INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONES: 224281/2/3/4/5/6
CABLE: CENTRATOM - TELEX 460392-I

SMR/107 - 10

WORKSHOP ON PATTERN RECOGNITION AND ANALYSIS OF SEISMICITY

(5 - 16 December 1983)

E X E R C I S E S

to pattern recognition of earthquake
prone areas.

Exercises to pattern recognition of earthquake prone areas

<u>Note</u>: Decisive part of pattern recognition is a set of numerical tests of the reliability of results.

After you leave this workshop questions may be addressed to Drs. A.D.Gvishiani, A.A.Soloviev or E.Ya. Rantsman (Institute of Physics of The Earth, B.Grouzinskaya 10, Moscow 123242 USSR).

I. THE CHOICE OF OBJECTS

We seek to discriminate objects where epicenters with $M \geq M_o$ may and may not be situated. Suppose we define objects as an intersections of the axes of lineaments.

<u>Exercise:</u> Identify such objects on the scheme of lineaments (Fig.1).

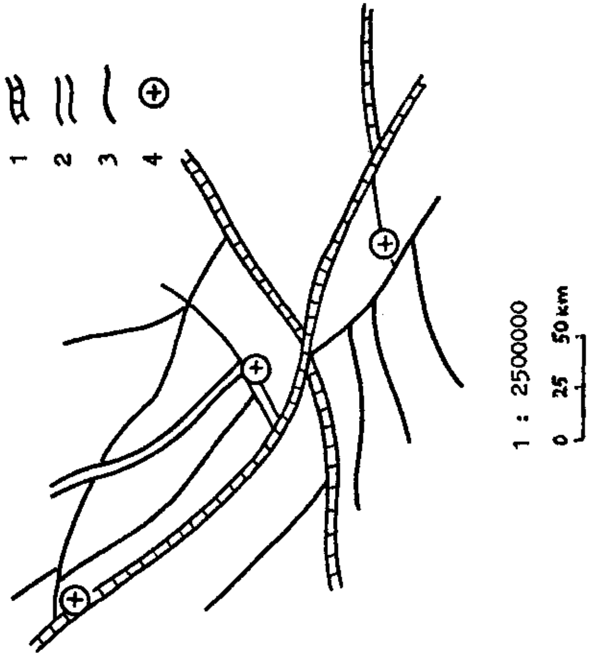


Fig. 1. 1, 2, 3 - lineaments; 1 - of the first rank; 2 - of the second rank; 3 - of the third rank; 4 - epicenters of earthquakes with M M .

II. DIVISION OF OBJECTS FOR LEARNING

CLASS D_o Intersections near which (within distance R_1) known epicentres with $M \geqslant M_o$ are situated

CLASS N_o Intersections near which (within distance R_2) no known epicentres with $M \geqslant M_o - \delta$ are situated. ($\delta = 0 - 0.2$)

CLASS X: Intersections which cannot be with certainty assigned to D_o nor to N_o (for example, because of uncertainty in some epicenters).

Exercise: Identify on Fig.1 the objects of Classes D_o, N_o and X for $R_1 = 25$ km, $R_2 = 50$ km, $\delta = 0$.

III. MEASUREMENT OF THE PARAMETERS

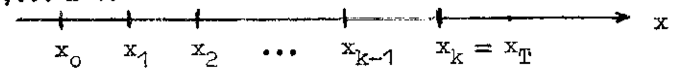
Exercise: Determine for intersections on Fig.1 two parameters:

r – the distance to the closest lineament of the first rank;

n – the number of lineaments, forming the intersections.

IV. DISCRETIZATION OF PARAMETRS

The values of each parameter x lie within certain range (x_o, x_T). We divide this range into k intervals by points x_i, $i = 1,2,\ldots k-1$.



Discretization means to consider not the value of the parameter but only the interval it belongs.

Usually we divide the range into two intervals ("large" and "small" values) or into three intervals ("large", "medium", and "small" values).

The purpose of discretization is to find such intervals where objects of one class occur more often than objects of another class.

Denote: P_D^i – % of objects of class D_o in the i-th interval,

P_N^i – % of objects of class N_o in the i-th interval,

Discretization is satisfactory, if

$$P_D^i - P_N^i > 15\%$$

at least for some i.

Objective discretization: Each interval has about equal number of all the objects together (or of the objects D_o and N_o together)

Exercise: Count P_D^i, P_N^i for parameters r and n from exercise III.

For r: $k = 3$, $r_1 = 5$km, $r_2 = 30$km

For n: $k = 2$, $n_1 = 2$.

Exercise: Find the boundaries of intervals for objective discretization of the same parameters. For r take $k=3$ and all objects. For n take $k=2$ and objects D_o and N_o only.

V. CODING OF PARAMETERS

Suppose a value of parameter belongs to l-th interval, $x_{l-1} < x \leq x_l$. Two ways of coding are the following:

i =	1	2	l-1	l	l+1	k-1	k	
I - coding:	0	0 ...	0	1	0 ...	0	0	(k digits)
S - coding:	0	0 ...	0	1	1 ...	1		(k-1 digits)

Exercise: The following table shows the value of three parameters (a,b,c)

Sequence number of object	Value of parameters a	b	c
1	-10	3	2,5
2	5	7	-4
3	- 2	6	-3
4	0	17	10
5	4	2	1

Discretization is the following:

Parameter	k	x_i
a	2	0.5
b	3	4.5, 10
c	3	0 , 5

Write the code of each object with I-coding for parameter a, S - coding for parameters b and c ;
another version - S - coding for parameters a and c, I - coding for parameter b.

VI . DEFINITION OF THE TRAIT IN ALGORITHM CORA-3

The trait is represented by the matrix

$$A = \begin{bmatrix} i_1 & i_2 & i_3 \\ \delta_1 & \delta_2 & \delta_3 \end{bmatrix}$$

i_1, i_2, i_3 - integers, $1 \leq i_1 \leq i_2 \leq i_3 \leq l$,
l - length of the binary code of the objects,
δ_j - 0 or 1.

Object with the code $(\omega_1, \omega_2, \ldots, \omega_l)$, ω_j = 0 or 1 has the trait A if

$$\omega_{i_1} = \delta_1 , \quad \omega_{i_2} = \delta_2 , \quad \omega_{i_3} = \delta_3$$

For example the trait

$$A = \begin{bmatrix} 1 & 3 & 4 \\ 0 & 1 & 0 \end{bmatrix}$$

means that the first and the fourth digits in the code of the object are 0 and the third digit is 1

Exercise: Consider objects

(0 1 0 1 1) and
(1 1 0 1 1);

find, whether they have the trait $\begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & 1 \end{bmatrix}$

VII. CHARACTERISTIC TRAITS

Denote:

$K_1(A)$ - the number of objects of class D_o which have the trait A

$K_2(A)$ - the number of objects of class N_o which have the trait A

A is a characteristic trait of class D ("D - trait") if

$$K_1(A) \geq K_1, \quad K_2(A) \leq \tilde{K}_1 .$$

A is a characteristic trait of class N ("N - trait") if

$$K_2(A) \geq K_2, \quad K_1(A) \leq \tilde{K}_2$$

Exercise: Given are the object of two classes:

Class D_o	Class N_o
1 1 1 0	0 0 1 0
0 1 1 0	0 1 1 1
1 1 1 1	0 0 0 0
1 1 0 0	0 0 0 1
1 1 0 1	0 1 0 0
	1 0 0 0

Find all D - traits and N - traits for $K_1=3$, $\tilde{K}_1=1$, $K_2=3$, $\tilde{K}_2=0$

VIII. EQUIVALENT AND WEAKER TRAITS

Consider two "D-traits".

Denote:

S_1 - the set of objects of class D_o which has the first trait.

S_2 - the set of objects of class D_o which has the second trait.

The traits are equivalent if S_1 and S_2 coincide.

The first trait is weaker than the second if S_1 is a subset of S_2.

Definition for N - traits is similar.

Exercise: Consider all characteristic traits from exercise VII. Eliminate all weaker traits. Leave only one trait from each group of equivalent traits.

IX. RECOGNITION IN CORA-3 ("VOTING")

Each object has some number n_D of "D-traits" and some number n_N of "N-traits"; n_D and n_N are $\geqslant 0$. The object is recognized

as D if $n_D - n_N \geqslant \Delta$, or

as N if $n_D - n_N < \Delta$.

Here Δ is a given constant

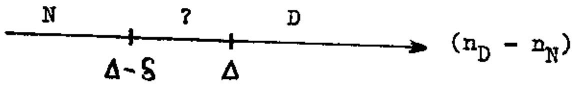
Exercise: Consider the characteristic traits left in the exercise VIII. Divide the objects from exercise VII into classes D and N, assuming $\Delta = 0$.

Note: It may be more reliable:

to assign to N only the object with $n_D - n_N \leq \Delta - \delta$;

and

to leave unassigned the objects with $\Delta > n_D - n_N > \Delta - \delta$



X. HAMMING ALGORITHM

Determination of the kernel. Each object is a binary vector, as in previous exercises. Kernel is a binary vector of the same length; each component of this vector is "typical" for class D

Denote:

$q_D(i|0)$ the number of objects of class D_o which have $\omega_i = 0$,

$q_D(i|1)$ the number of objects of class D_o which have $\omega_i = 1$,

$q_N(i|0)$ the number of objects of class N_o which have $\omega_i = 0$,

$q_N(i|1)$ the number of objects of class N_o which have $\omega_i = 1$.

Let us consider each component in turn. Count relative number of objects which have this component equal to 1 among class $\mathcal{D}_o$

$$\alpha_D(i|1) = \frac{q_D(i|1)}{q_D(i|0) + q_D(i|1)}$$

and among class N_o

$$\alpha_N(i|1) = \frac{q_N(i|1)}{q_N(i|0) + q_N(i|1)},$$

$i = 1, 2, \ldots, l.$

Kernel of class D $K = (k_1, k_2, \ldots k_l)$ is defined as follows:

$$k_i = \begin{cases} 1, & \text{if } \alpha_D(i|1) \geqslant \alpha_N(i|1), \\ 0, & \text{if } \alpha_D(i|1) < \alpha_N(i|1). \end{cases}$$

Exercise: Find the kernel for the following learning material

Class D_o	Class N_o
0 1 1 0 1	1 0 1 1 0
1 0 0 1 0	1 1 0 0 1
0 1 0 0 1	1 0 1 0 0
0 1 0 1 1	0 0 0 0 0
	1 0 0 0 0

Note: It may be more reliable to eliminate the parameters for which

$$|\alpha_D(i|1) - \alpha_N(i|1)| < \varepsilon,$$

where ε is a small constant.

XI. HAMMING ALGORITHM ("Voting" and recognition)

Hamming's distance from an object to the kernel of class D is

$$\rho = \sum_{i=1}^{\ell} p_i |\omega_i - k_i|.$$

Here p_i - are the weights of components.

Object is recognized as an object of class D, if $\rho \leq R$

or

as an object of class N, if $\rho > R$.

Exercise: Compute the distance between the kernel and all objects from exercise X. Find the value of R which will assign to class D all objects D_o. With this R divide the objects N_o into D and N classes.

ANSWERS

I. See fig. 2

II. D_o : 1, 2, 7, 8, 10, 13, 14

N_o : 3, 5, 9

X : 4, 6, 11, 12, 15

III.

	r, km	n
1	0	2
2	17,5	3
3	50	2
4	35	3
5	0	2
6	0	2
7	15	2
8	25	3
9	0	2
10	0	4
11	15	2
12	25	2
13	27,5	2
14	30	2
15	0	3

IV. 1) for r: $P_D^1 = 29\%$, $P_D^2 = 71\%$, $P_D^3 = 0\%$

$P_N^1 = 67\%$, $P_N^2 = 0\%$, $P_N^3 = 33\%$

for n: $P_D^1 = 57\%$, $P_D^2 = 43\%$

$P_N^1 = 100\%$, $P_N^2 = 0\%$

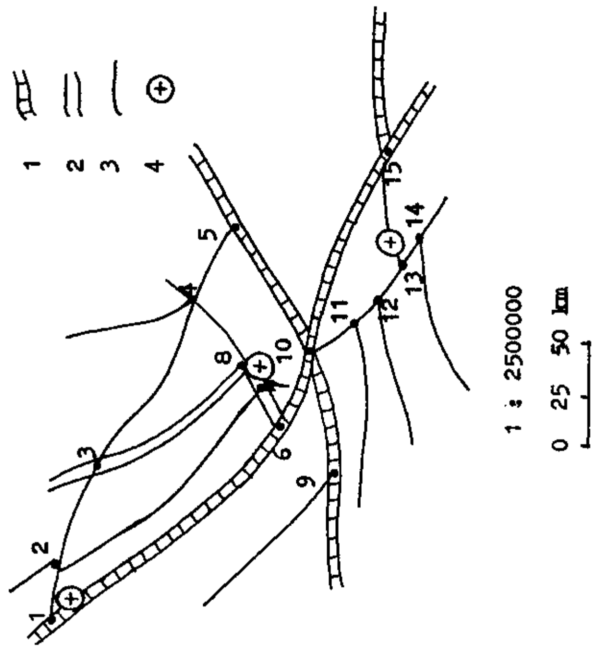


Fig. 2. 1,2,3 - Lineaments; 1 - of the first rank; 2 - of the second rank; 3 - of the third rank; 4 - epicenters of earthquakes with $M \geqslant M_o$

2) for r: $r_1 = 0$ $r_2 = 26$ km

for n: $n_1 = 2$

V. 1) 101101 2) 1 100 01

010111 0 010 11

100111 1 010 11

100000 1 001 00

011101 0 100 01

VI. (01011) has, (11011) has not

VII. D - traits: 111 112 223 224

111 , 111 , 111 , 110

N - traits: 112 113 222 223 224

000 , 000 000 , 000 , 000

VIII. D - traits: 111 223 224

111 , 111 , 110

N - traits: 113 222

000 , 000

IX.

	n_D	:	n_N	
D_o	3	:	0	D
	2	:	0	D
	2	:	0	D
	2	:	0	D
	1	:	0	D
N_o	0	:	1	N
	1	:	0	D
	0	:	2	N
	0	:	2	N
	1	:	1	D
	1	:	1	D

X. K = (0, 1, 0, 1, 1)

XI.

	ρ	
D_o	2	D
	3	D
	1	D
	0	D
N_o	4	N
	2	D
	5	N
	3	D
	4	N

R = 3