

A large, stylized silhouette of a person's head and shoulders, facing right. Inside the silhouette, a vibrant cityscape is visible, featuring the Eiffel Tower, the Oriental Pearl Tower, and the Empire State Building, suggesting a global or multicultural theme. The background is a warm, golden-yellow gradient.

SKEMA BUSINESS SCHOOL

**Introduction to
Artificial Intelligence**
Dmitry A. Zaitsev
<http://daze.ho.ua>



skema
BUSINESS SCHOOL

A woman's silhouette is shown from the back, with her hair in a bun. Inside her head and shoulders, a city skyline is visible, featuring the Eiffel Tower, Christ the Redeemer, and the Empire State Building. The background is a warm, orange-hued sky.

SKEMA BUSINESS SCHOOL

**Introduction to
Artificial Intelligence**
Dmitry A. Zaitsev
<http://daze.ho.ua>



skema
BUSINESS SCHOOL

Lesson 8

Unsupervised machine learning



Lesson 8

Unsupervised machine learning

Analyze and cluster unlabeled datasets

Types of unsupervised learning

Clustering, Association, Anomaly detection

K-Means Clustering Method

Initialize and move centroids

Orange Add-ons

Extend toolset of Orange

Unsupervised machine learning

Learns patterns from unlabeled data

- analyze and cluster unlabeled datasets, discover hidden patterns or data groupings (without the need for human intervention)
- can be used to discover the underlying structure of the data
- similar as a human learns to think by their own experiences
- combined application: unsupervised – clusters and helps to find appropriate labels for supervised machine learning



TRAINING

Supervised Learning | Analysis using Orange



= dog



= dog



= cat



= cat



= dog



= cat



= cat



= dog

classification



= dog



= cat

TESTING



= cat or dog?



= cat or dog?



Watch later



Share

Supervised learning



0:21 / 13:44 • unsupervised learning using orange



YouTube





Unsupervised Learning



A

B



MORE VIDEOS

Play (k)



0:35 / 13:44 • unsupervised learning using orange

clustering



YouTube



Unsupervised learning

Types of unsupervised learning

Clustering and Association

- Clustering is a method of grouping the objects into clusters such that objects with most similarities remains into a group and has less or no similarities with the objects of another group
- Association is an unsupervised learning method which is used for finding the relationships between variables in the large dataset (database)
- Anomaly detection finds unusual data patterns in dataset

Techniques of unsupervised learning

- K-means clustering
- KNN (k-nearest neighbors)
- Hierarchical clustering
- Anomaly detection
- Neural Networks
- Principle Component Analysis
- Independent Component Analysis
- Apriori algorithm
- Singular value decomposition

Advantages and disadvantages of Unsupervised Learning

Unsupervised vs Supervised Learning

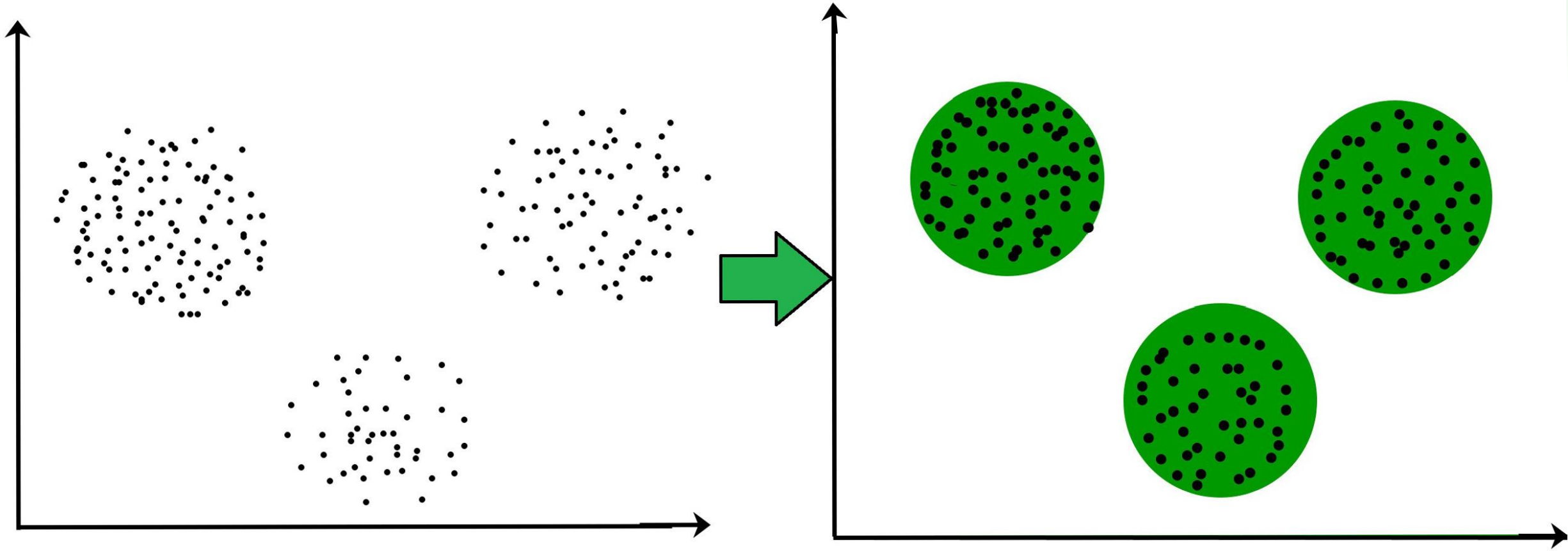
- Advantages
 - ✓ solve more complex tasks
 - ✓ it is easier to get unlabeled datasets
- Disadvantages
 - ✓ more difficult and time consuming
 - ✓ the result is not so accurate as for supervised learning

Cinderella fairytale: clustering of grains mixture

Cinderella was given the job but mice have done it instead of her

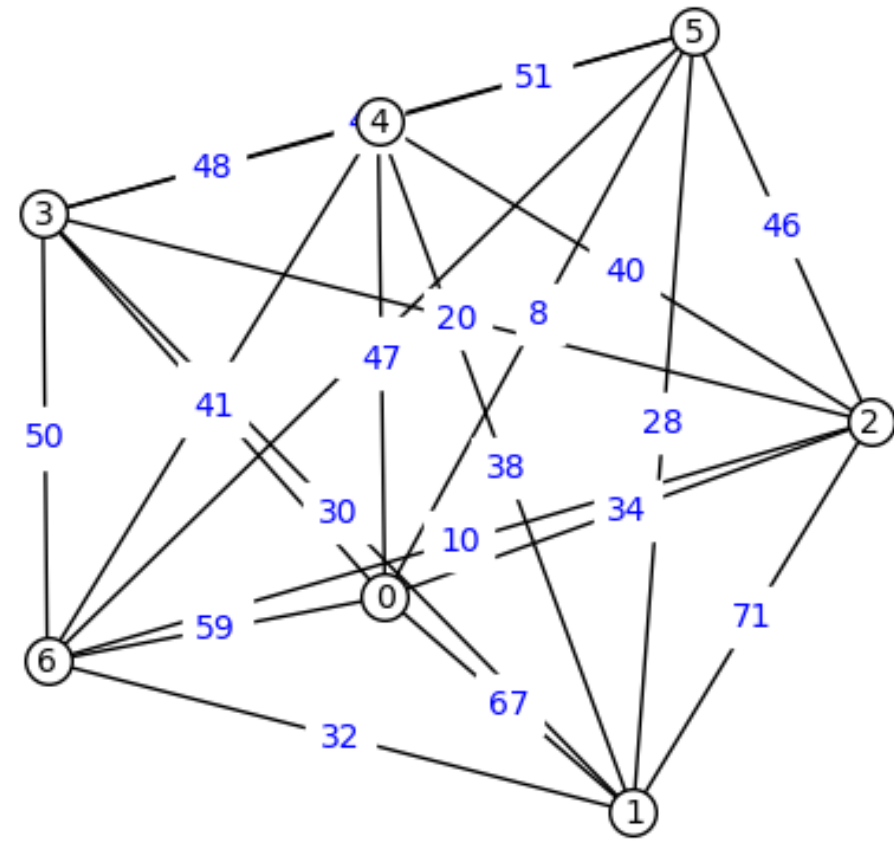
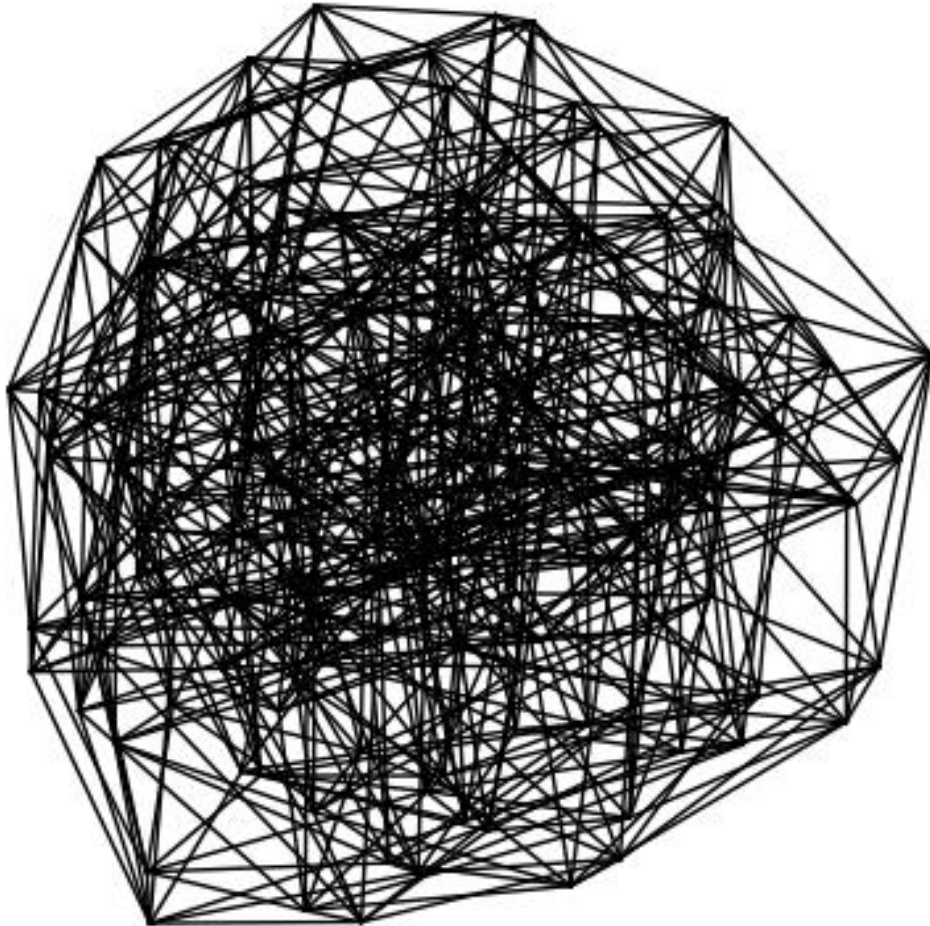


Clustering example



Clustering and aggregation of clusters example

Hypertorus communication grid verification: Zaitsev, Shmeleva, Luszczek, 2022



Clustering

Group similar objects

Clustering is the act of organizing similar objects into groups within a machine learning algorithm. Assigning related objects into clusters is beneficial for AI models. Clustering has many uses in data science, like image processing, knowledge discovery in data, unsupervised learning, and various other applications.

K-Means Clustering Method

- **Step-1:** Select the number K to decide the number of clusters.
- **Step-2:** Select random K points or centroids. (It can be other from the input dataset).
- **Step-3:** Assign each data point to their closest centroid, which will form the predefined K clusters.
- **Step-4:** Calculate the variance and place a new centroid of each cluster.
- **Step-5:** Repeat the third steps, which means reassign each datapoint to the new closest centroid of each cluster.
- **Step-6:** If any reassignment occurs, then go to step-4 else go to FINISH.
- **Step-7:** The model is ready.

Untitled * - Orange

File Edit View Widget Options Help

Data

File CSV File Import Datasets SQL Table

Data Table Paint Data Data Info Rank

Edit Domain Color Feature Statistics Save Data

Paint Data

Create data by painting data points on a plane.

[more...](#)

Paint Data Save Data

Save Data - Ora... ? X

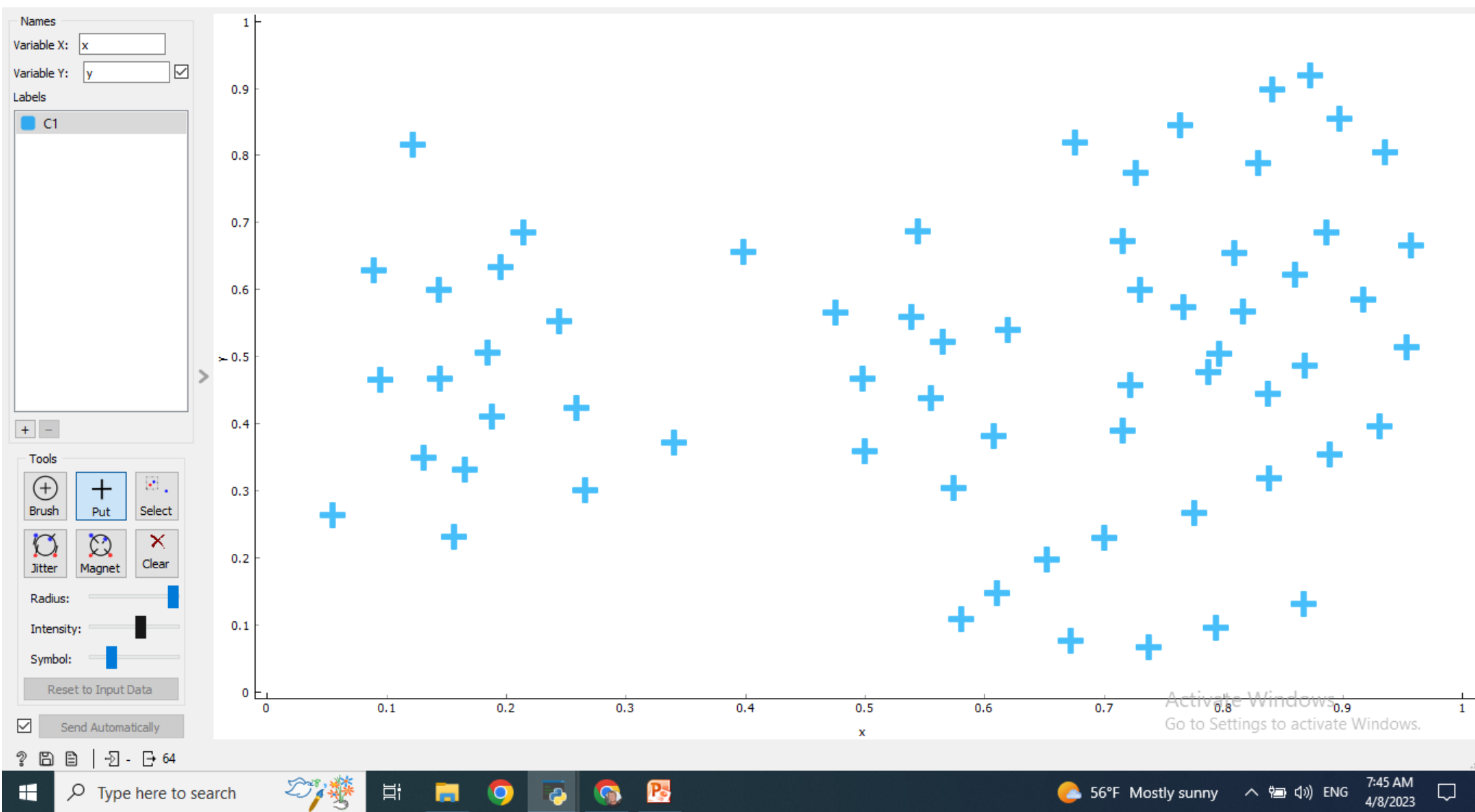
☒ Add type annotations to header

☐ Autosave when receiving new data

Save as mydots.tab Save as ...

? | 41

Draw and save your data



Type here to search



56°F Mostly sunny

ENG

7:45 AM
4/8/2023

Untitled * - Orange

File Edit View Widget Options Help

Unsupervised

Distance File Distance Matrix t-SNE Correlations

Distance Map Hierarchical Clustering k-Means Louvain Clustering

DBSCAN Manifold Learning Outliers PCA

Neighbors Correspondence Analysis Distances Distance Transformations

File Data Table Selected Data → Data Scatter Plot Source

Data Data Data

k-Means Scatter Plot Clustered

k-Means - Orange

Number of Clusters

☒ Fixed: 3

☐ From 2 to 8

Preprocessing

☒ Normalize columns

Initialization

Initialize with KMeans++

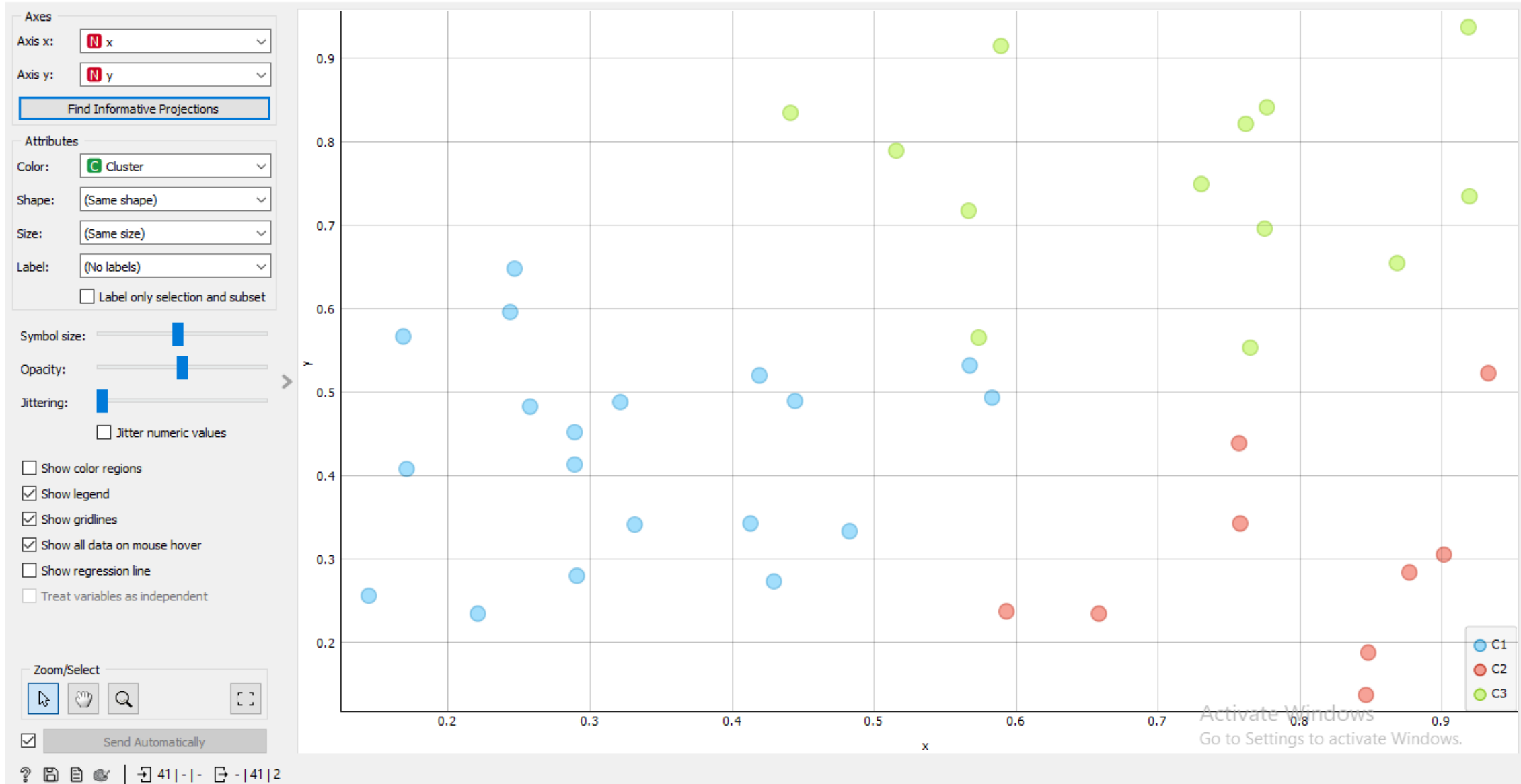
Re-runs: 10

Maximum iterations: 300

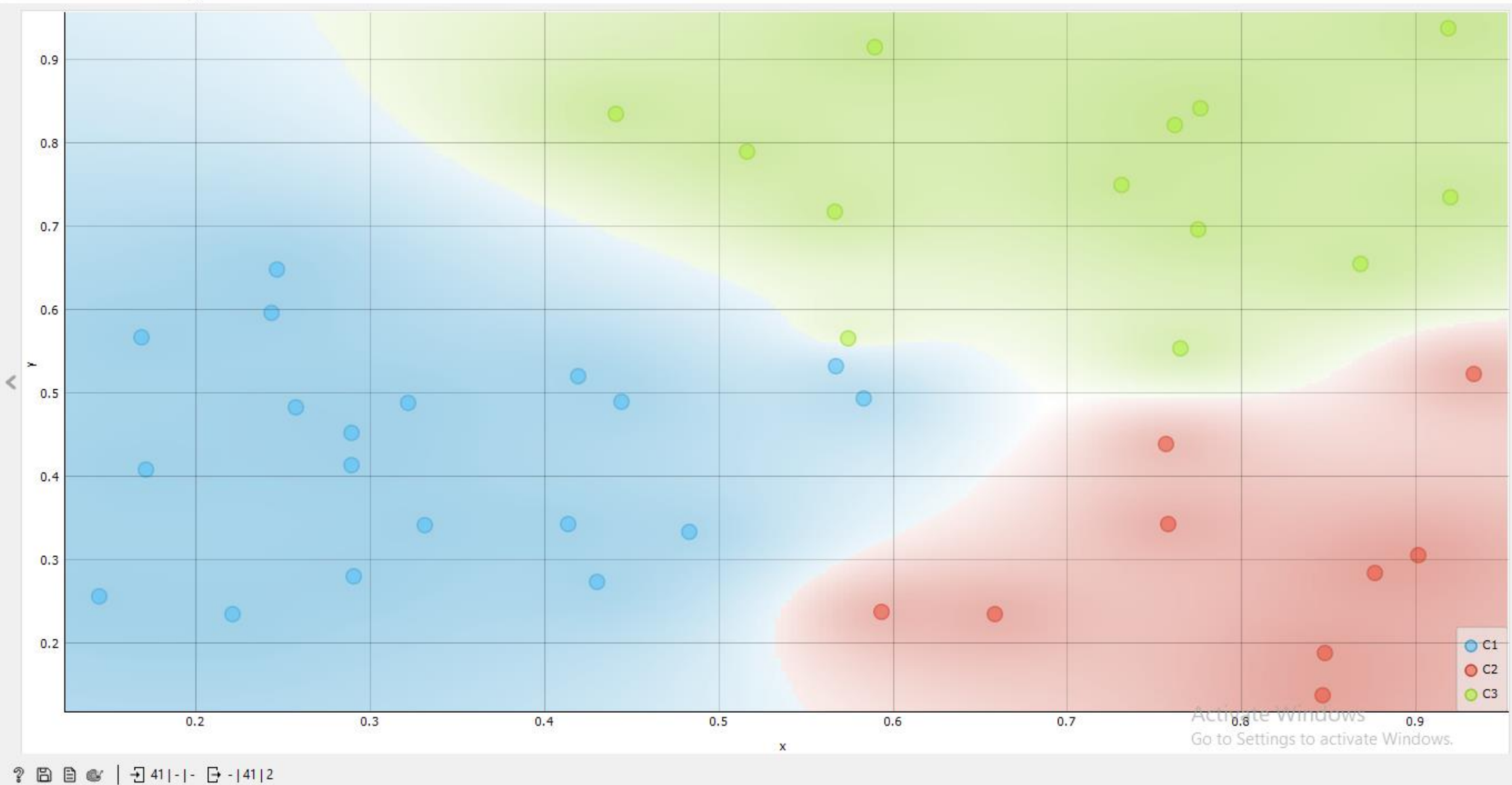
☒ Apply Automatically

? 41 41 | 3

Clustering in Orange



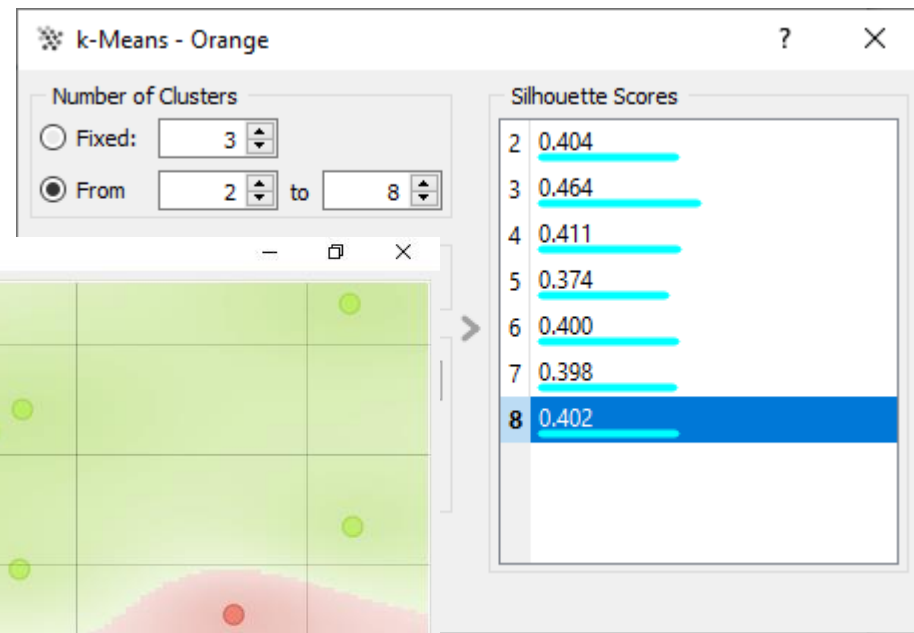
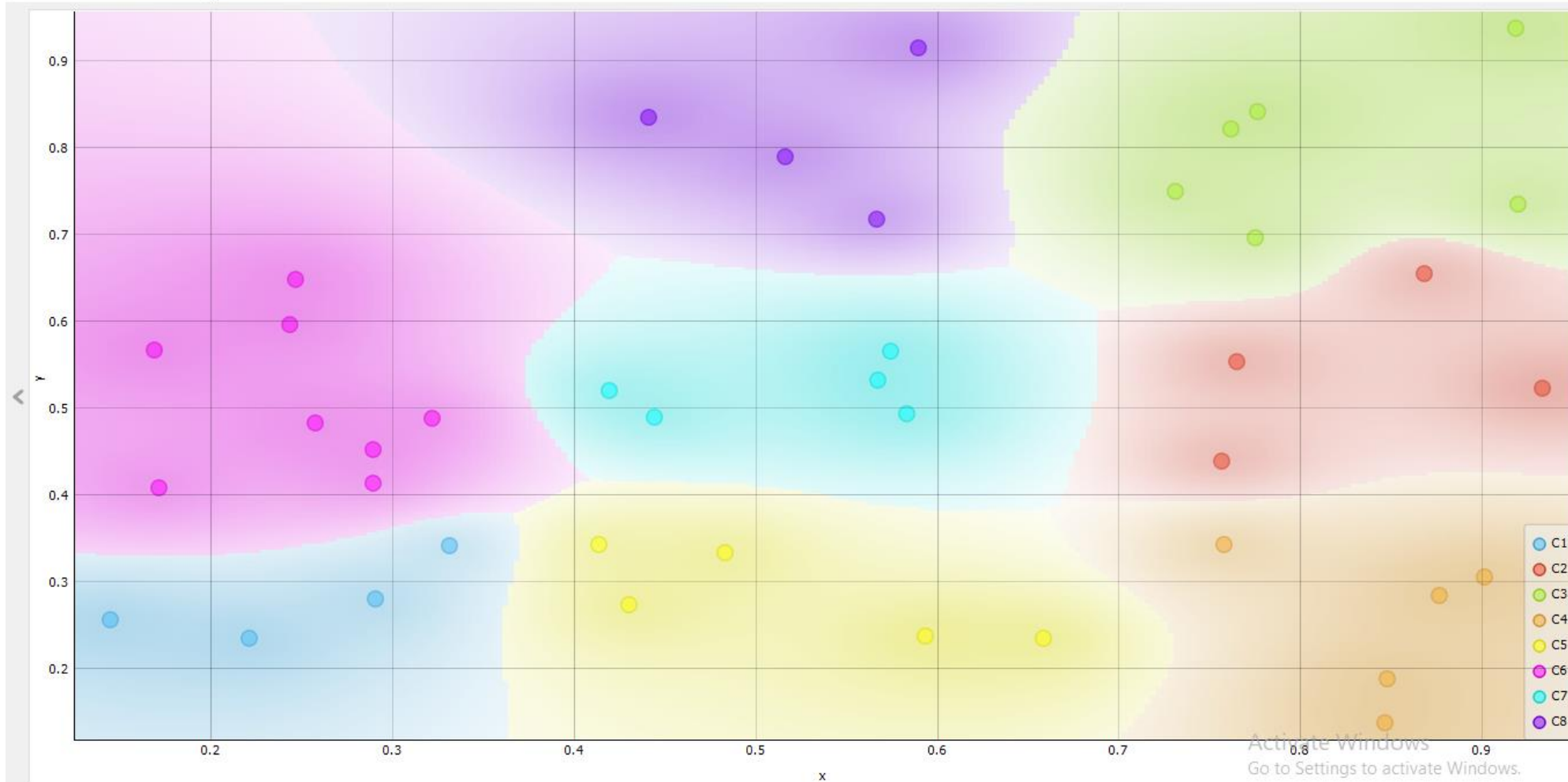
Clusters coloring

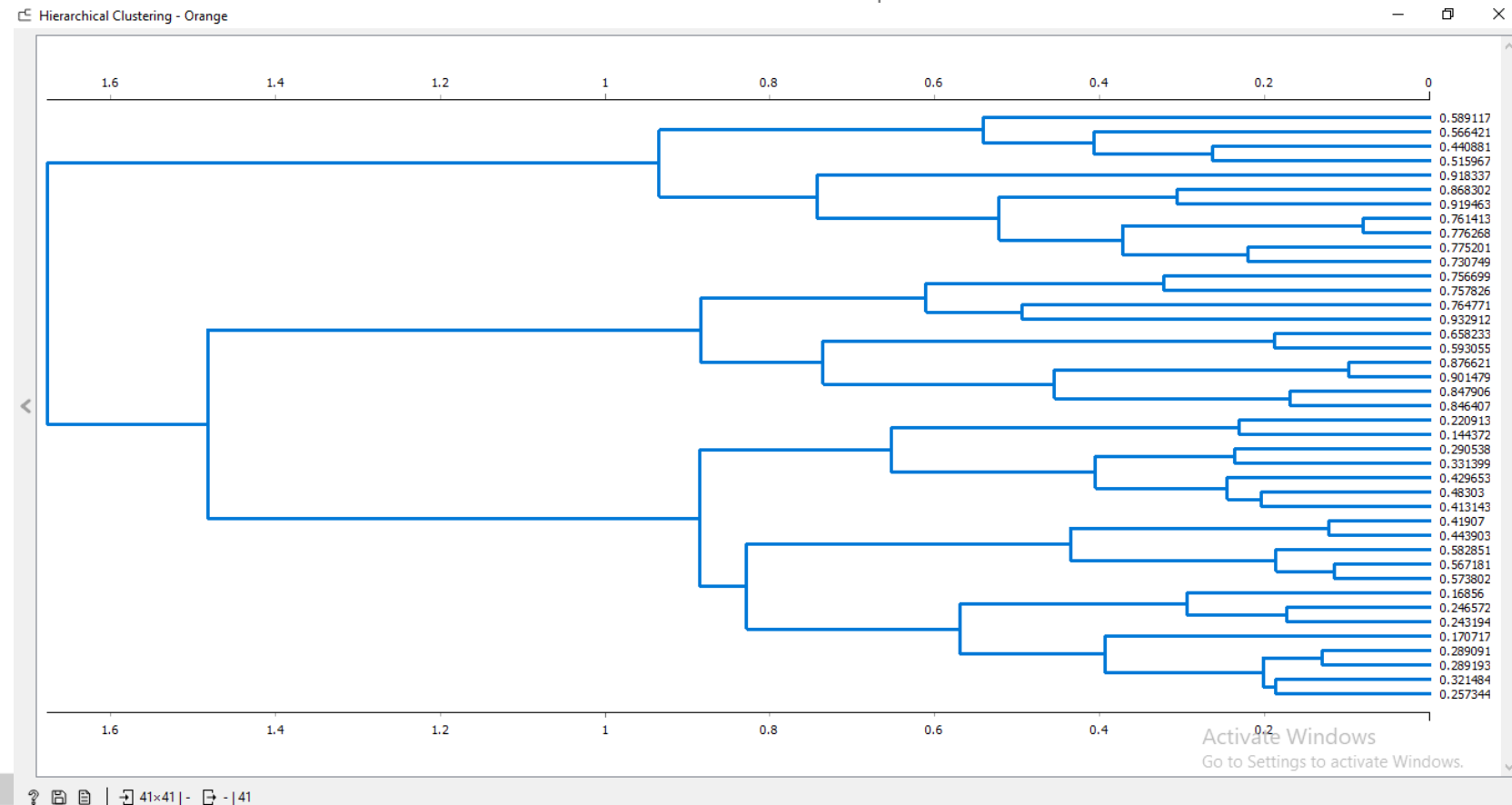
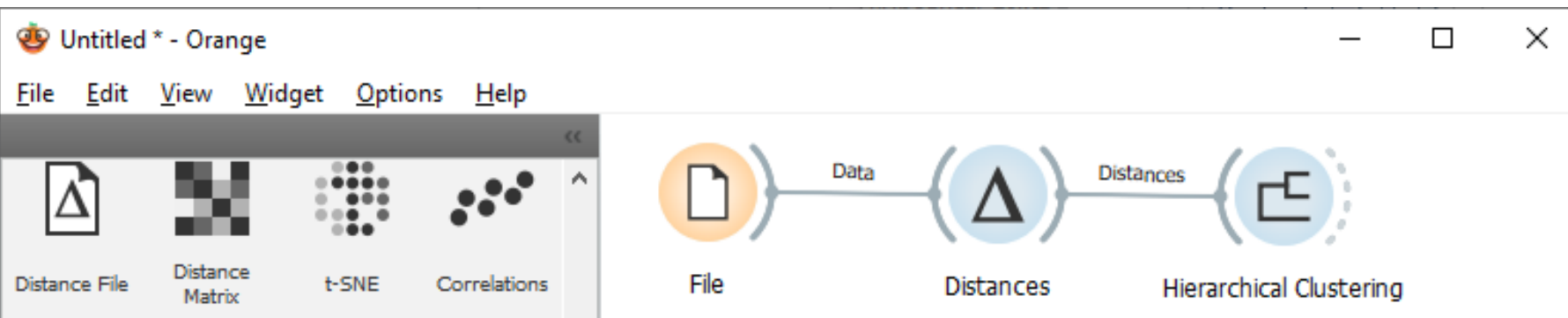


Color regions

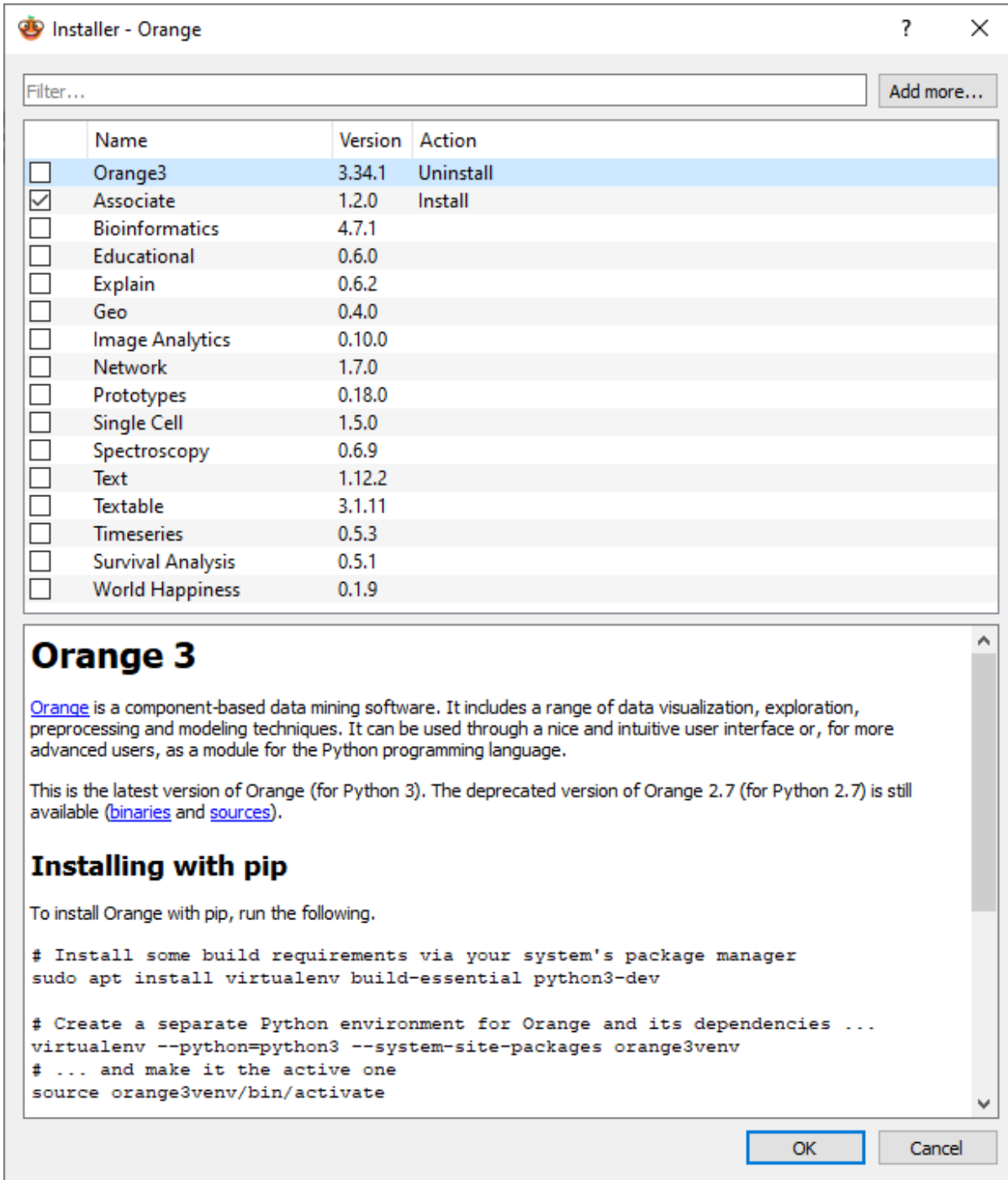
Adjust the number of clusters

Scatter Plot Clustered - Orange





Hierarchical clustering



Orange Add-ons

- Install and use additional sets of widgets
- Application domains:
 - Bioinformatics,
 - Education,
 - Geography and geo-location,
 - Image analysis,
 - Network analysis,
 - Gen and cell,
 - Spectroscopy,
 - Text analysis,
 - Happiness

Task 2

Unsupervised learning, clustering, and hierarchical clustering

1. Fish
2. Birds
3. Fruits
4. Vegetables
5. Whales
6. Mosquito
7. Viruses
8. Roses
9. Bushes
10. Trees

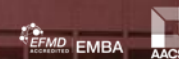
- Remove labels from your dataset
- Implement clustering in various number of clusters
- Adjust parameters of models
- Compare with the results of task 1
- Implement hierarchical clustering
- Try to find sub-classification, close to hierarchical clustering, in your domain

SKEMA BUSINESS SCHOOL

**Introduction to
Artificial Intelligence**

Dmitry A. Zaitsev

<http://daze.ho.ua>



Lesson 8

Unsupervised machine learning

