Experiment 9: Distance Matrix Construction using Various Distance Measures

Aim: Build a distance matrix for a given dataset using different distance measures such as Euclidean and Manhattan

DESCRIPTION:

Distance measures are mathematical formulas used to quantify the similarity or dissimilarity between data points. In machine learning and data mining, they are essential for clustering, classification, and pattern recognition.

A **distance matrix** is a square matrix where element *(i, j)* represents the distance between object *i* and object *j*.

Smaller values indicate higher similarity, and larger values indicate greater dissimilarity.

Commonly used distance measures:

1. **Euclidean Distance**: Straight-line distance between two points.

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

2. **Manhattan Distance**: Sum of absolute differences between coordinates.

$$d(x,y) = \sum_{i=1}^{n} | x_i - y_i |$$

ILLUSTRATION:

Build a decision tree for the following data

Object	X	Y
A	1	2
B	2	4
C	3	5

.

1. Euclidean Distance Calculations:

$$d(A,B) = \sqrt{(2-1)^2 + (4-2)^2} = \sqrt{1+4} = 2.236$$
$$d(A,C) = \sqrt{(3-1)^2 + (5-2)^2} = \sqrt{4+9} = 3.606$$
$$d(B,C) = \sqrt{(3-2)^2 + (5-4)^2} = \sqrt{1+1} = 1.414$$

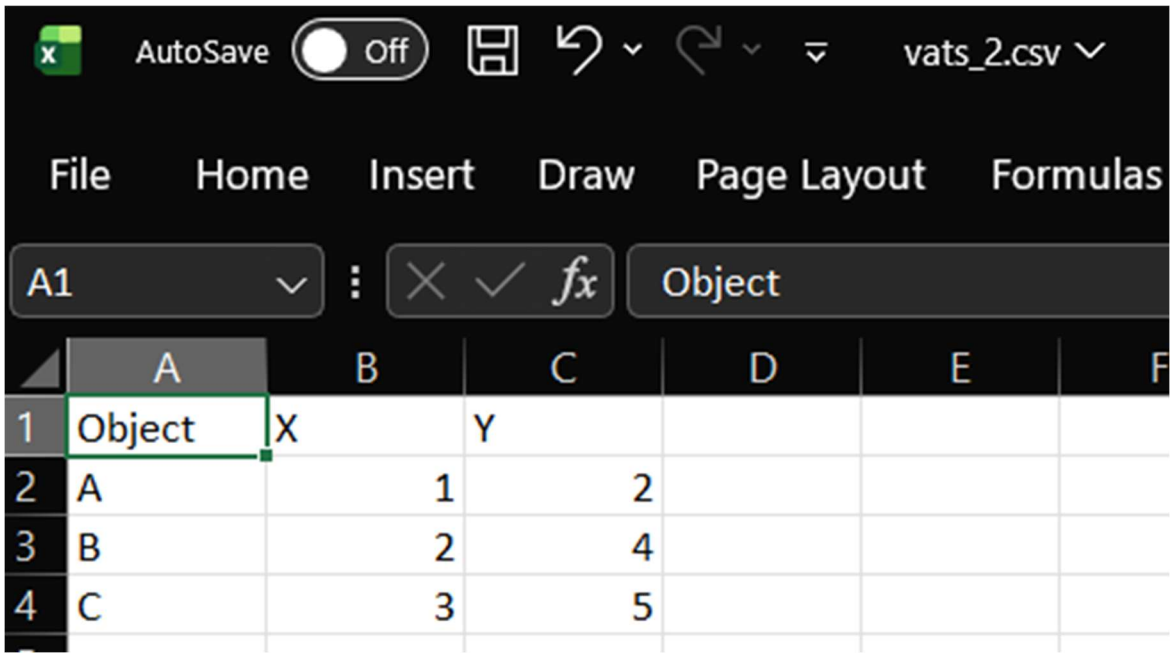
2. Manhattan Distance Calculations

$$d(A,B) = |\ 2-1\ | + |\ 4-2\ | = 3$$
$$d(A,C) = |\ 3-1\ | + |\ 5-2\ | = 5$$
$$d(B,C) = |\ 3-2\ | + |\ 5-4\ | = 2$$

Procedure for running in weka:

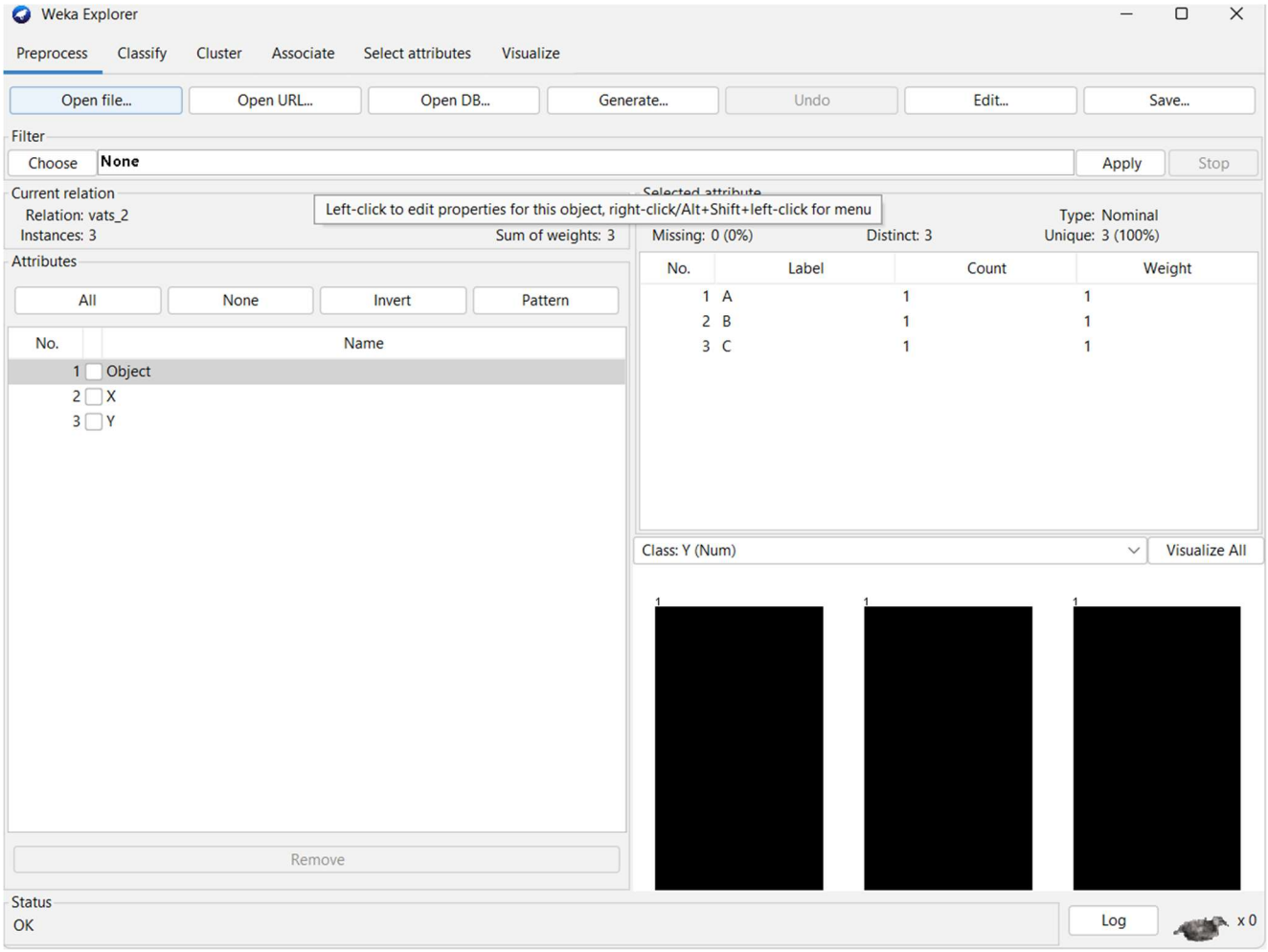
Step 1: Creating the ARRF file

Create the file and type the following data exactly as shown below.

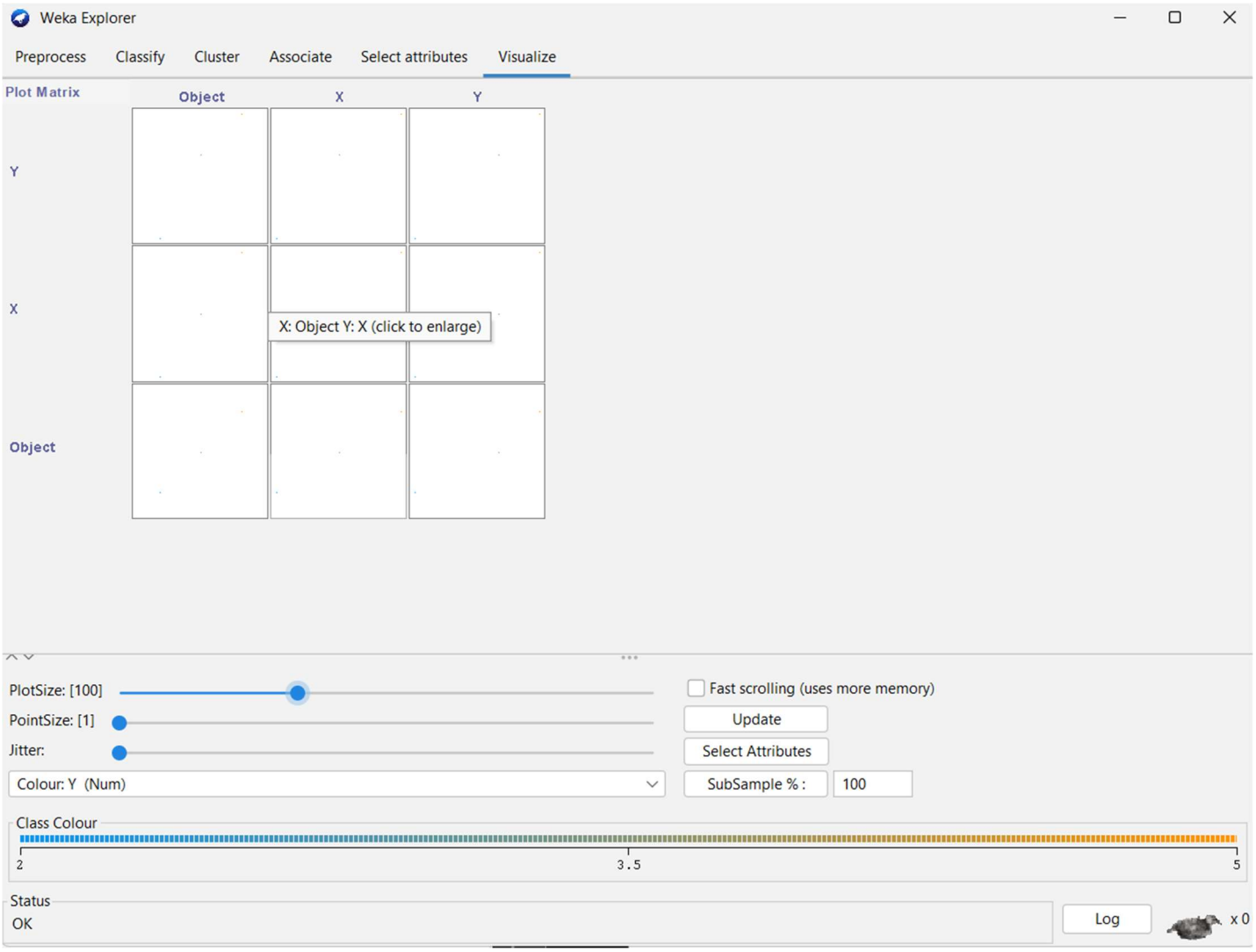


	A	B	C	D	E	F
1	Object	X	Y			
2	A	1	2			
3	B	2	4			
4	C	3	5			

Step2:
Open **WEKA Explorer → Preprocess Tab → Open File**

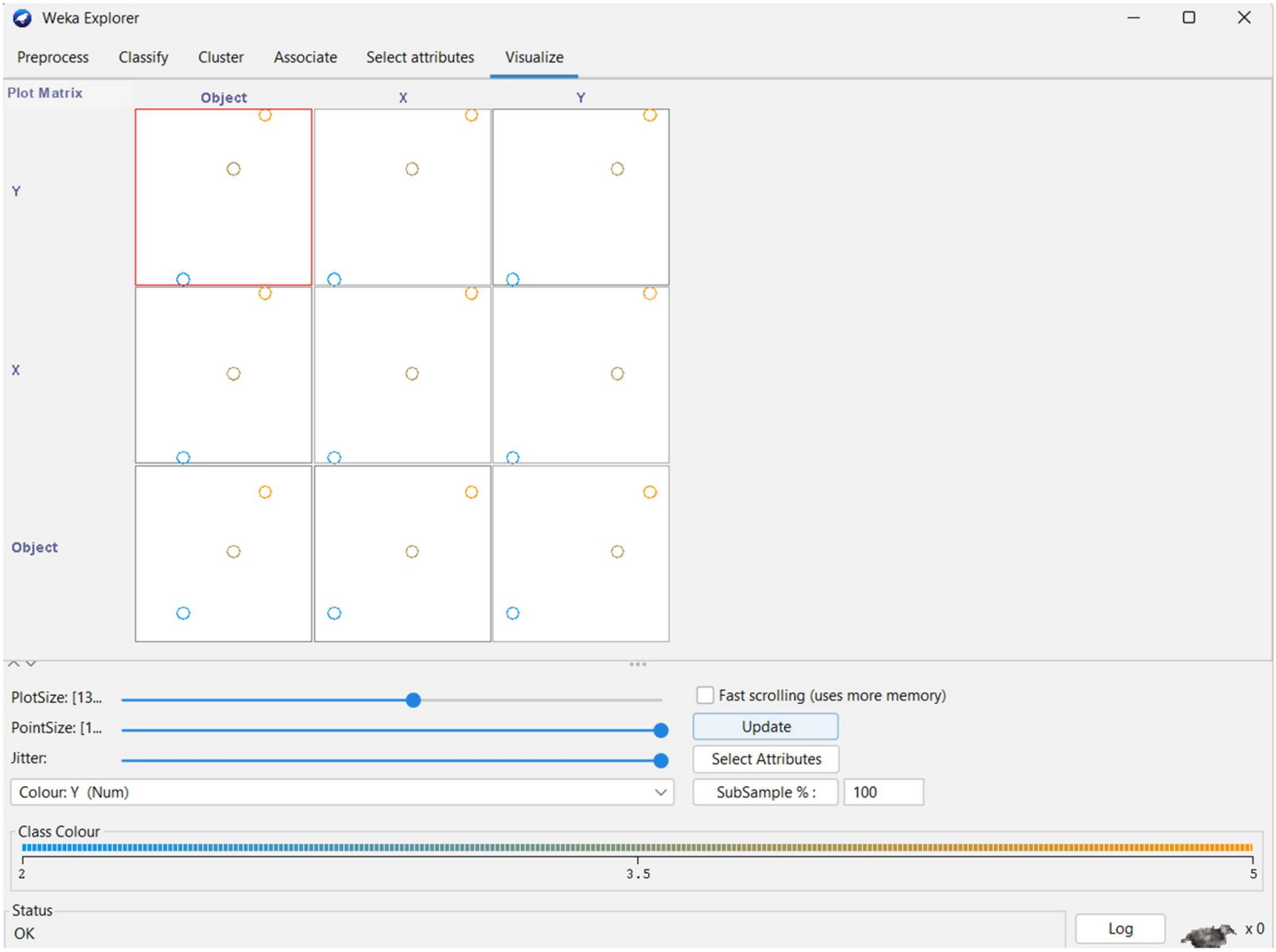


Step 3:
Go to the **Visualize Tab**



Step 4:

Make sure the sliders at the bottom are adjusted as in the image given below:



Conclusion:

From the above we can get the matrices as follows:

1) Euclidean Matrix:

	A	B	C
A	0.00	2.236	3.606
B	2.236	0.00	1.414
C	3.606	1.414	0.00

2) Manhattan Matrix:

	A	B	C
A	0	3	5
B	3	0	2
C	5	2	0

3)

The Euclidean and Manhattan distance matrices correctly represent pairwise dissimilarities between A, B, and C.
Euclidean measures direct straight-line distance, while Manhattan measures movement along grid-like paths.
Both matrices are symmetric, with diagonal elements equal to zero.

EXERCISE:

Add one more data point (D) with coordinates of your choice and recompute both matrices.