MACHINE LEARNING

EXERCISES

Elements of non-parametric techniques

All the course material is available on the web site

Course web site: http://pralab.diee.unica.it/MachineLearning

PART 5: Elements of non-parametric techniques

Exercise 1

Given the following patterns belonging to three different classes A, B, and C

A	1.1 1.3	1.7 1.4	1.2 2.0	1.6 1.9
B	2.7 1.4	2.6 1.2	2.2 2.0	2.2 1.3
C	1.4 2.5	1.2 2.4	1.8 2.6	1.5 2.9

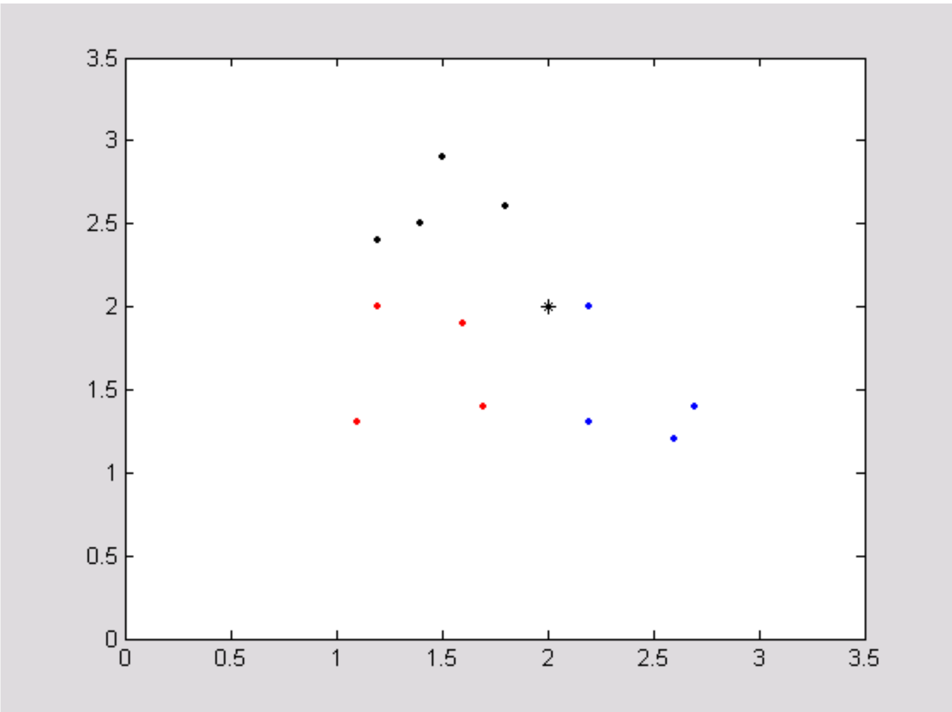
We want to classify the unknown pattern

x_t=(2; 2)’

but we do not know from which probability distribution the pattern has been generated. Then we can use a non-parametric method like *k*-nn.

A) Classify the pattern x_t with *k*= 1, …, 4 using the *Euclidean* and the *Manhattan* metric. An empirical choice for the *k* values could be $1..\sqrt{n}$

B) Choose the value of *k* through a *leave-one-out* error estimation (in a real case you can use just odd values of *k*, in the range 1 ... $\sqrt{(n)}$, where *n* is the size of the training set. In this case, for sake of simplicity, we are interested to choose between k=1 and k=4



Class A: red points; Class B: blue points; Class C: black points;

A) Classify the pattern

Squared *Euclidean* distances

	a1	a2	a3	a4
x_t	1.3000	0.4500	0.6400	0.1700

	b1	b2	b3	b4
x_t	0.8500	1.0000	0.0400	0.5300

	c1	c2	c3	c4
x_t	0.6100	0.8000	0.4000	1.0600

Classification result

k	A	B	C	Classification
1		1		B
2	1	1		A-B*
3	1	1	1	A-B-C*
4	2	1	1	A

* -> the result depends on the implementation

Manhattan distances

	a1	a2	a3	a4
x_t	1.60	0.90	0.80	0.50

	b1	b2	b3	b4
$\mathbf{x_t}$	1.30	1.40	0.20	0.90

	c1	c2	c3	c4
$\mathbf{x_t}$	1.10	1.20	0.80	1.40

k	A	B	C	Classification
1		1		B
2	1	1		A-B *
3	2	1		A *
4	2	1	1	A

* -> the result depends on the implementation
It is possible to see that the classification result depends on parameters choice.

B) Choose the value of *k*

How to choose the parameters?
In a *real* case we could follow these steps:

- Split the dataset in two parts: the first, called *training set*, will be used to 'design' the classifier (the *k*-nn, in the simplest implementation, memorize simply the patterns); the second, called *validation set*, will be used to evaluate the classifier performance.
- Design different classifiers using the training set and different parameter settings.
- Verify the performance classifying the validation patterns.
- Chose the classifier that obtained the best results on the validation set.

IMPORTANT! During the classifier design DO NOT use the test set.

If it is not available a big enough number of pattern to create a validation set (as in our example), it is possible to use the *leave-one-out* approach.

1) Given the training set D with *n* patterns
2) for i=1, …, *n* design the classifier using D-{x_i}, and classify x_i.
3) Repeat the point (2) varying the parameter values
4) Choose the values that allow to obtain the minimum error

Now, we are going to apply the above described method in order to decide if using k=1 or k=4 with the Euclidean metric. According to the chosen value we are going to classify the pattern x_t as belonging to the class A or B.

Evaluating the Euclidean squared distances between each pattern and all the other ones.

	b1	**b2**	**b3**	**b4**
a1	2.57	2.26	1.70	1.21
a2	1.0	0.85	0.61	0.26
a3	2.61	2.60	1.0	1.49
a4	1.46	1.49	0.37	0.72

	c1	**c2**	**c3**	**c4**
a1	1.53	1.22	2.18	2.72
a2	1.30	1.25	1.45	2.29
a3	0.29	0.16	0.72	0.90
a4	0.40	0.41	0.53	1.01

	c1	**c2**	**c3**	**c4**
b1	2.90	3.25	2.25	3.69
b2	3.13	3.40	2.60	4.10
b3	0.89	1.16	0.52	1.30
b4	2.08	2.21	1.85	3.05

	a1	**a2**	**a3**	**a4**
a1	0	0.37	0.50	0.61
a2	0.37	0	0.61	0.26
a3	0.50	0.61	0	0.17
a4	0.61	0.26	0.17	0

	b1	**b2**	**b3**	**b4**
b1	0	0.05	0.61	0.26
b2	0.05	0	0.80	0.17
b3	0.61	0.80	0	0.49
b4	0.26	0.17	0.49	0

	c1	**c2**	**c3**	**c4**
c1	0	0.05	0.17	0.17
c2	0.05	0	0.40	0.34
c3	0.17	0.40	0	0.18
c4	0.17	0.34	0.18	0

- k=1

pattern	**a1**	**a2**	**a3**	**a4**	**b1**	**b2**	**b3**	**b4**	**c1**	**c2**	**c3**	**c4**
Assigned class	**A**	**A**	**C**	**A**	**B**	**B**	**A**	**B**	**C**	**C**	**C**	**C**

Error = 2/12
- k=4

pattern	**a1**	**a2**	**a3**	**a4**	**b1**	**b2**	**b3**	**b4**	**c1**	**c2**	**c3**	**c4**
votes	**AA** **AB**	**AA** **AB**	**AA** **CC**	**AA** **BC**	**AB** **BB**	**AB** **BB**	**AB** **BC**	**AB** **BB**	**AC** **CC**	**AC** **CC**	**BC** **CC**	**AC** **CC**
Assigned class	**A**	**A**	**A-C***	**A**	**B**	**B**	**B**	**B**	**C**	**C**	**C**	**C**

Error = between 0/12 and 1/12

The results suggest to chose *k*=4.
IMPORTANT! Usually it is better choose an odd value for k in order to avoid an equal number of votes. It is clear that this choice does not avoid this situation when three or more classes are involved.

TABLES EXERCISE 1

Squared *Euclidean* distances

	a1	a2	a3	a4
x_t	1.3000	0.4500	0.6400	0.1700

	b1	b2	b3	b4
x_t	0.8500	1.0000	0.0400	0.5300

	c1	c2	c3	c4
x_t	0.6100	0.8000	0.4000	1.0600

Manhattan distances

	a1	a2	a3	a4
x_t	1.60	0.90	0.80	0.50

	b1	b2	b3	b4
$\mathbf{x_t}$	1.30	1.40	0.20	0.90

	c1	c2	c3	c4
$\mathbf{x_t}$	1.10	1.20	0.80	1.40

Evaluating the Euclidean squared distances between each pattern and all the other ones.

	b1	**b2**	**b3**	**b4**
a1	2.57	2.26	1.70	1.21
a2	1.0	0.85	0.61	0.26
a3	2.61	2.60	1.0	1.49
a4	1.46	1.49	0.37	0.72

	c1	**c2**	**c3**	**c4**
a1	1.53	1.22	2.18	2.72
a2	1.30	1.25	1.45	2.29
a3	0.29	0.16	0.72	0.90
a4	0.40	0.41	0.53	1.01

	c1	c2	c3	c4
b1	2.90	3.25	2.25	3.69
b2	3.13	3.40	2.60	4.10
b3	0.89	1.16	0.52	1.30
b4	2.08	2.21	1.85	3.05

	a1	a2	a3	a4
a1	0	0.37	0.50	0.61
a2	0.37	0	0.61	0.26
a3	0.50	0.61	0	0.17
a4	0.61	0.26	0.17	0

	b1	b2	b3	b4
b1	0	0.05	0.61	0.26
b2	0.05	0	0.80	0.17
b3	0.61	0.80	0	0.49
b4	0.26	0.17	0.49	0

	c1	c2	c3	c4
c1	0	0.05	0.17	0.17
c2	0.05	0	0.40	0.34
c3	0.17	0.40	0	0.18
c4	0.17	0.34	0.18	0

Exercise 2

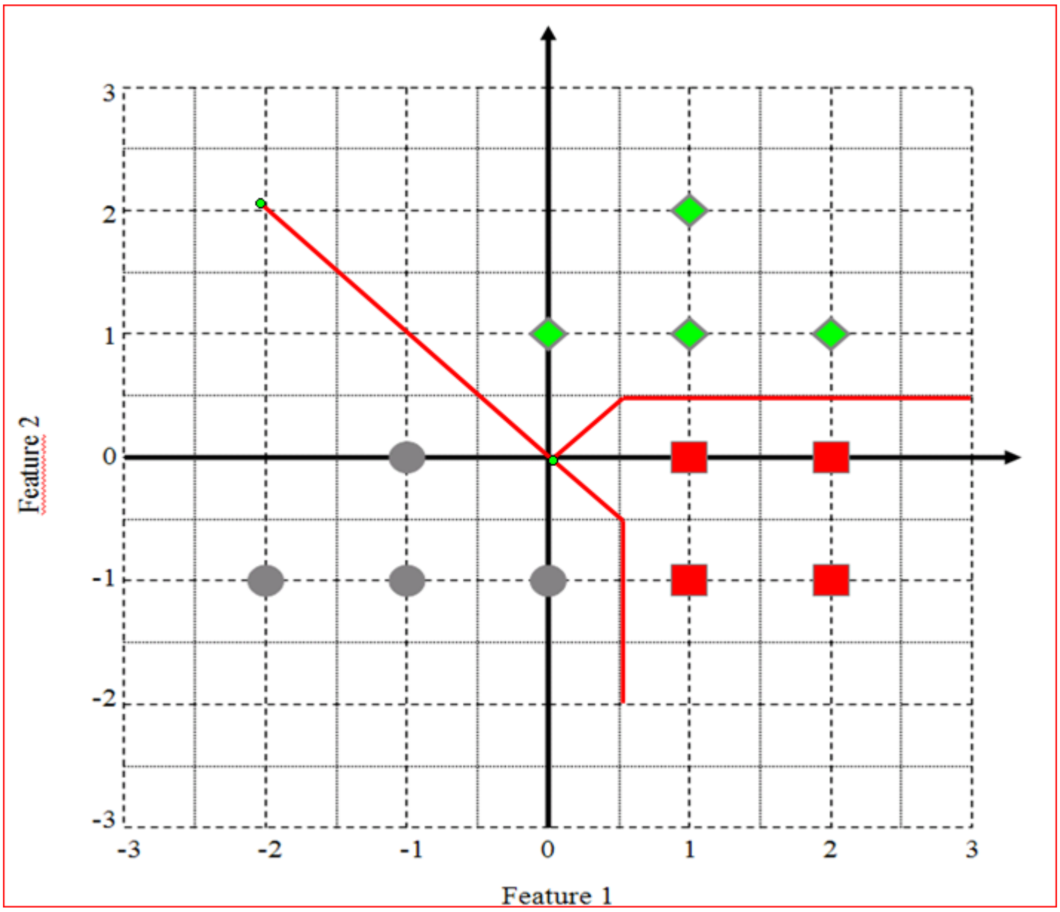
Given the patterns:

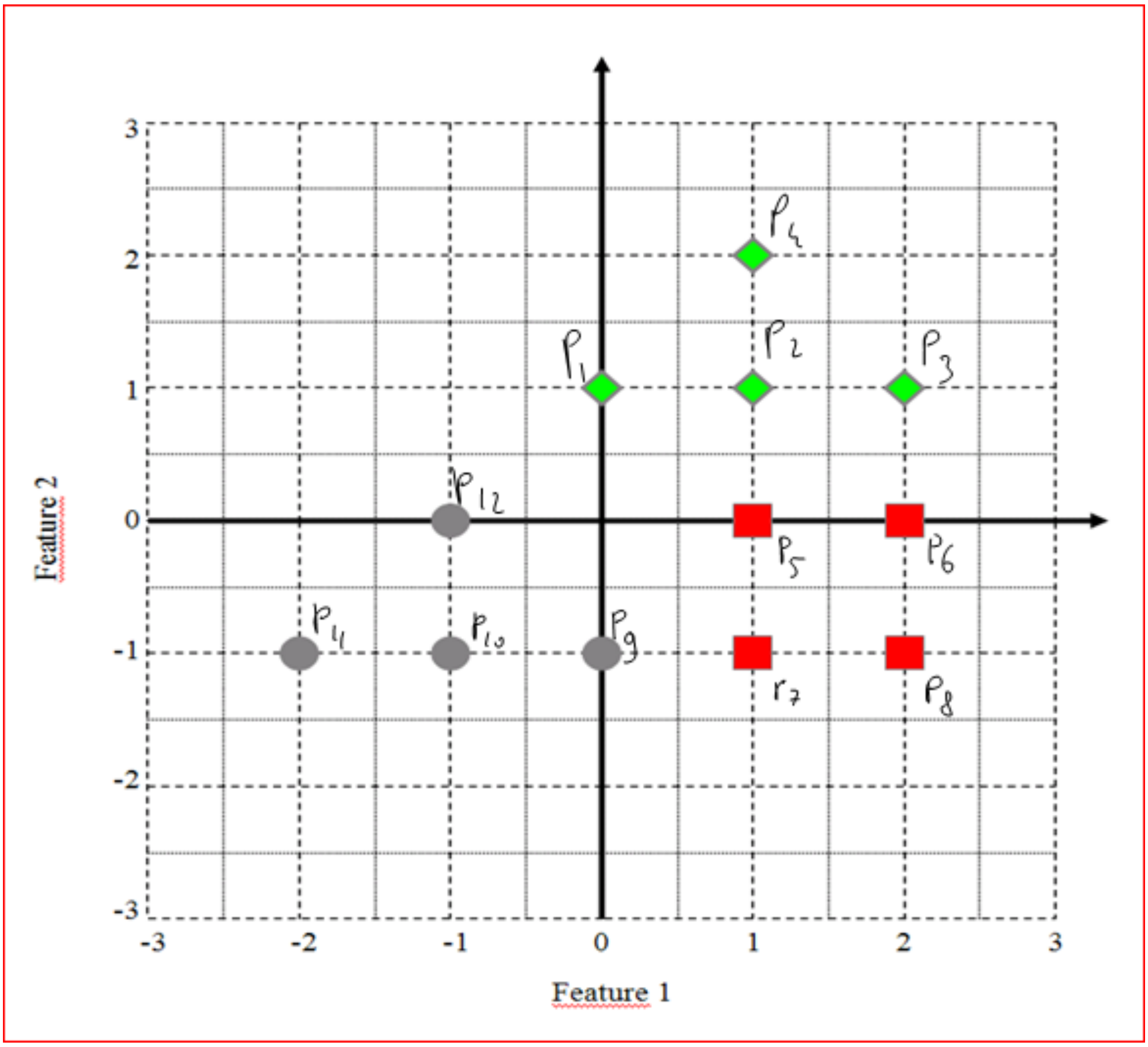
Class	ω_1				ω_2				ω_3			
	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12
	0.0 1.0	1 1	2 1	1 2	1 0	2 0	1 -1	2 -1	0 -1	-1 -1	-2 -1	-1 0

1) Classify the unknown pattern $X_t = (0.5, 0.4)^t$ using the K-nn classifier and the Euclidean distance for k=1 and k=3.
2) Estimate the optimal value of k, between k=1 and k=3, using the Manhattan metric. Explain the used method.

[In the case the unknown pattern is equidistant from different training patterns, suppose that the used algorithm choose first those patterns with a lower index.]

Below, the Voronoi Diagram





Solution:

Question (1):

Evaluate the euclidean distance between the unknown pattern and the training patterns

Class	ω_1				ω_2				ω_3			
	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12
X	**0.61**	**0.61**	2.61	2.81	**0.41**	2.41	2.21	4.21	2.21	4.21	8.21	2.41
	W1	W1			W2							

- K=1 $x \rightarrow \omega_2$
- K=3 $x \rightarrow \omega_1$

Question (2):

The optimal value of K can be estimated using the leave-one-out approach. According to this method it is possible classify each training pattern using all the remaining ones (a sort of N-fold Cross Validation).

Evaluate the distances between all the training patterns using the given metric:

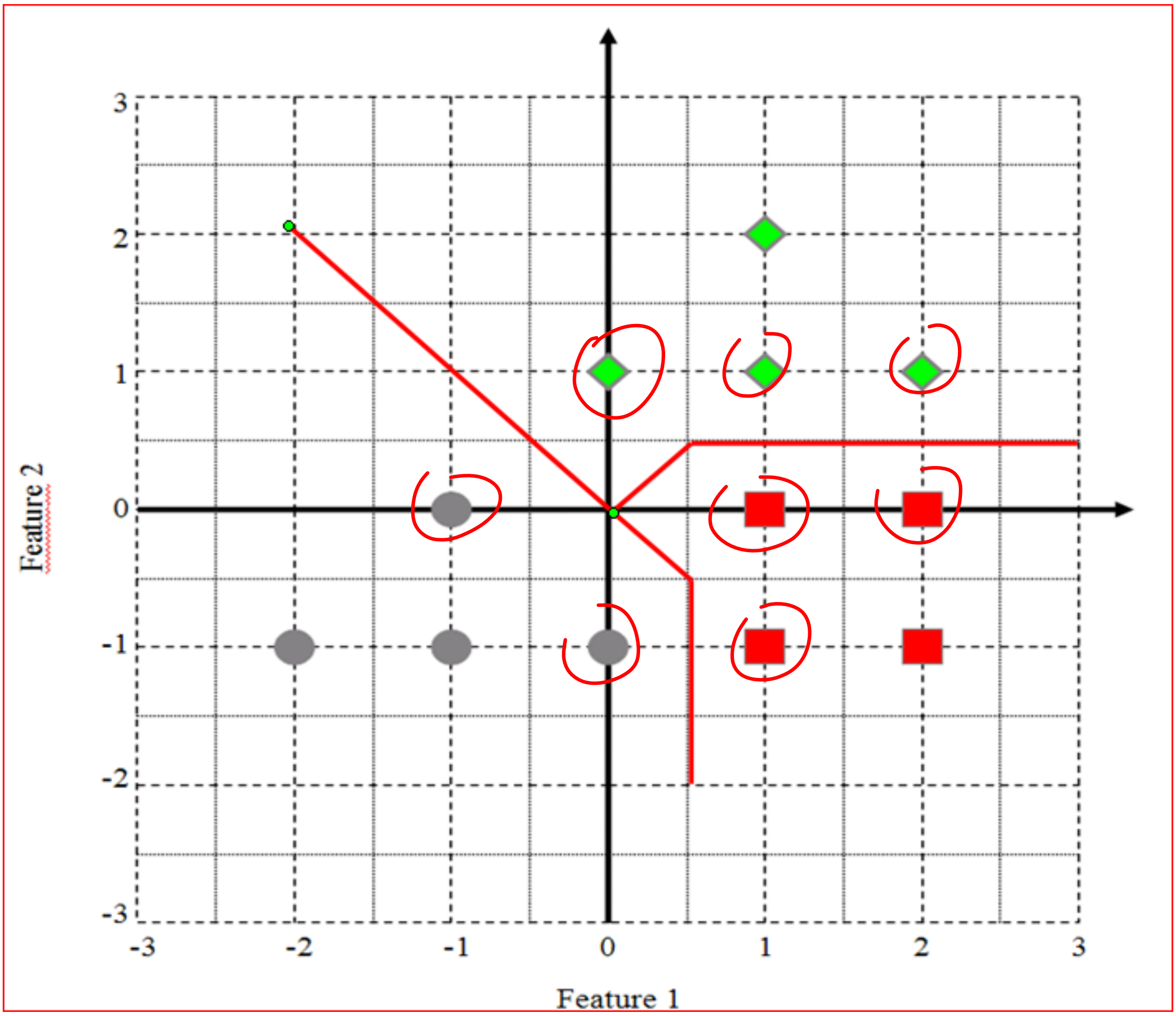
$$d(p_a, p_b) = |a_1 - b_1| + |a_2 - b_2|$$

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
P1	0	**1**	**2**	**2**	**2**	3	3	4	**2**	3	4	**2**
P2		0	**1**	**1**	**1**	2	2	3	3	4	5	3
P3			0	**2**	**2**	**1**	3	**2**	4	5	6	4
P4				0	**2**	3	3	4	4	5	6	4
P5					0	**1**	**1**	2	2	3	4	2
P6						0	2	**1**	3	4	5	3
P7							0	**1**	**1**	2	3	3
P8								0	**2**	3	4	5
P9									0	**1**	**2**	**2**
P10										0	**1**	**1**
P11											0	2
P12												0

Supposing that, in the case of equal distances or equal number of votes, the used algorithm chooses first the training patterns with a lower index, the result is:

True Class	Pattern	K=1 Nearest pattern	K=1 Assigned Class	K=3 Nearest pattern	K=3 Assigned Class
ω_1	P1	P2(ω_1)	ω_1	P2(ω_1), P3(ω_1), P4(ω_1)	ω_1
ω_1	P2	P1(ω_1)	ω_1	P1(ω_1), P4(ω_1), P5(ω_2)	ω_1
ω_1	P3	P2(ω_1)	ω_1	P2(ω_1) P6(ω_2) P1(ω_1)	ω_1
ω_1	P4	P2(ω_1)	ω_1	P2(ω_1) P1(ω_1), P3(ω_1)	ω_1
ω_2	P5	P2(ω_1)	$\boldsymbol{\omega_1}$	P2(ω_1) P6(ω_2) P7(ω_2)	ω_2
ω_2	P6	P3(ω_1)	$\boldsymbol{\omega_1}$	P3(ω_1) P5(ω_2) P8(ω_2),	$\boldsymbol{\omega_2}$
ω_2	P7	P5(ω_2)	ω_2	P5(ω_2) P8(ω_2) P9(ω_3)	ω_2
ω_2	P8	P6(ω_2)	ω_2	P6(ω_2) P7(ω_2) P3(ω_1),	ω_2
ω_3	P9	P7(ω_2)	$\boldsymbol{\omega_2}$	P7(ω_2) P10(ω_3) P1(ω_1)	$\boldsymbol{\omega_1}$
ω_3	P10	P9(ω_3)	ω_3	P9(ω_3) P11(ω_3) P12(ω_3)	ω_3
ω_3	P11	P10(ω_3)	ω_3	P10(ω_3) P9(ω_3) P12(ω_3)	ω_3
ω_3	P12	P10(ω_3)	ω_3	P10(ω_3) P1(ω_1), P5(ω_2)	$\boldsymbol{\omega_1}$
	ERR		3/12		2/12

The 'leave-one-out' method suggests to choose k=3



3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
Feature 1
Feature 2

TABLES EXERCISE 3

X_{A1}=(0.4 0.5 12 10)^T ; X_{A2}=(5 11 10 10)^T

X_{B1}=(1 1 0 6)^T ; X_{B2}=(5 2 0.6 0.6)^T

X_t= (0.4, 0.5, 0.4, 0.5)^T

Distances between X_t and all the other patterns.

	d(xa1,xt)^2	d(xa2,xt)^2	d(xb1,xt)^2	d(xb2,xt)^2
Feature = 1	0	21.16	0.36	21.16
Feature = 1,2	0	131.41	0.61	23.41
Feature = 1,2,3	134.56	223.57	0.77	23.45
Feature = 1,2,3,4	224.81	313.82	31.02	23.46