

Orange: Data Mining Fruitful and Fun

<https://orange.biolab.si/>

Installation

- Conda
 - conda config --add channels conda-forge
 - conda install orange3
- Pip
 - pip install orange3
- Run
 - python -m Orange.canvas

Class Notes

1. Data

- a. File (tab, URL), Data Table, Select Columns (and target), Select Rows
- b. Outliers
- c. Continuize, Discretize
- d. Merge Data, Concatenate
- e. Transpose
- f. Save Data

2. Visualize

- a. Box Plot, Distribution
- b. Scatter Plot, Linear Projection
- c. SelectedData

3. Ex. 1: Hierarchical Clustering with Distance

- a. iris, Distances, Distance Matrix, Hierarchical Clustering
- b. SelectedData, Data Table

4. Ex. 2: Titanic Sieve Diagram

- a. titanic, Sieve Diagram (A sieve diagram is a graphical method for visualizing the frequencies in a two-way contingency table and comparing them to the expected frequencies under assumption of independence.)

5. Ex. 3: Classification with PCA, Decision Tree

- a. iris, PCA, Scatter Plot w/ density, show factors by Data Table
- b. iris, Classification Tree, Tree Viewer, Scatter Plot + Box Plot (select a node in Tree Viewer)
- c. iris, Rank (Gini, Gain, X^2), Box Plot, Distribution

6. Ex. 4: Painted k-means and Silhouette

- a. Paint Data, k-means, Silhouette Plot; On iris data

7. Ex. 5-1: Test & Score

- a. Test & Score: (titanic, Tree, kNN, SVM) -> (Test & Score) -> ROC Analysis

8. Ex. 5-2: Test & Score & misclassification data

- a. Test & Score: (iris, Tree, kNN, SVM) -> (Test & Score) -> Confusion Matrix*3 (select misclassified) -> Venn Diagram + Data Table (see selected misclassified data)
- 9. Ex. 5-3: Prediction with data**
 - a. Predictions
- 10. Ex. 6: PCA problem**
 - a. titanic, Preprocess w/ and w/o PCA on same Classifier Method (LR), ROC; MDS
 - b. Explain why PCA does not perform better.
- 11. Ex. 7: Test on testing data**
 - a. Training and Testing, test on testing data
- 12. Ex. 8: Rules**
 - a. CN2 Rule and Association Rule

Workflow Examples

- Orange built-in workflow tutorial
 - in your Python's \Lib\site-packages\Orange\canvas\application\workflows
- Orange Youtube channel
 - <https://www.youtube.com/channel/UCIKKWBe2SCAEyv7ZNGhle4g>
- Orange workflow screenshot
 - <https://orange.biolab.si/screenshots/>
- Orange: Loading your Data
 - <http://orange-visual-programming.readthedocs.io/loading-your-data/index.html>
- Orange Visual Programming
 - <https://docs.orange.biolab.si/3/visual-programming/>