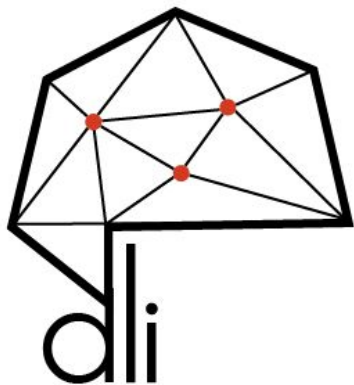


67616C696C656F2863726561746564286128636F6465642878617468286C656
16428746F28636875726368286F66286C6C756D696E6174696F6E28686528
62656C6965766528746861742863687572636828616E6428736369656E63652
865965656572656E74286C6516E67756167652874656C6C696E6728746865287
3616D652873746F727928746F2865696E648D887468652878617468286F6628
626C6C756D696E6174696F6E2874686528786C6163652877686572652874686
528696C6C756D696E61746928776F756C64286D65657428696E287365637265
7428796F75286D7573742866696E6428746865287365
676E6F286128736869
861742874686528
6865287861746
828

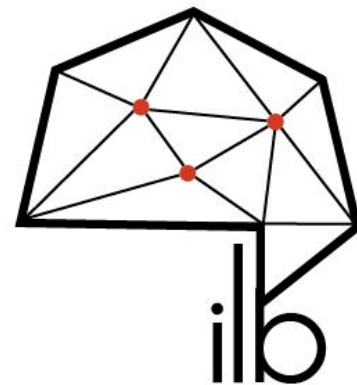
“Algoritmi non supervisionati per Time Series”

Nicola Procopio 20 - 10 - 2020



{CODEMOTION}
community

#Aperitech



Chi sono

Dove ho lavorato

UNIVERSITÀ
DELLA CALABRIA



Dove potremmo esserci già incontrati



FONDAZIONE



Nicola "Nico" Procopio
Senior Data Scientist @ **integr8**



Interessi "lavorativi"



PyData



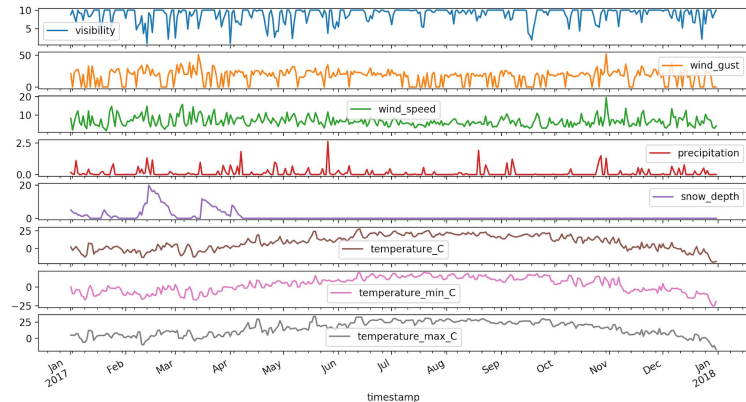
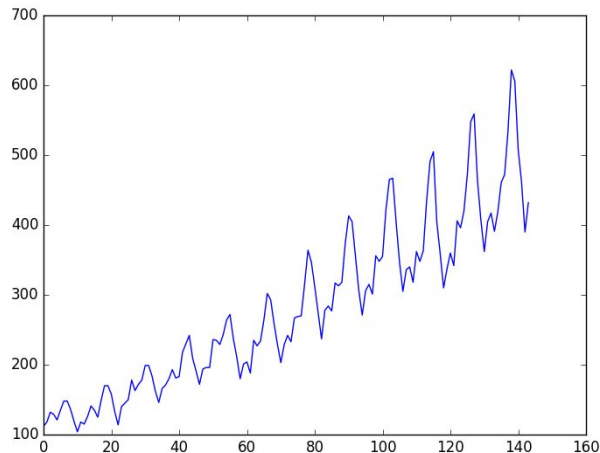
Altro

- Lettore bulimico
- Ciclista/Hikers molto amatoriale
- Indie-Rock fan
- Innamorato della mia terra

Time Series: Definizione

*“una **serie storica** si definisce come un insieme di variabili casuali ordinate rispetto al tempo”*
le cui osservazioni sono influenzate dal tempo stesso.

Dato un fenomeno Y si indica con Y_t l'osservazione di tale fenomeno al tempo $t=\{1, 2, \dots, T\}$



Time Series is Dynamic!

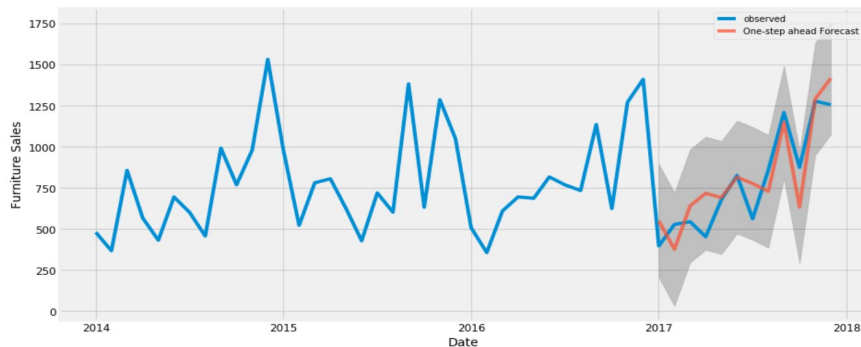
- **Cross Sectional Data:** una serie di variabili, rappresentative di un fenomeno, osservate in un particolare momento
 - Esempio: abitudini di acquisto durante i saldi, prezzi degli immobili in base alle caratteristiche in un dato periodo
- **Time Series Data:** una singola variabile, rappresentativa di un fenomeno, osservata su più periodi della stessa dimensione
 - Esempio: il PIL negli ultimi 30 anni, la fornitura elettrica bimestrale

E le serie storiche MULTIVARIATE?

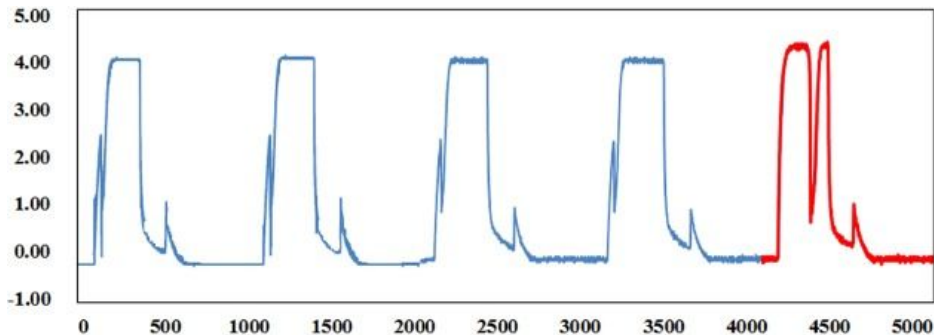


Problemi più comuni

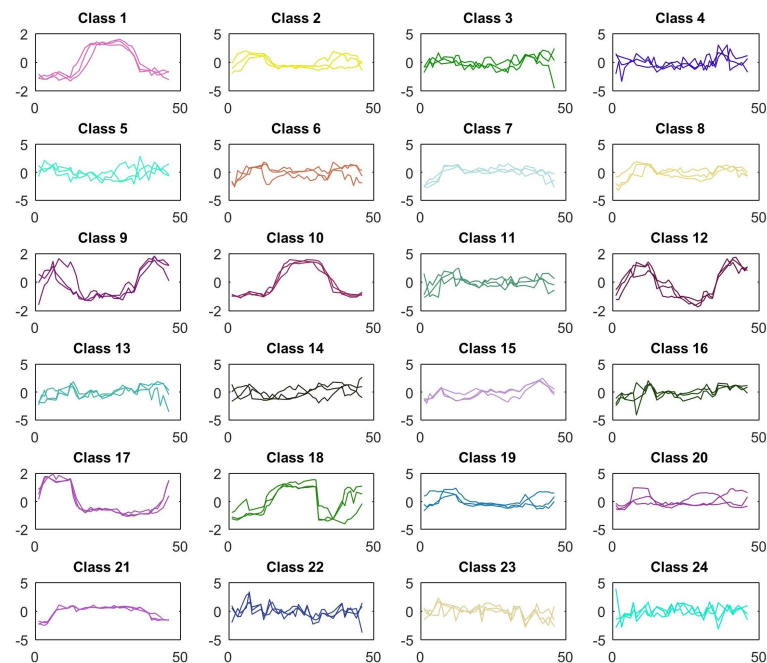
Forecasting



Anomaly Detection



Classification



In (MY) Real World



Approccio Supervisionato:

- infinita letteratura e codice on-line da studiare e riciclare
- poche aziende hanno uno storico di dati etichettati da utilizzare nella fase di training

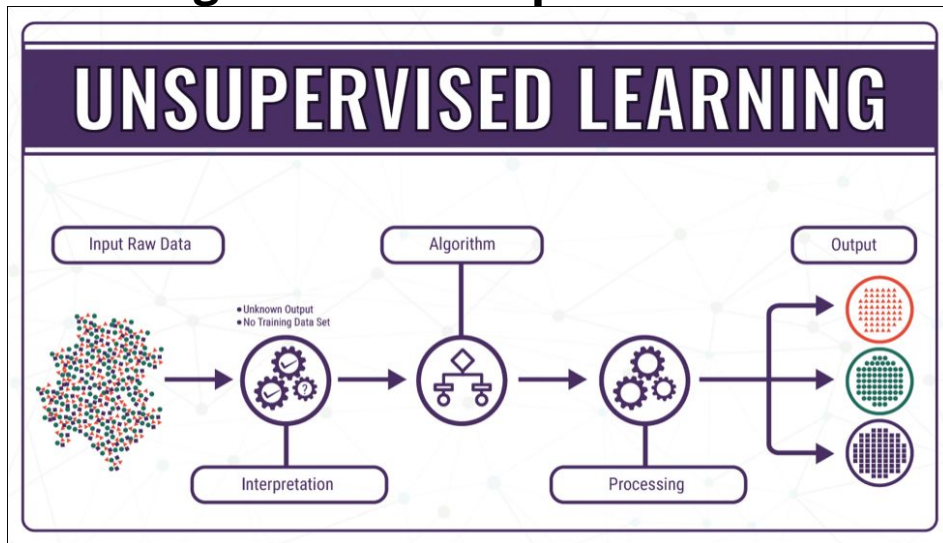
Possibili soluzioni

**Etichettare lo storico dati a
mano**



ATTENZIONE:
potrebbe nuocere alla salute

**Etichettare lo storico dati con
algoritmi non supervisionati**



ATTENZIONE:
coinvolgere l'esperto di dominio

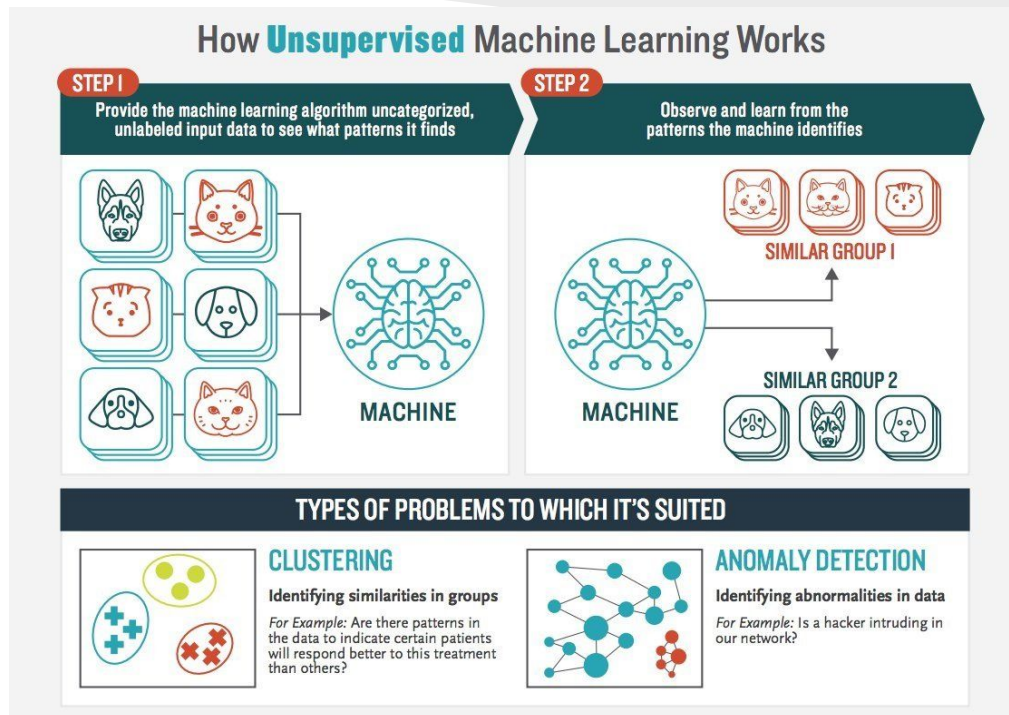
Basics

Approccio NON Supervisionato:

- non c'è bisogno di dati etichettati
- lavora sulla similarità, le distanze, la densità
- ha l'obiettivo di creare gruppi più omogenei possibile al loro interno ed eterogenei tra loro

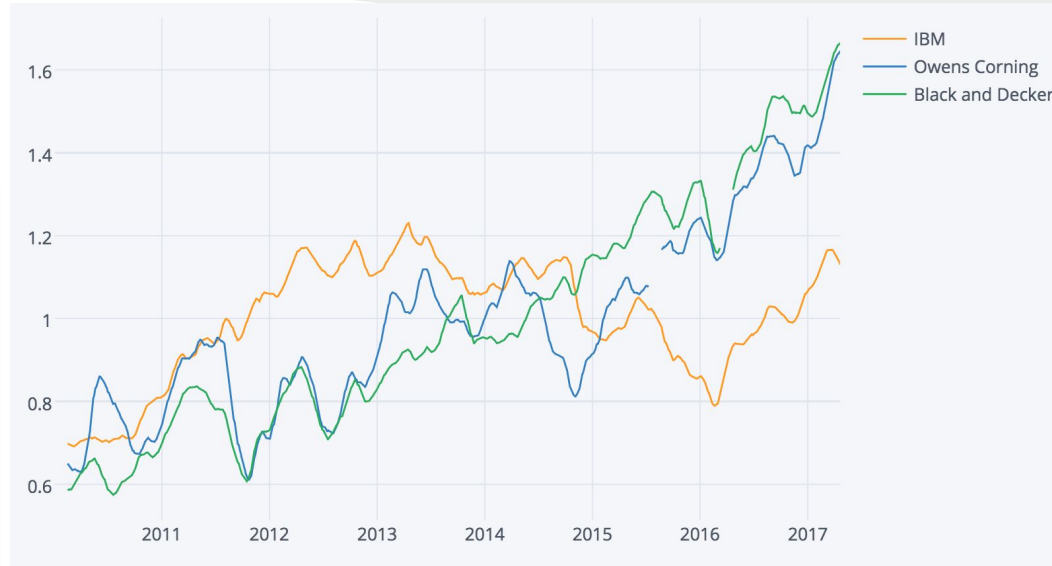
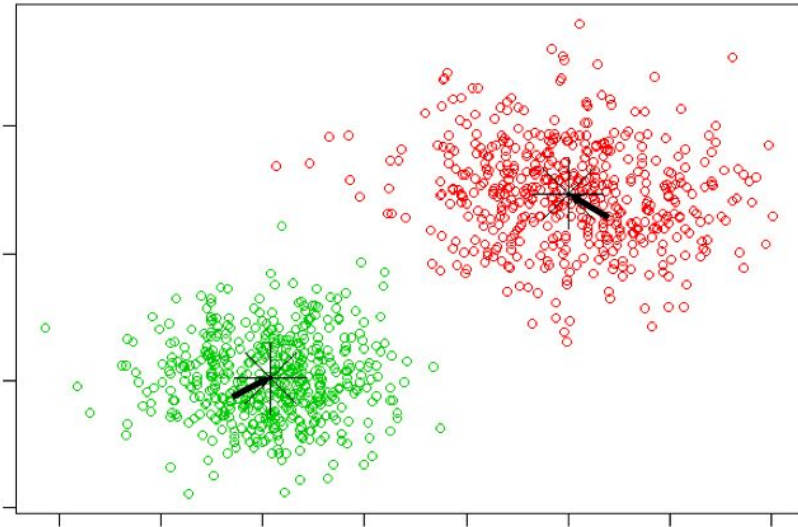
Tecniche più comuni:

- Clustering
- Dimensionality Reduction



Problema: misurare distanza

Calcolo della distanza tra osservazioni

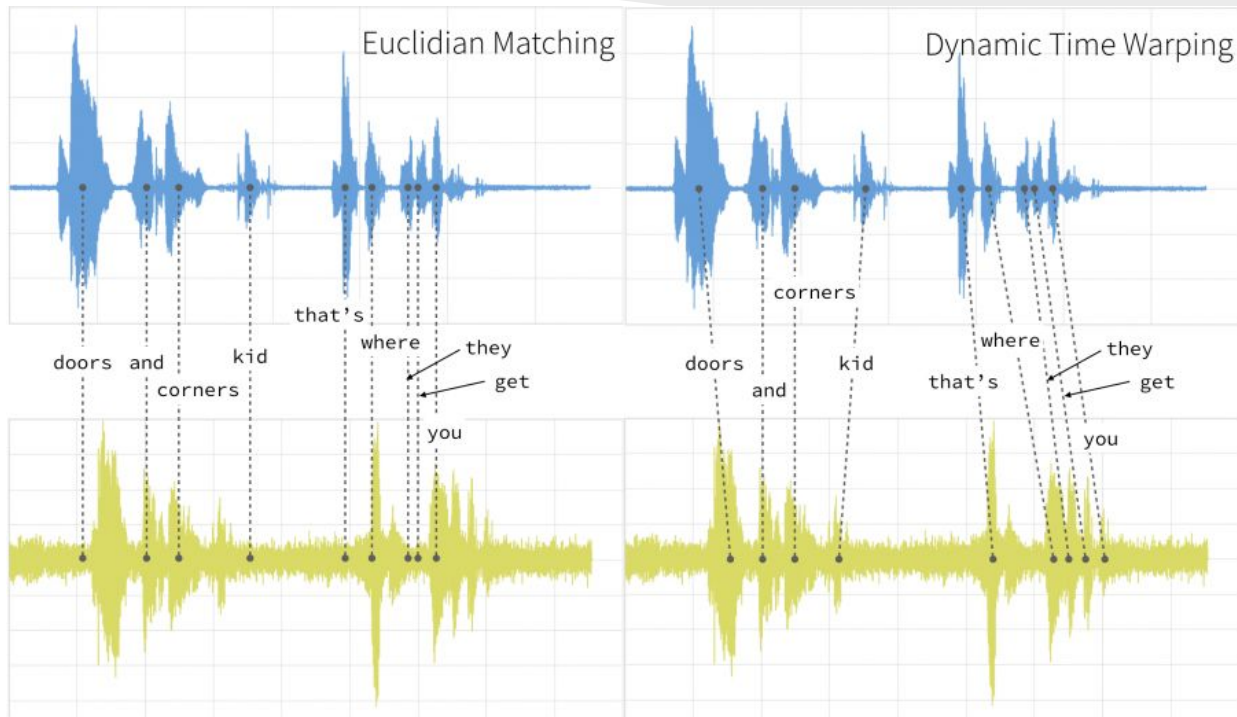


Ogni Time Series è una singola osservazione

Dynamic Time Warping

"Il **Dynamic Time Warping**, o DTW, è un algoritmo che permette l'allineamento tra due sequenze, e che può portare ad una misura di distanza tra le due sequenze allineate.

Tale algoritmo è particolarmente utile per trattare sequenze in cui singole componenti hanno caratteristiche che variano nel tempo, e per le quali la semplice espansione o compressione lineare delle due sequenze non porta risultati soddisfacenti."





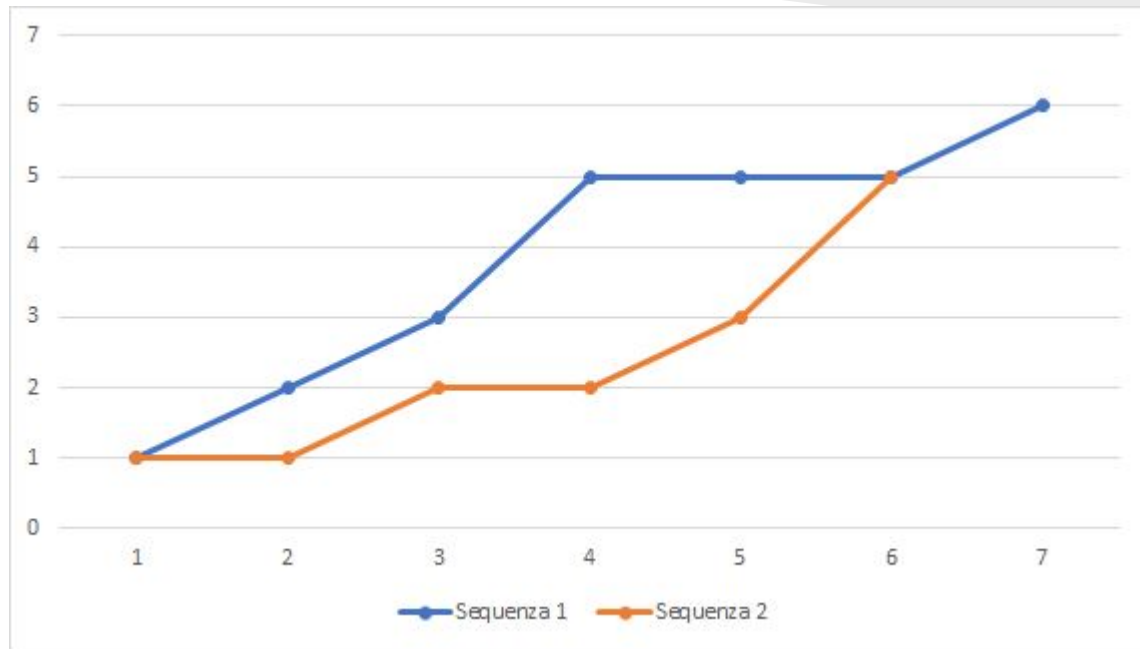
DTW: Prerequisiti

Per calcolare la distanza tra due sequenze, **indipendentemente dalla loro lunghezza**, col DTW devono essere presenti i seguenti requisiti:

- ogni indice della prima sequenza deve poter essere confrontato con tutti gli indici delle altre e viceversa;
- il primo indice della prima sequenza deve essere associato al primo indice dell'altra sequenza (ma non deve essere la sua unica corrispondenza);
- l'ultimo indice della prima sequenza deve corrispondere all'ultimo indice dell'altra sequenza (ma non deve essere la sua unica corrispondenza);
- la mappatura degli indici dalla prima sequenza agli indici delle altre sequenze deve essere monotonicamente crescente. A grandi linee, indipendentemente dalla lunghezza delle serie l'indice iniziale e finale devono coincidere.

DTW: Algoritmo

Date due sequenze di lunghezza diversa se volessimo calcolarne la similarità utilizzando il DTW per prima cosa bisogna creare una matrice $m \times n$ dove (m,n) sono le lunghezze delle serie + 1



DTW: Cold Start

Il **cold start** è il confronto con i periodi precedenti quando si è al primo passo dell'algoritmo.

Impostando una riga fittizia ad infinito non si incorre in situazioni anomale e si evita di implementare condizioni sullo start che appesantirebbero ulteriormente il calcolo.

	0	1	1	2	2	3	5
0	0	inf	inf	inf	inf	inf	inf
1	inf						
2	inf						
3	inf						
5	inf						
5	inf						
5	inf						
6	inf						

DTW: Calcolo della distanza

Partendo dalla cella [1,1] viene calcolata la distanza euclidea tra ogni punto delle due serie, a questa deve essere aggiunto il minimo tra:

- il valore precedente sulle righe: **cancellazione**;
- il valore precedente sulle colonne: **inserimento**;
- il valore precedente in diagonale: **corrispondenza**.

$$DTW[i,j] := distance[i,j] + \min(DTW[i-1, j], \\ DTW[i, j-1], \\ DTW[i-1, j-1])$$

DTW: Warping Path

Partendo da $DTW[0,0]$ costruire un percorso che tocchi il valore minimo muovendo da sinistra verso destra, dall'alto verso il basso.

- **spostamento orizzontale:**
serie_2 è accelerata
durante questo intervallo.
- **spostamento verticale:**
serie_2 è decelerata
durante questo intervallo.
- **spostamento diagonale:**
durante questo periodo le
serie camminano di pari
passo.

	0	1	1	2	2	3	5
0	0	inf	inf	inf	inf	inf	inf
1	inf	0	0	1	2	4	8
2	inf	1	1	0	0	1	4
3	inf	3	3	1	1	0	2
5	inf	7	7	4	4	2	0
5	inf	11	11	7	7	4	0
5	inf	15	15	10	10	6	0
6	inf	20	20	14	14	9	1

DTW: Miglioramenti

Window Constraint

Per rendere più performante l'algoritmo viene impostata una finestra temporale w su cui calcolare la distanza. Utile per sequenze lunghe che produrrebbero confronti inutili tra periodi molto distanti tra loro.

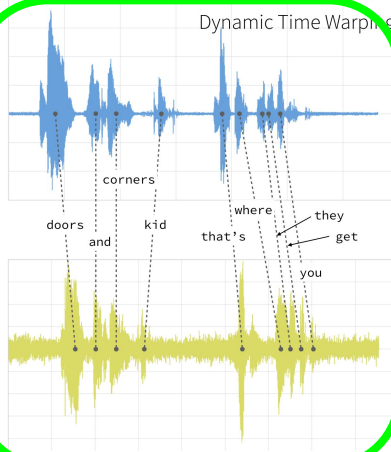
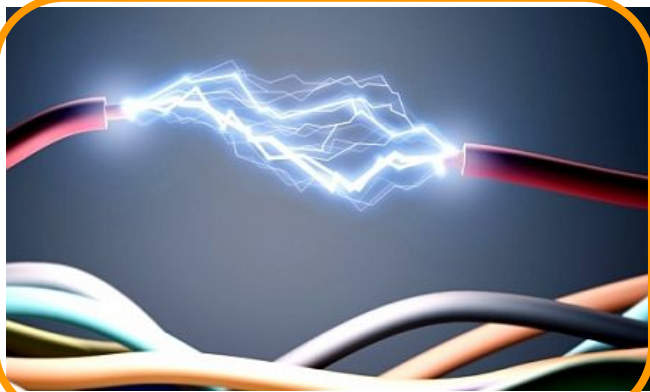
FastDTW

E' forse la versione più conosciuta dell'algoritmo. Si basa su alcune approssimazioni che velocizzano il calcolo. Molto studiato e discusso perchè non sempre la performance porta a risultati di qualità accettabile.

DTW: Applicazioni

Anomaly detection

Trovare pattern anomali in ambito health, ad esempio una persona che deambula a velocità diversa, o parla più lentamente/velocemente potrebbe aiutare nella diagnosi.

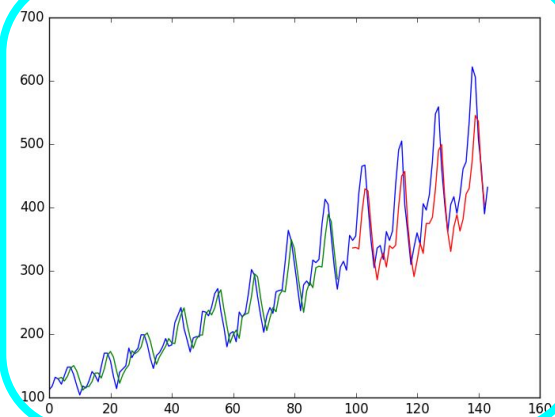


Customer Segmentation

Profilare le abitudini in ambito utilities per proporre nuove offerte o prevenire frodi.

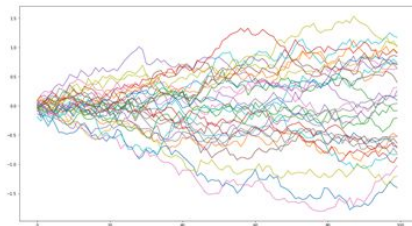
Forecasting Evaluation

Misurare la distanza tra dato reale e stime.

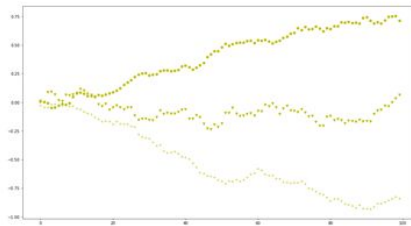


DTW e K-Means

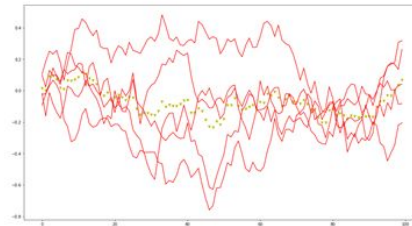
Sequenze originali



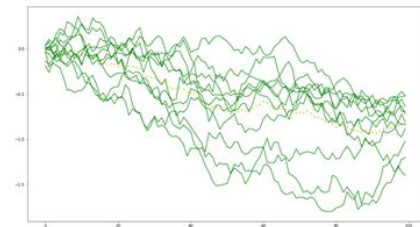
Centroidi



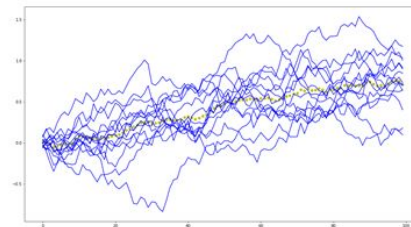
Cluster A: Trend costante



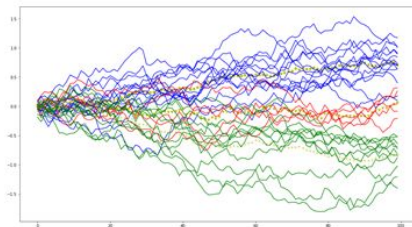
Cluster B: Trend decrescente



Cluster C: Trend crescente



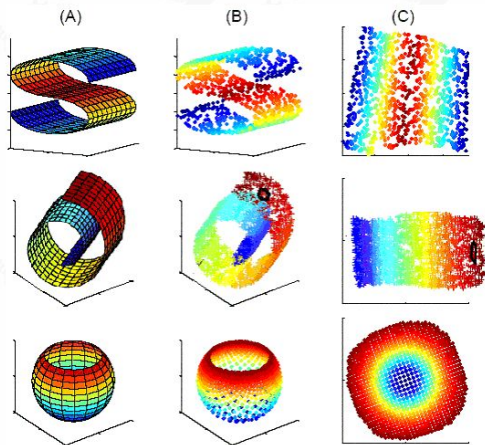
Sequenze etichettate



Dimensionality Reduction

In termini **estremamente semplicistici**:

“prendo un fenomeno molto complesso, condenso l'informazione (con una perdita accettabile) mediante un algoritmo, restituisco lo stesso fenomeno ma con una visione meno dettagliata”



N.B. Non siate estremisti
come Drake

Time Series?



PCA,
t-SNE,
Autoencoders

SAX
Encoding

SAX Encoding

Symbolic Aggregate approXimation Encoding:

- inventato nel 2002 da Keogh e Lin
- trasforma le time series in sequenze di simboli
- robusta ai valori mancanti
- non supervisionata
- **basata sia sui volumi che sugli andamenti**



Adolphe Sax inventore del sassofono.
 $\text{Correlazione}(\text{SAX}, \text{Sax}) = 0.00$

Preparazione del dataset

Il SAX Encoding ha bisogno di serie storiche disposte come segue:

- per ogni riga una serie storica
- per ogni colonna uno step temporale
- i dati devono essere standardizzati

Regione	1	2	...	57
Abruzzo	0	0	...	2067
Basilicata	0	245
...
Veneto	32	10077

Regione	1	2	...	57
Abruzzo	-1.067762	-1.067762	...	1.699956
Basilicata	-1.063952	1.041431
...
Veneto	-1.244880	10766

dati della protezione civile sul Covid-19

Piecewise Aggregate Approximation

Idea

“riassumere la sequenza in una serie di segmenti che ne riducano la lunghezza ma con la minima perdita di informazione”

Data una Time Series $Y = [Y_1, Y_2, \dots, Y_n]$ può essere ridotta in una sequenza $X = [X_1, X_2, \dots, X_m]$ con $m \leq n$ utilizzando l'equazione:

$$\bar{X}_i = \frac{m}{n} \cdot \sum_{j=n/N(i-1)+1}^{(n/M) \cdot i} x_j$$

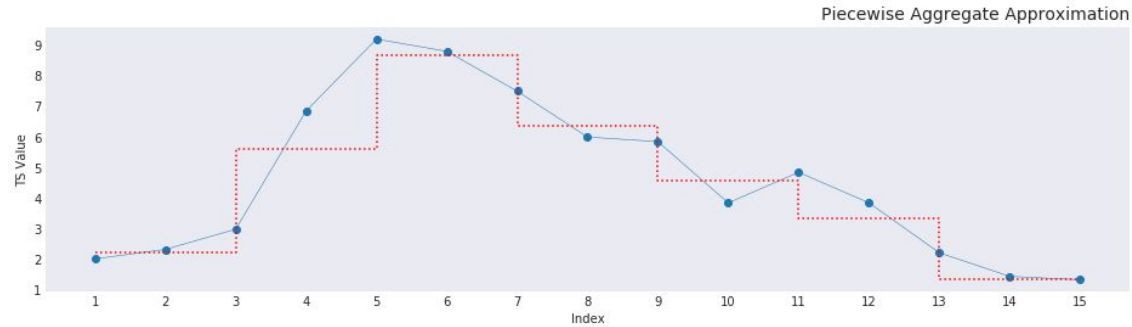
casi particolari:

- $m=n$ restituisce la serie originale
- $m=1$ la nuova sequenza avrà un solo valore pari alla media della time series originale

Piecewise Aggregate Approximation

Molto importante scegliere la finestra temporale di approssimazione w con un esperto di dominio.

Questo influisce sul numero di segmenti, se $\text{len}(TS)/w$ non è un intero arrotondare per eccesso così da non perdere informazioni.



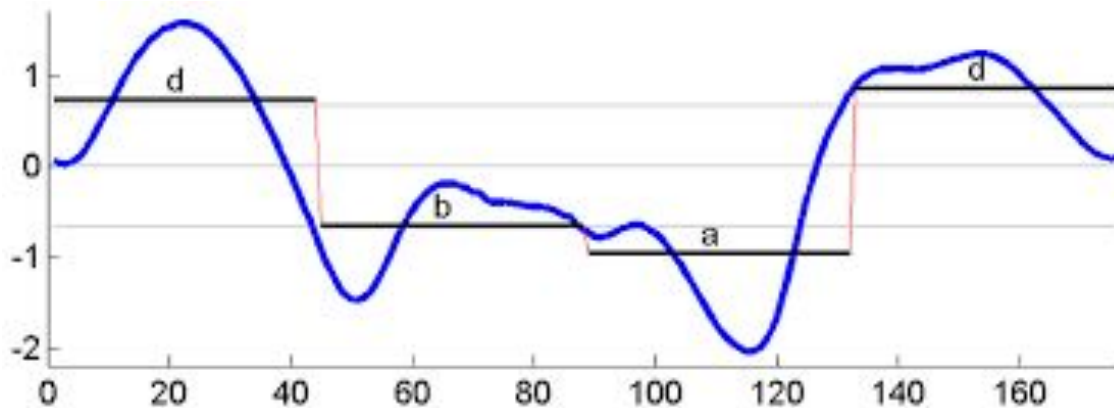
denominazione_region	0	1	2	3	4
Abruzzo	-1.060876	-0.785615	0.352059	1.252060	1.696609
Basilicata	-1.055972	-0.846662	0.516006	1.239693	1.028541
Calabria	-1.088706	-0.797959	0.449874	1.246324	1.333274

SAX String

Per creare una SAX String:

- scegliere il numero di livelli
- calcolare i livelli
- etichettare i periodi generati con la PAA

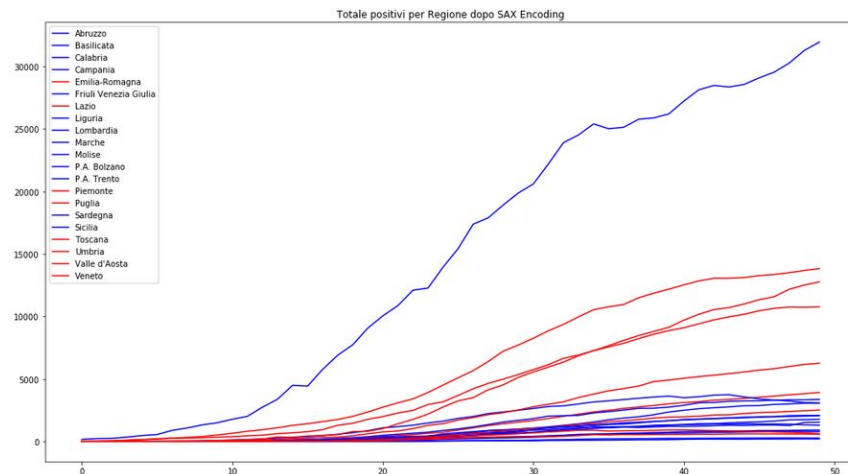
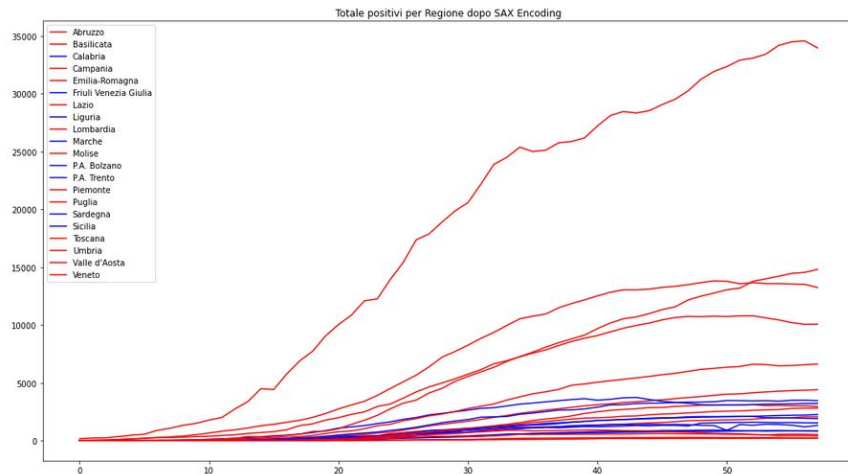
denominazione_regione	SAX_string
Abruzzo	CAABC
Basilicata	CABBA
Calabria	BABBB
...	...



Anomaly Detection

- **L'anomalia non è una sola osservazione ma un'intera serie storica**

1. Fissare un limite entro il quale etichettiamo il pattern come anomalo
2. Calcolare la frequenza per ogni SAX String



Grazie a tutti!



Nicola Procopio

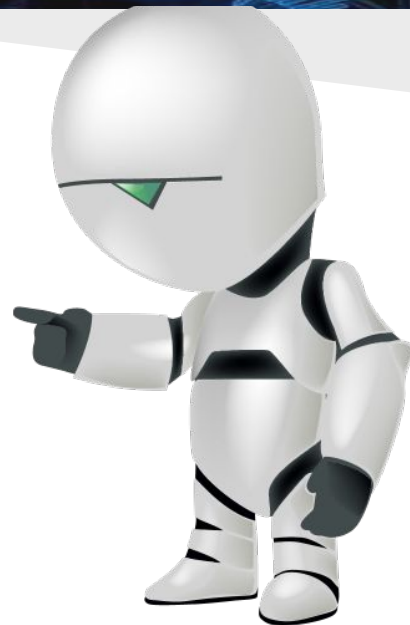


@nickprock



Breve introduzione al DTW

SAX Encoding



Domande?