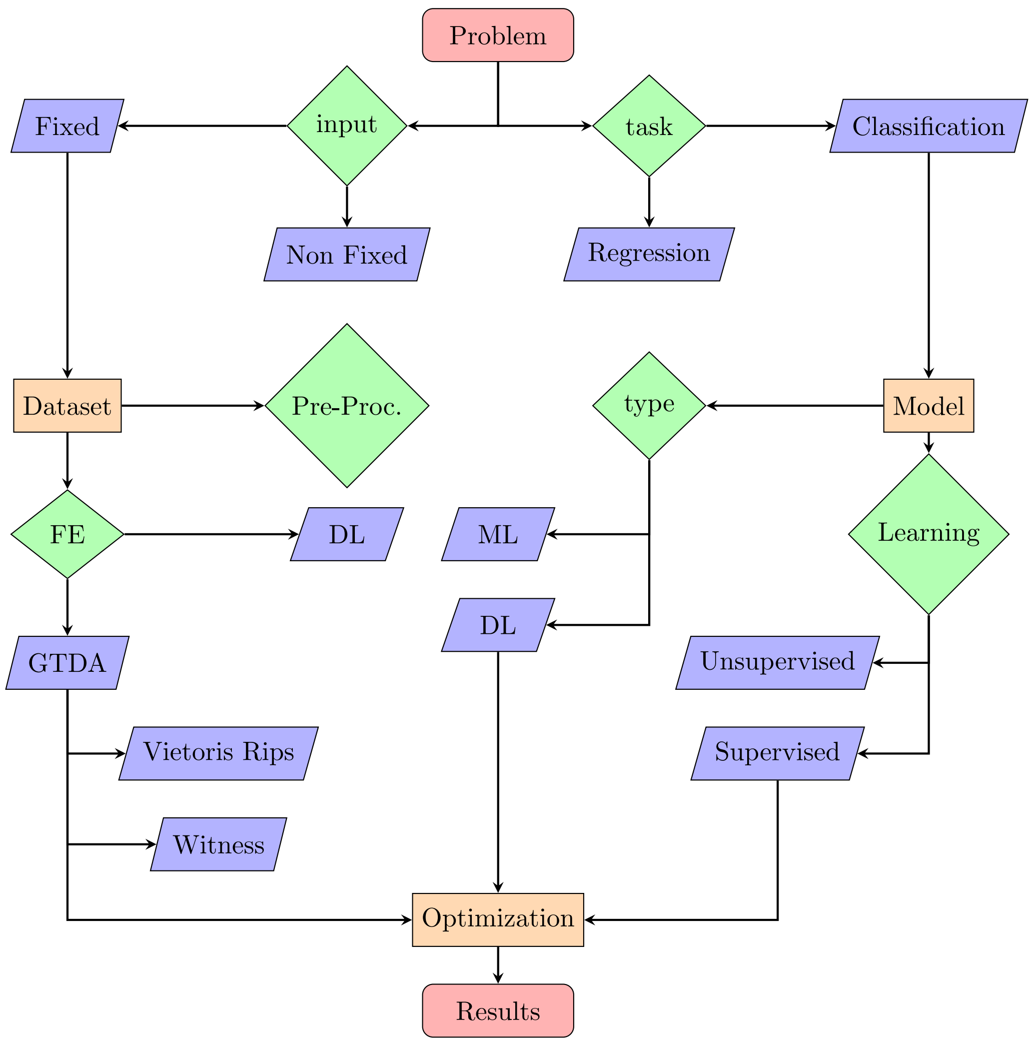
Chapter 1

How to model a Problem



Chapter 2

Naive Approach to TS classification

2.1 First approach

Definition 1 (**Time series**).
A time series (TS) is a finite part of a realisation from a stochastic process $\{X_t, t \in T\}$.

Given a dataset $\mathcal{D} := \{(\underline{x}_i, l_i)\}_i$ of labelled time series where $\underline{x}_i$ is a TS and l_i is the corresponding label. A naive approach to TS classification is to cluster the $\underline{x_i}$ in $|L|$-classes where L is the set of classes in $\mathcal{D}$, by computing the distance among them.

If $|\mathcal{D}| = k$ then computing the distance between all the times series requires $O(k^2 T(S))$ where $T(S)$ is the complexity of computing the distance between TS.

2.1.1 DTW

In theory, computing the distance between two functions is easy. It is sufficient to choose the best L^p and then it is done. However, in practice, it is not so easy since the two functions, the two time series can have different lengths and so it is not possible to compute the distance step-by-step.

Definition 2 (**Alignment path**).
Consider 3 sequence of indexes $I = \{0, \cdots, n-1\}$, $J = \{0, \cdots, m-1\}$ *and* $K = \{0, \cdots, k-1\}$. *An alignment path between* I *and* J *is a function* $\pi : K \to I \times J$ *such that*

- $\pi(0) = (0, 0)$.
- $\pi(k-1) = (n-1, m-1)$
- *The sub sequence of index created by* π, *i.e.* $\{i_k\}_{k \in K} \subset I$ *and* $\{j_k\}_{k \in K} \subset J$, *is monotonically increasing, i.e.* $\forall l \in K$ *it holds that*

$$i_{l-1} \leq i_l \leq i_{l+1}$$
$$j_{l-1} \leq j_l \leq j_{l+1}$$

In order to make it simple π will represent $\pi(K)$ in $I \times J$.

Definition 3 (**Dynamic Time Warping (DTW)**).
Let $S_1 = \{x_i\}_{i \in I}, S_2 = \{x'_j\}_{j \in J}$ *be two sequence of elements in* $\mathbb{R}^d$ *and let* $\mathcal{A}(S_1, S_2)$ *be the set of all path alignment between* I *and* J. *Then the DTW of* S_1 *and* S_2 *is defined as*

$$DTW_q(S_1, S_2) = \min_{\pi \in \mathcal{A}(S_1, S_2)} \left(\sum_{(i,j) \in \pi} \|x_i - x'_j\|_{l^q}^q \right)^{\frac{1}{q}}$$

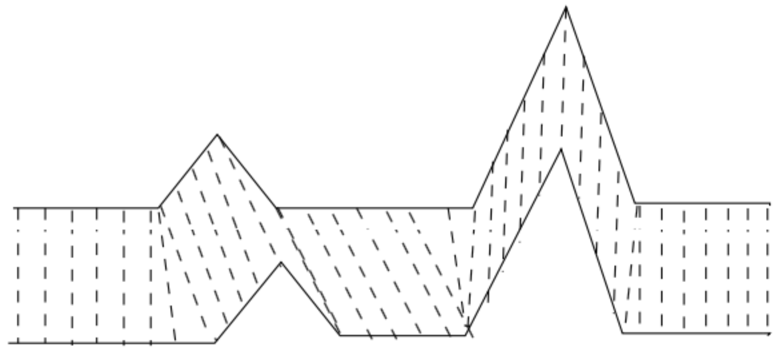


Figure 2.1: The dotted line represents the path alignment between the two-time series given by the two continuous lines

The problem of DTW is caused by the path alignment because the cardinality of $\mathcal{A}(S_1, S_2)$ is highly influenced by $n = |S_1|$ and $m = |S_2|$.
It is possible to see that the problem of the cardinality of $\mathcal{A}(S_1, S_2)$ can be solved by the following dynamic relation

$$D(n,m) = \begin{cases} 0 & m < 0 \vee n < 0 \\ 1 & m = 0 \vee n = 0 \\ D(n-1,m) + D(n-1,m-1) + D(n,m-1) & otherwise \end{cases}$$

It is possible to see that the computational cost of filling the Matrix D is $O(nm)$ if we consider every arithmetic operation to be $O(1)$ but since the number increases exponentially then the whole cost becomes exponential.
This means that this problem is an NP problem.

Chapter 3

GTDA

3.1 From TS to Cloud Point

Previously it was shown how the naive approach cannot be pursued due to the high computational cost. For this reason, now it will be presented the GTDA approach that allows us to extremely reduce the computational cost.

Definition 4 (**Smooth function**).
A function f is said to be smooth on an interval $I \subset \mathbb{R}^n$ if

$$\exists m \in \mathbb{N}| \quad f^{(k)} \in C^0(I) \qquad \forall k \leq m$$

Moreover, f is said to be infinitely smooth or C^∞ if it has continuous derivatives of all orders on I.

For example, the function $f(x) = \sin(x)$ is infinitely smooth or C^∞ because it has derivatives of all orders and each derivative is continuous.

Definition 5 (**Homeomorphic**).
Let X, Y be two topological spaces. It is said that X and Y are homeomorphic if

$$\exists f \in C(X,Y)|\exists f^{-1} \in C(Y,X)$$

Definition 6 (**Embedding**).
An injective continuous map $f : X \to Y$ between topological spaces X and Y is a topological embedding if f yields a homeomorphism between X and $f(X)$ (where $f(X)$ carries the subspace topology inherited from Y).

Definition 7 (**Manifold**).
A d-dimensional differentiable manifold is a subset $M \subset R^d$ if there exists a collection $\{(W_i, h_i)\}_{i \in I}$ such that:

- *$\{W_i\}_{i \in I}$ is cover of M , i.e., $M = \bigcup_i W_i$*
- *$\forall i \in I, h_i : W_i \to U_i \subset R^s$ is a diffeomorphism and the set is called coordinate maps*
- *$\forall j, i \in I, h_j \circ h_i^{-1} : U_i \to U_j$ is smooth*

The inverses, $h_i^{-1} : U_i \to W_i$, are called parametrizations. The collection $\{(W_i, h_i)\}_{i \in I}$ is called atlas of the manifold.

Definition 8 (**Diffeomorphism**).
A map $f : M \to N$ is said to be a diffeomorphism if

- *f is a smooth bijective*
- *M, N are two smooth manifolds.*
- *f^{-1} is smooth*

Sometimes it is useful and more practical to study not the real flow but a reconstruction of it. In particular, it could be useful to study the shadow one:

Theorem 3.1.1 (Takens' embedding theorem).
Let M be a compact manifold of dimension m, ϕ is a smooth vector field on M and y a smooth function then the map

$$\begin{aligned} \Phi_{(\phi,y)} : M &\to \mathbb{R}^{2m+1} \\ x &\mapsto \Phi_{(\phi,y)}(x) = \big(y(x), y(\phi(x)), \cdots, y\left(\phi^{2m}\right)\big) \end{aligned}$$

is an embedding and ϕ is the flow on M

Observation 1. *This theorem, which will not be proved, is very important because it allows to transform a data stream into a cloud point and it preserves the topology of the original flow.*

3.2 GTDA

The core idea of persistent homology is to analyse how holes appear and disappear, as simplicial complexes are created.

Definition 9 (Simplex).
A point $x \in \mathbb{R}^d$ is a convex combination of the points in S if it is an affine combination with non-negative coefficients. The convex hull of S is the set of all convex combinations of S, i.e.,

$$conv(S) := \left\{ \sum_{i=0}^{k} \lambda_i u_i \,\middle|\, \sum_{i=0}^{k} \lambda_i = 1, \lambda \geq 0 \right\}$$

A $k-$simplex is the convex hull of $k+1$affinely independent points in $\mathbb{R}^d$. We say that the simplex is spanned by S and it has dimension k. If S is the set of the $d+1$ unit coordinate points of $\mathbb{R}^{d+1}$, we call it standard $d-$simplex.

Definition 10 (Face).
Let σ be a k-simplex defined by a set $S \subset \mathbb{R}^d$ of $k+1$ affinely independent points. The $k'-$simplex, σ', generated by $S' \subset S$ is said to be a face of σ and

- *If $S' \subseteq S$ then the face is said to be improper, $\sigma' \leq \sigma$.*
- *If $S' \subsetneq S$ then the face is said to be proper, $\sigma' \lneq \sigma$.*

Moreover the boundary of σ is the set of all the proper face of σ, i.e.

$$\partial\sigma := \{\sigma' | \sigma' \lneq \sigma\}$$

Definition 11 (Simplicial complex).
A simplicial complex K is a finite collection of simplexes such that

1. *if $\sigma \in K$ and $\tau \leq \sigma$ then $\tau \in K$.*
2. *if $\sigma_1, \sigma_2 \in K$ then $\sigma_1 \cap \sigma_2 \in K$*

In order to use simplicial complexes it is necessary to define a practical way to construct it.
One of the most used simplicial complexes is the Vietoris-Rips-complexes.

Definition 12 (Vietoris-Rips (VR) complexes).
Let S be a set of points in R^d. The Vietoris-Rips complex of S with radius r is the collection of subsets of S of diameter at most $2r$, i.e.,

$$VR_r(S) = \{\sigma \subset S : diam(\sigma) \leq 2r\}$$

Observation 2. *It is possible to demonstrate that the VR complex is actually a simplicial complex by using the property of the Euclidean norm induced by the space $\mathbb{R}^d$.*

It is possible to see that this kind of topological construction requires all the points in the set S and for this reason it can become very expensive and slow from a computational point of view.
In order to get around this problem it is possible to build a simplicial complex from a point cloud Z and chose a sub-set of vertices $L \subset Z$ which is called Landmark points. In particular, the construction depends only on the matrix $D := D(L, Z)$ of distances between the landmark points and the data points.

Observation 3. *The metric that is used for D is not fixed but it must be suitable for the problem. For example in a graph structure the distance can be given by the shortest path distance.*

Definition 13 (Witness complex).
Let D be $n \times N$ matrix of non-negative entries regarded as the matrix of distances between a set of n landmarks and N data points. The (strict) witness complex $W_\infty(D)$ with vertices $\{1, \cdots, n\}$ is defined as

$$\sigma := [a_0, \cdots, a_p] \in W_\infty(D) \iff \exists i \in \{1, \cdots, N\} | D(a_j, i) \leq D(b, i) \qquad \forall b \in \{1, \cdots, N\} \setminus \{a_1, \cdots, a_n\}$$

Moreover, i is called the witness.

Definition 14 (Lazy Witness complex).
Let D be $n \times N$ matrix of non-negative entries regarded as the matrix of distances between a set of n landmarks and N data points. The lazy witness complex $W_1(D) \supset W_\infty(D)$ with vertices $\{1, \cdots, n\}$ is defined as

- *$W_1(D)$ has the same $1-$skeleton as $W_\infty(D)$*
- *$\sigma = [a_0 a_1 \cdots a_p] \in W_1(D) \iff$ all the edges are in $W_1(D)$*

3.2.1 Homology

Up to now, nothing give us numerical information about the cloud point.

Definition 15 (p-chain).
Let K be a simplicial complex and A an abelian group. A p-chain is a formal sum of p-dimensional faces of K, i.e.

$$C_p(K) = \left\{ \sum_{\sigma \in F_p(K)} a_\sigma \sigma | a_\sigma, \in A \right\}$$

At this point, it is possible to define the boundary function between two p-chain, $C_p(K) \xrightarrow{\partial_p} C_{p-1}(K)$.

Definition 16. *Let K be a simplex. A p-cycle $Z_p(K)$ is a chain with empty boundary*

$$Z_p(K) := \{\sigma \in Ker(\partial_p(C_p(K)))\}$$

A p-boundary $B_p(K)$ is a chain which is a boundary of a (p+1)-chain

$$B_p(K) := \{\sigma \in C_p(K) | \exists \sigma' \in C_{p+1}(K), \partial_{p+1}\sigma' = \sigma\}$$

From a result, it is known that $\partial_p \circ \partial_{p+1} = 0$, for this reason from the definition of p-boundary it holds that

$$\forall \sigma \in B_p(K), \exists \sigma' \in C_{p+1}(K) | \partial_{p+1}\sigma' = \sigma$$

thus it holds that

$$0 = \partial_p(\partial_{p+1}\sigma) = \partial_p \sigma \implies \sigma \in Z_p(K)$$

so $B_p(K) \subset Z_p(K)$. This relationship makes the following definition make sense

Definition 17. *Let G be a space then*

$$\mathrm{rank}(G) = \min\{|X| \, | X \subset G, \langle X \rangle = G\}$$

Definition 18 (p-th homology group).
Let K be a simplicial complex. The p-th homology group is

$$H_p(K) := {}^{Z_p(K)}\!\big/_{B_p(K)}$$

and $\beta_p(K) := \mathrm{rank}\,(H_p(K))$ is the p-th Betti number.

Definition 19 (Filtration).
A filtration $\mathcal{F}$ is an indexed family $\{S_i\}_{i \in I}$ of sub-objects of a given algebraic structure S such that

$$S_i \subseteq S_j \qquad \forall i, j \in I, i \leq j$$

Let K be a simplicial complex and let $f : K \to \mathbb{R}$ be a monotonic function, i.e. if $\sigma \subset \tau$ then $f(\sigma) \leq f(\tau)$. The induced filtration of K is $\{K_i\}_{i=0}^{n}$ where $K_0 = \phi, K_n = K$ and

$$a_i := \min\left\{r \,\middle|\, K_i = f^{-1}((-\infty, r])) \right\}$$

By construction it holds that $\forall i \leq j, K_i \subset K_j$ then it is possible to define $f_p^{i,j} : H_p(K_i) \to H_p(K_j)$ where $f_p^{i,j}$ is the induced map.

Definition 20. *The p-th persistent homology groups are the images of the homomorphism induced by inclusion so*

$$H_p^{i,j}(K) = imm f_p^{i,j}$$

Moreover, the ranks of these groups are the p-th persistence Betti numbers

$$\beta_p^{i,j} = \mathrm{rank}\left(H_p^{i,j}(K)\right)$$

Definition 21 (life-and-death).
Give the p-th homology group $H_p(K_i)$ and $\gamma \in H_p(K_i)$, it is said that

- *γ is born at K_i if $\gamma \notin H_p^{i-1,i}$*
- *γ dies in K_j if it merges with an older class as we go from K_{j-1} to K_j.*

Moreover if a_i, a_j previously defined, the persistence of γ is

$$pers(\gamma) = a_j - a_i$$

and (a_j, a_i) is called the persistent pair.

Observation 4. *Given a certain filtration, the persistence homology is given, but if the filtration is not, then it is possible to create it by increasing the balls' radius.*

The number of generated simplicial complexes can be exponential in the number of input points for VR filtration. For this reason, it will be used the Lazy witness Algorithm. The Lazy one is because it does not use all the points to compute the complexes. The points not used are witnesses to the edges or simplices spanned by a combination of landmarks.

3.3 Features selection

Using Takens' embedding theorem it is possible to generate a cloud point and in this case, it has been taken as a step of the embedding theorem 3.
In order to take a constant number of features for each point cloud, it will be consider the following one:

- The number of holes in each dimension (in the 0-th dimension the hols actually indicate connected component)
- Maximum holes life-time in each dimension
- Number of relevant holes which is the fraction of points that have at a similar distance
- Average lifetime of all hoes in dimension d.
- sum of all lifetime. The integral of the PD graph

In a dataset, the sum close to zero in a particular dimension indicate that it has practically no holes in that dimension.

Observation 5. *This part is both the easiest to understand but at the same time is the most time-consuming from a practical/computational point of view since calculating all the various features for each cloud point is computationally expensive.*
Here below it is possible to see how much time did it take to compute the feature extraction from the training dataset using a multi-core implementation on Processore Intel Xeon Gold 6226R

```
(gtda) achelini@frontend:~/Projects/GTDA$ cat err.txt
  0%|          | 300/76610 [05:05<21:36:16,  1.02s/it](gtda)
```

3.4 Classifier

Since the features have been selected it is necessary to define a way to classify all the time series. In order to do so it will be used a simple feed-forward neural network.

Definition 22 (Neural network (NN)).
We call a sequence $\mathbb{N} = \left(D_0, D_1, \cdots, D_L, W^1, \theta^1, W^2, \theta^2, \cdots, W^L, \theta^L\right)$ *neural network where:*

- $L \in \mathbb{N} \setminus \{0\}$ *and it represent the depth of* $\mathbb{N}$.
- $\left(D^0, \cdots, D^L\right) \in \mathbb{N}^{L+1}$ *is the layout of the network and* D_0 *is the number of features of the input and* D_L *is the number of the features of the output.*
- $W^l = \left(W^l_{j,k}\right) \in \mathbb{R}^{D_l \times D_{l-1}}$ *matrices whose entries are referred to be the network's weights and* j *is referred to the* $j-th$ *nodes in the* D^l *layer while* k *to the* $k-th$ *node of the* D^{l-1} *layer.*
- $(\theta^l) = \left(\theta^l_j\right) \in \mathbb{R}^{D_l}$ *where* $l \in \{1, \cdots, L\}$ *and it is called the bias.*

Definition 23 (Forward propagation).
Given a neural network $\mathbb{N}$ *and* $\rho = \{\rho_i\}_{i=1}^{L}$ *the set of activation function associated at each level of* $\mathbb{N}$. *The function defined by* $\mathbb{N}$ *equipped with* ρ *is defined as*

$$< \mathcal{N} >^{\rho} \colon \mathbb{R}^{D_0} \to \mathbb{R}^{D_L}$$
$$x \mapsto < \mathbb{N} >^{\rho} (x) = y^{[L]}$$

where

$$\begin{cases} y^{[0]}(x) = 0 \\ y^{[l]}(x) = \rho\left(W^l y^{[l-1]}(x) + \theta^l\right) \quad l \in \{1, \cdots, L\} \end{cases}$$

and this system defines forward propagation.