

Uvod v strojno učenje

Jure Žabkar

jure.zabkar@fri.uni-lj.si

30.3.2021



Predavanja

Torek, 15:15 - 18:00

Marec

30

April

6

April

13

April

20

Maj

4

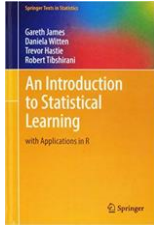
Maj

11

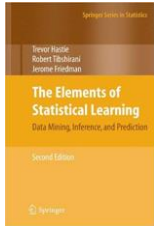
Maj

18

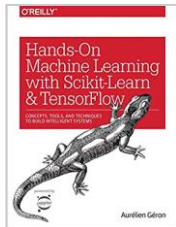
Literatura



James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). ***An introduction to statistical learning*** (Vol. 6). NY: Springer.



Friedman, J., Hastie, T., & Tibshirani, R. (2009). ***The Elements of Statistical Learning: Data Mining, Inference, and Prediction***. Springer Series in Statistics.



Geron, A. (2017). ***Hands-on machine learning with Scikit-Learn and TensorFlow***. O'Reilly.

Ocenjevanje

2 domači nalogi (2 x 15%)

- Kaggle playground (april, maj)
- Največ 4 dni zamude pri oddaji, vsak dan izgubite 15%
- Priznane točke = (ocenjene točke)* $0.85^{\text{dni_zamude}}$
- Primer: $90 * 0.85^4 = 46.98$

Projekt oz. izpitna naloga (70%):

- Tema po izbiri (npr. izziv na Kaggle, članek, snov izven učnega načrta)
- Ustni zagovori v okviru razpisanih izpitnih rokov (**prijava na izpit**):
 - 9.6.2021, 23.6.2021, 19.8.2021

Podatkovne zbirke

[UC Irvine Machine Learning Repository](#)

[Kaggle datasets](#)

[Amazon's AWS datasets](#)

<http://dataportals.org/>

<http://opendatamonitor.eu/>

<http://quandl.com/>

Strojno učenje

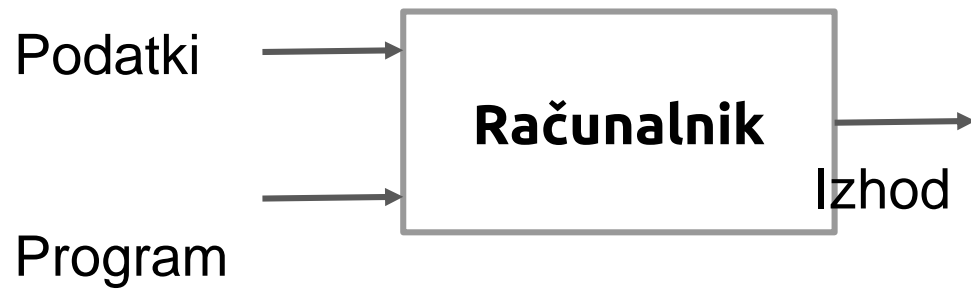
(angl. Machine Learning)

Kaj je strojno učenje?

**Umetna
inteligenca**

**Strojno
učenje**

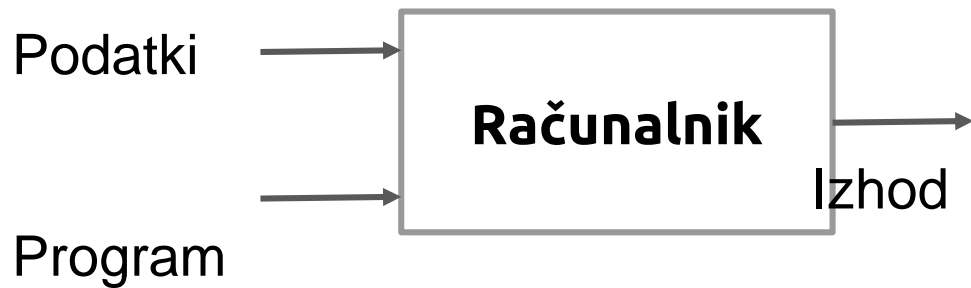
Klasično programiranje



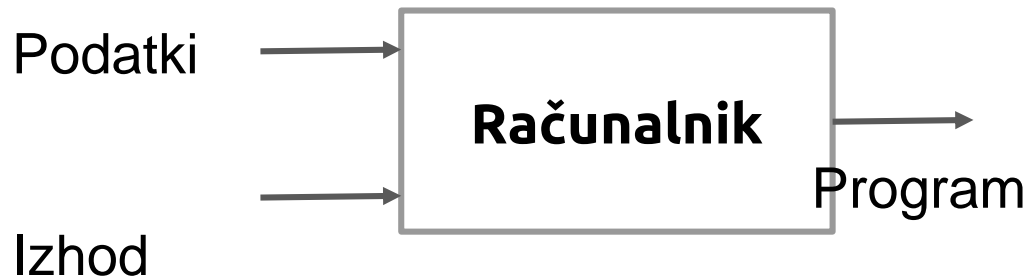
Cenitev nepremičnin **klasično**

$$\text{\$} = 1.2 \times \text{house icon} + 3.14 \times \text{map icon}$$

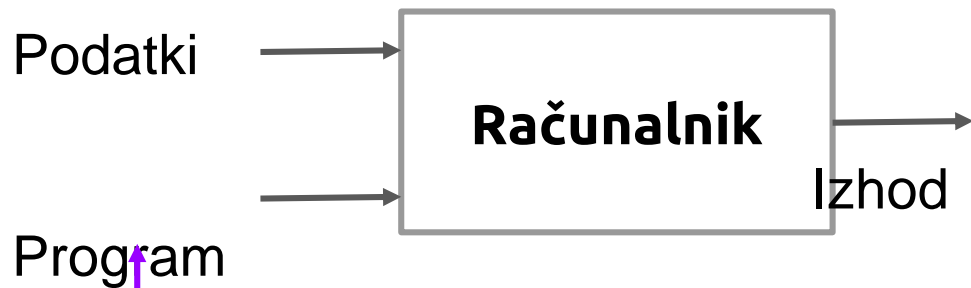
Klasično programiranje



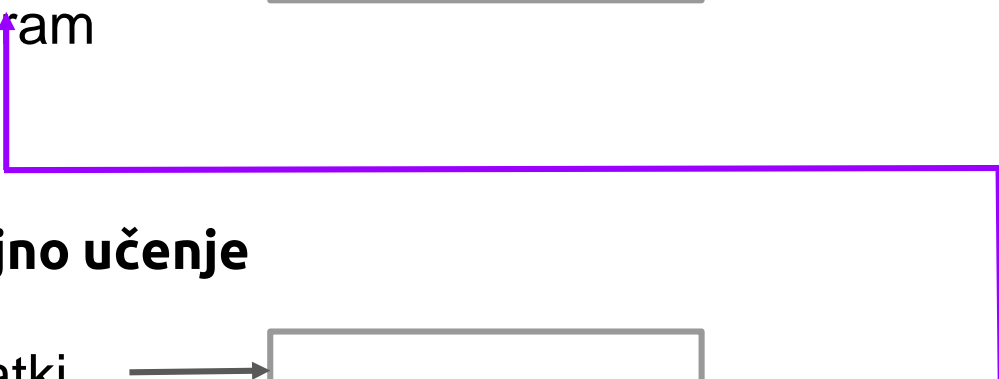
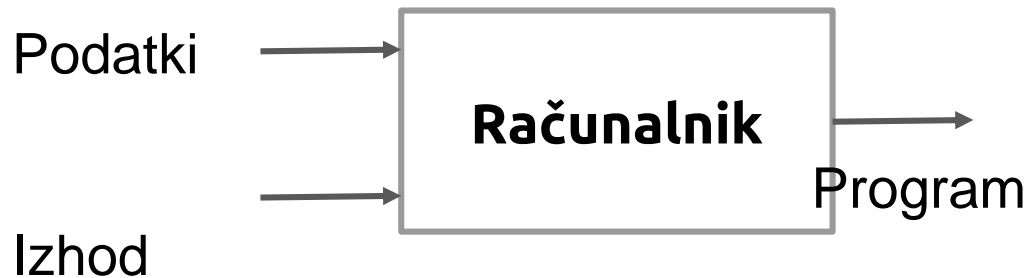
Strojno učenje



Klasično programiranje



Strojno učenje

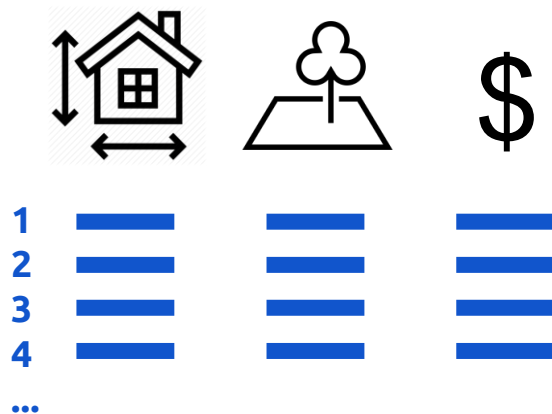


Cenitev nepremičnin s strojnim učenjem

$$\text{\$} = A \times \text{house_size} + B \times \text{land_area}$$

The equation represents a linear model for real estate pricing. The first term, $A \times \text{house_size}$, uses a house icon with vertical and horizontal double-headed arrows to represent the size of the building. The second term, $B \times \text{land_area}$, uses a tree icon on a plot of land to represent the area of the land. In both terms, A and B are orange, and the multiplication signs are blue.

Cenitev nepremičnin s **strojnim učenjem**



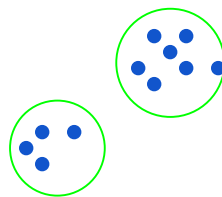
"Field of study that gives computers
the ability to learn without being
explicitly programmed"

- Arthur Samuel, 1959

Strojno učenje

**Nadzorovano
Spodbujevalno**

Nenadzorovano



učenje

Gručenje, povezov



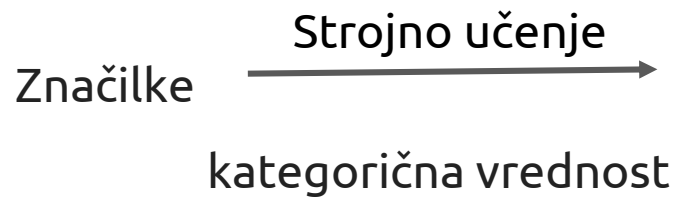
&



Regresija



Klasifikacija



Predstavitev značilk

Značilke nikoli popolnoma ne opišejo domene.

“All models are wrong, but some are useful.” - George Box

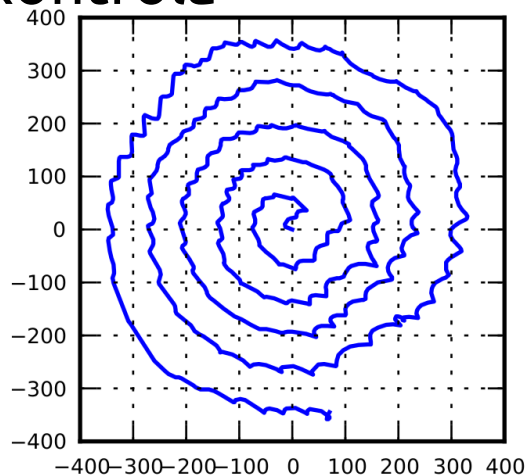
Predstavitev značilk

- Vektorji značilk predstavljajo učne primere.
- Na podlagi katerih značilk bi lahko napovedovali uspešnost študentov pri "Uvodu v strojno učenje" na začetku semestra?
- Npr. "povp. ocena že opravljenih izpitov", "znanje programiranja", ... se zdita smiselni značilki?

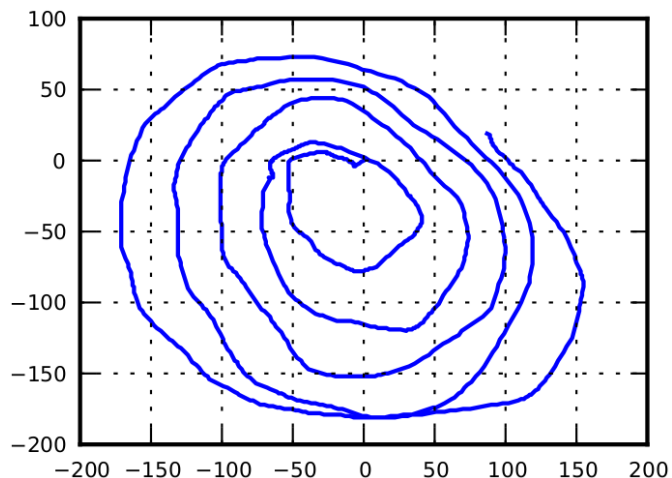
Predstavitev značilk

- Kaj pa "mesec rojstva" ali "barva oči"?
- Pretirano prileganje podatkom ali ...
- ... napačna interpretacija
- "some features matter, others don't":
zakaj ne vržemo kar vseh noter in vidimo, kaj pade ven?

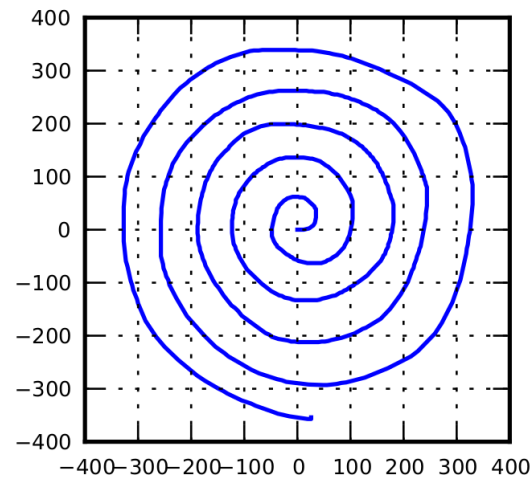
Esencialni tremor
kontrola

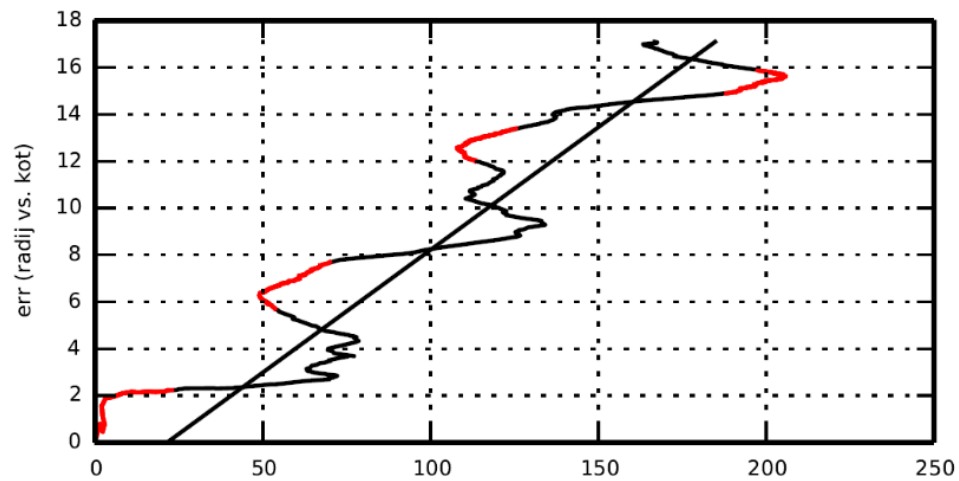
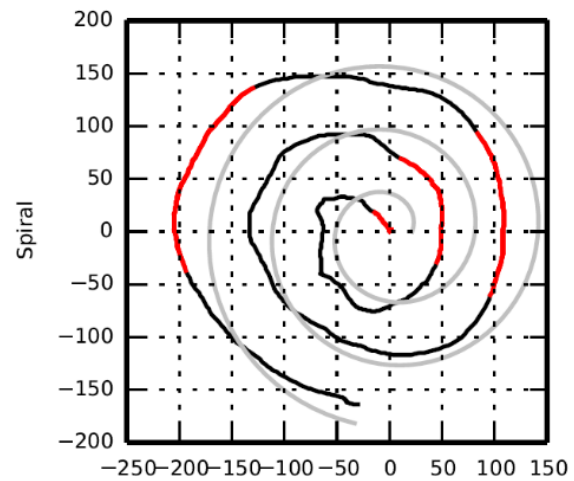


Parkinsonski tremor



Zdrava





Attribute	Importance	General description
radSp.avgP.min	1.13	radial speed variability
tangSp.avgP.min	1.02	tangential speed variability
absSp.avgP.min	0.78	absolute speed variability
plrErrComCnt.avg	0.70	level of curvature/smoothness of the spiral
radSp.percNeg005.min	0.69	percentage of time the patient drew towards the centre
plrErrComCnt.max	0.67	level of curvature/smoothness of the spiral
plrErrFit.avg	0.65	general misfit from the ideal spiral (template)
tangSp.avgP.rng	0.65	tangential speed variability
tangSp.avgP.max	0.63	tangential speed variability
rot.avgP.min	0.62	number of times the spiral crosses itself

Problemi

- kako preprečiti pretirano prilagajanje?
- kako ocenjevati model?
- kako izbrati najboljše značilke?



ParkinsonCheck™

- A smartphone app for (early) detection of motoric signs of Parkinson's disease and some other tremors
- Freely available in Slovenia
- A built-in expert system enables users to use it in their home environment
- Fully standalone, no need to communicate with an outside server or sensor
- Based on spirometry, but enhanced with other sensors, e.g. accelerometry



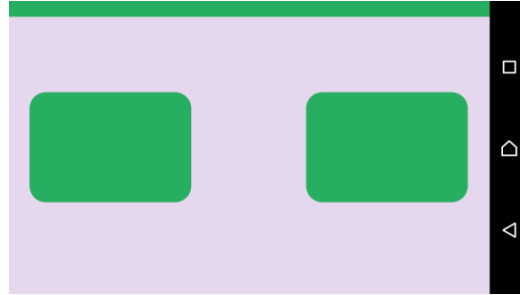
<http://www.parkinsoncheck.net/>

Early detection, Patient monitoring



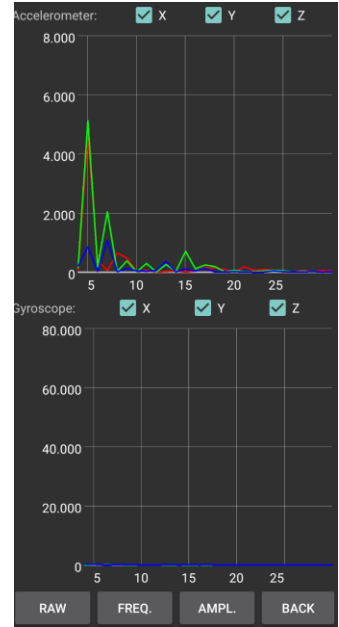
Digitizing UPDRS

Tapping test



Accelerometry


Finger tapping



QUIERO

Dislex.AI, Screening for Dyslexia



Projekt ŠIPK



Aplikacija za odkrivanje disleksije
s pomočjo naprave za sledenju
pogledu

STAROST

SPOL

  ☐ Drugo

ODLOČBA

☒ ☐

Začni ?

Iskani študenti, razpisni, razpisni in priložnosti
Iskani študenti, razpisni, razpisni in priložnosti

University of Ljubljana
Faculty of Computer and
Information Science

EVROPSKA UNIJA
evropski
skladniški sklad
NALOŽBA V VAŠO PRIHODNOST

Screening for Dyslexia

AIM:

To develop a set of digital tests for assessing dyslexia in primary school children.

Medvedek Medo se je nekega jutra prebudil in zaupano vstal iz postelje. Najprej si je umil zobe, nato obraz, šele potem pa se je preoblekel iz pižame v svoja dnevna oblačila. Odklenil je vrata in odšel iz svojega stanovanja.

Dyslectic dyslectic

Medvedek Medo je sedel v svoj bleščeče rdeč
avtomobil in se odpeljal skozi velik zelen park do
mesta, ki je bilo sivo in je imelo visoke stolpnice.
Skozi mesto se je pripeljal do dolge podezelske poti
in nadaljeval vožnjo vse do majhnega jezera, ki je
bilo obraščeno z visokimi smrekami.

vs.

non-

Medvedek Medo je sedel v svoj bleščeče rdeč
avtomobil in se odpeljal skozi velik zelen park do
mesta, ki je bilo sivo in je imelo visoke stolpnice.
Skozi mesto se je pripeljal do dolge podezelske poti
in nadaljeval vožnjo vse do majhnega jezera, ki je
bilo obraščeno z visokimi smrekami.

Priporočilni sistemi



Prilagajanje uporabniku



Prepoznavanje obraza



Prepoznavanje govora



amazon alexa

Robotika



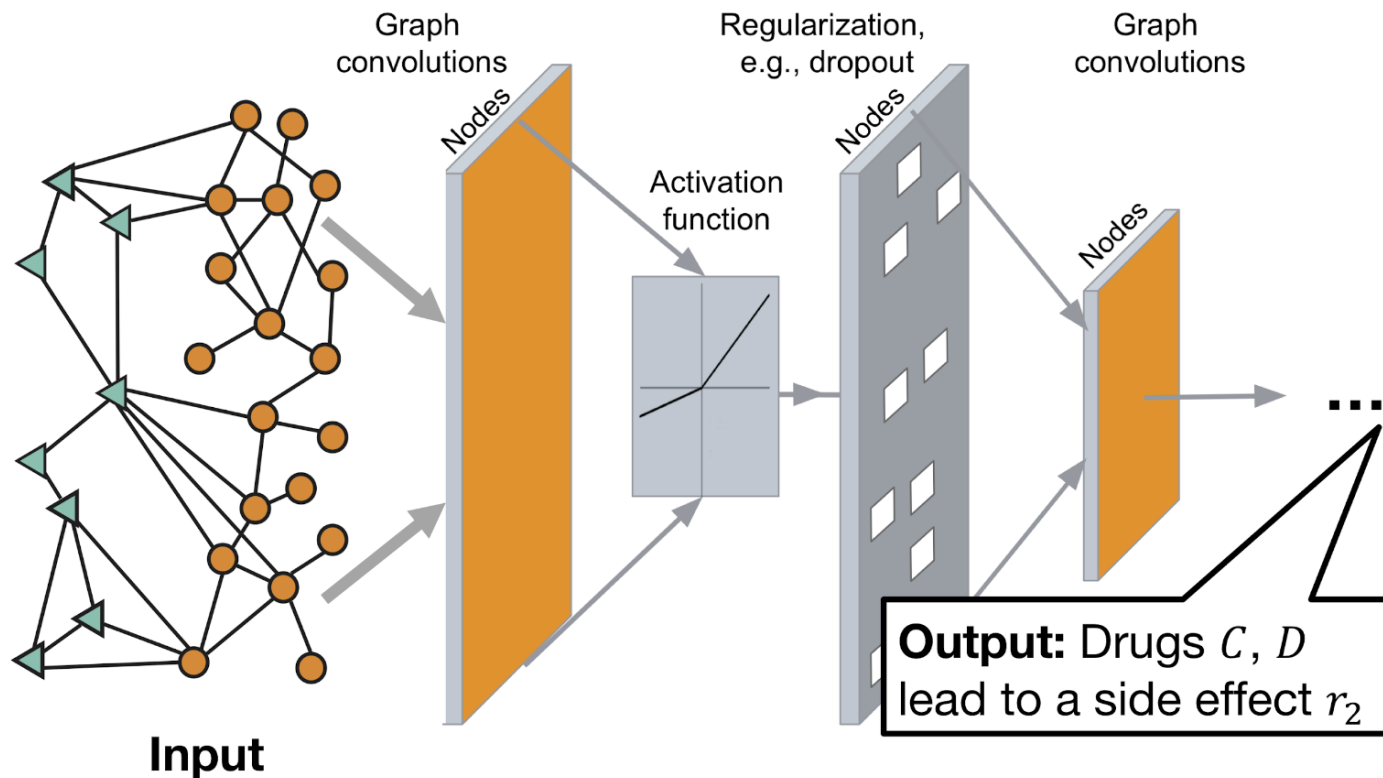


BostonDynamics

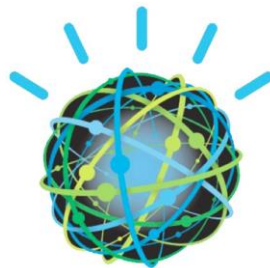




Odkrivanje zdravil



Igre, delnice



IBM WATSON



Orodja

Orange



Scikit-learn



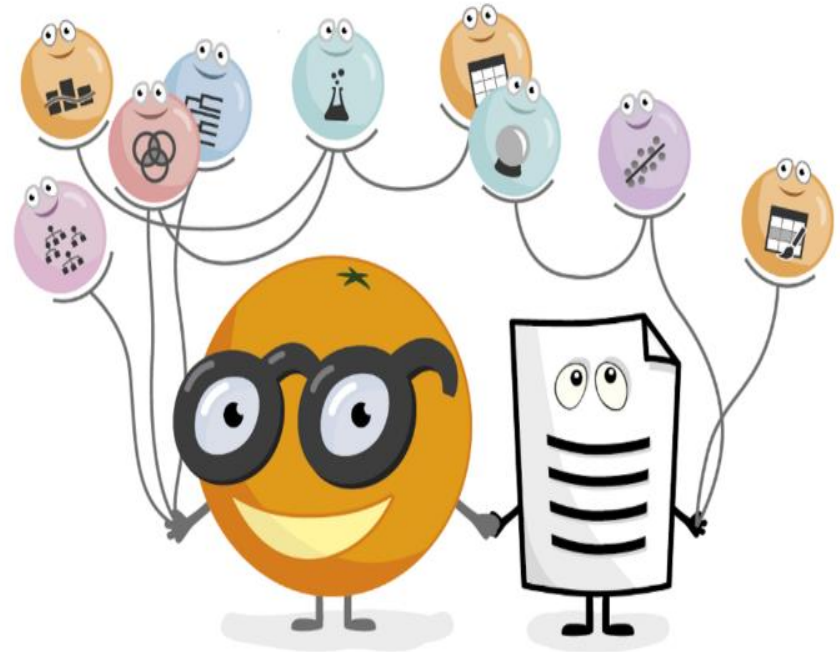
[Screenshots](#)[Download](#)[Docs](#)[Blog](#)

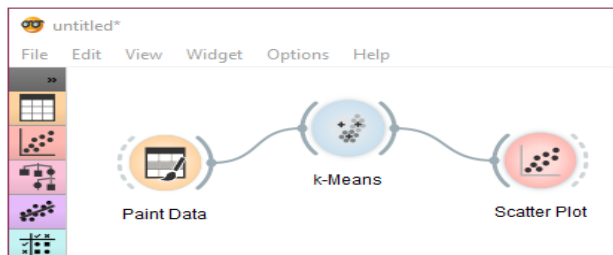
Data Mining Fruitful and Fun

Open source machine learning and data visualization for novice and expert. Interactive data analysis workflows with a large toolbox.

[Download Orange](#)

The **old version**, Orange 2.7, is still available.





Scatter Plot

Axis Data

Axis x:

Axis y:

Jittering:

☐ Jitter continuous values

Points

Color:

Label:



Paint Data

Names

Variable X:

Variable Y:

Labels

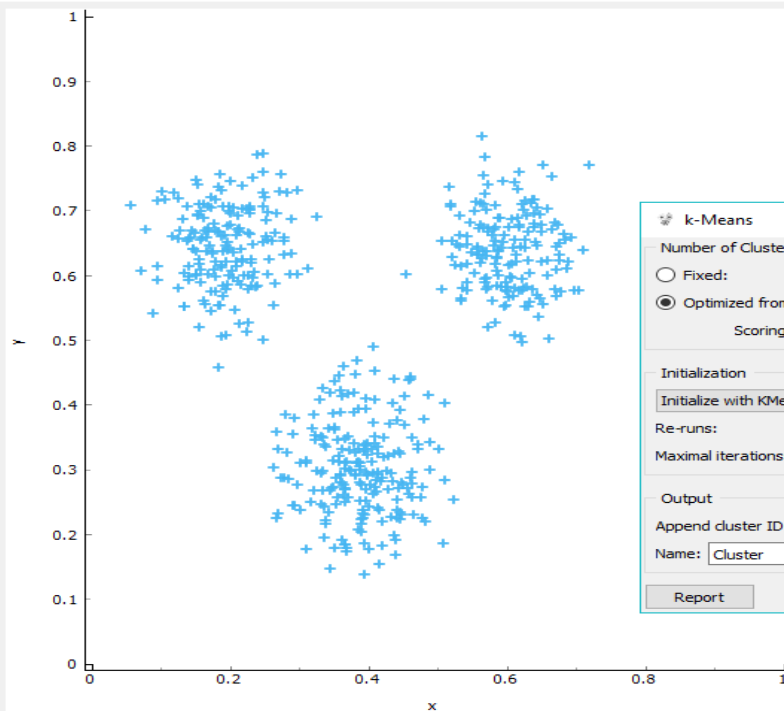
☒ C1

Tools

Radius:

Intensity:

☒



k-Means

Number of Clusters

☐ Fixed:

☒ Optimized from to

Scoring:

Initialization

Re-runs:

Maximal iterations:

Output

Append cluster ID as:

Name:

☒

Scoring (bigger is better)

k	Score
2	0.49
3	0.72
4	0.57
5	0.46
6	0.35
7	0.34
8	0.34