Colloque
Apprentissage statistique : Théorie et application

A statistical approach for separability of classes

Djamel A. ZIGHED
Stéphane LALLICH
Fabrice MUHLENBACH

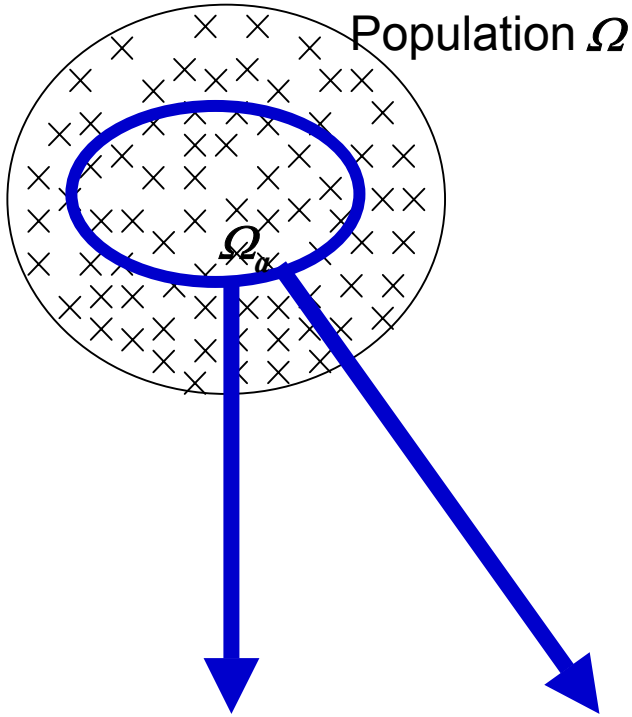
Laboratoire ERIC - Lyon
Université Lumière Lyon 2
5, avenue Pierre Mendès-France
69676 BRON Cedex
FRANCE

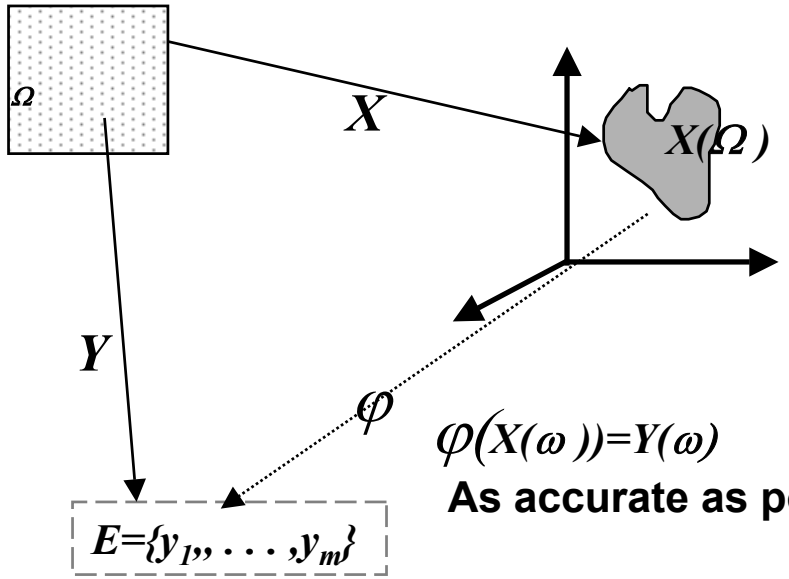
zighed@univ-lyon2.fr
lallich@univ-lyon2.fr
muhlenba@eric.univ-lyon2.fr

Overview

- General framework
- Class separability
- Neighborhood graph and clusters
- Cut weighted edges statistic
- Experiments
- Conclusions and future work

General framework





Representation space

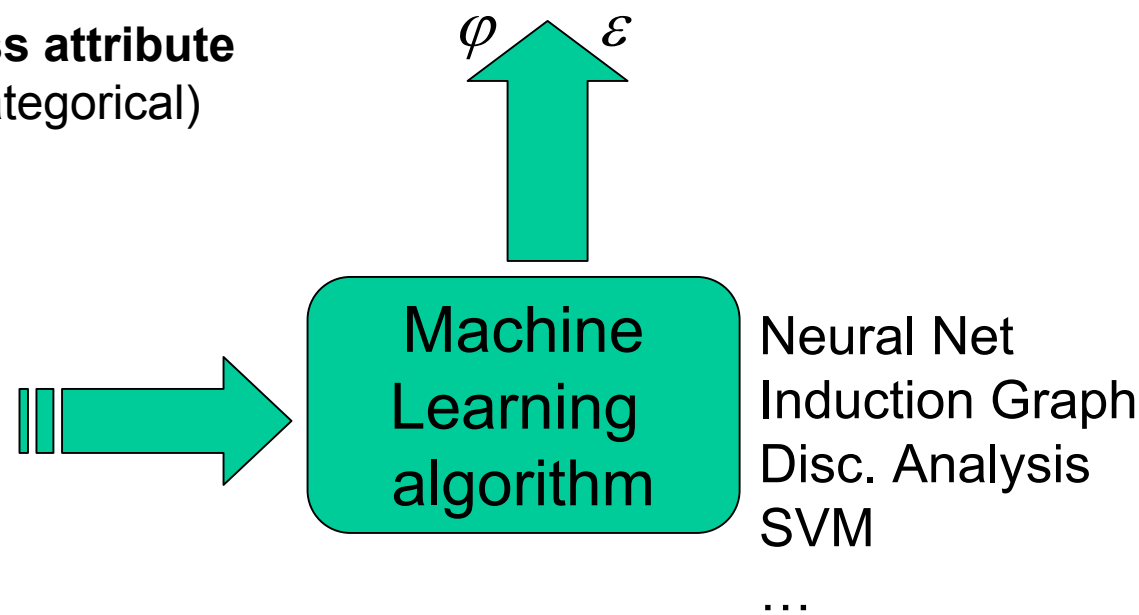
Class attribute
(categorical)
Y

$(X_1, X_2, X_3, \ldots, X_p)$

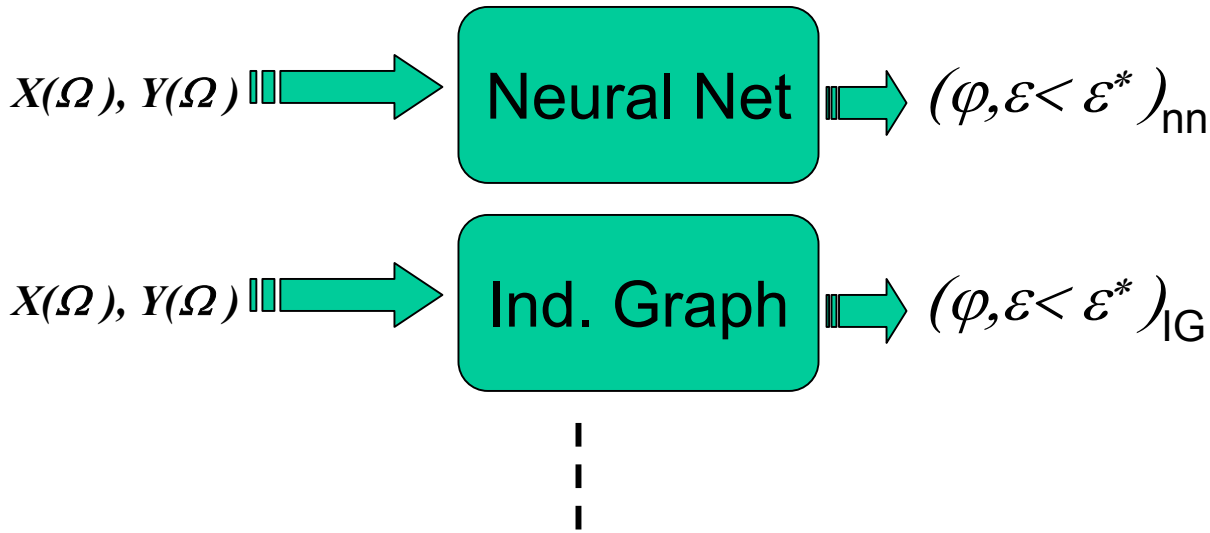
Labeled examples

70	1	4	130	322	0	2	109	0	2,40	2	3	3	2
67	0	3	115	564	0	2	160	0	1,60	2	0	7	1
57	1	2	124	261	0	0	141	0	0,30	1	0	7	2
64	1	4	128	263	0	0	105	1	0,20	2	1	7	1
74	0	2	120	269	0	2	121	1	0,20	1	1	3	1
65	1	4	120	177	0	0	140	0	0,40	1	0	7	1
56	1	3	130	256	1	2	142	1	0,60	2	1	6	2
59	1	4	110	239	0	2	142	1	1,20	2	1	7	2
60	1	4	140	293	0	2	170	0	1,20	2	2	7	2

Predictive attributes
(continuous)



Assume that we aim to find a prediction model φ that fits the data with an accuracy rate $\varepsilon > \varepsilon^*$, whatever the machine learning algorithm.

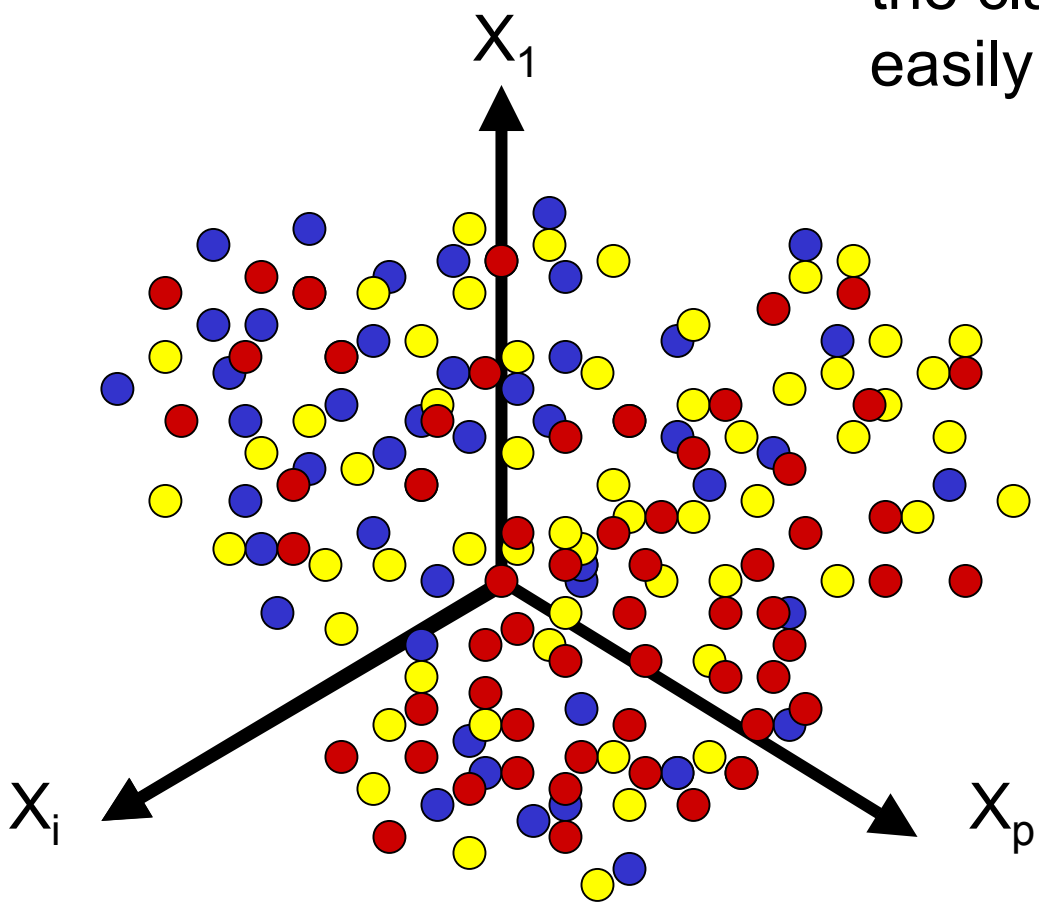


We failed finding out such prediction model

- All the MLA aren't suitable for this specific problem, so we have to look for a new MLA,… until…

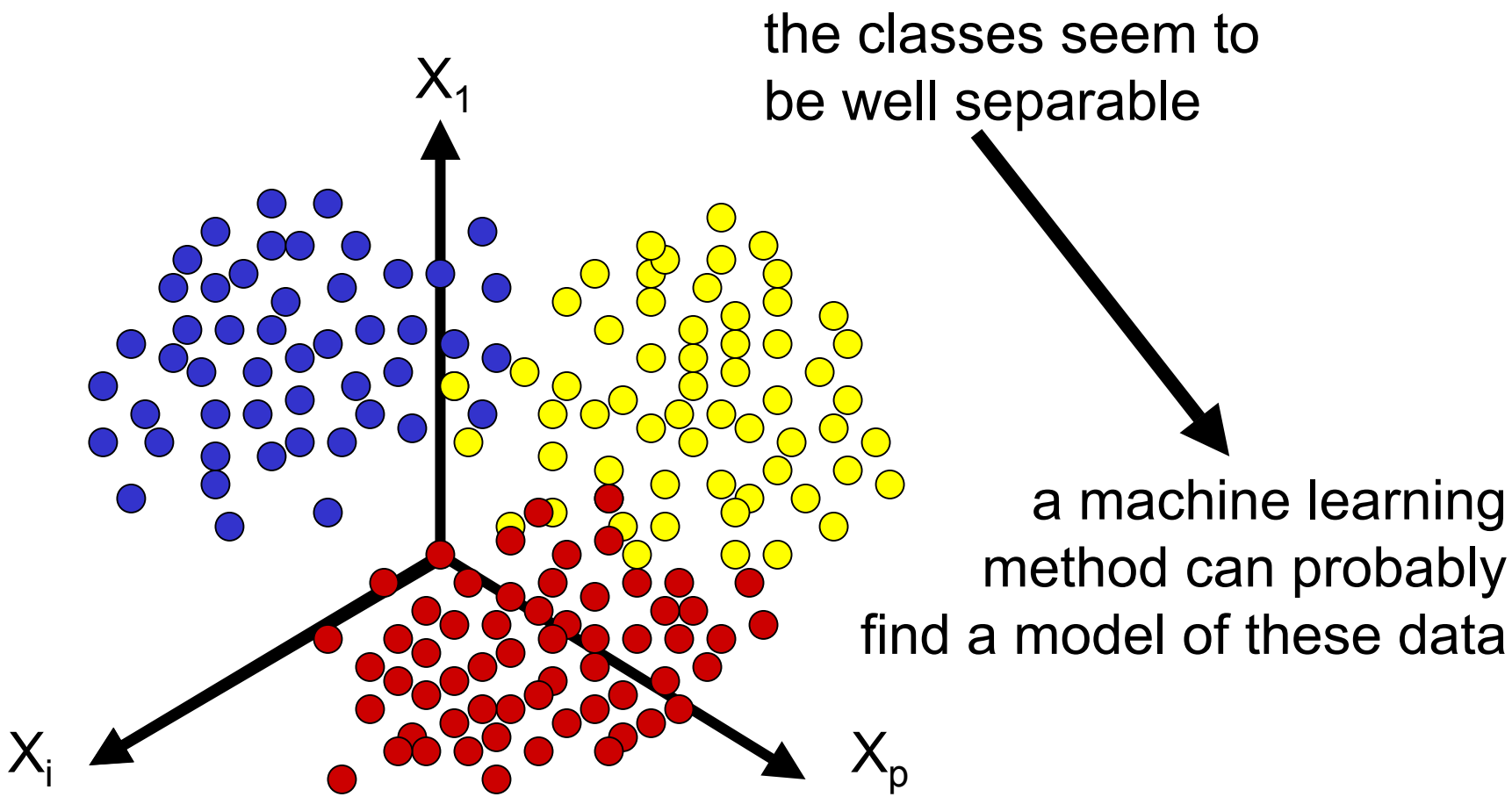
- The classes aren't separable.

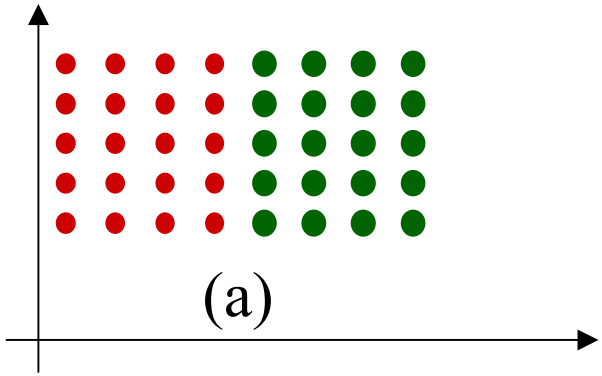
Class separability



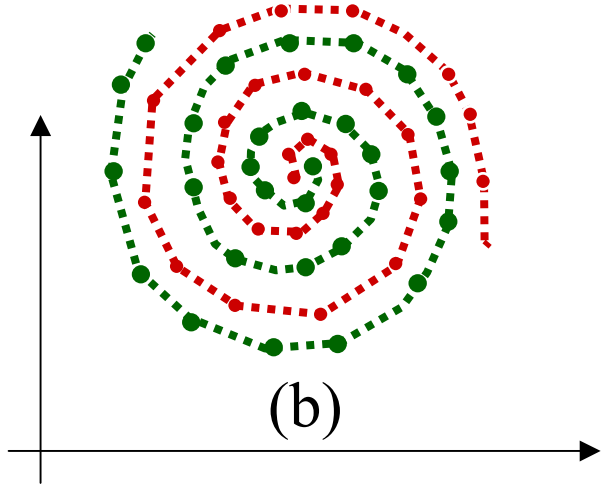
the classes are not easily separable

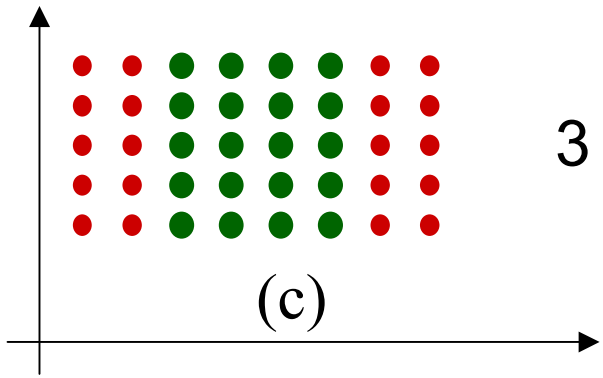
it will be difficult to find a method that can produce a reliable model


the classes seem to
be well separable
X_1
X_i
X_p
a machine learning
method can probably
find a model of these data

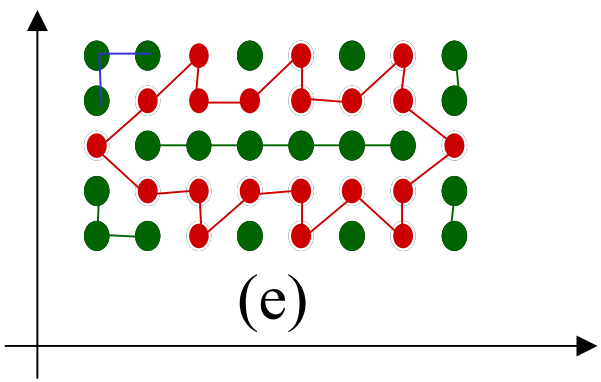


2 clusters

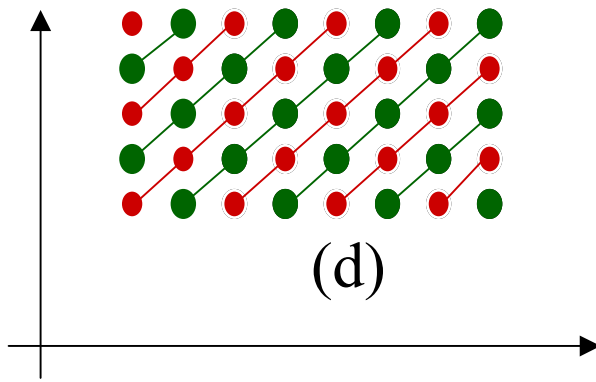




3 clusters

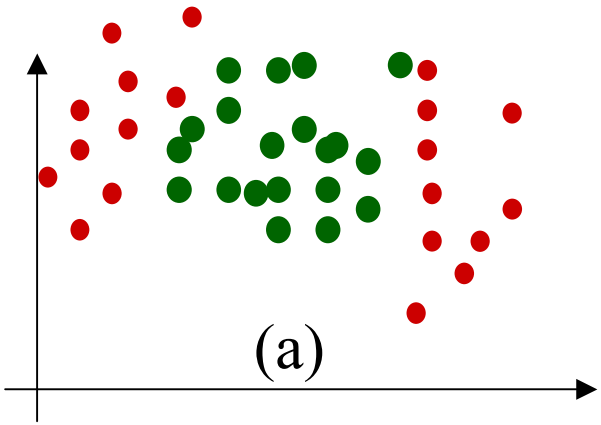


12 clusters

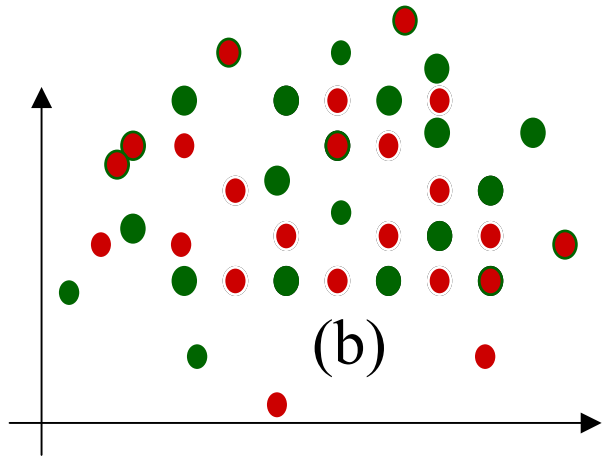


The separability of classes seems to be linked to

- The number of homogenous clusters
- The size of each cluster

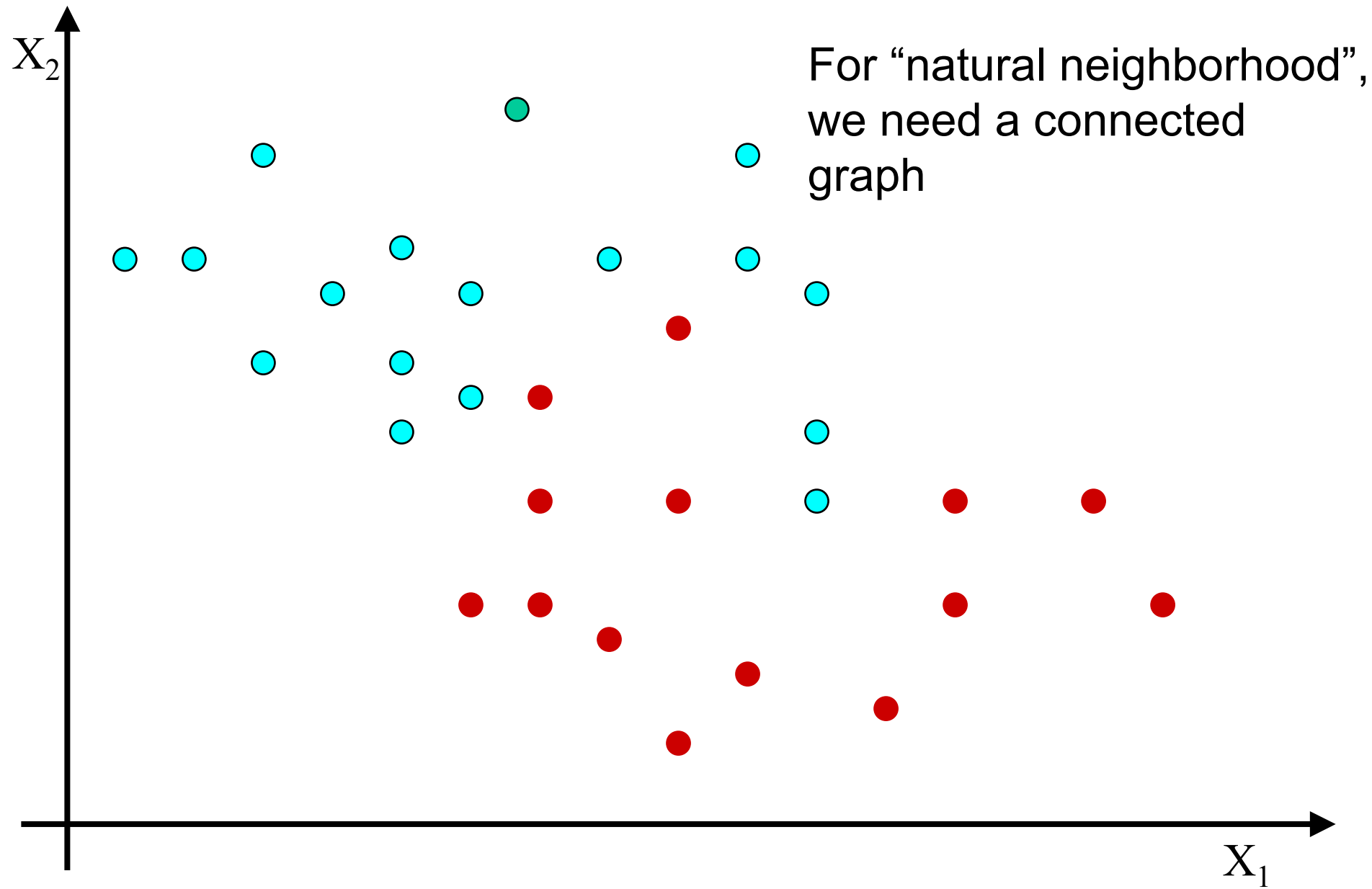


Few homogenous clusters

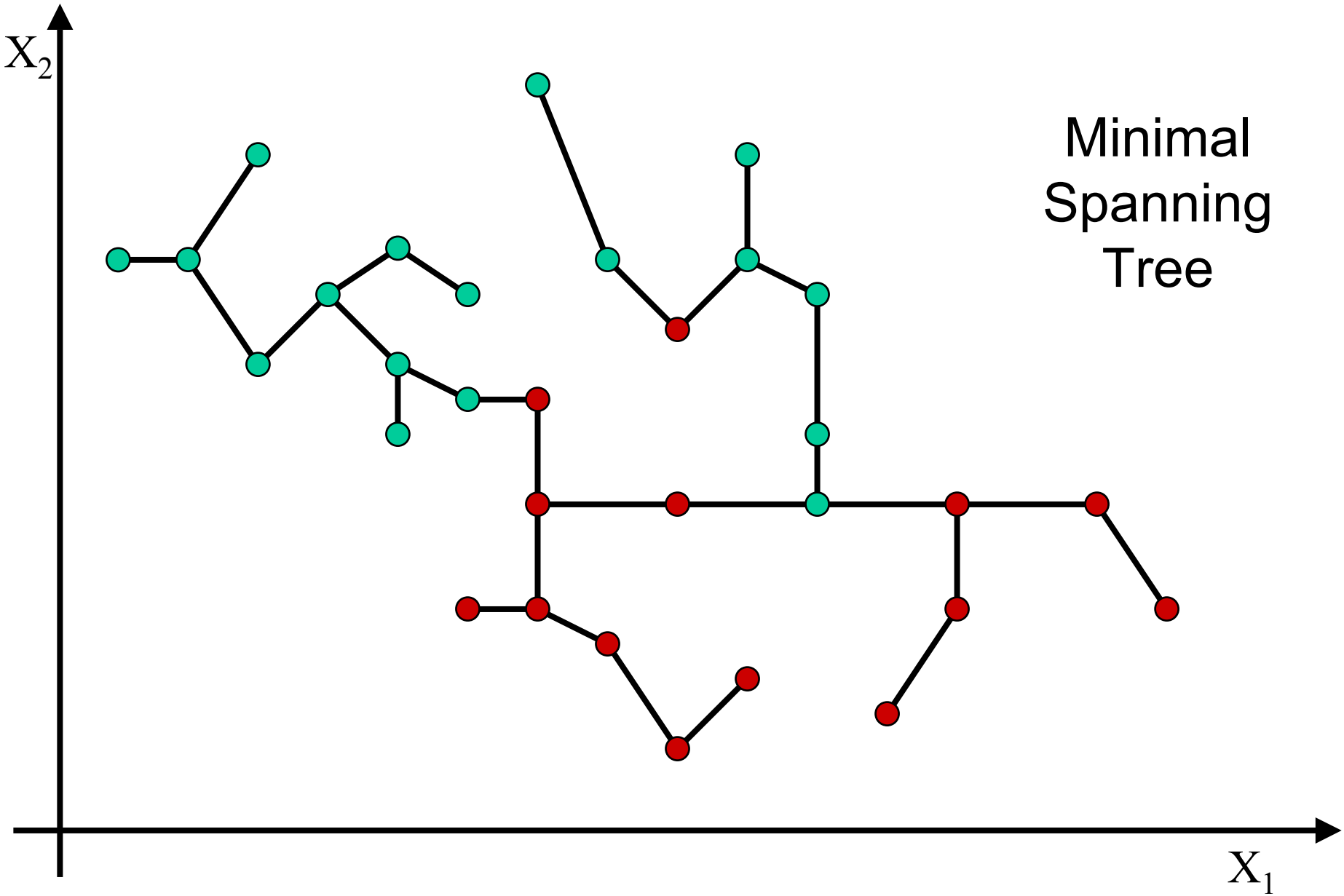


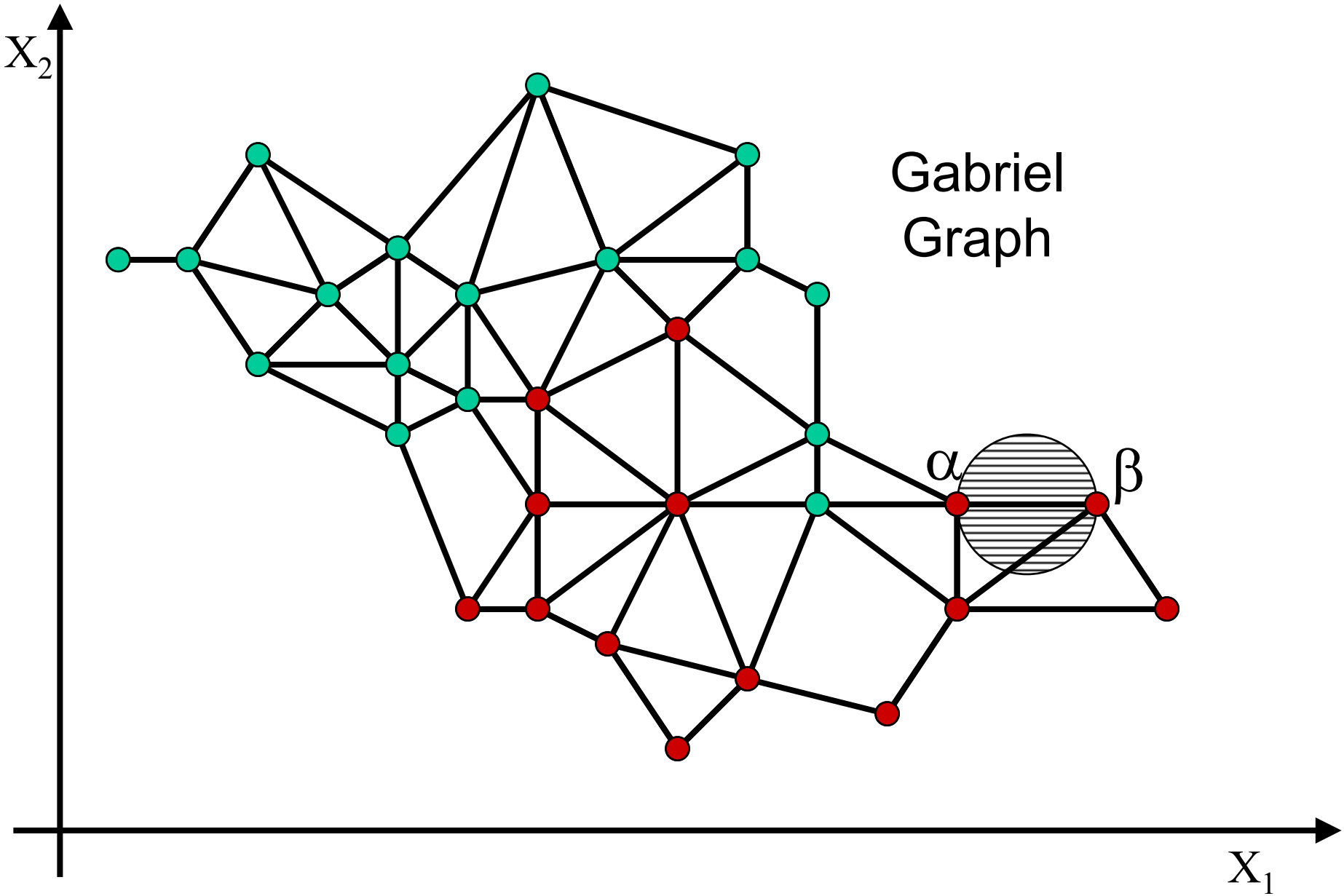
Many homogenous clusters

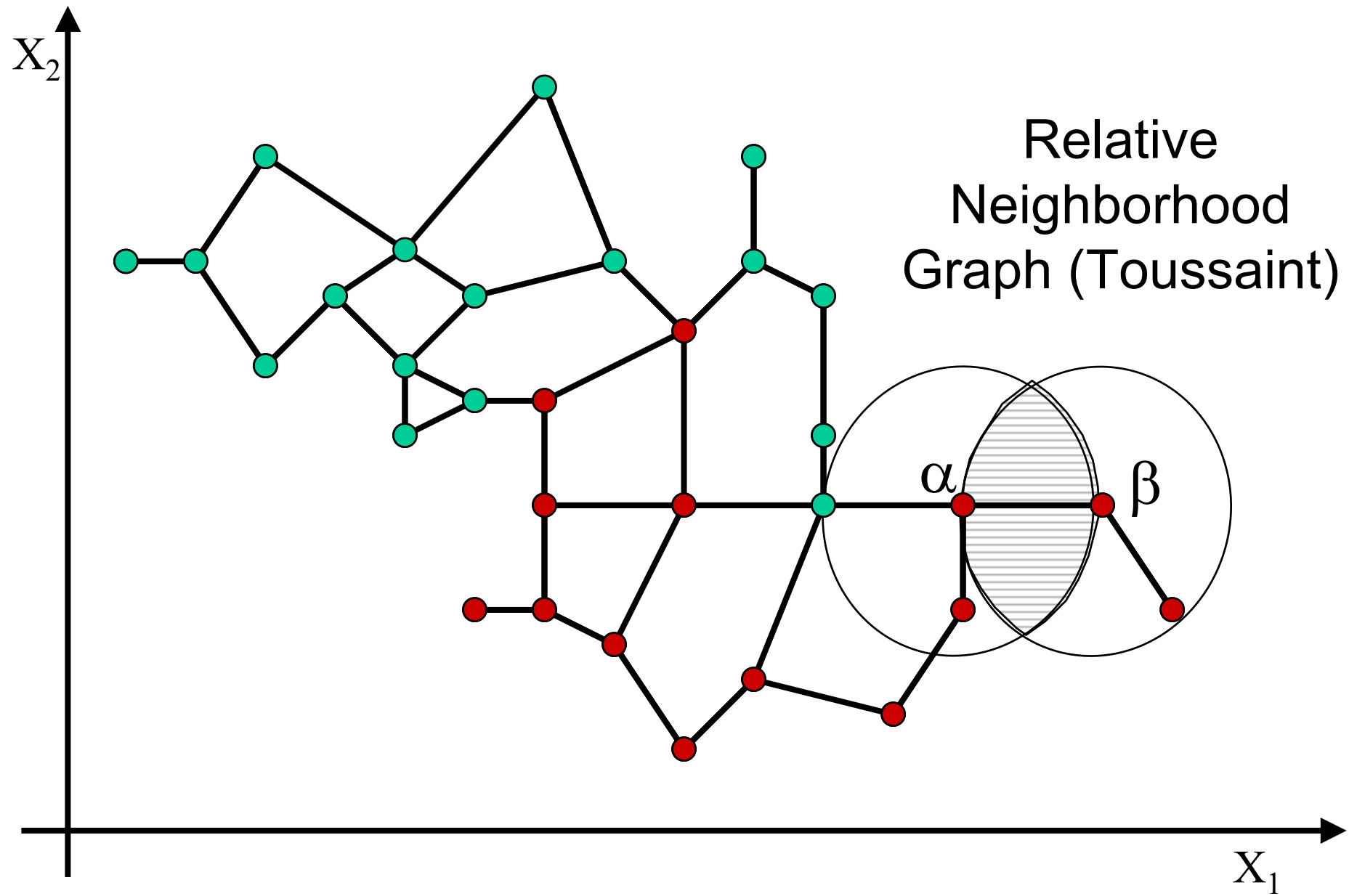
Neighborhood graph and clusters



For “natural neighborhood”, we need a connected graph

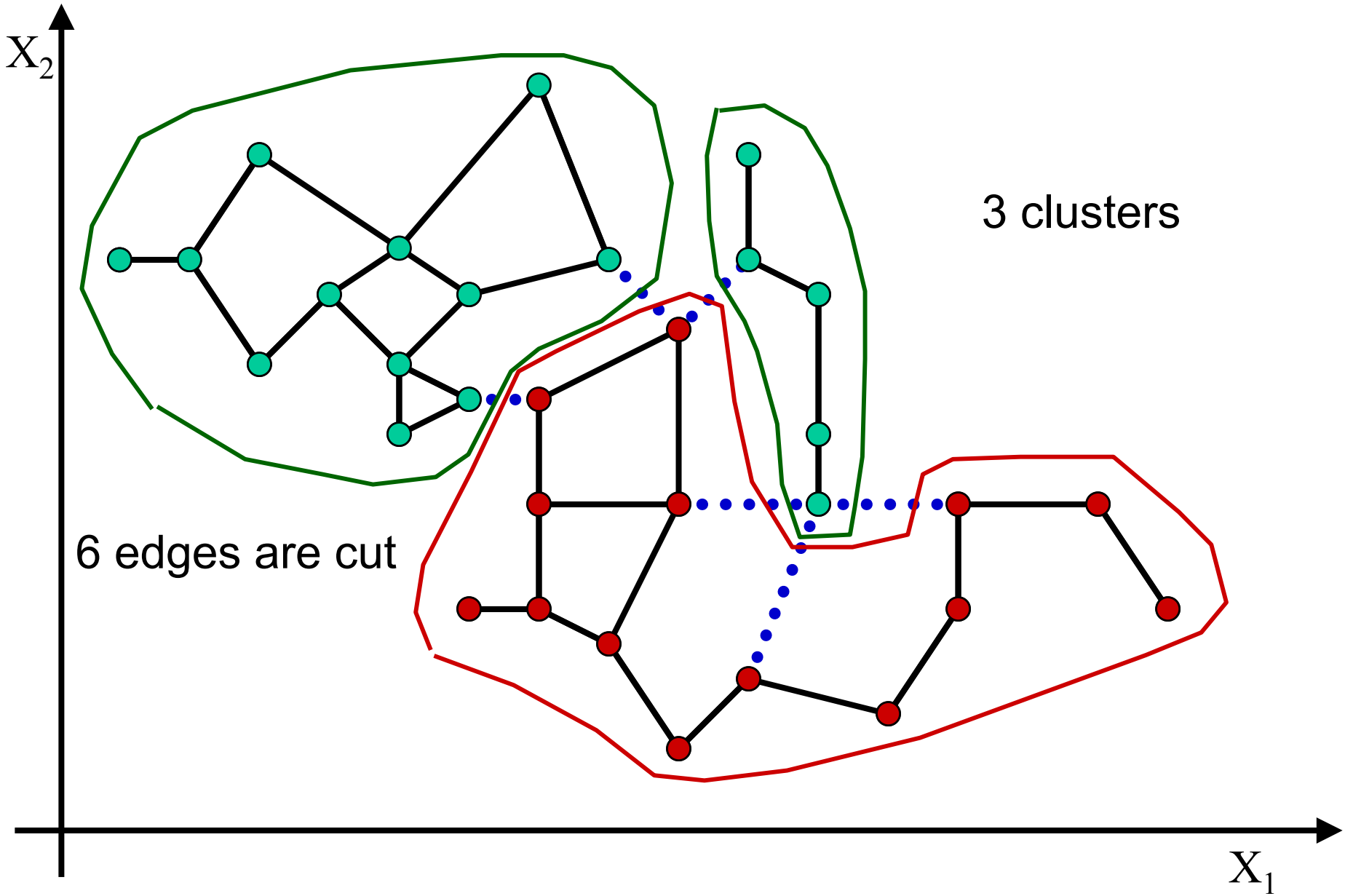

X2
Minimal
Spanning
Tree
X1


X2
Gabriel
Graph
α
β
X1


X2
Relative
Neighborhood
Graph (Toussaint)
α
β
X1

There are many other possibilities to build such neighborhood graphs, for instance :

Voronoi Diagram,
Delaunay triangulation,

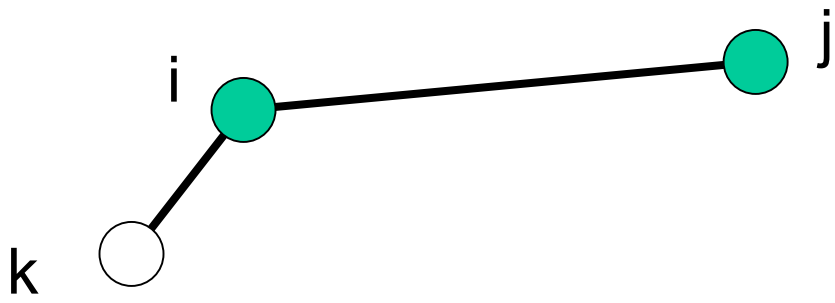

X2
3 clusters
6 edges are cut
X1

Cut weighted edge statistic

Weighted edges

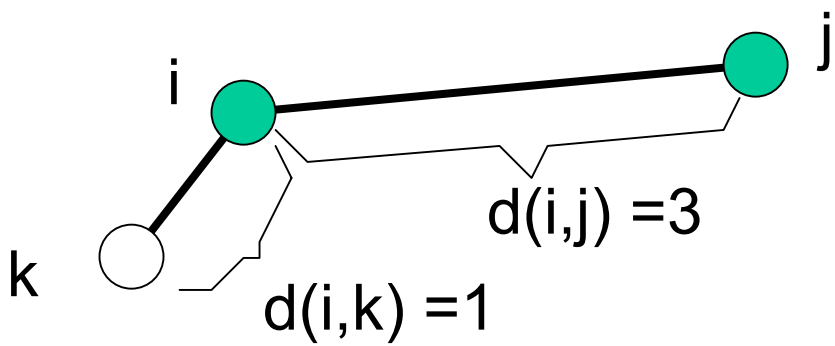
- Connection:

$w_{i\,j} = 1$
$w_{i\,k} = 1$
$w_{j\,k} = 0$ (no edge)



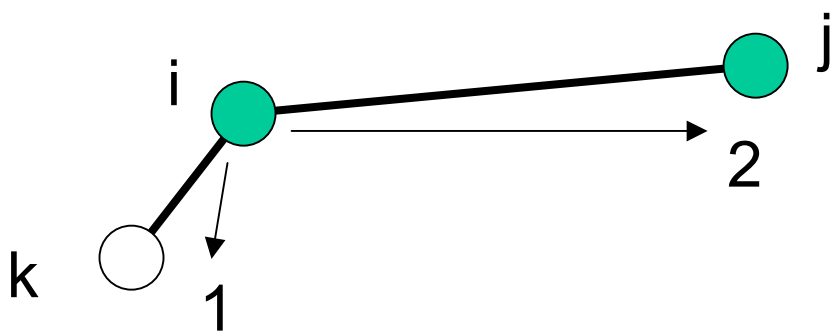
- Similarity :

$w_{i\,j} = 1/(1+d(i,j))$
$w_{i\,k} = 1/(1+d(i,k))$



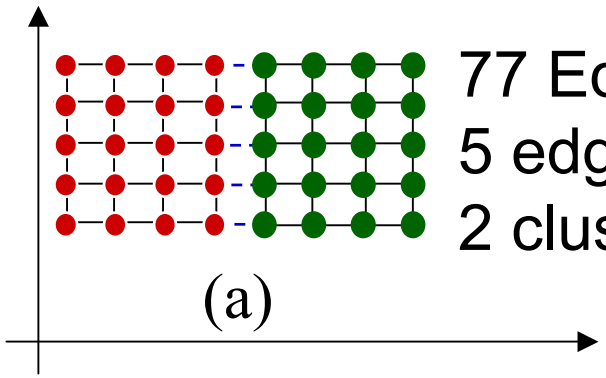
- Rank:

$w'_{i\,j} = 1/2 = 0.5$
$w'_{i\,k} = 1/1 = 1$

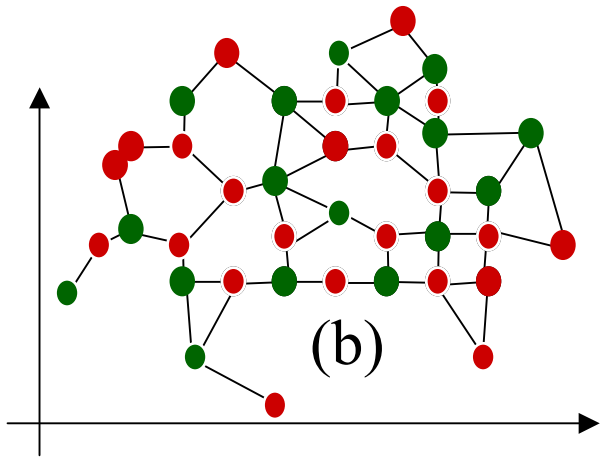


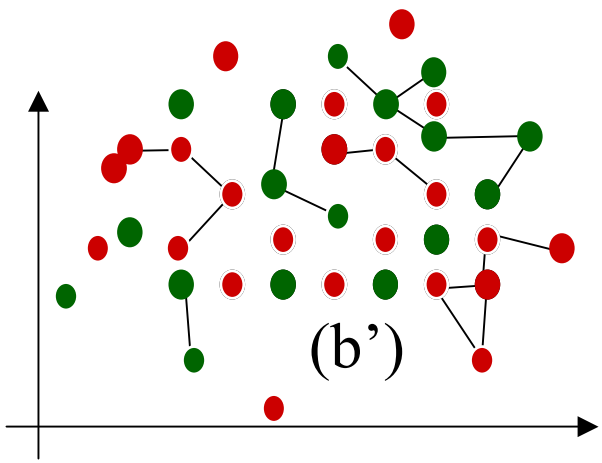
Principle:

The classes are well separable if the global weight of the cut edges (for generating the clusters) is very small.



77 Edges
5 edges cut
2 clusters





57 Edges
39 edges cut
9G+13R = 22 clusters

Notation:

Let ***I*** be the sum of the non cut edges weight

Let ***J*** be the sum of the cut edges weight

Null hypothesis:

H_0: the vertices of the graph are labeled independently of each other, according to the same probability distribution of the labels.

Reject H_0: the classes are not independently distributed or the probability distribution of the classes is not the same for the different vertices.

➡ Study the statistic ***J***

Law of *J* :

•Boolean case (Moran 1948, Cliff & Ord 1986):

2 classes:
1 and 2

$$J_{1,2} = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} Z_{ij}$$

where Z_{ij} equals 1 if *Y(i)* = *Y(j)* and 0 if not

•Multiple classes :

we consider all the pairs of different labels *r* and *s*

→ Mean of ***J*** is calculated from $\boldsymbol{J_{r,s}}$

$$J = \sum_{r=1}^{k-1} \sum_{s=r+1}^{k} J_{r,s} \qquad J_{r,s} = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} Z_{ij}$$

Decision:

H_0 is rejected if the p-value of J^s is lower than 5% (for example)

Experiments

Domain name	n	p	k	clust.	edges	J / (I + J)	J^s	p-value
Waves-20	20	21	3	6	25	0.400	-0.44	0.6635
Waves-50	50	21	3	11	72	0.375	-4.05	5.0E-05
Waves-100	100	21	3	12	156	0.301	-8.44	3.3E-17
Waves-1000	1000	21	3	49	2443	0.255	-42.75	0

Breiman L., Friedman J.H., Olshen R.A. and Stone J. Classification and regression trees, 1984 Wadsworth Int.

- 13 benchmarks of the UCI Machine Learning Repository
- Graph: Relative Neighborhood Graph of Toussaint
- Weights: connection, distance and rank

General information							weighting: connection			weighting: distance			weighting: rank		
Domain name	n	p	k	clust.	edges	error r.	J / (I + J)	J^s	p-value	J / (I + J)	J^s	p-value	J / (I + J)	J^s	p-value
Wine recognition	178	13	3	9	281	0.0389	0.093	-19.32	0	0.054	-19.40	0	0.074	-19.27	0
Breast Cancer	683	9	2	10	7562	0.0409	0.008	-25.29	0	0.003	-24.38	0	0.014	-25.02	0
Iris (Bezdek)	150	4	3	6	189	0.0533	0.090	-16.82	0	0.077	-17.01	0	0.078	-16.78	0
Iris plants	150	4	3	6	196	0.0600	0.087	-17.22	0	0.074	-17.41	0	0.076	-17.14	0
Musk "Clean1"	476	166	2	14	810	0.0650	0.167	-17.53	0	0.115	-7.69	2E-14	0.143	-18.10	0
Image seg.	210	19	7	27	268	0.1238	0.224	-29.63	0	0.141	-29.31	0	0.201	-29.88	0
Ionosphere	351	34	2	43	402	0.1397	0.137	-11.34	0	0.046	-11.07	0	0.136	-11.33	0
Waveform	1000	21	3	49	2443	0.1860	0.255	-42.75	0	0.248	-42.55	0	0.248	-42.55	0
Pima Indians	768	8	2	82	1416	0.2877	0.310	-8.74	2E-18	0.282	-9.86	0	0.305	-8.93	4E-19
Glass Ident.	214	9	6	52	275	0.3169	0.356	-12.63	0	0.315	-12.90	0	0.342	-12.93	0
Haberman	306	3	2	47	517	0.3263	0.331	-1.92	0.054	0.321	-2.20	0.028	0.331	-1.90	0.058
Bupa	345	6	2	50	581	0.3632	0.401	-3.89	1E-04	0.385	-4.33	1E-05	0.394	-4.08	5E-05
Yeast	1484	8	10	401	2805	0.4549	0.524	-27.03	0	0.512	-27.18	0	0.509	-28.06	0

nb of instances

nb of predictive attributes

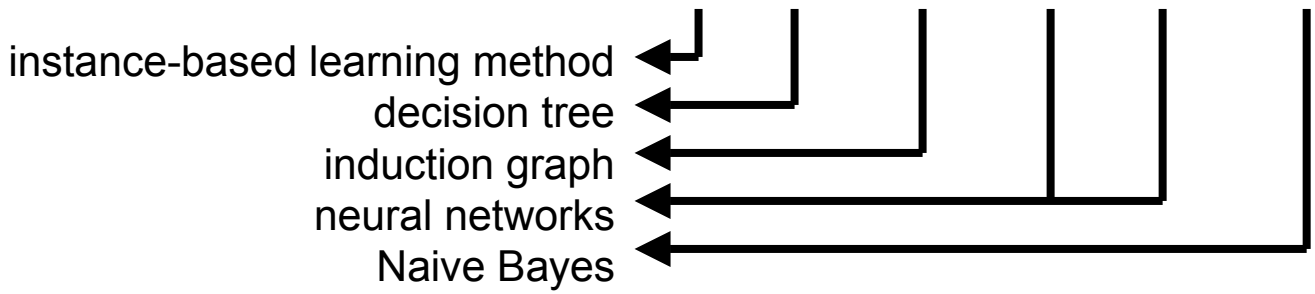
nb of classes

error rate on a 1-NN
(in a 10-fold cross validation)

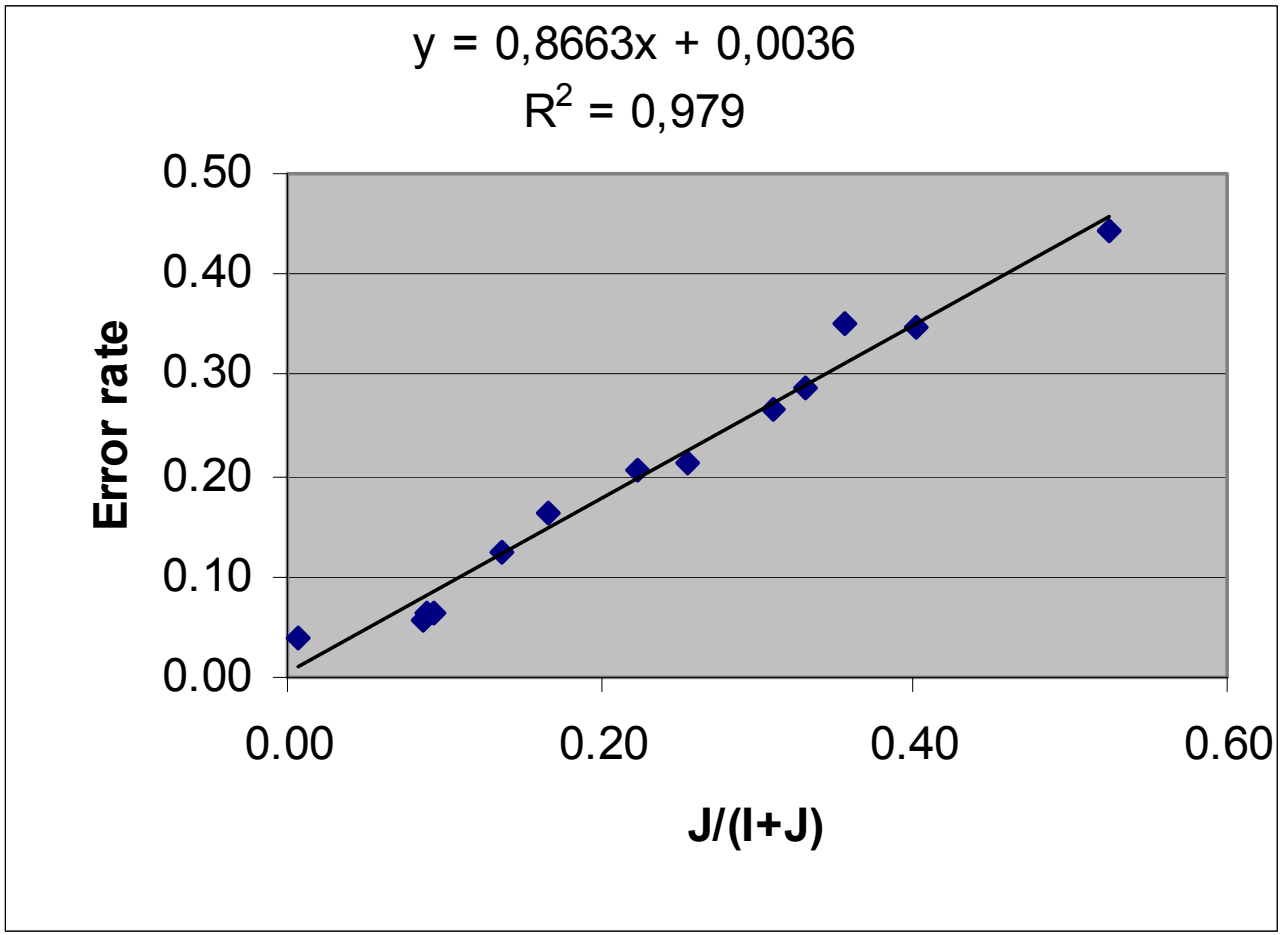
J/(I+J): relative cut edge weight
J^s: standardized cut edge weight

- Error rate in machine learning and weight of the cut edges

	General information					Statistical value			Error rate						
Domain name	**n**	**p**	**k**	**clust.**	**edges**	**J / (I + J)**	**J^s**	**p-value**	**1-NN**	**C4.5**	**Sipina**	**Perc.**	**MLP**	**N. Bayes**	**Mean**
Breast Cancer	683	9	2	10	7562	0.008	-25.29	0	0.041	0.059	0.050	0.032	0.032	0.026	0.040
BUPA liver	345	6	2	50	581	0.401	-3.89	0.0001	0.363	0.369	0.347	0.305	0.322	0.380	0.348
Glass Ident.	214	9	6	52	275	0.356	-12.63	0	0.317	0.289	0.304	0.350	0.448	0.401	0.352
Haberman	306	3	2	47	517	0.331	-1.92	0.0544	0.326	0.310	0.294	0.241	0.275	0.284	0.288
Image seg.	210	19	7	27	268	0.224	-29.63	0	0.124	0.124	0.152	0.119	0.114	0.605	0.206
Ionosphere	351	34	2	43	402	0.137	-11.34	0	0.140	0.074	0.114	0.128	0.131	0.160	0.124
Iris (Bezdek)	150	4	3	6	189	0.090	-16.82	0	0.053	0.060	0.067	0.060	0.053	0.087	0.063
Iris plants	150	4	3	6	196	0.087	-17.22	0	0.060	0.033	0.053	0.067	0.040	0.080	0.056
Musk "Clean1"	476	166	2	14	810	0.167	-17.53	0	0.065	0.162	0.232	0.187	0.113	0.227	0.164
Pima Indians	768	8	2	82	1416	0.310	-8.74	2.4E-18	0.288	0.283	0.270	0.231	0.266	0.259	0.266
Waveform	1000	21	3	49	2443	0.255	-42.75	0	0.186	0.260	0.251	0.173	0.169	0.243	0.214
Wine recognition	178	13	3	9	281	0.093	-19.32	0	0.039	0.062	0.073	0.011	0.017	0.186	0.065
Yeast	1484	8	10	401	2805	0.524	-27.03	0	0.455	0.445	0.437	0.447	0.446	0.435	0.444
								Mean	0.189	0.195	0.203	0.181	0.187	0.259	0.202
								R^2 (J/(I+J) ; error rate)	0.933	0.934	0.937	0.912	0.877	0.528	0.979
								R^2 (J^s ; error rate)	0.076	0.020	0.019	0.036	0.063	0.005	0.026



•Relative cut edge weight and mean of the error rates



Conclusion

The cut weighted edge statistic is a good class separability index.

The cut weighted edge statistic gives an appropriate information of the *a priori* ability of a database to be learnt.

Related work:

- Local version of the test to detect outliers (Lallich, Muhlenbach, Zighed, ISMIS 2002)
- Feature selection based on the best separability of classes (in progress)