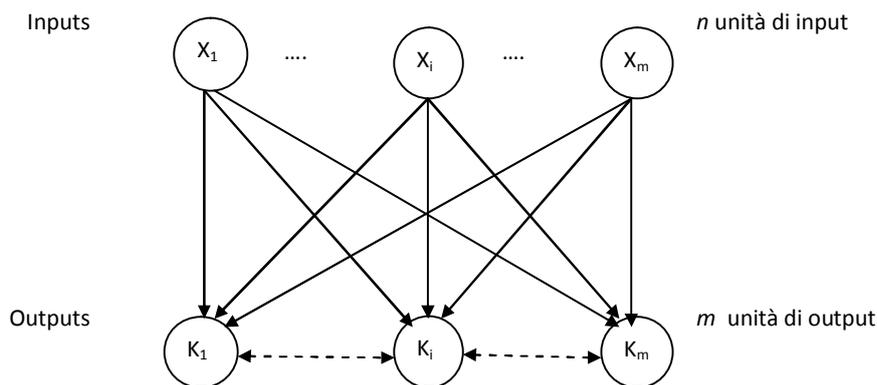
a cura di Mario Lucchini e Daniele Zaccaria

La Self Organizing Map (SOM) è una tecnica di raggruppamento (*clustering*) che realizza una mappatura di uno spazio multidimensionale dei dati di input X su di un piano bi-dimensionale di output K (lattice), che si compone di un certo numero di unità (definite anche nodi o micro-cluster), ciascuna delle quali esprime un particolare profilo, ovvero un particolare vettore di pesi \mathbf{w}_k o vettore prototipico.

Nella mappa SOM che proponiamo ciascun vettore di input che compone la base dati - vale a dire ciascuna osservazione i misurata sui 33 indicatori di qualità del lavoro selezionati - viene assegnato ad uno specifico nodo k ($k=1, \dots, m$) dello spazio di output, definito unità vincitrice (*the best matching unit*) e, in quanto tale, presenta il vettore di pesi \mathbf{w}_k che si approssima meglio al vettore di input \mathbf{x}_i in termini di distanza euclidea.

Ciascuna unità k che compone lo spazio di output esprime un vettore di pesi \mathbf{w}_k che rappresenta il centroide di una regione di Voronoi - m - dello spazio multidimensionale di input \mathcal{R}^d all'interno della quale vengono attratte tutte quelle osservazioni che si collocano nelle vicinanze. Le osservazioni poste in lontananza verranno invece attratte da nodi concorrenti. Ne consegue la proprietà topologica per cui i nodi di output o micro-cluster che si collocano nelle vicinanze tendono a riconoscere osservazioni tendenzialmente simili mentre i nodi che si collocano in lontananza tendono ad attrarre osservazioni tendenzialmente dissimili.

Figura 10.3 Architettura di una Self Organizing Map



Le unità di output k sono spazialmente continue e preservano la struttura contenuta nello spazio multidimensionale dei vettori di input, quella che in gergo specialistico si definisce *topologia*. Ne consegue che, se la partizione è stata effettuata in modo appropriato, le osservazioni che sono vicine nello spazio di input risulteranno vicine anche nello spazio di output.

Una SOM viene addestrata attraverso un processo che comporta un certo numero di iterazioni o passi di apprendimento ℓ , in occasione dei quali i vettori dei pesi vengono aggiustati come segue:

1) in primo luogo viene computata la distanza D_{ik} tra il vettore di input \mathbf{x}_i e ciascun vettore di pesi $\mathbf{w}_k(\ell-1)$. In genere D_{ik} è rappresentata dalla distanza euclidea $\|\mathbf{x}_i - \mathbf{w}_k(\ell-1)\|$;

2) viene poi identificata l'unità vincitrice (la *best matching unit*) riferita al vettore di input \mathbf{x}_i , vale a dire l'unità di output che esibisce la minor distanza D_{ik} . L'unità vincitrice viene denotata in pedice con la lettera b ;

3) infine, il vettore dei pesi che esprime ciascuna unità di output k viene modificato in accordo alla seguente equazione:

$$\mathbf{w}_k(\ell) = \mathbf{w}_k(\ell-1) + \alpha(t) \nu_{kb}(t) [\mathbf{x}_i - \mathbf{w}_k(\ell-1)]$$

Da tale equazione si evince che il vettore dei pesi \mathbf{w}_k al passo di apprendimento ℓ è espresso come funzione del vettore dei pesi \mathbf{w}_k allo *step* precedente $\ell-1$ a cui si aggiunge un fattore di aggiustamento $\alpha(t) \nu_{kb}(t) [\mathbf{x}_i - \mathbf{w}_k(\ell-1)]$ finalizzato a rendere il vettore dei pesi \mathbf{w}_k più simile al vettore di input. Dunque, nel corso delle epoche di apprendimento (t) una data unità di *output* vincitrice k_b diventerà sempre più specializzata a riconoscere quelle osservazioni che si addensano nel proprio campo o regione di attrazione m .

Nell'equazione dell'aggiustamento dei pesi compare un primo parametro definito *neighbourhood kernel* $\nu_{kb}(t)$ che rappresenta una funzione della distanza spaziale tra ciascuna unità di output k e l'unità vincitrice k_b . Tale parametro assicura che l'aggiustamento del vettore di pesi \mathbf{w}_k dell'unità k in direzione del vettore di input \mathbf{x}_i sarà tanto più consistente quanto maggiore è la sua vicinanza all'unità vincitrice k_b .

Il valore della *neighbourhood kernel* è regolato da un secondo parametro, definito *raggio di vicinato* o *neighbourhood radius* $\sigma(t)$, il quale stabilisce l'ampiezza della *kernel* ed è funzione decrescente di t .

Va inoltre precisato che il grado di aggiustamento complessivo del vettore dei pesi è espresso come funzione di un terzo parametro denominato *learning rate* $\alpha(t)$ il quale regola la velocità di aggiustamento del vettore dei pesi ed è espresso come una funzione decrescente di t . Al termine dell'apprendimento, ciascuna osservazione i viene assegnata alla rispettiva unità vincitrice k_b che minimizza la distanza D_{ik} tra \mathbf{x}_i e $\mathbf{w}_k(L)$. Il risultato finale sarà una tassonomia delle n osservazioni in $g \leq m$ gruppi mutualmente esclusivi.¹

¹ Tale approccio di raggruppamento e proiezione è stato già applicato al tema della deprivazione multipla da Pisati *et al.* [2010]