

Machine Learning: uno strumento diagnostico per l'individuazione di anomalie di mercato

Alessia Cafferata
Università degli Studi di Genova – DIEC

Pier Giuseppe Giribone
CARIGE – Financial Engineering

XIII Convention AIFIRM – Milano, 28 Novembre 2017

Motivation and Outline

Scopo del presente lavoro è la presentazione di un metodo automatico per il *clustering* delle informazioni osservate sui mercati secondari finalizzato al riconoscimento di potenziali anomalie.

Al fine di evidenziare opportunità di trading, si impiega una rete neurale non supervisionata, nota in letteratura come *Kohonen Network* o *Self-Organizing Map* (SOM).

La presentazione è suddivisa in tre parti:

- Architettura della rete e l'algoritmo di funzionamento
- Validazione del codice
- Presentazione di un'applicazione pratica:
il riconoscimento di anomalie nel mercato dei bonds



SOM – Self Organizing Map

L'obiettivo di una SOM è organizzare un input complesso ad elevate dimensioni (N) in aree spaziali, partizionate da celle (G), caratterizzate da avere una minore dimensione ($G \ll N$).

Questo processo genera una mappa rappresentativa che evidenzia le caratteristiche principali con un minor grado di complessità.

Per generare questa partizione, la SOM inizialmente colloca i suoi nodi all'interno del sistema vettoriale complesso di riferimento a N dimensioni secondo una griglia iniziale a G dimensioni, che può essere rettangolare, esagonale o casuale.

Durante l'algoritmo di apprendimento (*training*), i nodi della rete modificano la loro posizione nello spazio vettoriale al fine di concentrarsi nelle regioni del dominio sperimentale di maggiore interesse.



Network structure and algorithm

Inizialization

La mappa distribuisce i centri $c_k \in \mathcal{R}$ in accordo con la griglia di partenza.

Pattern creation

Uno stimolo p viene selezionato in \mathcal{R}^N .

Distance estimation

Per ciascun neurone k della rete viene stimata la distanza.

Learning-rule

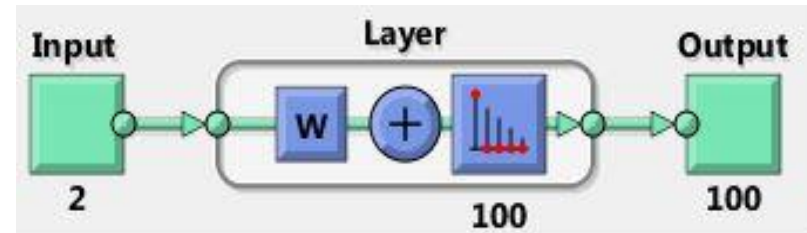
Determinazione del *winner neuron* i , caratterizzato dalla minima

distanza euclidea: $c_k^*(t) = \min_{w_k(t) \in M} d_E[p(t), w_k(t)]$

Updating rules of the centers

$$\Delta c_k = \eta(t) \cdot h(i, k, t) \cdot (p - c_k)$$

$$c_k(t + 1) = c_k(t) + \Delta c_k(t)$$

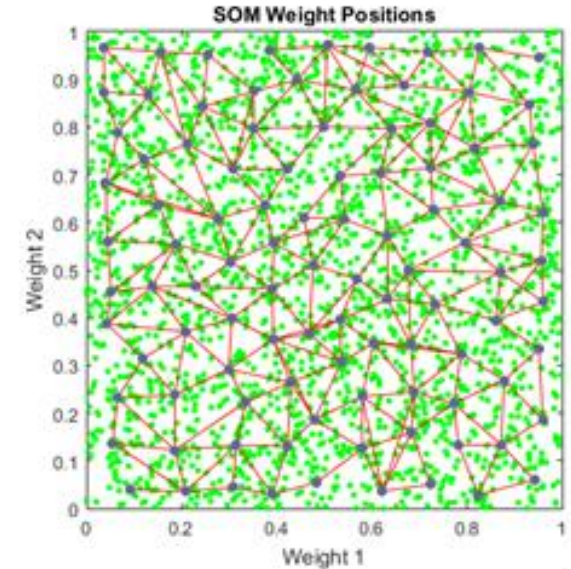
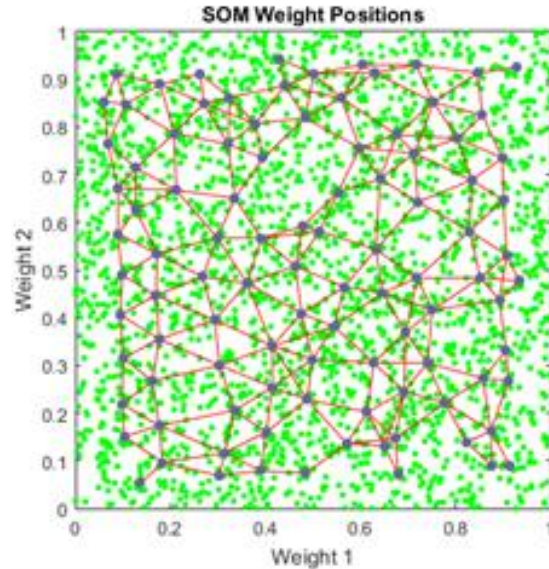
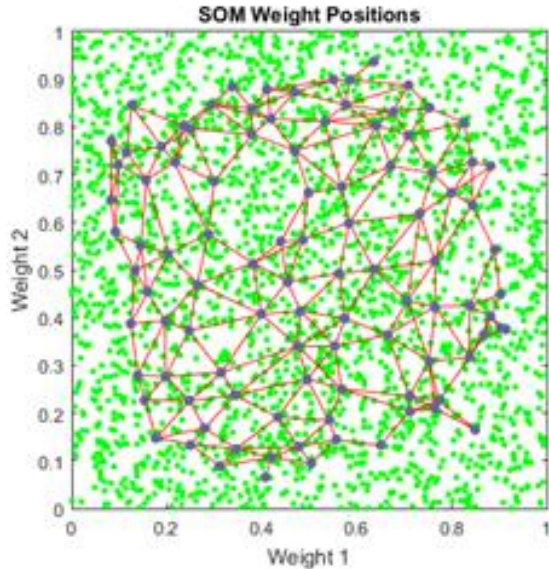


Code Validation

Il codice è scritto in linguaggio Matlab.

La prima validazione si basa sulla generazione di 5000 punti uniformemente distribuiti nello spazio $[0,1] \times [0,1]$.

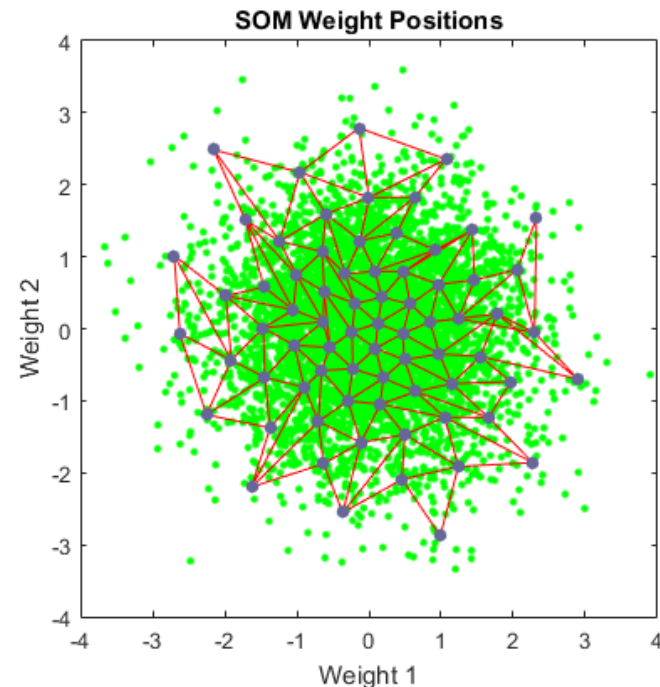
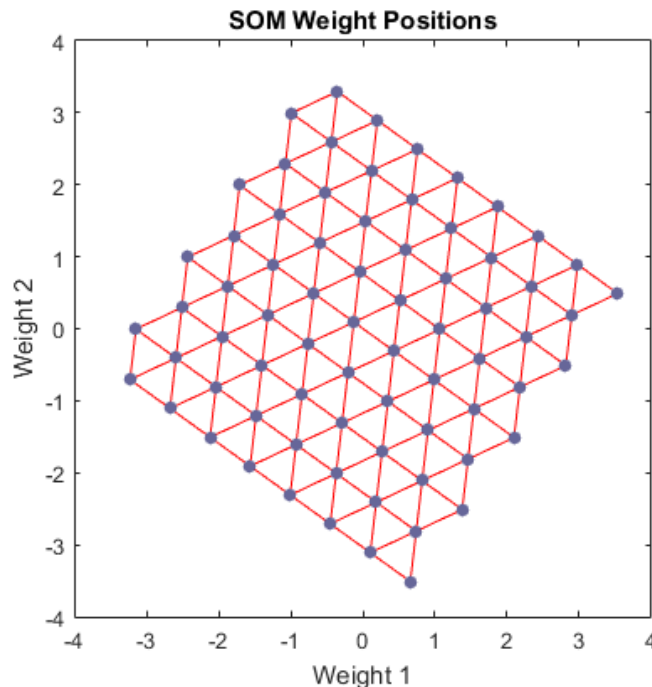
La mappa, costituita da 100 nodi, partendo da una configurazione random iniziale, ha originato un buon grado di ricoprimento dell'intero dominio.



Code Validation

Il secondo test è consistito nella generazione di 5000 punti casuali, attribuiti secondo una distribuzione normale standard.

La rete, costituita da 72 nodi, è stata in grado di isolare gli *outliers*, identificati come quei punti disposti fuori dal dominio supposto circolare con centro nell'origine e raggio superiore a 3σ



Field of Application

Uno tra i problemi più sfidanti nel trading è quello di determinare il *fair value* per gli strumenti finanziari.

Per un singolo prodotto, diversi contributori di mercato possono fornire quotazioni differenti.

L'abilità per un trader è quella di analizzare i livelli di prezzo, prendendo in considerazione i volumi negoziati dei deals oggetto di analisi: maggiori sono le quantità e maggiore è la probabilità che la transazione possa essere portata a termine.

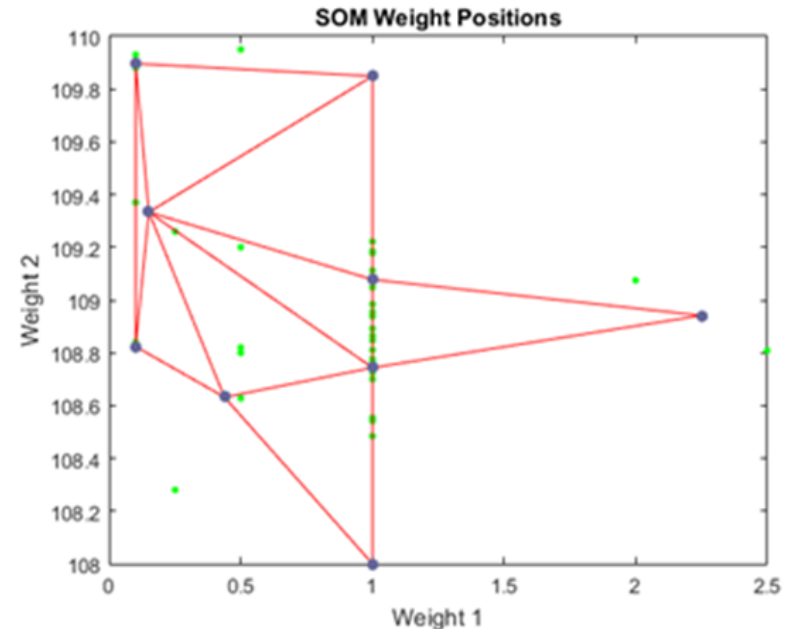
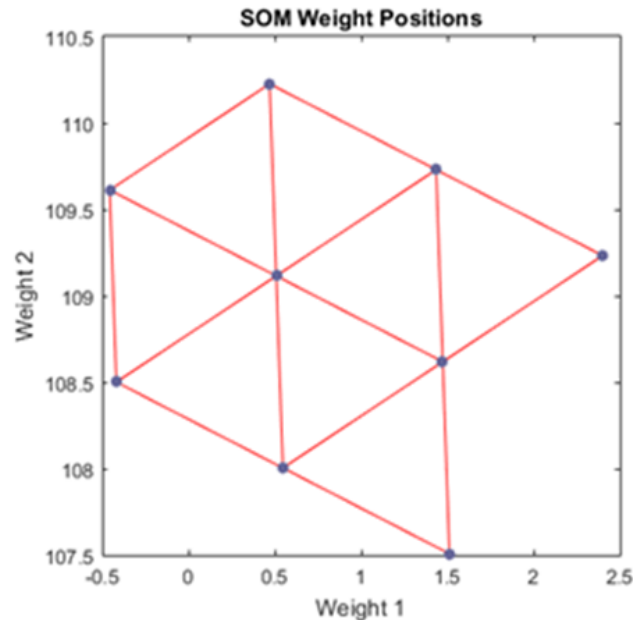
In tale contesto le SOM possono essere considerate una buona metodologia per evidenziare i possibili outliers, dando luogo a potenziali interessanti inferenze sulle negoziazioni.

Si procede a mostrare un'applicazione della metodologia di *Machine Learning* in applicazione al mercato obbligazionario.



Practical Application

Partendo da una griglia con topologia esagonale formata da 9 nodi, si procede all'addestramento non-supervisionato della rete finalizzato a mappare i possibili outliers nel piano prezzi/volumi.



La figura di sinistra mostra la SOM iniziale, quella di destra è la mappa dopo un training di 1000 epoche.

Le quantità sono espresse in milioni e il prezzo è in percentuale.

Practical Application

I punti blu sono i neuroni della rete ($G = 9$), mentre quelli verdi rappresentano i dati sperimentali osservati sul mercato ($N = 42$).

Le linee rosse caratterizzano la Self Organizing Map e ne determinano i cluster.

Si considerano outliers – e quindi potenziali anomalie di mercato – i punti non inclusi all'interno della mappa.

La partizione spaziale generata dalla mappa di Kohonen è stata in grado di identificare correttamente i potenziali outliers – ovvero le proposte di trading aventi, per esempio, un prezzo maggiore di 109.9% ed inferiore al 108.4%, in corrispondenza di una quantità uguale od inferiore al mezzo milione.



Conclusions

L'articolo ha evidenziato come le *Self-Organizing Maps* possano essere impiegate per progettare procedure automatiche di individuazione di anomalie di mercato e, conseguentemente, potenziali opportunità di trading.

L'algoritmo impiegato si è dimostrato una metodologia efficiente sia dal punto di vista teorico che nella sua applicazione esemplificativa nel mercato obbligazionario.

Le reti di Kohonen non dovendo, a differenza di molti altri algoritmi di clustering (ad esempio il *K-means*, la *Fuzzy C-means*) specificare aprioristicamente il numero di suddivisioni da effettuare, si sono rivelate uno strumento flessibile di organizzazione delle informazioni finanziarie.



Further Development

Si reputa un'interessante estensione del presente lavoro l'introduzione di mappe di calore (*heating maps*), che consentono di analizzare più approfonditamente le informazioni dei mercati.

Al fine di rendere ancor più robusto il risultato, si potrebbe testare l'affiancamento alle SOM con altri algoritmi di *clustering*, che consentirebbero l'organizzazione dei dati senza la necessità di specificare a-priori il numero di partizioni da effettuare.

Tra questi si citano: *t-SNE* (t-distributed Stochastic Neighbor Embedding) e *Kernel methodologies*.

Contacts

Alessia Cafferata: alessia.cafferata@economia.unige.it

Pier Giuseppe Giribone: piergiuseppe.giribone@carige.it



Further Readings

[A] A. Cafferata, P. G. Giribone – “I paradigmi di apprendimento non supervisionato per reti neurali in campo finanziario: progettazione di self-organizing maps per il rintracciamento di anomalie di mercato” – AIFIRM Magazine (Associazione Italiana Financial Industry Risk Managers) (Settembre 2017).

[B] A. Cafferata, P. G. Giribone, M. Neffelli, M. Resta – “Yield curve estimation under extreme conditions: do RBF networks perform better?” – WIRN 2017, 27th Italian Workshop on Neural Networks Proceedings (14-16 Giugno 2017).

[C] P. G. Giribone, O. Caligaris, S. Fioribello – “L’algoritmo Fuzzy C-Means clustering come tecnica automatica per l’individuazione di anomalie di mercato” – AIFIRM Magazine (Associazione Italiana Financial Industry Risk Managers) (Marzo 2017).

[D] P. G. Giribone, O. Caligaris, S. Fioribello, S. Ligato – “Implementazione della Fuzzy Logic per la gestione ottimale del portafoglio: la modellizzazione dell’avversione al rischio di un investitore attraverso tecniche di soft-computing” – AIFIRM Magazine (Associazione Italiana Financial Industry Risk Managers) (Settembre 2016).

[E] P. G. Giribone, S. Ligato, O. Caligaris – “ Applicazione delle reti neurali feed-forward per la ricostruzione di superfici di volatilità ” – AIFIRM Magazine (Associazione Italiana Financial Industry Risk Managers) (Giugno 2015).

