Mohamed Boudiaf University of M'sila

Faculty of Mathematics and Computer Science

Department: Computer Science/ICT

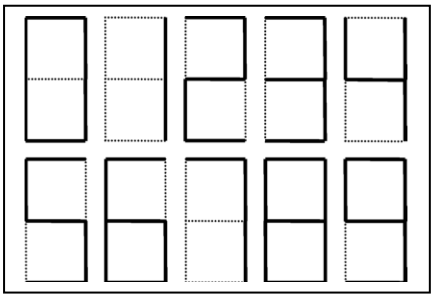
Subject: AI for M1-AI/M2-ISSI

SERIES NO. 04

Exercise 01

To display decimal digits: 0 ...9, we use

A 7 segment display as shown in
The figure opposite.



1. Propose a binary encoding of each digit
2. Propose an RNA architecture to decide whether the number is even or not
3. Find the learning threshold ε and the weights correspondents *wi*
4. Run the proposed model through examples.

Exercise 2

A telecommunications company organized a recruitment competition for computer and electronics engineers. Candidates must have tests in several teaching units. The results of deliberations are injected into an intelligent system which gives its decision based on a learning base and a test base. The decision taken by the system depends essentially on the following three grades: the mark of unit U1, the mark of the fundamental unit UF and the general average G_Avg. The table below illustrates the content of the learning base of the decision-making system of this company.

Candidate	Mark_U1	Mark_UF	G_Avg	Decision
Ali	4	8	9	Postponed
Mohamed	7	11	10.50	Admitted
Omar	3.50	10	12	Postponed
Farid	7	9	11.50	Postponed
Fatma	15	11	9.80	Postponed
Amira	12.50	13	13.75	Admitted
Salim	3.50	10.50	10	Postponed
Souad	6	14	12	Admitted

1. From the learning base given above, build a prediction model in the form of a decision tree.
2. Reformulate the predictive model in the form of decision rules.

Consider the following test base:

Candidate	Mark_U1	Mark_UF	G_Avg	Decision
Samira	5	10.50	9.99	?
Mourad	7	11	11.25	?
Nassim	2	7	10.50	?
Chahinez	6	12	11.16	?

3. Based on the predictive model constructed in (1), predict the decision for the candidates given in the test frame.
4. What type of learning does the decision system use? (supervised, unsupervised)

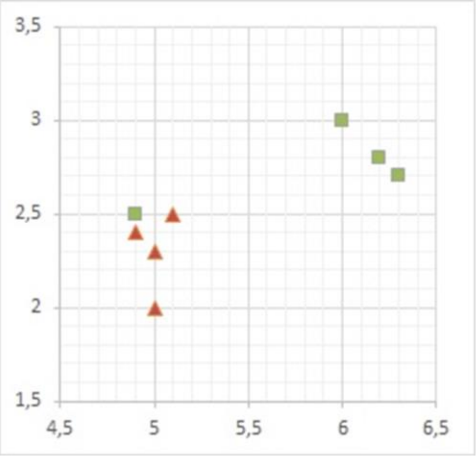
Exercise 03 (Principle of the K-NN algorithm)

The IRIS flower has several variations, among which three are particularly difficult to differentiate. These three variations are: Iris setosa, Iris versicolor and Iris virginica. Recent scientific research suggests that the three flowers can be distinguished by measuring certain elements of the flower, particularly the length and width of the sepal and petal. After taking and analyzing the samples in the lab, here are the results obtained:

Iris	Class	Sepal Length	Sepal Width	Petal Length
1	*Iris setosa*	4.6	3.6	1.0
2	*Iris setosa*	4.3	3.0	1.1
3	*Iris setosa*	5.0	3.2	1.2
4	*Iris setosa*	5.8	4.0	1.2
5	*Iris versicolor*	5.1	2.5	3.0
6	*Iris versicolor*	5.0	2.3	3.3
7	*Iris versicolor*	4.9	2.4	3.3
8	*Iris versicolor*	5.0	2.0	3.5
9	*Iris virginica*	4.9	2.5	4.5
10	*Iris virginica*	6.2	2.8	4.8
11	*Iris virginica*	6.0	3.0	4.8
12	*Iris virginica*	6.3	2.7	4.9

Consider the following graph:

1. Visualize the point with coordinates (4.8, 2.7)
2. What is the closest point to the added one?
3. Which three points are closest?
4. What is the majority color (shape) among the three? closest points?

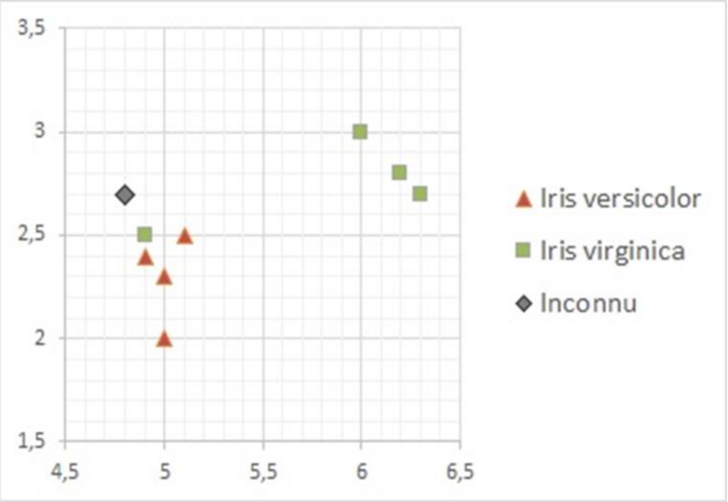


Now let's assume that the Red Triangles represent Iris versicolor and the Green Squares represent Iris virginica.

5. What would be the class of the flower that we just added? For what ?

Now a counter is associated with each flower class, in our case we have a counter for Iris setosa, one for Iris versicolor and one for Iris virginica. Each counter starts at 0 and increases by 1 each time a neighbor of the counter's class is encountered.
Let's reuse the previous graph in this exercise, considering that the red triangles represent Iris versicolor and the green squares represent Iris virginica. The gray diamond-shaped marker is an unknown Iris.

6. Determine the value of the counters for the 4 closest neighbors.
7. Give a general conclusion regarding the K-NN algorithm



Exercise 04 (Calculating distances)

Let the two points be: A(xA, yA), B(xB, yB)

We define the following distance metrics between A and B:

1. The distance defined by the inner product (Inner Product) between A and B is given by the following formula:

$$Inner(A,B) = x_A * x_B + y_A * y_B$$

2. Euclidean distance

$$Euc(A,B) = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}$$

3. Distance from Manhattan (taxi-distance)

$$Manh(A,B) = |x_A - x_B| + |y_A - y_B|$$

4. Minkowski distance

$$Mink(A,B) = \sqrt[q]{(|x_A - x_B|)^q + (|y_A - y_B|)^q}$$

with: q=1 Manhattan distance, q=2 Euclidean distance, q>2, Minkowski distance

5. The Cosine distance

$$Cos(A,B) = \frac{x_A * x_B + y_A * y_B}{\sqrt{(x_A{}^2 + y_A{}^2)}.\sqrt{(x_B{}^2 + y_B{}^2)}}$$

6. The Dice Distance

$$Dice(A,B) = \frac{2 * (x_A * x_B + y_A * y_B)}{(x_A{}^2 + y_A{}^2) + (x_B{}^2 + y_B{}^2)}$$

6. The Jaccard distance

$$Jacc(A,B) = \frac{(x_A * x_B + y_A * y_B)}{(x_A{}^2 + y_A{}^2) + (x_B{}^2 + y_B{}^2) - (x_A * x_B + y_A * y_B)}$$

Numerical Applications: Calculate the previous distances for: A(-1, -2), B(2, 2)

Exercise 5 (Naive Bayes Algorithm)

Consider the dataset of 1000 fruits. We have three types: Banana, Orange, and “other”. For each fruit, we have 3 characteristics:

- Whether the fruit is long or not
- Whether it is sweet or not
- Whether its color is yellow or not

Our dataset looks like this:

Kind	Long	small (not long)	sugar	Unsweetened	YELLOW	Not yellow	Total
Banana	400	100	350	150	450	50	500
Orange	0	300	150	150	300	0	300
Other fruit	100	100	150	50	50	150	200
Total	500	500	650	350	800	200	1000

Suppose someone asks us to tell him the type of a fruit he has. Its characteristics are as follows:

- ✓ It is yellow
- ✓ It is long
- ✓ It is sweet

1. Apply the Naïve Bayes algorithm to find out if it is a banana, or an orange or something else fruit ?
2. From the previous example, deduce the advantages and limitations of the NB algorithm.

Exercise 4 (SVM Algorithm)

The table below contains data on individuals in a population described according to two attributes: attribute 1 and attribute 2. We want to use the SVM method to classify this data into two classes: C1 and C2.

No.	Attribute 1	Attribute 2	Class
1	0	3	C1
2	3	3	C1
3	1	4	C1
4	2	5	C1
5	1	6	C2
6	2	7	C2
7	3	7	C2
8	3	8	C2
9	2	9	C2
10	4	9	C2
11	8	9	C2

1. Represent the data from the previous table on an orthogonal reference, symbolizing the class C1 by the symbol (+) and class C2 by a circle (o).
2. Represent the optimal separating hyperplane by clearly showing the vector supports that you use.
3. We wish to classify a twelfth individual having Attribute1 equal to 4 and Attribute2 equal to 5.6. Redraw all the points, showing the class (+ or o) of the element that was just added.
4. Represent the new optimal hyperplane.

   We add another point (2, 3) whose class we know, which is C1.

5. Are the data always linearly separable? If so, give the equation of the new optimal hyperplane. Otherwise, say what SVM predicts in this case.

Good luck