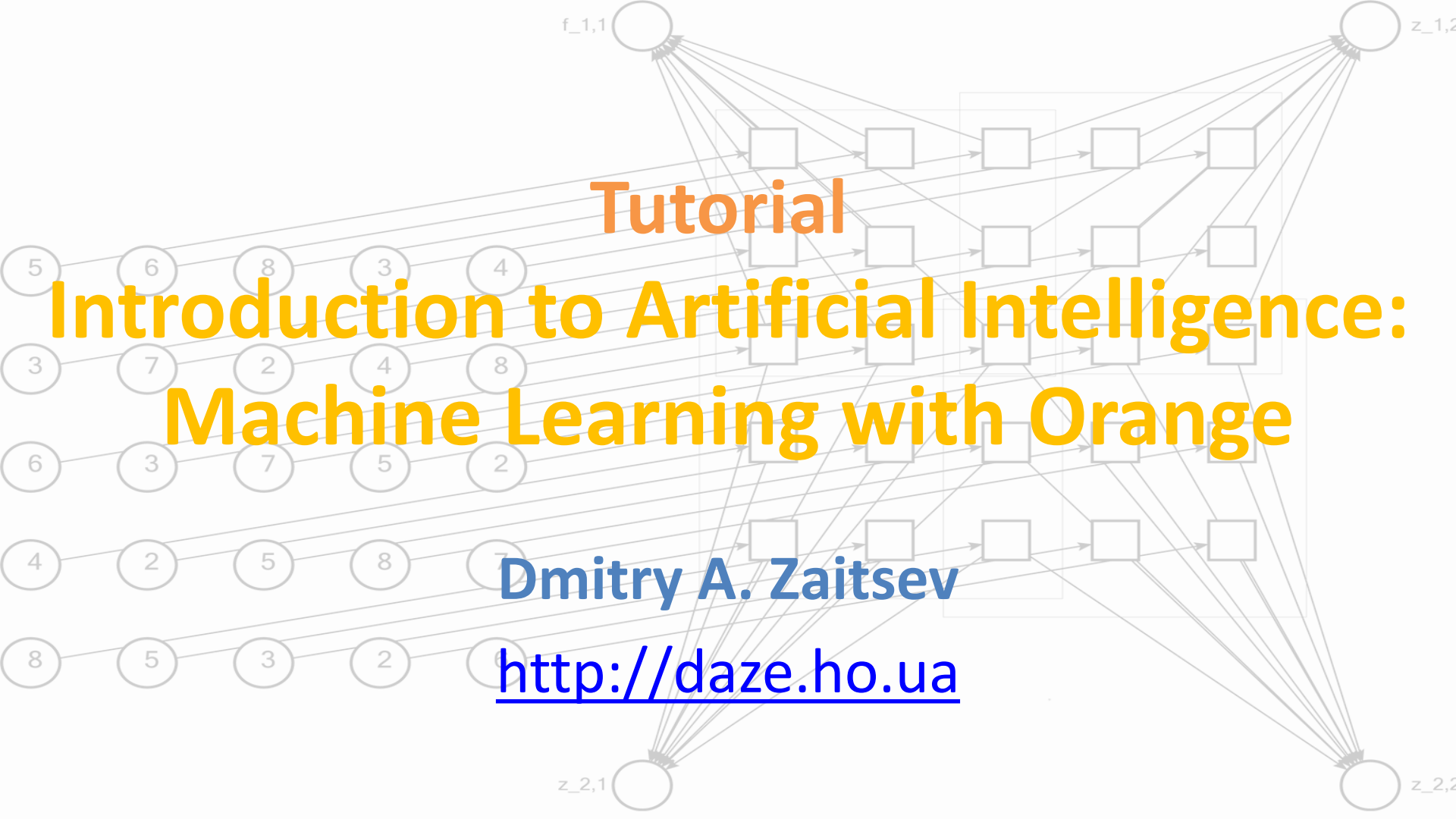




Tutorial

Introduction to Artificial Intelligence: Machine Learning with Orange

Dmitry A. Zaitsev

<http://daze.ho.ua>

Basic sources

- Lecture course “Introduction to AI”
<http://daze.ho.ua/iai.zip>
- Orange Data Mining Fruitful and Fun: Open source machine learning and data visualization
<https://orangedatamining.com>

Artificial Intelligence around us

Recently widespread

- advanced web-search engines (Google Search)
- content recommendation systems of YouTube, Amazon, Netflix
- understanding human speech – Siri, Alexa
- self-driving cars – Waymo
- automatic decision-making in business
- strategic game systems – Chess, Go

Definition of Artificial Intelligence

John McCarthy, 1955

Artificial Intelligence (AI) is the science and engineering of making intelligent machines.

Definition of Artificial Intelligence

John McCarthy, 1997

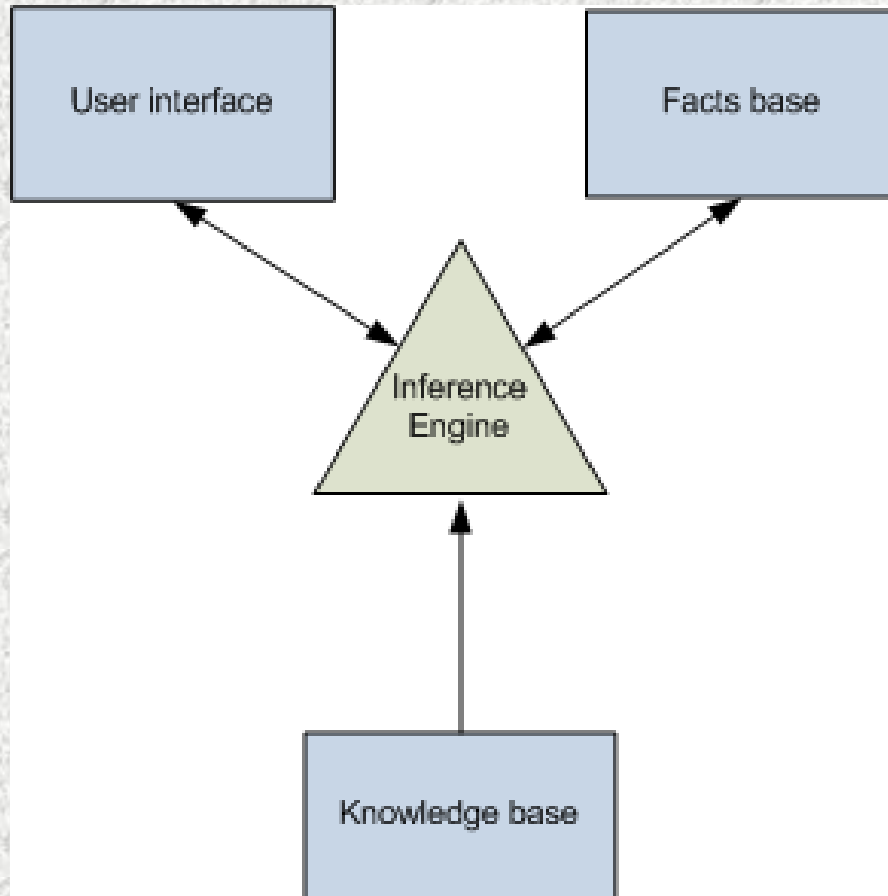
Intelligence is the computational part of the ability to achieve goals in the world.

Intellectual approach

John McCarthy

- goal
- knowledge
- facts
- solution

To achieve a given goal AI system applies knowledge over facts to infer a solution

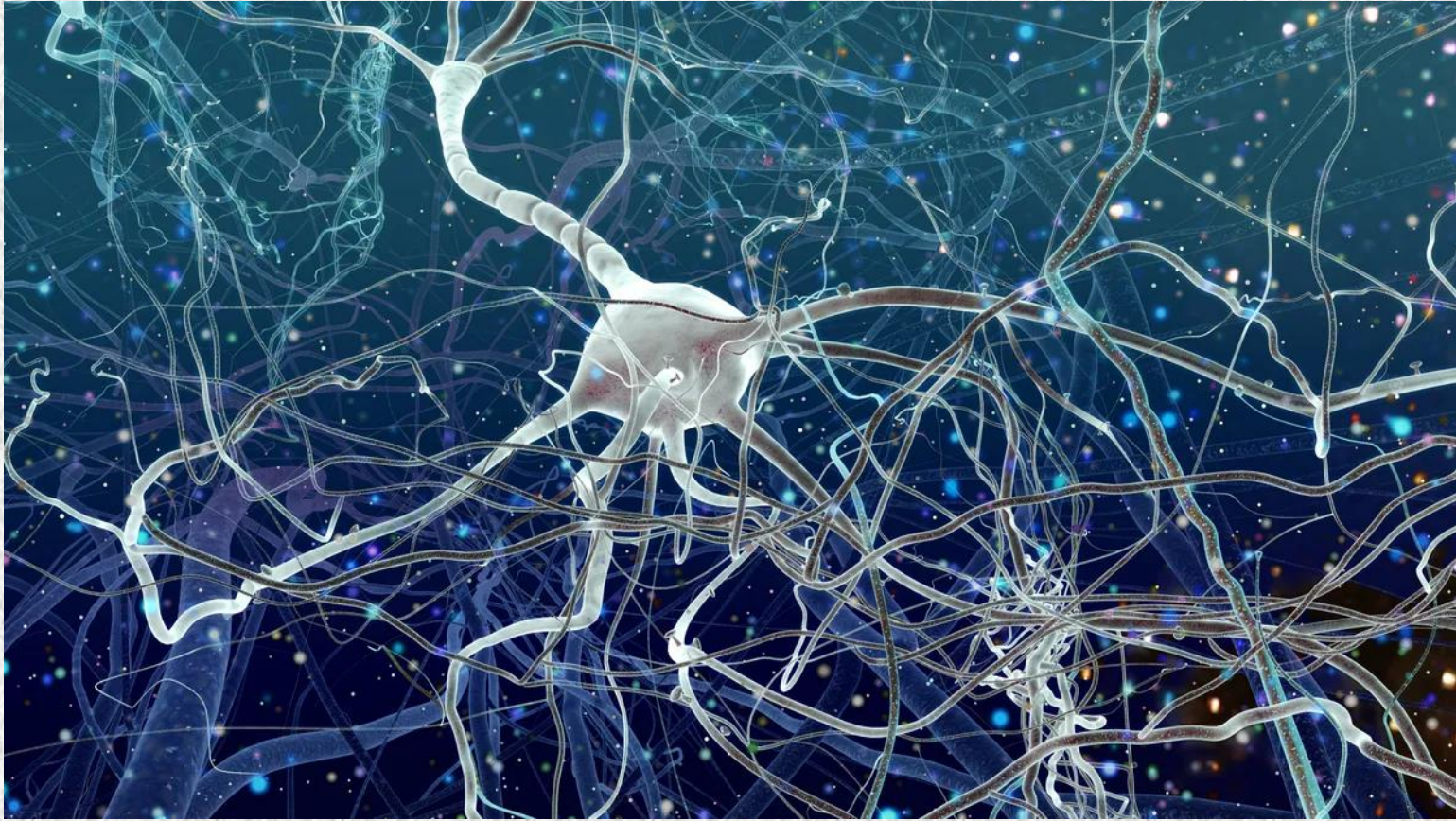


**Knowledge based
system**

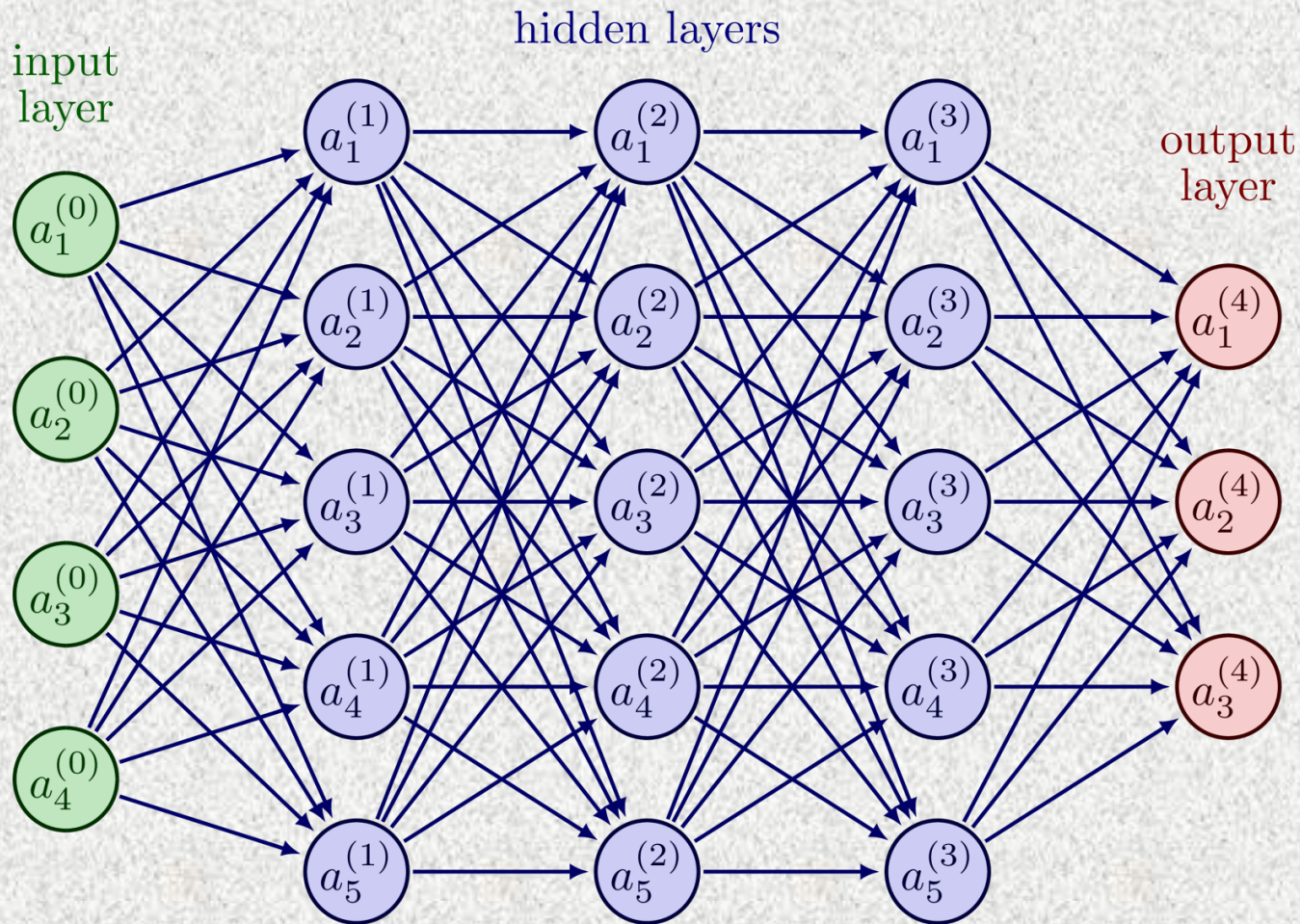
Basic approaches to AI

Psychological vs biological

- simulation of intellectual processes of *reasoning* and solving problems based on *knowledge*
- simulation of human brain structure and functioning – *neuron networks*



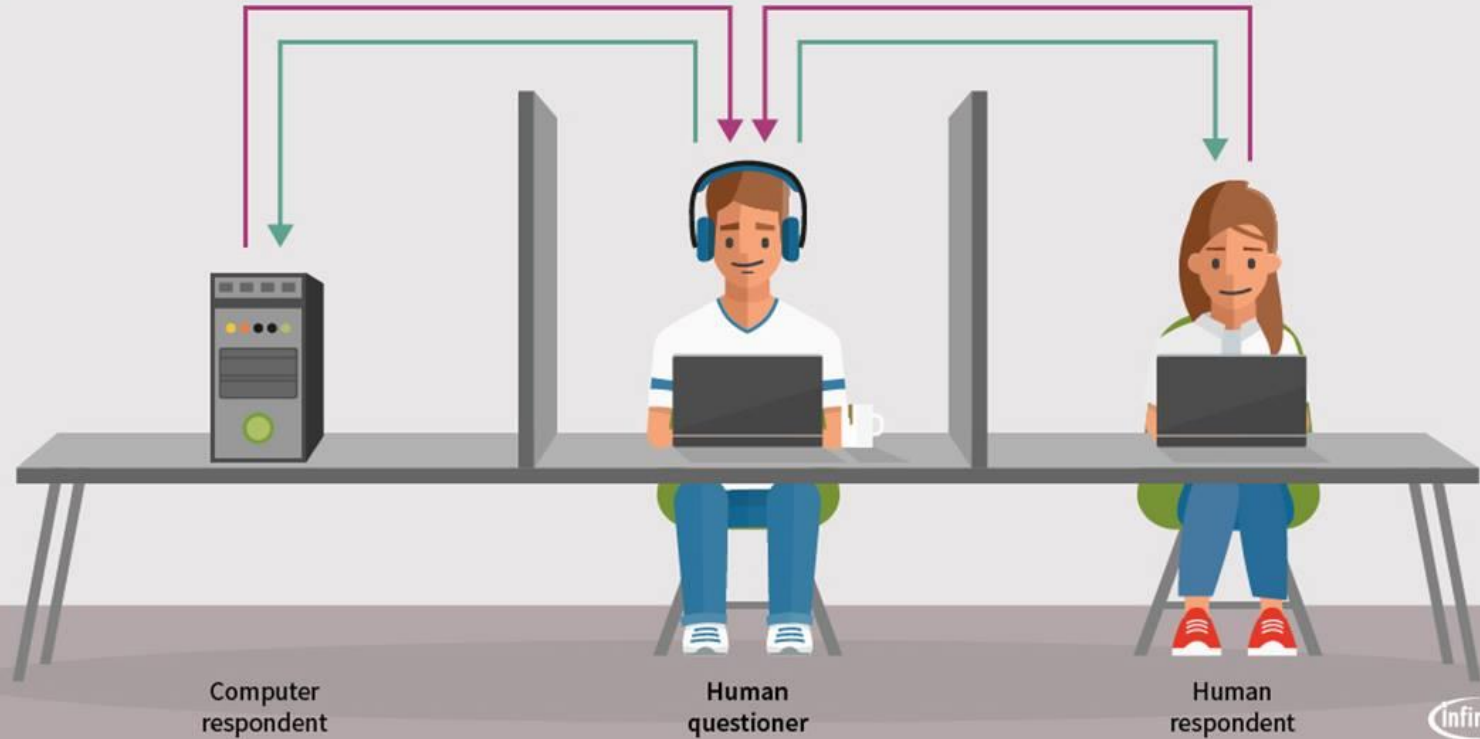
**Super
macro
close-up
view of
neurons
inside of
human
brain**



**Machine
learning
neural
network
architecture**

Turing Test

- Question to respondents
- Answers to questioner



Building AI to help keep the oceans clean

<https://studio.code.org/s/oceans/lessons/1/levels/2>

Building AI to help keep the oceans clean.
Teach NN to recognize garbage in ocean.
Please be careful to avoid hurting creatures.
Please be thorough to catch each possible
kind of garbage. Encounter problems AI
raises and think on them.

When playing, think about

And collect info to prove your point of view

- What are advantages of ML approach?
- Can we guarantee that AI recognizes each given object precisely?
- How we increase precision of AI decisions?
- Can AI do harm to nature?
- How to avoid harmful AI decisions?
- Can we entrust to AI automated control?

Data science

Data Science is a broad, multidisciplinary field that aims to make sense of raw data

- **extract essential information from raw data**
- **identify trends, patterns, connections and correlations in large data sets**
- **tools and techniques: computer programming, predictive analysis, statistics and artificial intelligence, especially machine learning**

Yahoo finance

<https://finance.yahoo.com/>

- financial info on world leading companies
- brief description – Profile, Holders
- current situation description – Summary, Financials
- statistical info on recent periods of time – Statistics, Chart
- data on specified period of time – Historical Data
- forecasts – Analysis





54°F Sunny ^ 🔋 🔊 ENG 10:46 AM 3/3/2023

finance.yahoo.com/quote/1810.HK/financials?p=1810.HK

Search for news, symbols or companies

Finance Home Watchlists My Portfolio Crypto Yahoo Finance Plus News Screeners Markets Videos Personal Finance

Show: **Income Statement** Balance Sheet Cash Flow Annual Quarterly

Income Statement Currency in CNY. All numbers in thousands [Expand All](#)

Breakdown	TTM	12/30/2021	12/30/2020	12/30/2019
> Total Revenue	299,571,815	328,309,145	245,865,633	205,838,682
Cost of Revenue	248,712,678	270,048,204	209,113,771	177,284,649
Gross Profit	50,859,137	58,260,941	36,751,862	28,554,033
> Operating Expense	41,465,510	38,229,595	26,980,465	20,668,672
Operating Income	9,393,627	20,031,346	9,771,397	7,885,361
> Net Non Operating Interest Inc...	138,394	-1,611,631	-2,401,297	402,429
Pretax Income	4,014,712	24,417,033	21,633,432	12,162,646
Tax Provision	2,211,308	5,133,798	1,320,722	2,059,696
> Net Income Common Stockhold...	1,808,449	19,339,321	20,355,504	10,044,164
Average Dilution Earnings	0	-	-	0
Diluted NI Available to Com Stock...	1,808,449	19,339,321	20,355,504	10,044,164
Basic EPS	-	0.78	0.85	0.42

People Also Watch

Symbol	Last Price	Change	% Change
9988.HK Alibaba Group Holding Limited	88.000	+0.400	+0.46%
3690.HK Meituan	142.300	+1.000	+0.71%
0700.HK Tencent Holdings Limited	367.600	+4.800	+1.32%
9618.HK JD.com, Inc.	184.700	+5.600	+3.13%
0175.HK Geely Automobile Holdings Limited	10.620	+0.140	+1.34%

Earnings > [Consensus EPS](#)

54°F Sunny 11:11 AM 3/3/2023

finance.yahoo.com/quote/1810.HK/key-statistics?p=1810.HK

Search for news, symbols or companies

Summary Chart Conversations **Statistics** Historical Data Profile Financials Analysis Options Holders Sustainability

Currency in HKD

Valuation Measures⁴

Market Cap (intraday)	296.53B
Enterprise Value	250.67B
Trailing P/E	150.17
Forward P/E	19.46
PEG Ratio (5 yr expected)	1.48
Price/Sales (ttm)	0.82
Price/Book (mrq)	1.86
Enterprise Value/Revenue	0.84
Enterprise Value/EBITDA	47.38

Financial Highlights

Currency in CNY.

Fiscal Year

Fiscal Year Ends	Dec 31, 2021
------------------	--------------

Trading Information

Stock Price History

Beta (5Y Monthly)	N/A
52-Week Change ³	-10.32%
S&P500 52-Week Change ³	-8.72%
52 Week High ³	15.480
52 Week Low ³	8.310
50-Day Moving Average ³	12.118
200-Day Moving Average ³	11.317

Share Statistics

Avg Vol (3 month) ³	86.31M
Avg Vol (10 day) ³	69.12M
Shares Outstanding ⁵	20.39B
Implied Shares Outstanding ⁶	N/A

See [Statistics Help](#) for definitions of terms used.

Abbreviation Guide:

mrq = Most Recent Quarter
ttm = Trailing Twelve Months
yoy = Year Over Year
lfy = Last Fiscal Year
fye = Fiscal Year Ending

Footnotes

¹ Data provided by Refinitiv.
² Data provided by EDGAR Online.
³ Data provided by Refinitiv.
⁴ Data provided by Refinitiv.
⁵ Data provided by Refinitiv.
⁶ Data provided by Refinitiv.

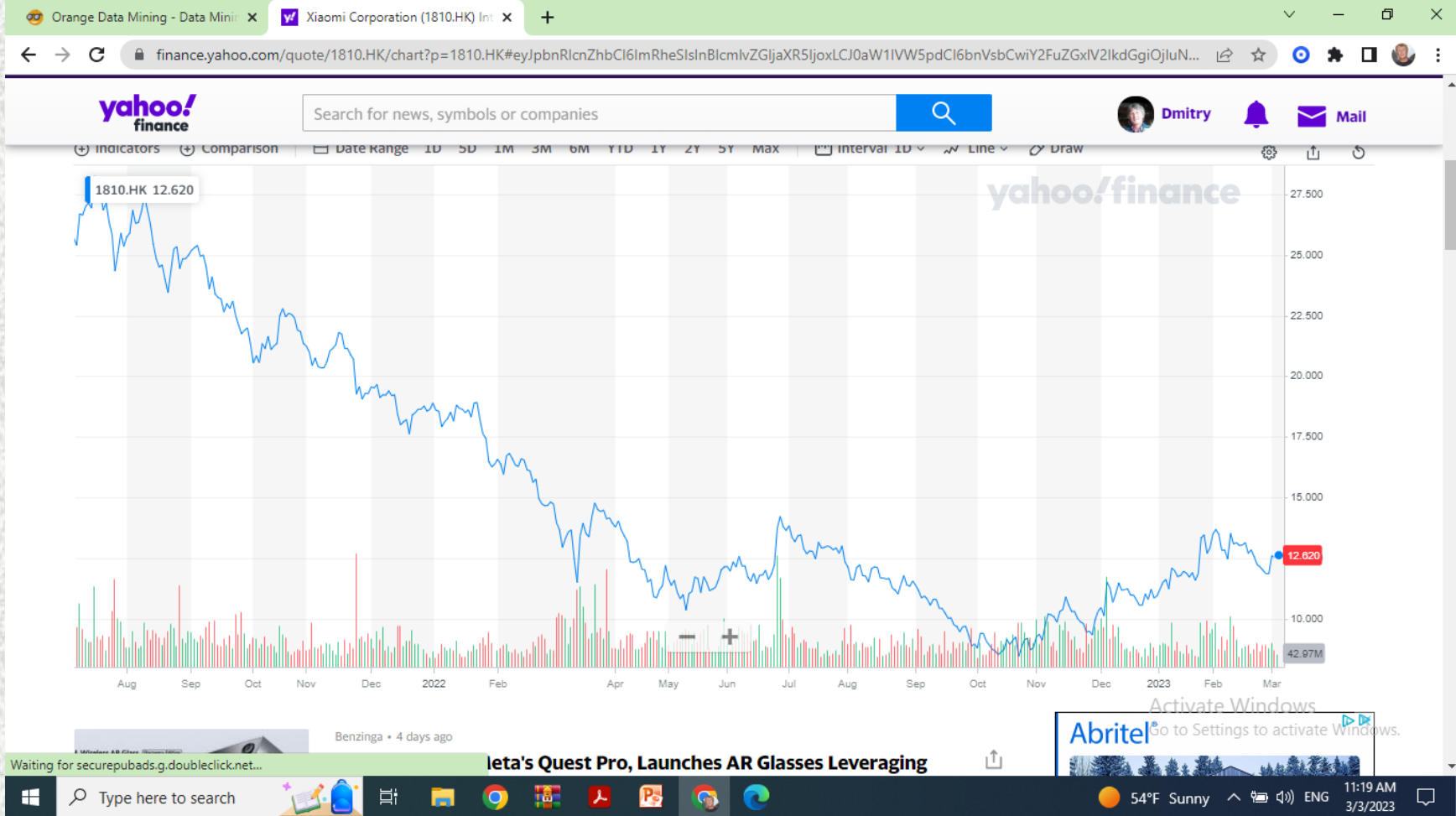
Gagne ton voyage
10 séjours linguistiques. 50 destinations. Gagne un voyage pour toi & ton.ta BFF !
Je participe

Activate Windows
Go to Settings to activate Windows.

Construction 0.7 mile...

11:13 AM
3/3/2023

Chart



Historical prices

Orange Data Mining - Data Mining | Xiaomi Corporation (1810.HK) Stock Price History

finance.yahoo.com/quote/1810.HK/history?p=1810.HK

Search for news, symbols or companies

Time Period: Mar 03, 2022 - Mar 03, 2023 Show: Historical Prices Frequency: Daily Apply

Currency in HKD Download

Date	Open	High	Low	Close*	Adj Close**	Volume
Mar 03, 2023	12.840	12.840	12.440	12.620	12.620	42,967,550
Mar 02, 2023	12.360	12.720	12.360	12.520	12.520	59,850,698
Mar 01, 2023	11.880	12.680	11.860	12.600	12.600	85,423,517
Feb 28, 2023	11.940	12.200	11.760	11.860	11.860	67,656,036
Feb 27, 2023	11.860	12.000	11.640	11.860	11.860	78,381,408
Feb 24, 2023	12.060	12.340	11.960	12.000	12.000	76,140,115
Feb 23, 2023	12.420	12.440	12.100	12.140	12.140	53,332,904
Feb 22, 2023	12.560	12.660	12.080	12.240	12.240	86,113,597
Feb 21, 2023	12.900	13.020	12.500	12.560	12.560	62,268,840
Feb 20, 2023	12.640	13.080	12.640	12.840	12.840	46,286,725
Feb 17, 2023	13.100	13.220	12.660	12.700	12.700	75,713,458
Feb 16, 2023	13.120	13.480	13.020	13.140	13.140	74,662,669
Feb 15, 2023	13.020	13.220	12.800	12.940	12.940	64,885,344

TRY

yahoo/finance

NEW: EXPERIENCE OUR BEST CHARTS YET.

Explore new charts

People Also Watch

Symbol	Last Price	Change	% Change
9988.HK	88.000	+0.400	+0.46%
Alibaba Group Holding Limited			
3690.HK	142.300	+1.000	+0.71%
Meituan			
0700.HK	367.600	+4.800	+1.32%
Tencent Holdings Limited			
9618.HK	184.700	+5.600	+3.13%
JD.com, Inc.			

Windows

Type here to search

54°F Sunny 11:21 AM 3/3/2023

Format CSV – Comma Separated Values

Delimited text file that uses a comma to separate values

Date,Open,High,Low,Close,Adj Close,Volume

2022-03-04,14.240000,14.300000,13.700000,13.960000,13.960000,172115195

2022-03-07,13.400000,13.920000,13.040000,13.680000,13.680000,186408382

2022-03-08,13.700000,13.980000,12.900000,12.940000,12.940000,179112275

2022-03-09,13.080000,13.540000,12.740000,13.300000,13.300000,168374161

2022-03-10,13.800000,13.900000,13.380000,13.580000,13.580000,81509483

2022-03-11,12.980000,13.880000,12.860000,13.660000,13.660000,137860658

...

Excel imports CSV file

1810.HK - Microsoft Excel

Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид

Вставить Буфер обмена Шрифт Выравнивание Число Стили

Общий

Условное форматирование Форматировать как таблицу Стили

Вставить Удалить Формат Ячейки

Сортировка и фильтр Найти и выделить Редактирование

1810.HK

	A	B	C	D	E	F	G	H	I	J
1	Date	Open	High	Low	Close	Adj Close	Volume			
2	4-Mar-2022	14.24	14.3	13.7	13.96	13.96	172115195			
3	7-Mar-2022	13.4	13.92	13.04	13.68	13.68	186408382			
4	8-Mar-2022	13.7	13.98	12.9	12.94	12.94	179112275			
5	9-Mar-2022	13.08	13.54	12.74	13.3	13.3	168374161			
6	10-Mar-2022	13.8	13.9	13.38	13.58	13.58	81509483			
7	11-Mar-2022	12.98	13.88	12.86	13.66	13.66	137860658			
8	14-Mar-2022	13.44	13.48	12.36	12.48	12.48	169050905			
9	15-Mar-2022	11.86	12.92	11.38	11.5	11.5	267858299			
10	16-Mar-2022	12.02	13.46	11.84	13.4	13.4	277546915			
11	17-Mar-2022	14.5	14.5	13.42	13.94	13.94	209742078			
12	18-Mar-2022	13.82	13.82	13.04	13.52	13.52	154366597			

Activate Windows
Go to Settings to activate Windows.

Готово

Type here to search

54°F Sunny 11:37 AM 3/3/2023

Orange: Open source machine learning and data visualization

<https://orangedatamining.com/>

- Install Python
- Download and install Orange
- YouTube tutorials:
<https://www.youtube.com/channel/UClKKWBe2SCAEyv7ZNGhle4g>
- Run Orange:
 - cmd
 - `python3 -m Orange.canvas`

Example of Orange Use

testa-ex1.ows* - Orange

File Edit View Widget Options Help

Data

File CSV Import Datasets SQL Table

Data Table Paint Data Data Info Rank

Edit Domain Color Feature Statistics Save Data

Transform

Visualize

Model

Evaluate

Unsupervised

Data Table

View the dataset in a spreadsheet.

[more...](#)

CSV File Import Data Table Scatter Plot

Data Table - Orange

Info

13 instances (no missing data)
2 features
No target variable.
No meta attributes.

Variables

☒ Show variable labels (if present)

☐ Visualize numeric values

☒ Color by instance classes

Selection

☒ Select full rows

	Date	Open
1	2022-02-01 00:00...	311.737
2	2022-03-01 00:00...	289.893
3	2022-04-01 00:00...	360.383
4	2022-05-01 00:00...	286.923
5	2022-06-01 00:00...	251.72
6	2022-07-01 00:00...	227
7	2022-08-01 00:00...	301.277
8	2022-09-01 00:00...	272.58
9	2022-10-01 00:00...	254.5
10	2022-11-01 00:00...	234.05
11	2022-12-01 00:00...	197.08

Scatter Plot - Orange

Axes

Axis x: Date

Axis y: Open

Find Informative Projections

Attributes

Color: (Same color)

Shape: (Same shape)

Size: (Same size)

Label: (No labels)

Zoom/Select

☒ Send Automatically

$r = -0.83$

Activate Windows
Go to Settings to activate Windows.

56°F Mostly sunny 12:07 PM 3/3/2023

Basic Orange concepts

Draw your workflow of data processing and visualization

- **data analysis workflow**
- **widgets – computational units of Orange:**
 - read data
 - process data
 - visualize data
- **basic functions:**
 - data clustering
 - build predictive models

Create workflow for Xiaomi stock prices

Analyze and visualize historical data

- **File widget – load data from file or URL**
- **Data Table widget – print table**
- **Connect widgets – data flow connecting of input-output channels – just drag a line connecting widgets**
- **Scatter Plot widget – represent (visualize) data values by dots**

Attributes and options of widgets

DoubleClick to open widget specification

- **File widget**
 - Path or URL
 - Import options – encoding, delimiter, ignore columns
- **Data Table widget**
 - visualize numeric values
- **Scatter Plot widget**
 - axes
 - attributes: color, shape, size, label; show legend, show gridlines, show regression line

Principles of Orange work

DoubleClick to open widget specification

- I. Set of widget, attributes, and options**
- II. Draw workflow as a network of connected widgets**
- III. Widgets communicate with each other**
- IV. Changes in some widget are immediately propagated through the workflow**

Statistical values

Basic statistical values to calculate and analyze

- minimum, maximum, range
- mean
- median, mode
- dispersion, variance
- higher rank statistical moments
- covariance and correlation

Machine learning procedure

Neuron network learns and applies results of learning

- collect dataset or get data in real time
- train neuron network – adjust NN parameters for your task
- test neuron network – check how it recognizes new data
- use neuron network – practical use for prediction, recognition, classification, recommendation, control
- upload NN onto FPGA – Field Programmable Gate Arrays – AI in a box

Examples of FPGA



Types of machine learning

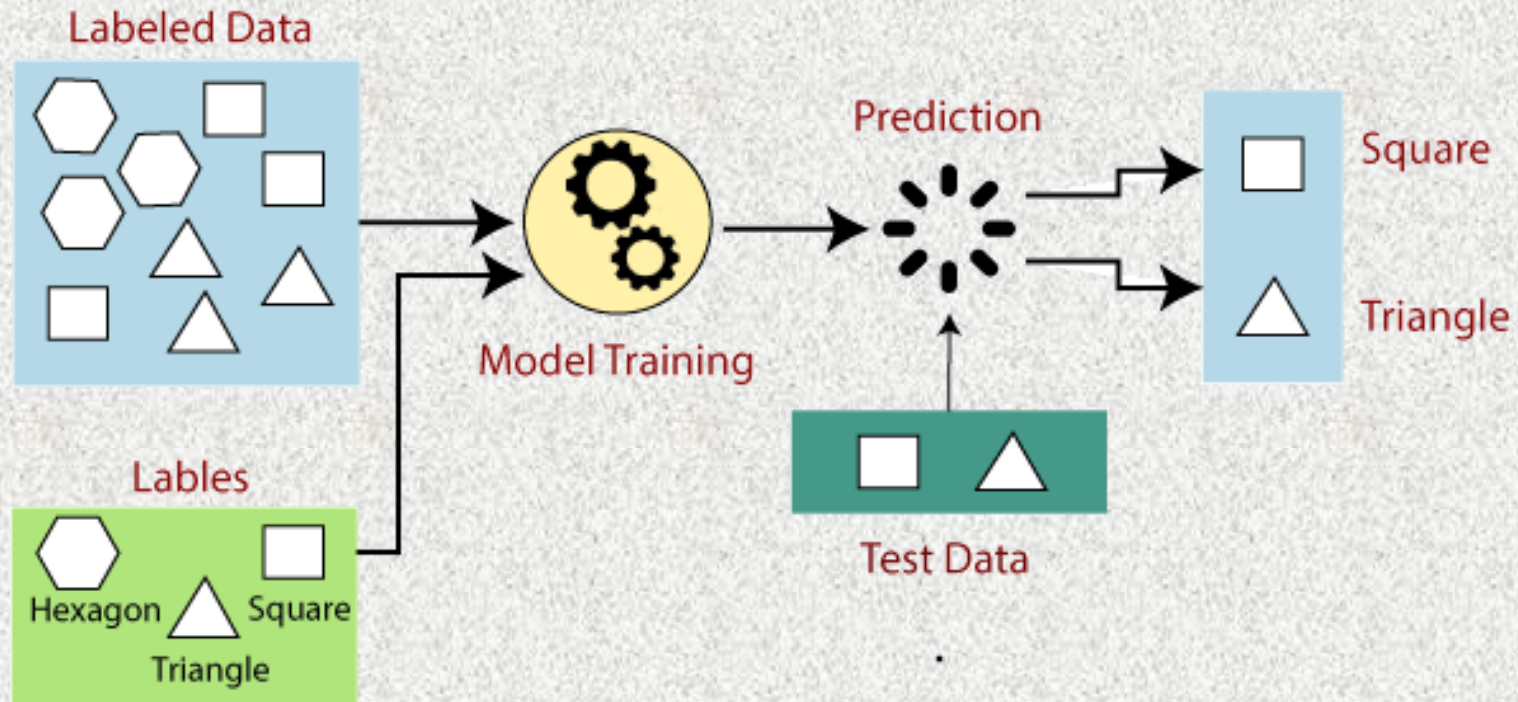
Learn and make predictions, find patterns, or classify data

- supervised – labelled training sets and data – recognize and classify data (fruits or vegetables, cats or dogs)
- unsupervised – unlabelled data – identify patterns and make decision – clustering (customer preferences)
- reinforced – learn by interacting with its environment and getting a positive or negative reward (traffic control)

Supervised learning

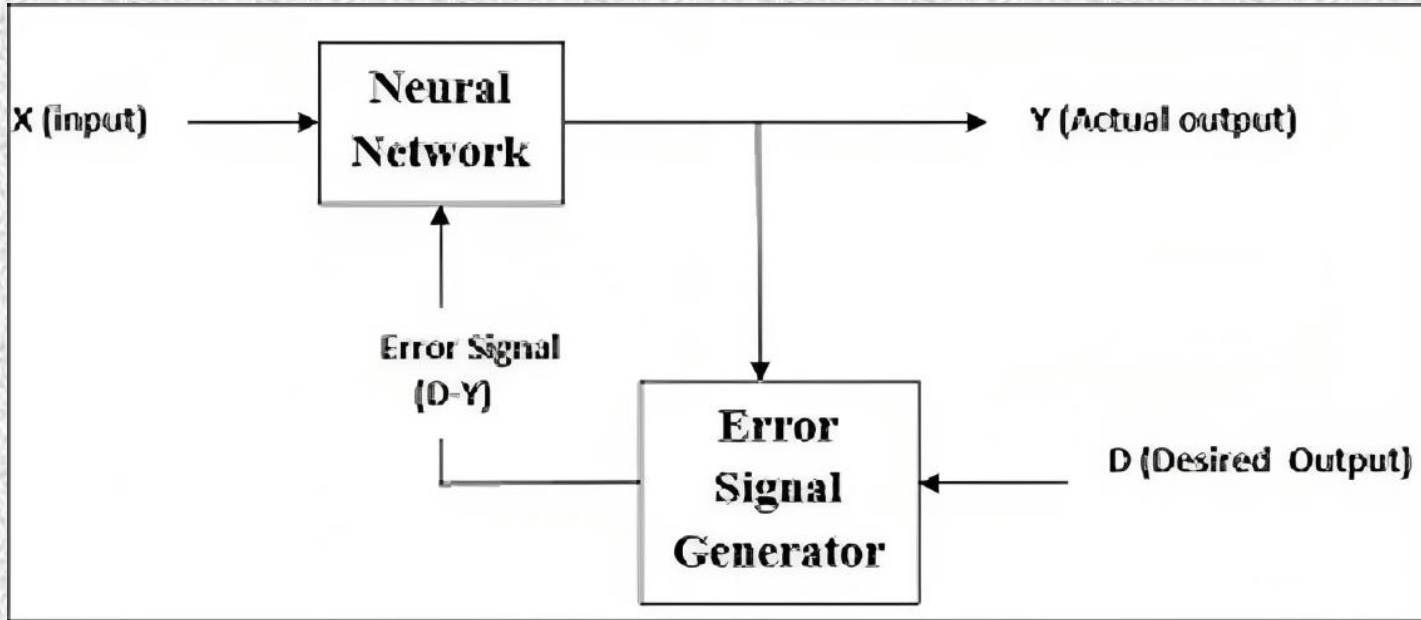
- Supervised learning describes a class of problem that involves using a model to learn a mapping between input examples and the target variable.
- Models are fit on training data comprised of inputs and outputs and used to make predictions on test sets where only the inputs are provided and the outputs from the model are compared to the target variables and used to estimate the skill of the model.
- Classification: Supervised learning problem that involves predicting a class label.
- Regression: Supervised learning problem that involves predicting a numerical label.

Supervised learning example scheme

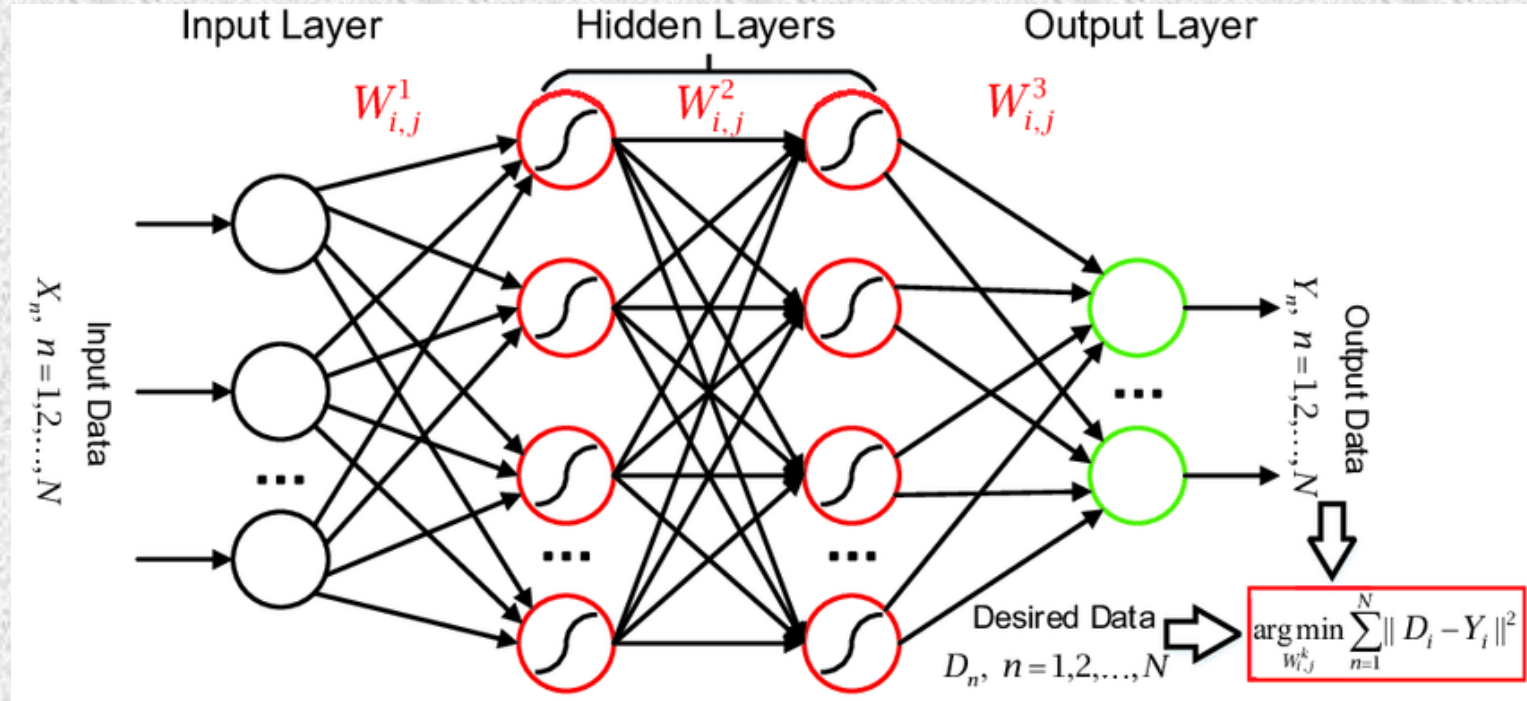


How to train NN

Error signal based correction



Neuron network for deep learning



Models for supervised learning

Basic models

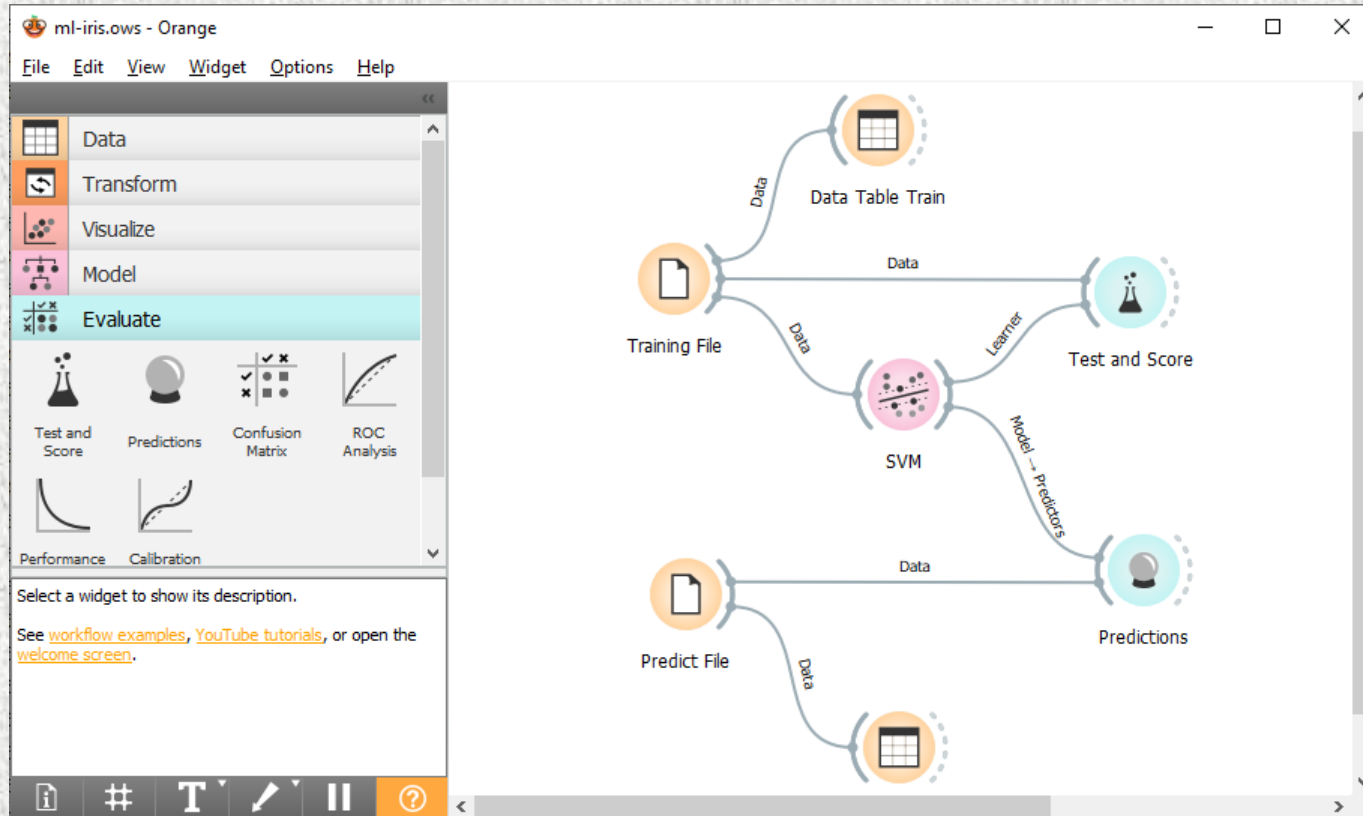
- Nearest Neighbor
- Naive Bayes
- Decision Trees
- Linear Regression
- Support Vector Machines (SVM)
- Neural Networks

Machine learning with Orange

Draw workflow, attach dataset, and use

- Model:
- Calibrated Learning, kNN, Tree, Random Forest, Gradient Boosting, SVM, Linear Regression, Logistic Regression, Naïve Bayes, Ada Boost, Curve Fit, Neural Network, Stochastic Gradient descend
- Evaluate:
- Test and Score, Prediction, Confusion Matrix, ROC Analysis, Performance Curve, Calibration Plot

Classifying with Orange



Training data set - irises

Data Table Train - Orange

Info
150 instances (no missing data)
4 features
Target with 3 values
No meta attributes.

Variables

☒ Show variable labels (if present)

☐ Visualize numeric values

☒ Color by instance classes

Selection

☒ Select full rows

	iris	sepal length	sepal width	petal length	petal width
1	Iris-setosa	5.1	3.5	1.4	0.2
2	Iris-setosa	4.9	3.0	1.4	0.2
3	Iris-setosa	4.7	3.2	1.3	0.2
4	Iris-setosa	4.6	3.1	1.5	0.2
5	Iris-setosa	5.0	3.6	1.4	0.2
6	Iris-setosa	5.4	3.9	1.7	0.4
7	Iris-setosa	4.6	3.4	1.4	0.3
8	Iris-setosa	5.0	3.4	1.5	0.2
9	Iris-setosa	4.4	2.9	1.4	0.2
10	Iris-setosa	4.9	3.1	1.5	0.1
11	Iris-setosa	5.4	3.7	1.5	0.2
12	Iris-setosa	4.8	3.4	1.6	0.2
13	Iris-setosa	4.8	3.0	1.4	0.1
14	Iris-setosa	4.3	3.0	1.1	0.1
15	Iris-setosa	5.8	4.0	1.2	0.2
16	Iris-setosa	5.7	4.4	1.5	0.4
17	Iris-setosa	5.4	3.9	1.3	0.4
18	Iris-setosa	5.1	3.5	1.4	0.3
19	Iris-setosa	5.7	3.8	1.7	0.3
20	Iris-setosa	5.1	3.8	1.5	0.3
21	Iris-setosa	5.4	3.4	1.7	0.2
22	Iris-setosa	5.1	3.7	1.5	0.4
23	Iris-setosa	4.6	3.6	1.0	0.2
24	Iris-setosa	5.1	3.3	1.7	0.5
25	Iris-setosa	4.8	3.4	1.9	0.2
26	Iris-setosa	5.0	3.0	1.6	0.2
27	Iris-setosa	5.0	3.4	1.6	0.4
28	Iris-setosa	5.2	3.5	1.5	0.2
29	Iris-setosa	5.2	3.4	1.4	0.2
30	Iris-setosa	4.7	3.2	1.6	0.2
31	Iris-setosa	4.9	3.1	1.6	0.2

Restore Original Order

☒ Send Automatically

150 150 | 150

Training File - Orange

Source

☒ File:

☐ URL:

File Type

Automatically detect type

Info

Iris flower dataset
Classical dataset with 150 instances of Iris setosa, Iris virginica and Iris versicolor.

150 instances
4 features (no missing values)
Classification; categorical class with 3 values (no missing values)
0 meta attributes

Columns (Double click to edit)

	Name	Type	Role	Values
1	sepal length	numeric	feature	
2	sepal width	numeric	feature	
3	petal length	numeric	feature	
4	petal width	numeric	feature	
5	iris	categorical	target	Iris-setosa, Iris-versicolor, Iris-virginica

150

Activate Windows
Go to Settings to activate Windows.

Type here to search

62°F Partly sunny 8:50 AM
4/1/2023

Prediction

Predictions - Orange

Show probabilities for (None)

Restore Original Order

	Random Forest	sepal length	sepal width	petal length	petal width
1	Iris-setosa	5.1	3.5	1.4	0.2
2	Iris-setosa	4.9	3.0	1.4	0.2
3	Iris-setosa	4.7	3.2	1.3	0.2
4	Iris-setosa	4.6	3.1	1.5	0.2
5	Iris-setosa	5.0	3.6	1.4	0.2
6	Iris-setosa	5.4	3.9	1.7	0.4
7	Iris-setosa	4.6	3.4	1.4	0.3
8	Iris-setosa	5.0	3.4	1.5	0.2
9	Iris-setosa	4.4	2.9	1.4	0.2
10	Iris-setosa	4.9	3.1	1.5	0.1
11	Iris-setosa	5.4	3.7	1.5	0.2
12	Iris-setosa	4.8	3.4	1.6	0.2
13	Iris-setosa	4.8	3.0	1.4	0.1
14	Iris-setosa	4.3	3.0	1.1	0.1
15	Iris-setosa	5.8	4.0	1.2	0.2
16	Iris-setosa	5.7	4.4	1.5	0.4
17	Iris-setosa	5.4	3.9	1.3	0.4
18	Iris-setosa	5.1	3.5	1.4	0.3
19	Iris-setosa	5.7	3.8	1.7	0.3
20	Iris-setosa	5.1	3.8	1.5	0.3
21	Iris-setosa	5.4	3.4	1.7	0.2
22	Iris-setosa	5.1	3.7	1.5	0.4
23	Iris-setosa	4.6	3.6	1.0	0.2
24	Iris-setosa	5.1	3.3	1.7	0.5
25	Iris-setosa	4.8	3.4	1.9	0.2
26	Iris-setosa	5.0	3.0	1.6	0.2
27	Iris-setosa	5.0	3.4	1.6	0.4
28	Iris-setosa	5.2	3.5	1.5	0.2

Activate Windows
Go to Settings to activate Windows.

150 | 150 | -

Type here to search

69°F Mostly cloudy

ENG

12:30 PM
4/1/2023

Prediction

Predictions - Orange

Show probabilities for (None) [Restore Original Order](#)

	SVM	KNN	Random Forest	Neural Network	sepal length	sepal width	petal length	petal width
1	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	5.1	3.5	1.4	0.2
2	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	4.9	3.0	1.4	0.2
3	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	4.7	3.2	1.3	0.2
4	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	4.6	3.1	1.5	0.2
5	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	5.0	3.6	1.4	0.2
6	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	5.4	3.9	1.7	0.4
7	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	4.6	3.4	1.4	0.3
8	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	5.0	3.4	1.5	0.2
9	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	4.4	2.9	1.4	0.2
10	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	4.9	3.1	1.5	0.1
11	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	5.4	3.7	1.5	0.2
12	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	4.8	3.4	1.6	0.2
13	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	4.8	3.0	1.4	0.1
14	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	4.3	3.0	1.1	0.1
15	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	5.8	4.0	1.2	0.2
16	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	5.7	4.4	1.5	0.4
17	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	5.4	3.9	1.3	0.4
18	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	5.1	3.5	1.4	0.3
19	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	5.7	3.8	1.7	0.3
20	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	5.1	3.8	1.5	0.3
21	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	5.4	3.4	1.7	0.2
22	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	5.1	3.7	1.5	0.4
23	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	4.6	3.6	1.0	0.2
24	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	5.1	3.3	1.7	0.5
25	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	4.8	3.4	1.9	0.2
26	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	5.0	3.0	1.6	0.2
27	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	5.0	3.4	1.6	0.4
28	Iris-setosa	Iris-setosa	Iris-setosa	Iris-setosa	5.2	3.5	1.5	0.2

Activate Windows
Go to Settings to activate Windows.

Test and Score

Test and Score - Orange

☒ Cross validation

Number of folds: 5

☒ Stratified

☐ Cross validation by feature

☐ Random sampling

Repeat train/test: 10

Training set size: 66 %

☒ Stratified

☐ Leave one out

☐ Test on train data

☐ Test on test data

Evaluation results for target (None, show average over classes)

Model	AUC	$\hat{CA}$	F1	Precision	Recall
Neural Network	0.993	0.947	0.947	0.948	0.947
SVM	0.997	0.960	0.960	0.960	0.960
kNN	0.987	0.967	0.967	0.967	0.967
Random Forest	0.991	0.967	0.967	0.967	0.967

Compare models by: Classification accuracy

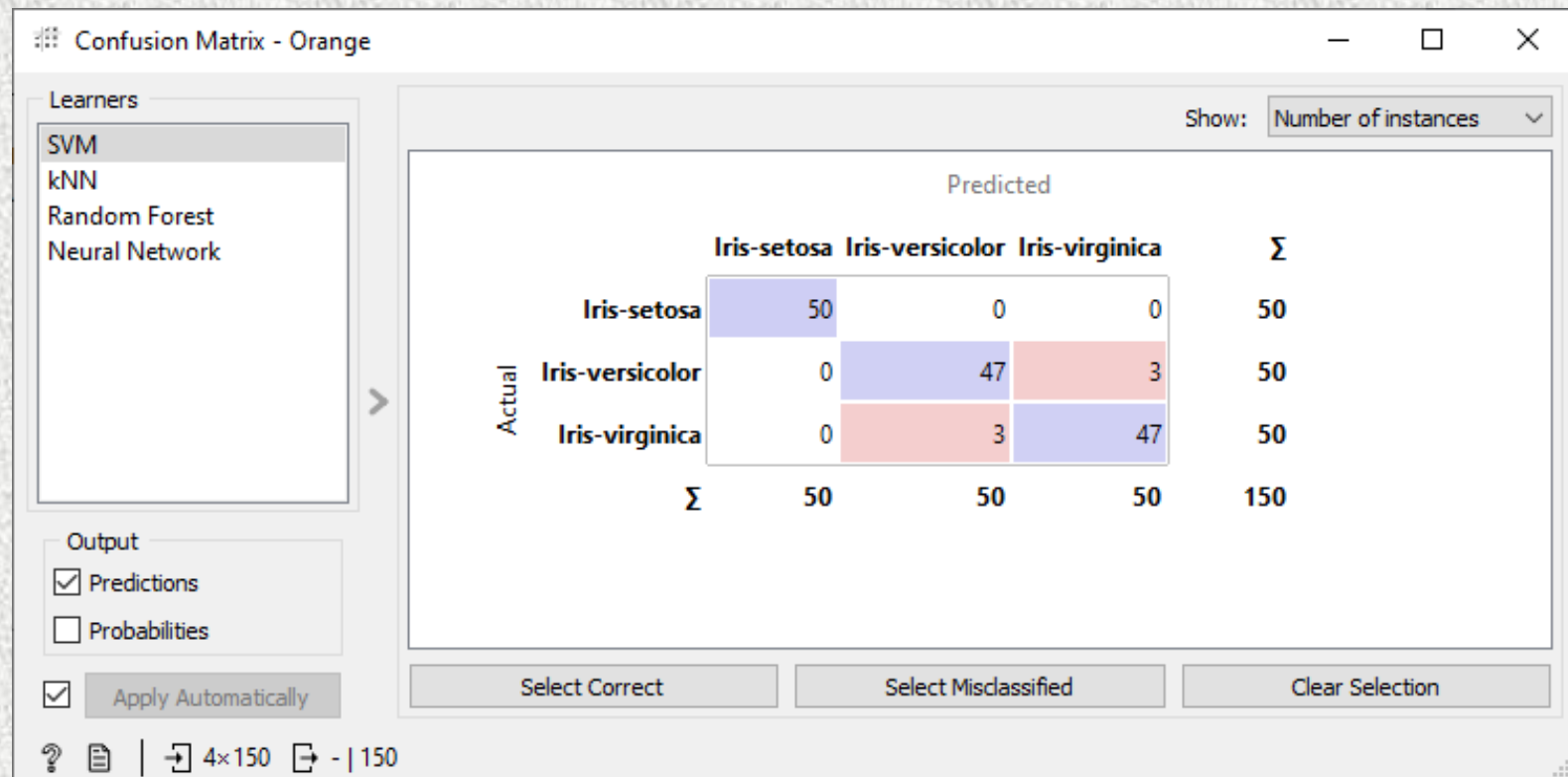
☐ Negligible diff.: 0.1

	Neural Network	SVM	kNN	Random Forest
Neural Network		0.169	0.187	0.187
SVM	0.831		0.271	0.271
kNN	0.813	0.729		0.500
Random Forest	0.813	0.729	0.500	

Table shows probabilities that the score for the model in the row is higher than that of the model in the column. Small numbers show the probability that the difference is negligible.

150 | - | 150 | 4x150

Confusion matrix



Save the best trained model

ml-iris-multi.ows* - Orange

File Edit View Widget Options Help

Data
Transform
Visualize
Model

Constant CN2 Rule Induction Calibrated Learner kNN
Tree Random Forest Gradient Boosting SVM
Linear Regression Logistic Regression Naive Bayes AdaBoost
Curve Fit Neural Network Stochastic Gradient Descent Stacking
Save Model Load Model

Save Model
Save a trained model to an output file.
[more...](#)

Training File

Data Table Train data

Data

SVM

kNN

Random Forest

Neural Network

Data

Predict File

Data Table Predict

Test and Score

Evaluation Results

Confusion Matrix

Model

Save Model

Save Model - Ora... ? X

☒ Autosave when receiving new data

Save as iris-best.pkcds Save as ...

Activate Windows
Go to Settings to activate Windows.

65°F Partly sunny 11:35 AM 4/1/2023

Use saved model

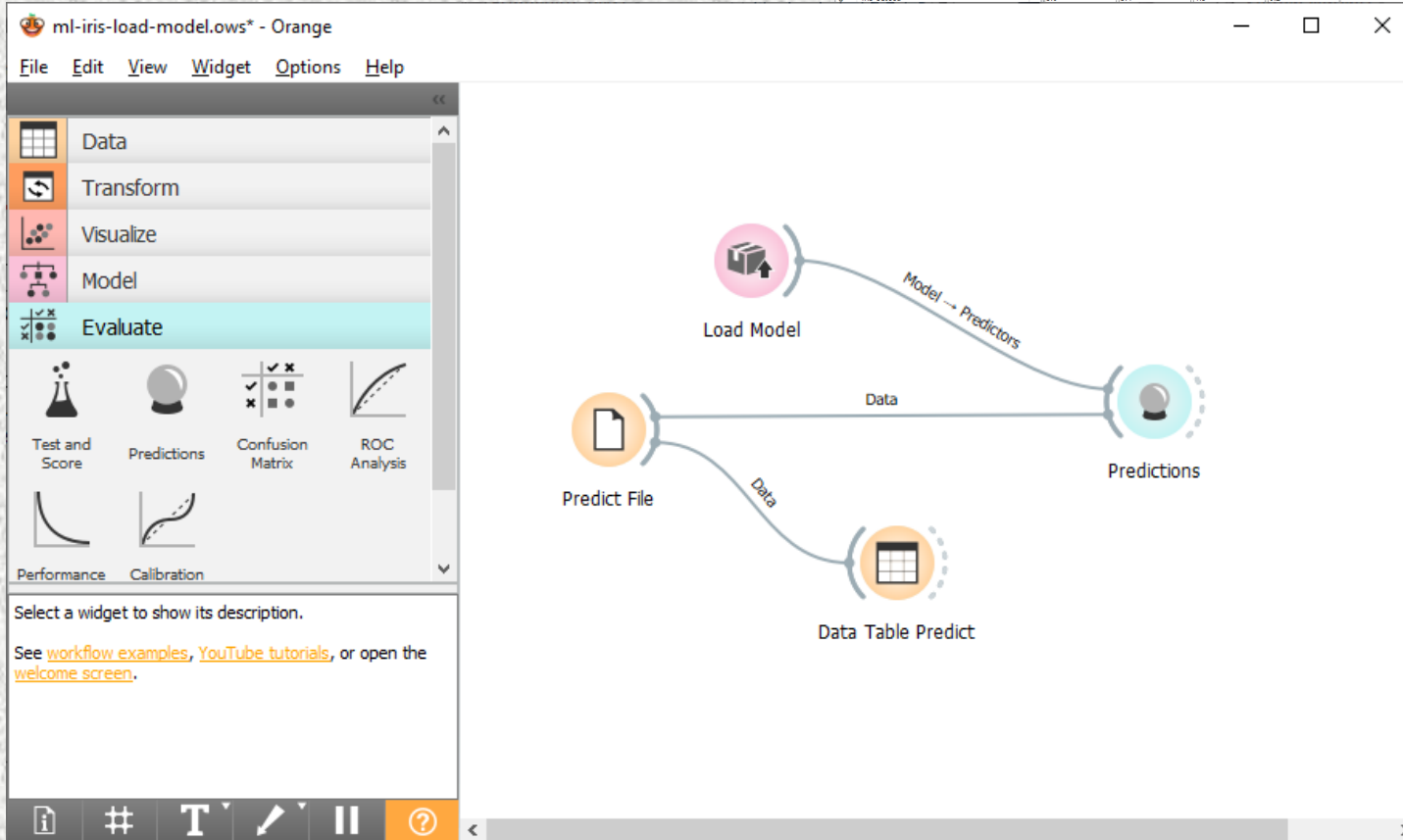
Predictions - Orange

Show probabilities for (None)

Random Forest

	sepal length	sepal width	petal length	petal width	
1	Iris-setosa	5.1	3.5	1.4	0.2
2	Iris-setosa	4.9	3.0	1.4	0.2
3	Iris-setosa	4.7	3.2	1.3	0.2
4	Iris-setosa	4.6	3.1	1.5	0.2
5	Iris-setosa	5.0	3.6	1.4	0.2
6	Iris-setosa	5.4	3.9	1.7	0.4
7	Iris-setosa	4.6	3.4	1.4	0.3
8	Iris-setosa	5.0	3.4	1.5	0.2

Restore Original Order



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

Unsupervised Learning | Analysis using Orange



= dog



= cat



= cat



= dog



= cat



= cat



= dog



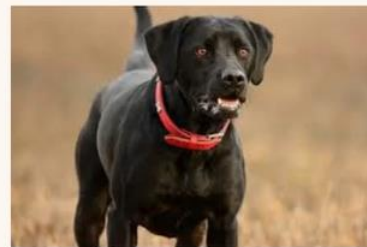
= dog



= cat

classification

TESTING



= cat or dog?



= cat or dog?



Watch later



Share

Supervised learning



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



YouTube





Unsupervised Learning



A



MORE VIDEOS

Play (k)



B



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 remain into a group and have 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

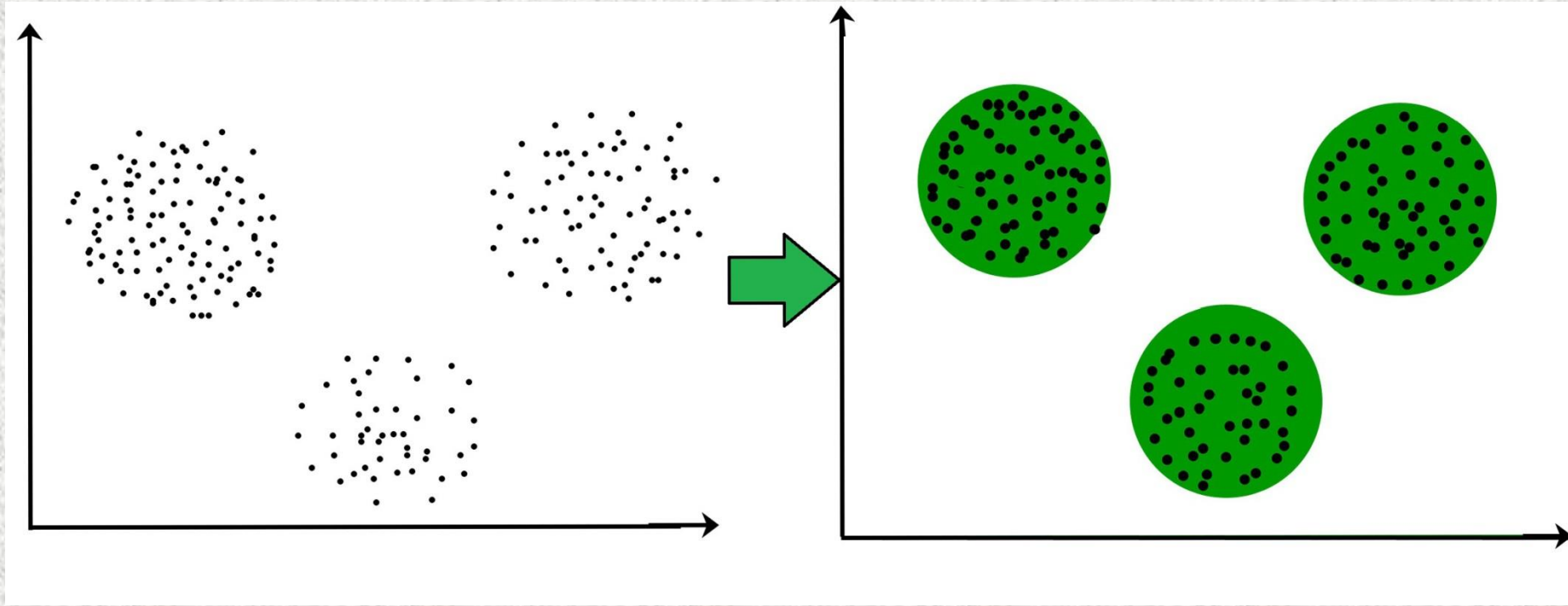
- Advantages
 - ✓ solve more complex tasks
 - ✓ it is easier to get unlabeled datasets
- Disadvantages
 - ✓ more difficult and time consuming
 - ✓ the result is not so accurate as for supervised learning

Cinderella fairytale: clustering of grains mixture

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

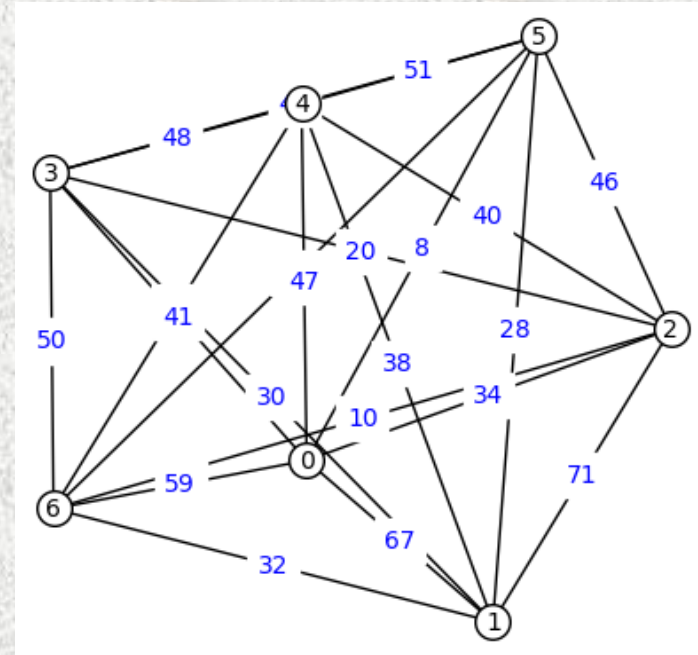
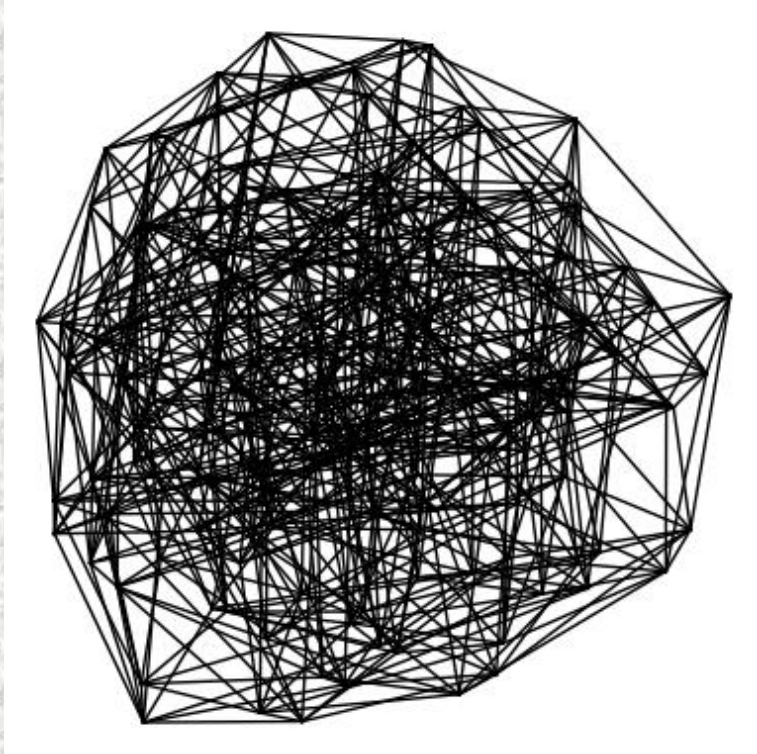


Clustering example



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 → 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

Names

Variable X:

Variable Y: ☒

Labels

☒ C1

+ -

Tools

☒ Brush ☒ Put ☐ Select

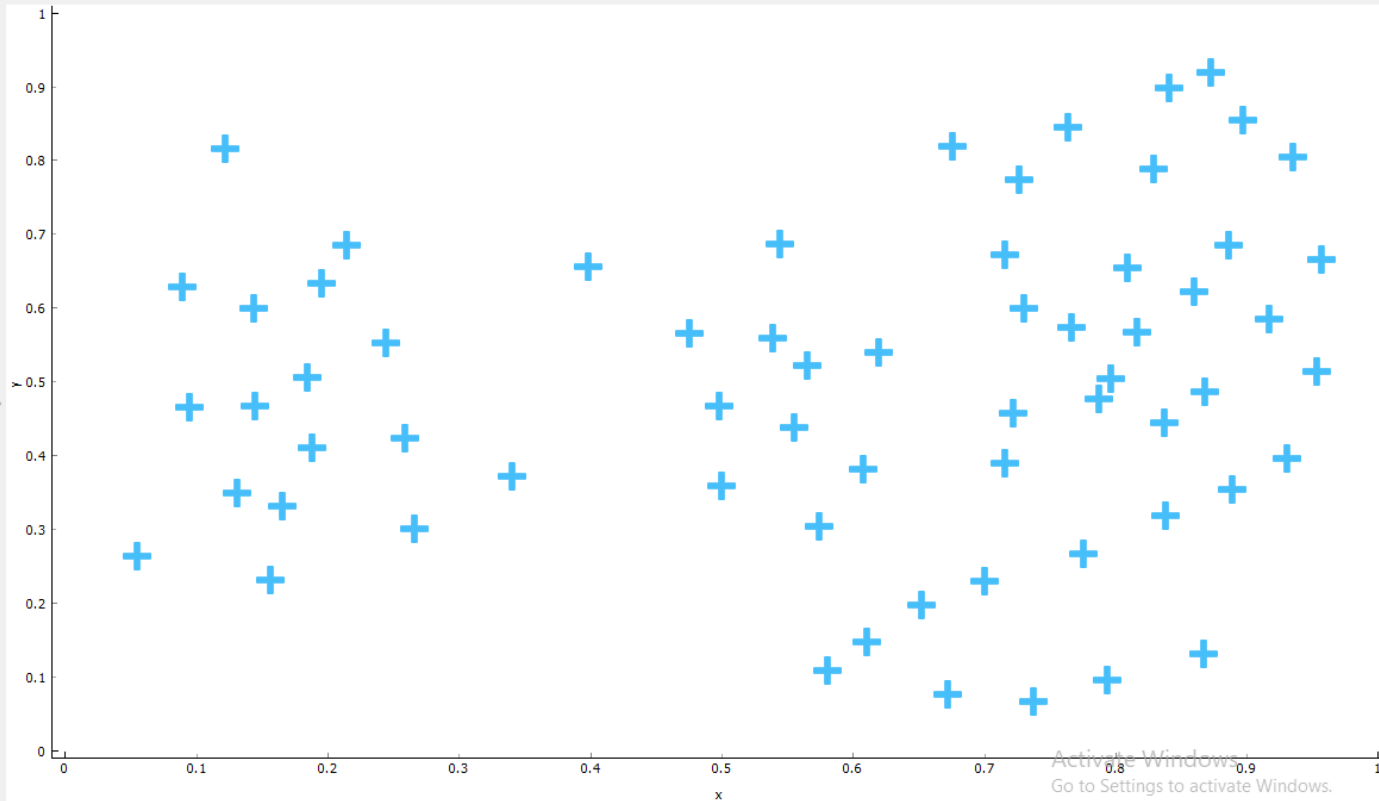
☒ Jitter ☒ Magnet ☐ Clear

Radius:

Intensity:

Symbol:

Reset to Input Data



Activate Windows
Go to Settings to activate Windows.



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

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

Scatter Plot Clustered - Orange

Axes

Axis x:

Axis y:

Find Informative Projections

Attributes

Color:

Shape:

Size:

Label:

☐ Label only selection and subset

Symbol size:

Opacity:

Jittering:

☐ Jitter numeric values

☐ Show color regions

☒ Show legend

☒ Show gridlines

☒ Show all data on mouse hover

☐ Show regression line

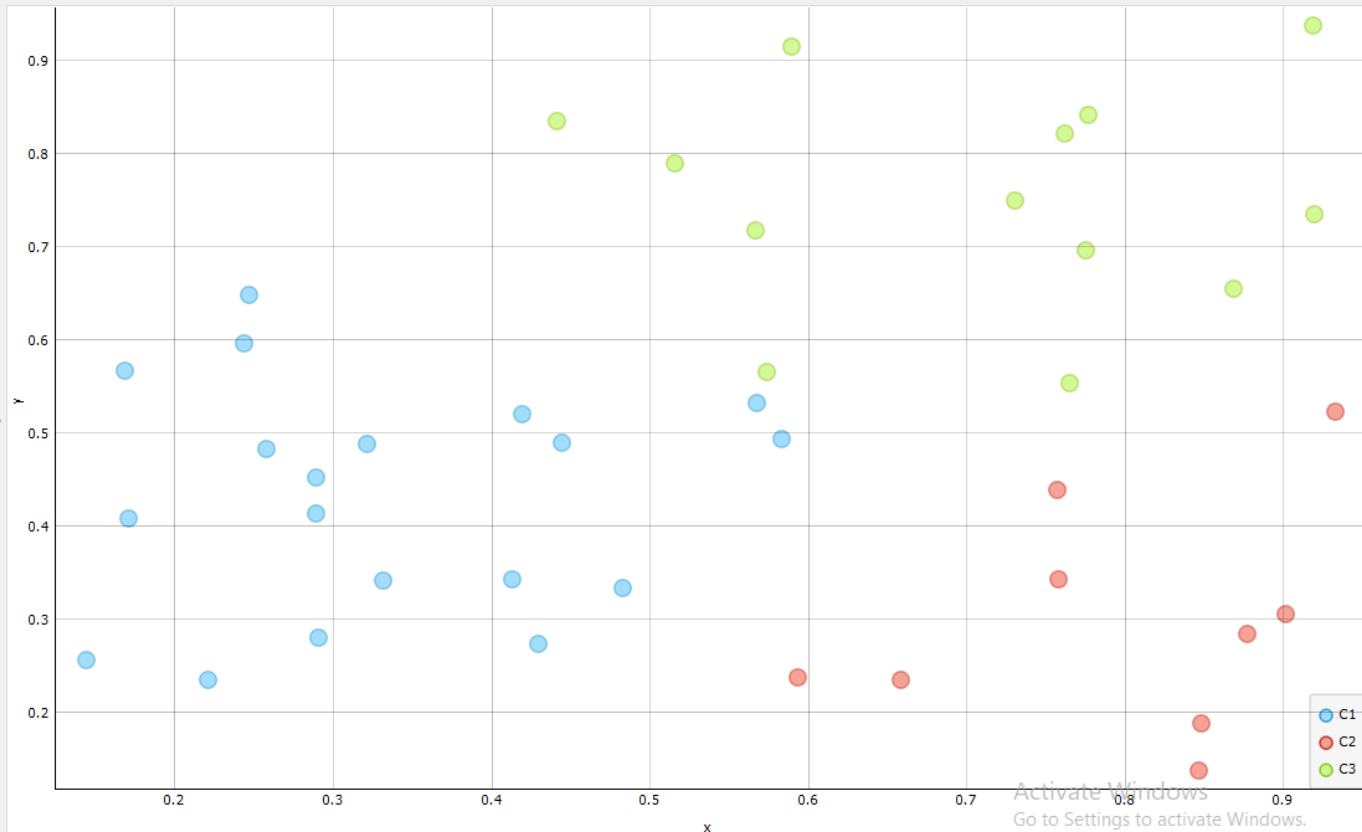
☐ Treat variables as independent

Zoom/Select

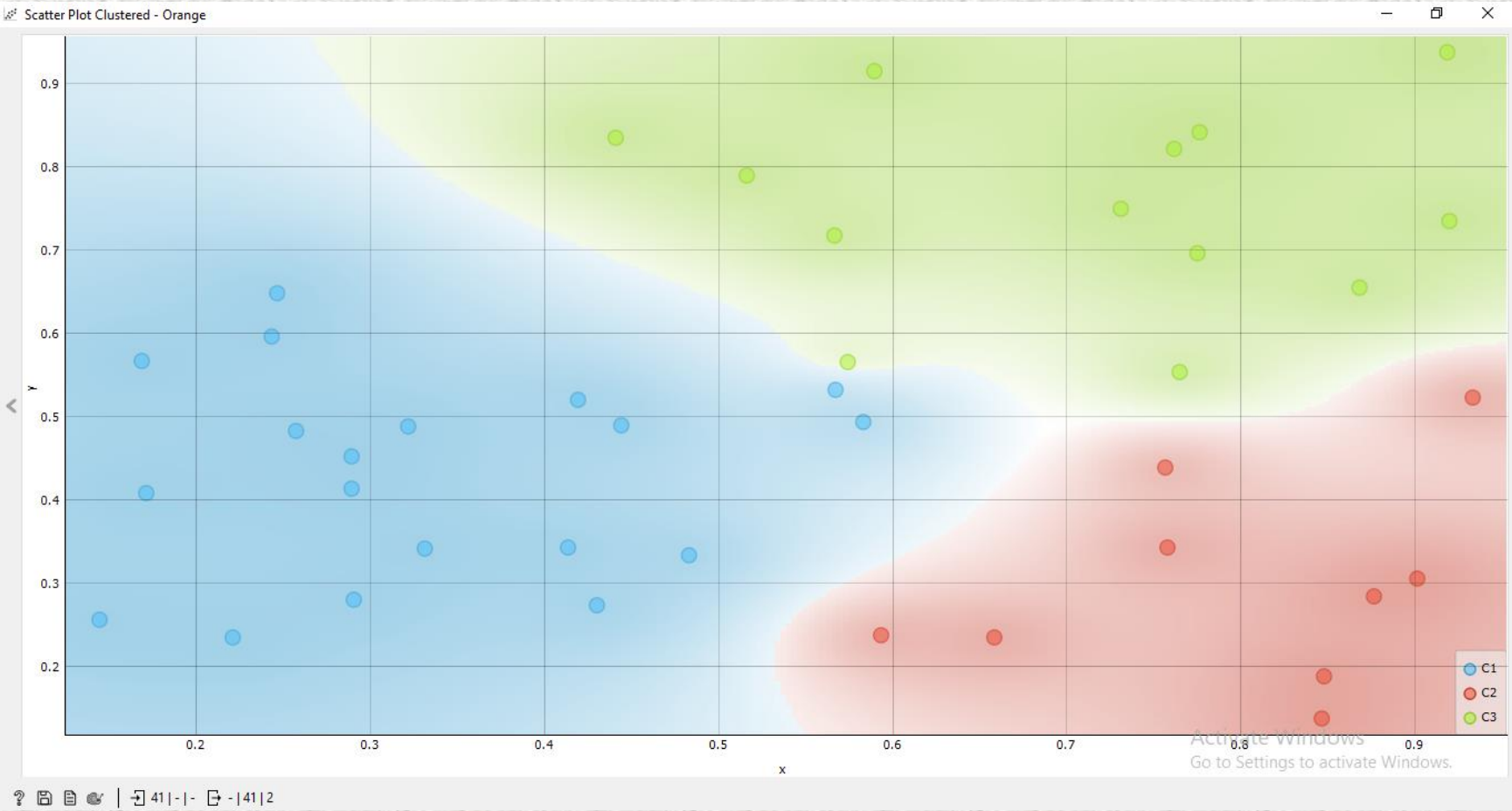


☒ Send Automatically

41 | - | - | 41 | 2



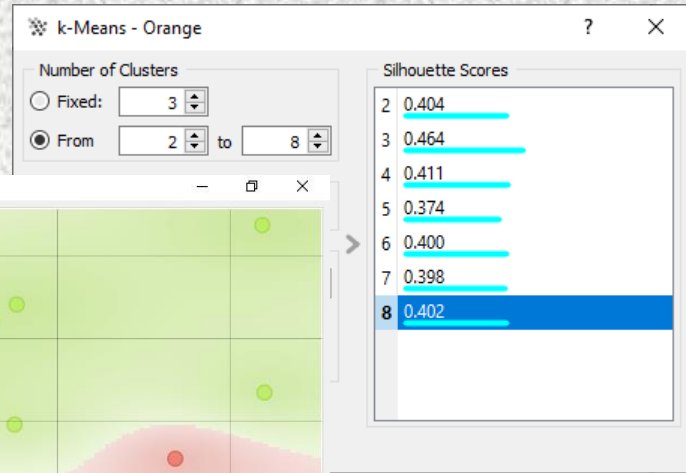
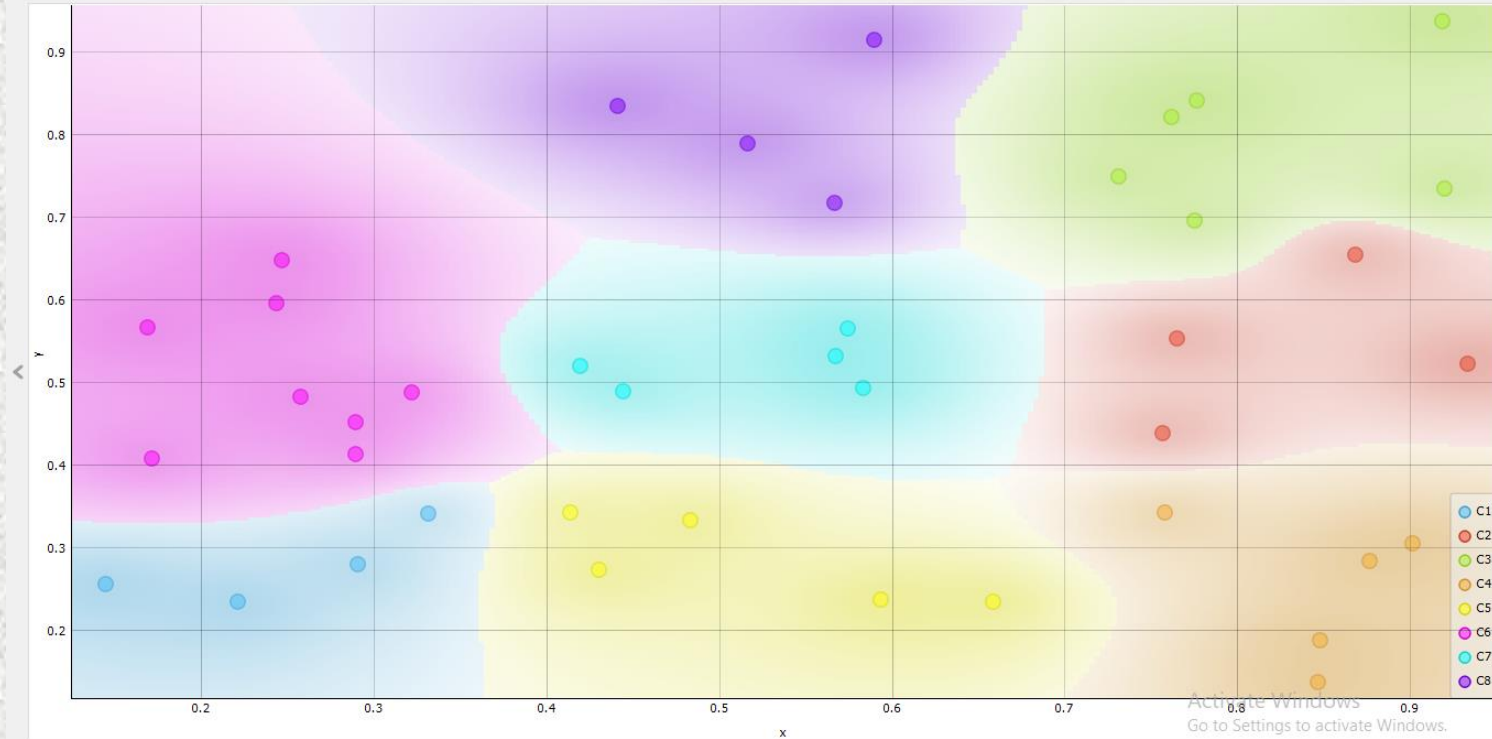
Clusters coloring

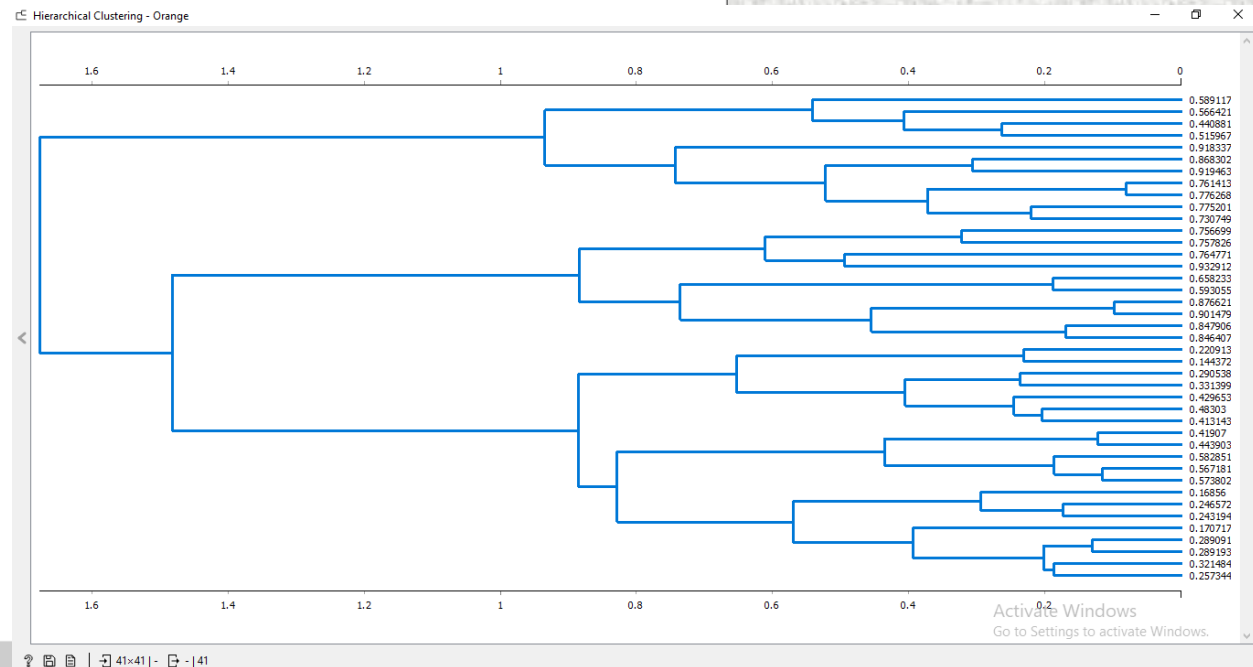
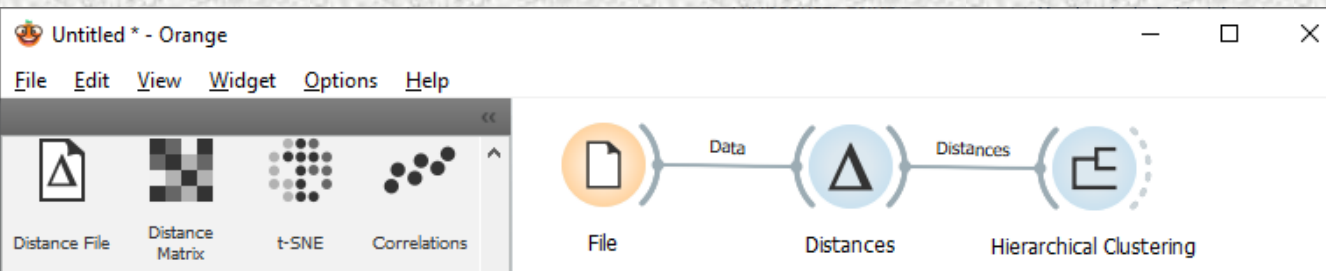


Color regions

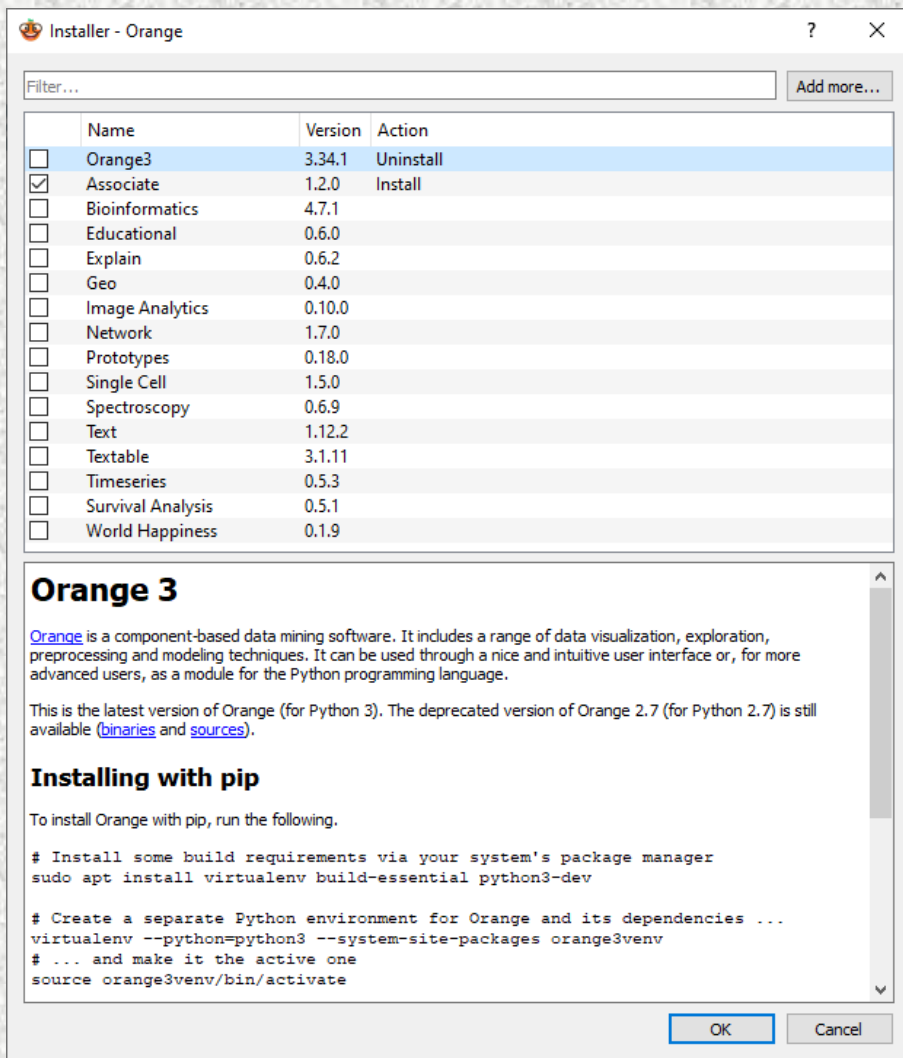
Adjust the number of clusters

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

Further training

- Recent courses of Dmitry Zaitsev
- <http://daze.ho.ua/rcz.htm>
- “Introduction to AI”
- “Python Bootcamp”
- AI powered self-driving vehicle JetBot – buy, assemble, train, and enjoy!

JetBot +

NVIDIA Jetson Nano

