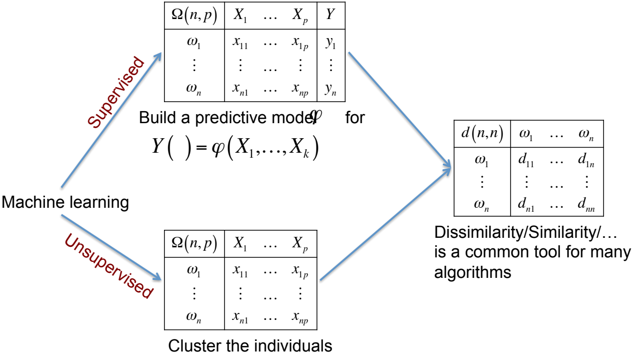
New insights in topological learning

D. A. Zighed, F. Ricco, R. Abdeslam

University of Lyon (Lumière Lyon 2) - CNRS

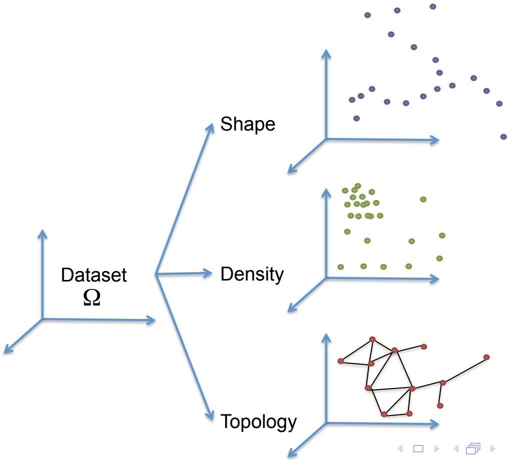
Beijing - China - 27/29 October 2011

Machine learning



Topological learning

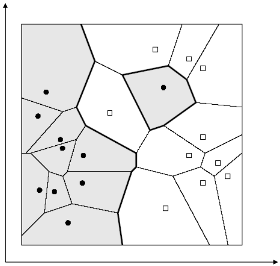
Focuses on other aspects that only density.



Framework : Topological Graphs

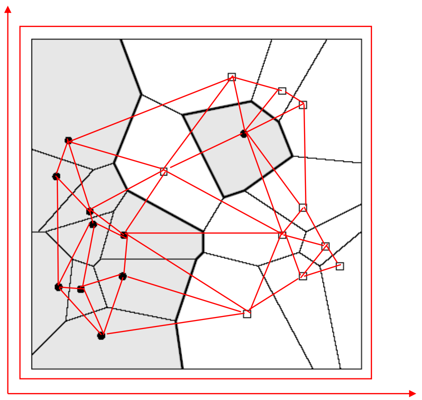
Let us assume that the feature space is $R = IR^p$ and we have 2 class-problem.
There are plenty of ways to define the topology of the learning the dataset.

Topology of Voronoi's Diagram



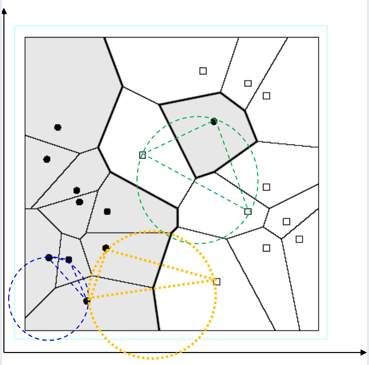
- Feature space is partitioned by the dataset; each part defines the area of influence;
- Two points are neighbors if they share a common border;
- the graph brought about by the links between neighbors is the Delaunay's Polyhedron.

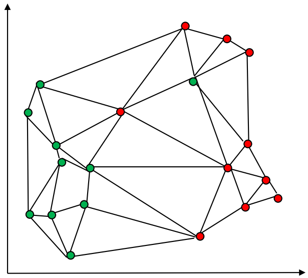
Topology of Delaunay's polyhedron



Property: all set of $P + 1$ neighbors of the p-dimensional space are on tangents of an empty hypersphere.

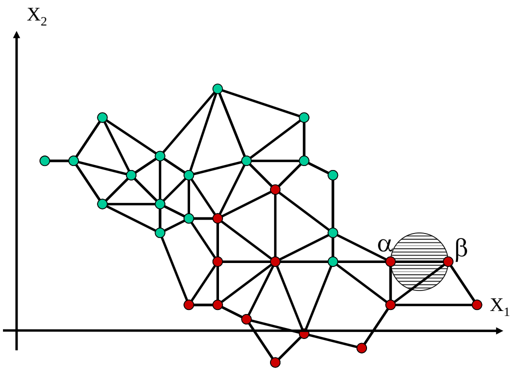
Topology of Delaunay's polyhedron





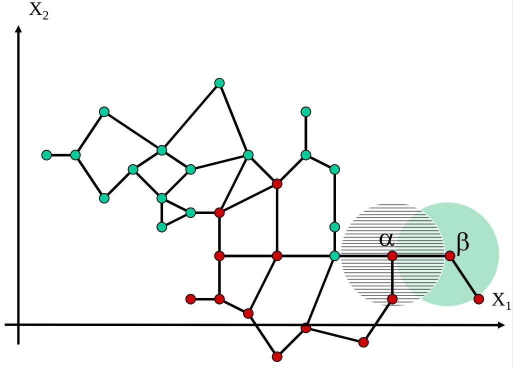
- Building Delaunay's graph or Vornoi's Diagram is intractable in high dimension feature space
- Delaunay's Graph is a **related graph**

Gabriel's Graph (GG)



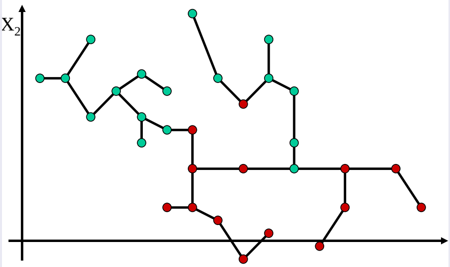
- Gabriel's Graph is a **related graph**
- It feasible $O(n^2)$ even in high dimension space

Relative Neighborhood Graph (RNG)



- Relative Neighborhood Graph is a **related graph**
- RNG ⊂ GG ⊂ DG

Minimum Spanning Tree (MST)



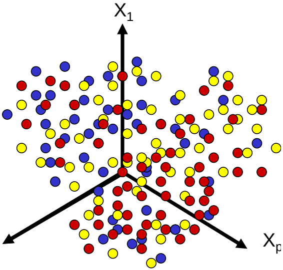
- MST is a **related graph**
- MST $\subset$ RNG $\subset$ GG $\subset$ DG

Learnability

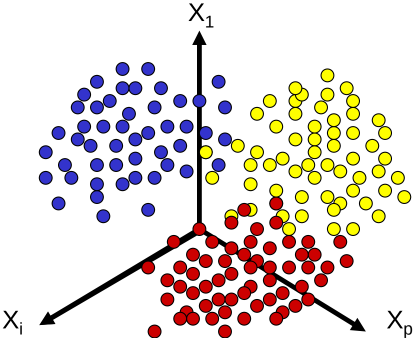
- Definition

The classes are not LEARNABLE if the learning data set in the feature space have been randomly labeled: $P(c_i/X) = P(c_i)$

Example :

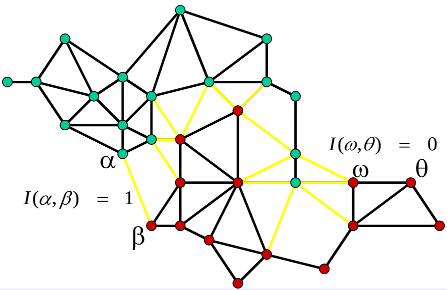


In such case, the underlying problem of machine learning is not



In that case, the classes are separable, therefore There exists, potentially, a machine learning algorithm capable to produce a reliable model φ, consequently, we can launch the screening process.

Statistic of the cut edges



- $I = 14$ couples belonging to two different classes
- $J = 61$ couples belonging to the same class
- $P_J = \frac{I}{I+J} = 18,6\%$; $1 \leq P_J < 7n$

What would be this proportion in random labeling ?

Distribution of I and J under the null hypothesis

H_0: The vertices of the graph are randomly labeled according to the same probability π_k for the class k, $k = 1, \ldots, K$. We have established in

- Zighed et al. (2002) "Separability Index in Supervised Learning", LNAI 2431, pp. 475-487, .
- Zighed et al. (2005) "A statistical approach of class separability", App. Stochastic Models in Bus. and Ind., Vol. 21, No. 2, , pp. 187-197.

the law of I and J for K classes.

Comparing proximity measures

- Proximity measure = dissimilarity/similarity/ressemblance/...

- In many domains, such as information retrieval, clustering, classification... the choice of a proximity measure plays a key role in the final result.

- There are dozens of proximity measures
- Are they all equivalent ?
- How can we differentiate them ?

Comparison based on preordonnance/topology

Definition of equivalence in preordonnance

Let us consider two proximity measures u_i and u_j to compare. If for any quadruple (x, y, z, t), we have:
$u_i(x, y) \leq u_i(z, t) \Rightarrow u_j(x, y) \leq u_j(z, t)$ then, the two measures are considered equivalent.

$S(u_i, u_j)$ is an index of similarity between proximity measures.

$S(u_i, u_j) = \frac{1}{n^4} \sum_x \sum_y \sum_z \sum_t \delta_{ij}(x, y, z, t)$

where $\delta_{ij}(x, y, z, t) =$

$$\begin{cases} 1 \text{ if } [u_i(x, y) - u_i(z, t)] \times [u_j(x, y) - u_j(z, t)] > 0 \\ \text{or } u_i(x, y) = u_i(z, t) \text{ and } u_j(x, y) = u_j(z, t) \\ 0 \text{ otherwise} \end{cases}$$

$S \in [0, 1]$ and the complexity : $O(n^4)$

Definition based on topological graphs

To each proximity measure u_i we can associate a neighborhood graph V_i from which we can say that two proximity measures u_i and u_j are equivalent if the topological graphs V_i and V_j induced are the same.

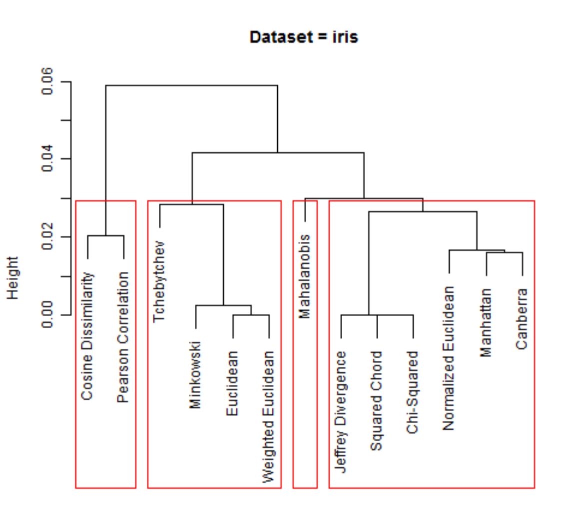
$$S(u_i, u_j) = \frac{1}{n^2} \sum_{x \in \Omega} \sum_{y \in \Omega} \delta_{ij}(x, y)$$

where $\delta_{ij}(x, y) = \begin{cases} 1 \text{ if } V_{u_i}(x, y) = V_{u_j}(x, y) \\ 0 \text{ otherwise} \end{cases}$

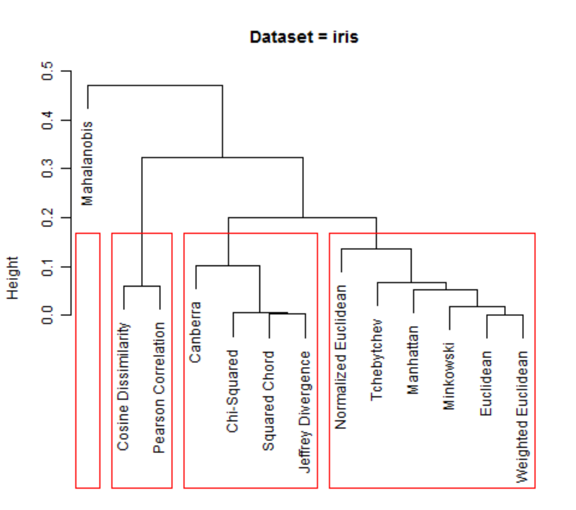
Some results

- If it exists a strictly monotonic function f such that $u_i = f(u_j)$ then if the preorder is preserved this implies that the topology is preserved and vice versa.
- In the context of topological structures induced by the graph of relative neighbors, if two proximity measures u_i and u_j are equivalent in preordonnance, they are necessarily topologically equivalent.
- Both approaches give different results and they are, generally, sensitive to the dataset.

Dendogram for topological comparison



Dendogram for preordonance comparison



Some references

- Batagelj, V., Bren, M.: *Comparing resemblance measures*. In Journal of classification 12 (1995) 73-90
- Lerman,I.C.:*Indice de similarité et prtéordonnance associtée, Ordres*. In Travaux du stéminaire sur les ordres totaux finis, Aix-en-Provence (1967)
- Djamel Abdelkader Zighed, Rafik Abdesselam, Ahmed Bounekkar: *Equivalence topologique entre mesures de proximité*. EGC 2011: 53-64
- Djamel Abdelkader Zighed, Rafik Abdesselam, Ahmed Bounekkar: *Topological comparisons of proximity measures*, **Submitted**

Topological random classification

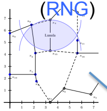
Learning sample

K-random sampling on the learning sample (n,p)

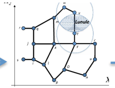
K-Classifier based on the neighborhood graph (RNG)

Aggregation Function (Decision Template)

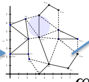
$\Omega(n,p)$	X_1	…	X_p	Y
ω_1	x_{11}	…	x_{1p}	y_1
$\vdots$	$\vdots$	…	$\vdots$	$\vdots$
ω_n	x_{n1}	…	x_{np}	y_n

$\Omega(n_1, p_1)$ → 

φ_1

$\Omega(n_k, p_k)$ → 

φ_k

$\Omega(n_K, p_K)$ → 
φ_K

→ Φ

$$Y(\omega) = \Phi(\varphi_1(\omega), \ldots, \varphi_k(\omega), \ldots, \varphi_K(\omega))$$

Evaluation

TRC has been compared to
- *kNN* with k = 1, 2, 3.
- Decision tree/CART : random forests (RFs),
- SVM : K support vector machines (KSVMs),
- Adaboost,
- Discriminant analysis (DA),
- logistic regression (RegLog)
- C4.5.

All was done with R software.
- We used 14 quantitative data sets from UCI repository.
- We ran the same protocol over all the methods mentioned
- For each experiment, we applied 10-Cross Validations

Results

Algorithm	Average rank / X validation
TRC	2.88
Random Forest	3.19
Ksvm	4.04
1-NN	4.15
3-NN	4.58
AdaBoost	5.06
LDA	6.58
2-NN	7.04
C4.5	7.46
Log. Reg	7.56

Some references

- Fabien Rico, Djamel Abdelkader Zighed: *Classificateurs aléatoires topologiques à base de graphes de voisinages*. EGC 2011: 83-88
- Fabien Rico, Djamel Abdelkader Zighed and D. Azzedine: *Neighborhood Random Classification*, **submitted**

- Working in the topological framework generates new issues and provide some efficient tools to address some basic question in machine learning
- We are just opening the door : many works are undergoing : feature selection, building an efficient representation space, discrimination without an explicit raws/Columns data (social networks), testing other definitions of topology, working on the shape of data...

Thank you