



Library Academy

Innovate your education

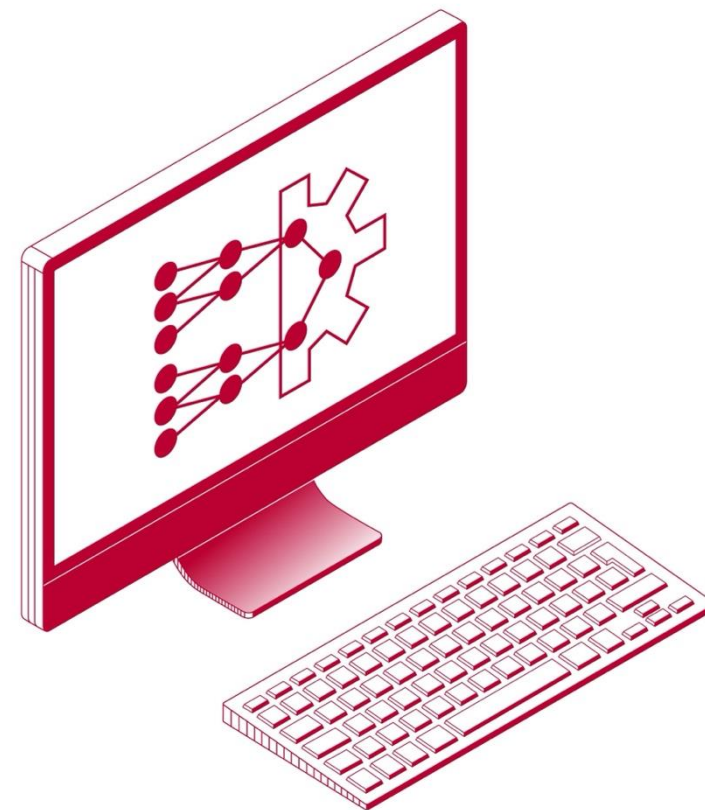
21-01-2026





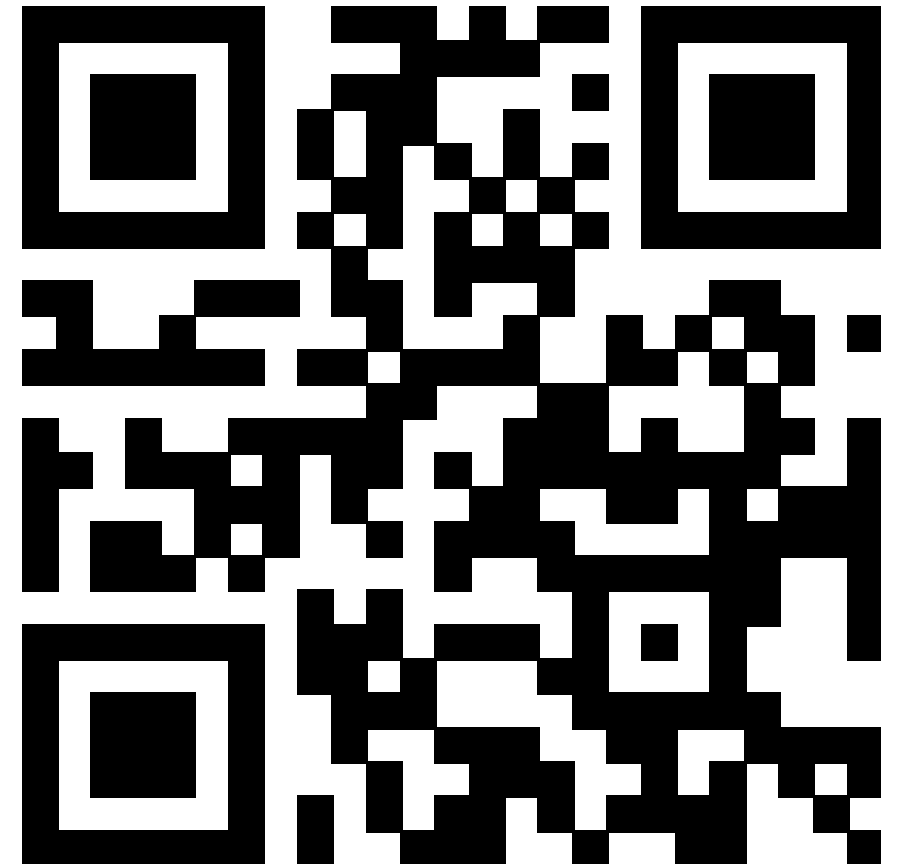
Ontdek wat data vertellen: Aan de slag met Machine Learning in JASP

Don van den Bergh & Johnny van Doorn
Library Academy 2026



Outline

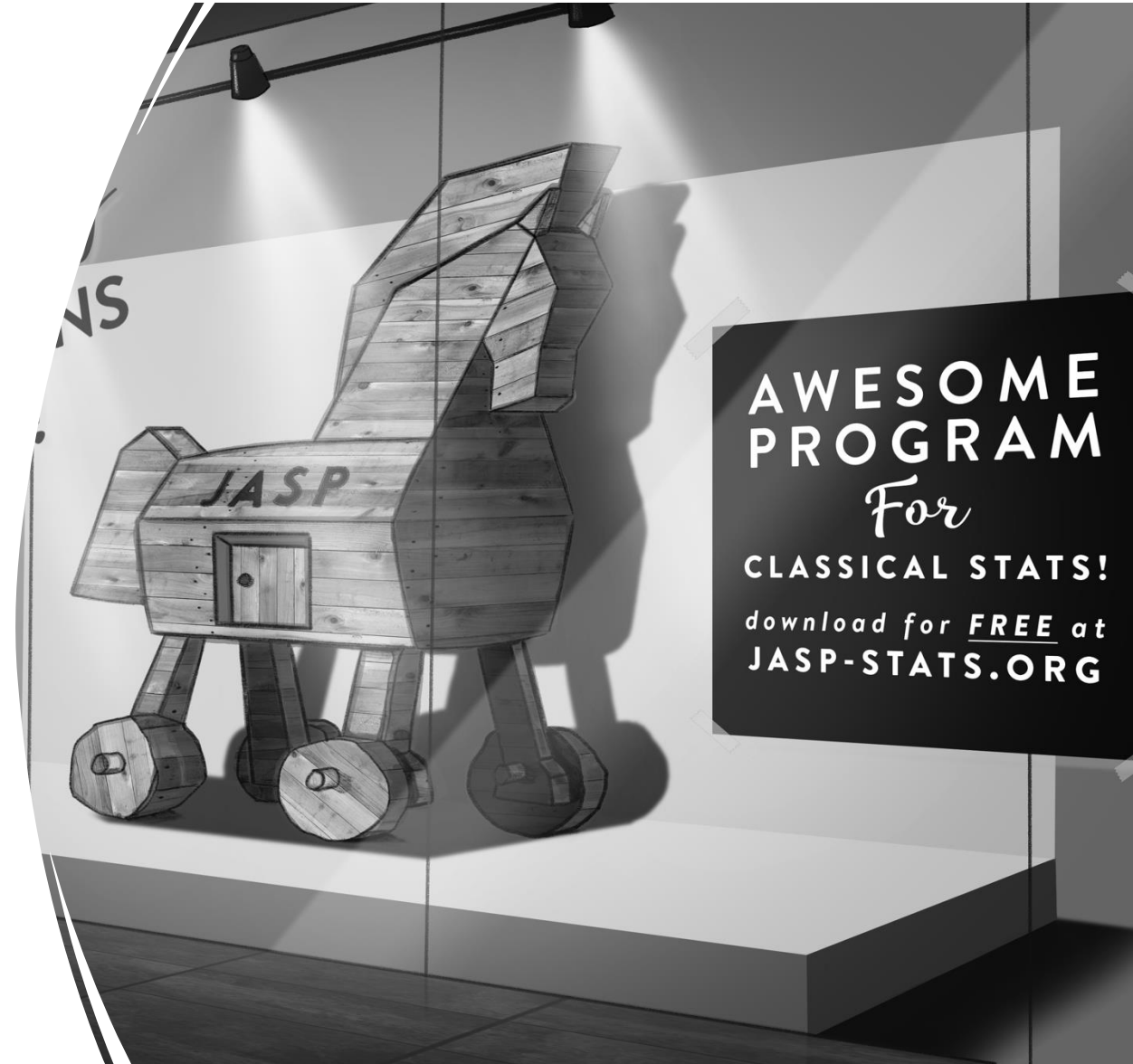
- Basics
 - What is JASP?
 - Regression
- Machine Learning
 - General philosophy
 - K-nearest neighbors
- Exercises



edu.nl/aa8ch

What is JASP?

- Developed at UvA over the past 10 years, funded by research grants (NWO/EU)
- Graphical user interface for conducting frequentist and Bayesian statistics
- <https://jasp-stats.org/>
- Open-source → forever free!



What is JASP?

The screenshot displays the JASP software interface. The top toolbar includes icons for Edit Data, Descriptives, T-Tests, ANOVA, Mixed Models, Regression, Frequencies, and Factor. The main window is titled 'ANOVA' and shows the following configuration:

- Dependent Variable:** Score
- Fixed Factors:** Group
- Display:**
 - ☐ Descriptive statistics
 - ☒ Estimates of effect size
 - ☒ ω^2 ☐ partial ω^2
 - ☐ η^2 ☐ partial η^2
 - ☒ Confidence intervals 95 %
 - ☐ Vovk-Sellke maximum p-ratio
- Model:** (empty)
- Assumption Checks:** (empty)
- Contrasts:** (empty)
- Order Restricted Hypotheses:** (empty)
- Post Hoc Tests:** (empty)

The right panel shows the ANOVA results:

ANOVA - Score

Cases	Sum of Squares	df	Mean Square	F	p	ω^2	95% CI for ω^2	
							Lower	Upper
Group	24.420	3	8.140	2.894	.036	0.025	0.000	0.069
Residuals	613.139	218	2.813					

Note. Type III Sum of Squares

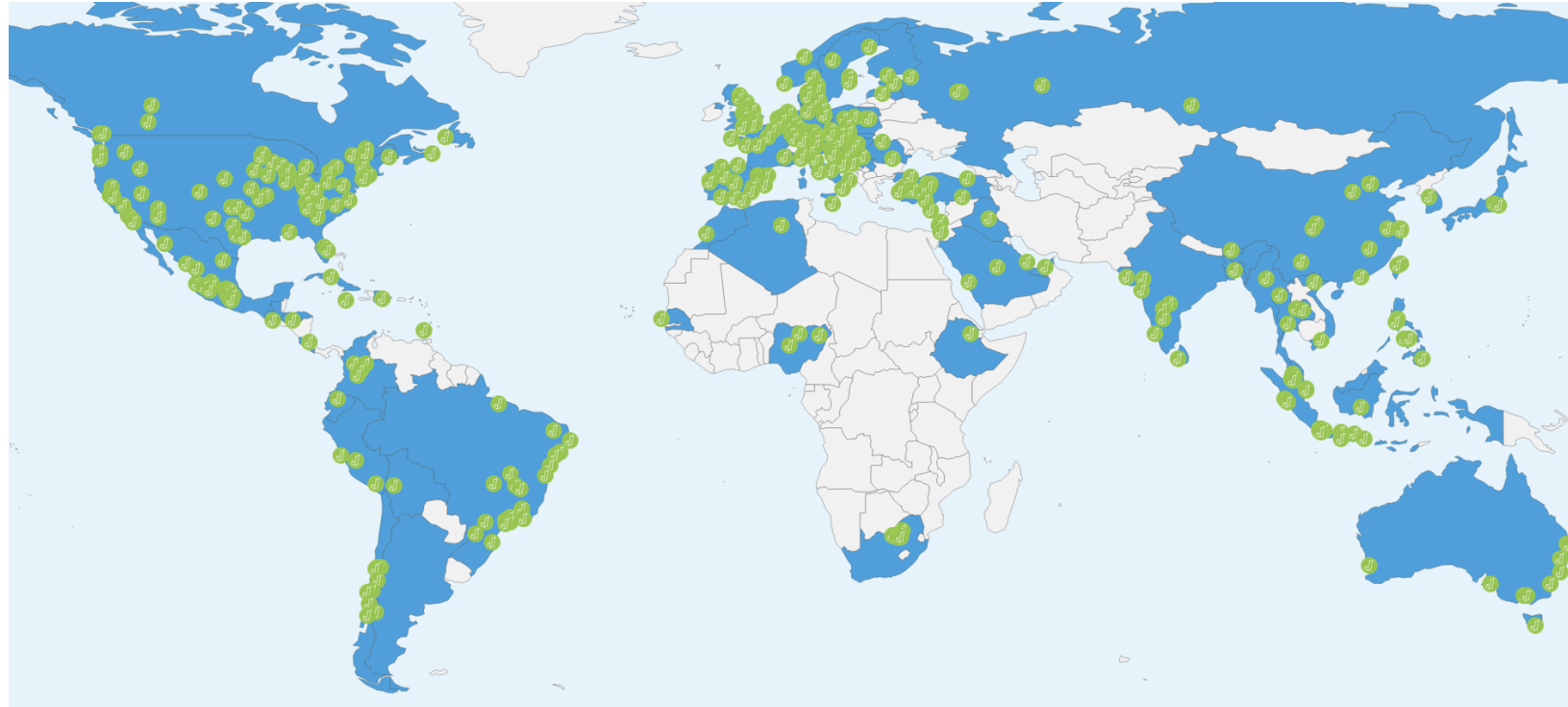
Descriptives

Descriptives plots

Group	Mean Score	Lower CI	Upper CI
Down	3.1	2.6	3.6
Blue	3.2	2.8	3.6
Brown	3.7	3.2	4.2
Green	3.8	3.5	4.1

What is JASP?

- Used at 374 universities across 76 countries
- 100,000 monthly downloads



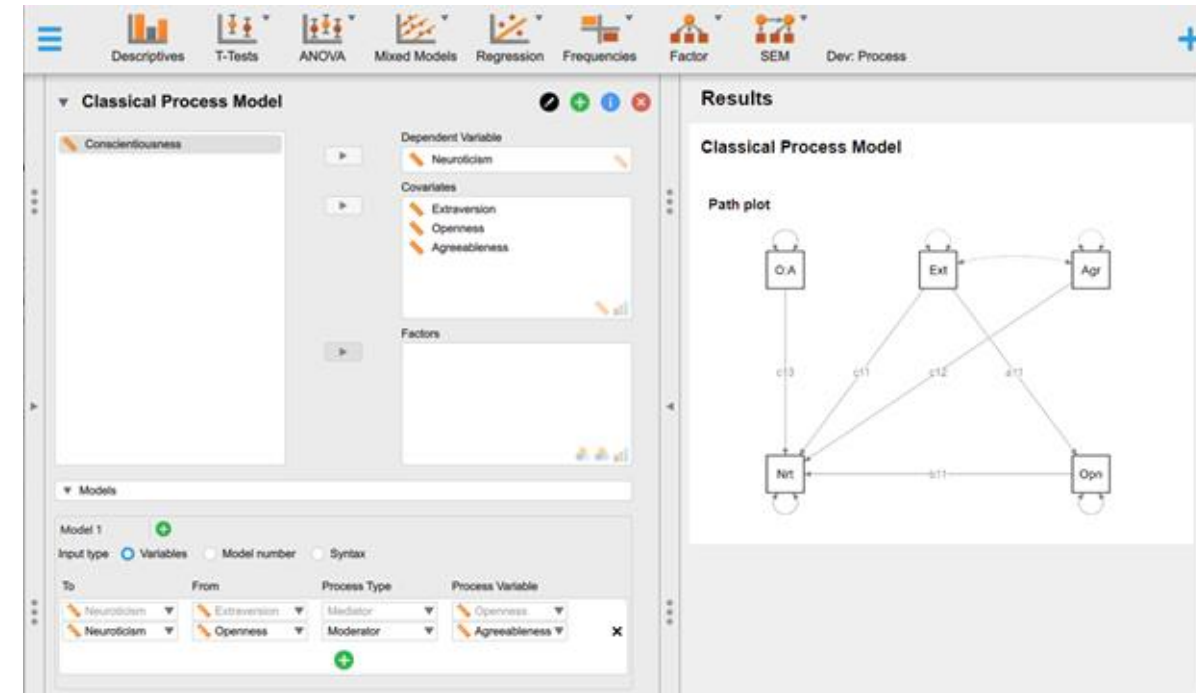
<https://jasp-stats.org/teaching-with-jasp/#worldmap>

Features

- [Website overview](#)
- [JASP vs. SPSS feature comparison](#)
- Data formats: .sav, .xls, .txt, .csv, .ods, .tsv, .dta, .por, .sas7bdat, .sas7bcat, and the .jasp format
- APA tables
- OSF integration
- R console
- Compute columns
- Filtering

Feature Roadmap

- Full syntax mode ([blog about the first implementation](#))
- Better data manipulation
- Select filters



Other Handy Resources

- [The JASP Video Library](#)
- [How to Use JASP – Inventory of blogs/videos/gifs for frequentist and Bayesian analyses](#)
- [JASP YouTube page](#)
- [Step By Step Guide: 1. Bayesian One-Way ANOVA](#) and the [full playlist](#)
- JASP on Bluesky - <https://bsky.app/profile/jaspstats.bsky.social>
- JASP forum - <https://forum.cogsci.nl/index.php?p=/categories/jasp-bayesfactor>
- Found a bug? Please report on Github: <https://github.com/jasp-stats/jasp-issues/issues>
- [JASP Verification Project](#)
- More JASP workshops: <https://jasp-stats.org/workshop/>

JASP Literature

- [The JASP Data Library](#)
- [Discovering Statistics Using JASP](#)
- [Learning Statistics with JASP: A Tutorial for Psychology Students and Other Beginners by Danielle J. Navarro, David R. Foxcroft, and Thomas J. Faulkenberry](#)
- [Statistics of Doom by Erin Buchanan](#)
- [Statistical Analysis in JASP. A Guide for Students by Mark Goss-Sampson](#)
- [Quantitative Analysis with JASP open-source software by Chris Halter](#) (amazon)




Data Management

<






The Variable View

Name: Long name:




Column type:  Nominal ▼ Description:

Computed type: Not computed ▼

Label editor Missing values


	Filter	Value	Label
	✓	Lars Ulrich	Lars Ulrich
	✓	James Hetfield	James Hetfield
	✓	Kirk Hammett	Kirk Hammett
	✓	Rob Trujillo	Rob Trujillo
	✓	Jason Newsted	Jason Newsted

Variable Types

- Scale 
 - Numbers (e.g., 7, 0, 120, 8.5)
- Nominal 
 - Categories (e.g., 'Control group', 'Experimental group')
- Ordinal 
 - Ordered values (e.g., 'Dislike', 'Neutral', 'Like')

Variable Settings

Name:

Column type:  Ordinal ▼

Computed type: ▼

Label editor Missing values

1
N ↓

1
N ↺

↑
↓

▲

▼

Filter	Value	Label
✓	1	Light
✓	2	Moderate
✓	3	Heavy

Computing a New Variable

Name: Long name:

Column type: Description:

Computed type:

Computed column definition Label editor Missing values

$+ - * \div / ^ \sqrt \% = \neq < \leq > \geq \wedge \vee | \neg$

☐ Name
☐ Instrument
☐ Current member
☐ Headbangi... ntensity

☐ Net worth... million)
☐ Songs

Computed columns code applied

☐ Converting types Compute column

$|y|$
 σ_y
 σ^2_y
 Σy

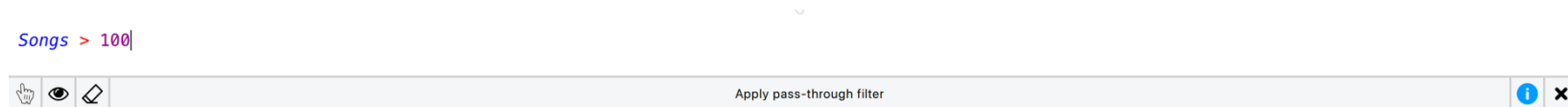


Filtering Data

- Using Variable Settings
- Using the Filter functionality
 - Drag and drop






- R-mode



Descriptives – output window

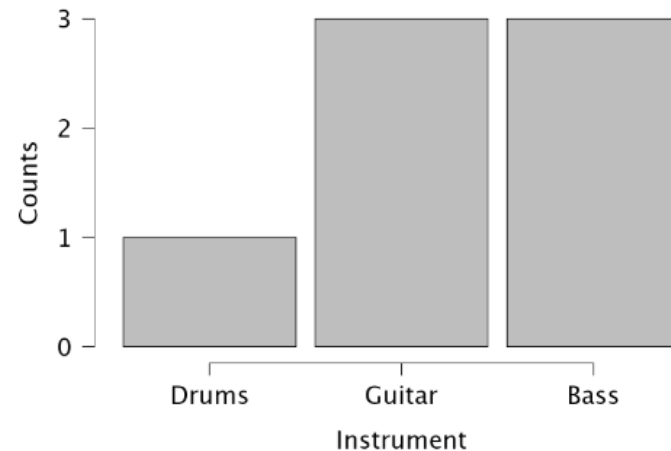
Distribution Plots

B I U f_x \langle / \rangle   Normal x_2 x^2 '' '' '' Normal \mathcal{I}_x

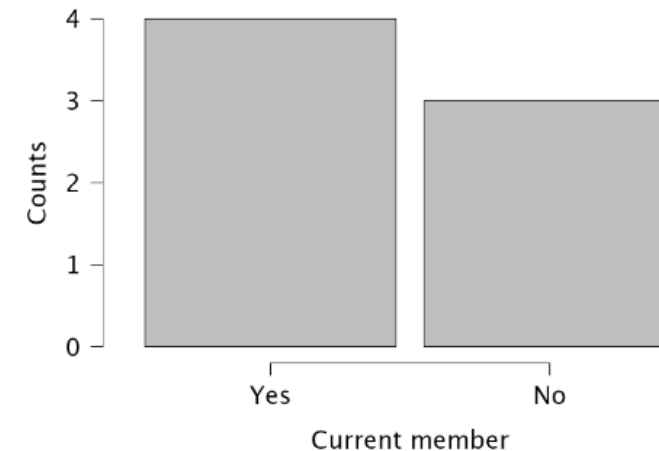
We can have fancy $LaTeX$ formula's in here, [link cute cat video's](#), or insert a drumkit 

Below are two distribution plots outlining the members of Metallica. On the left, we see the various instruments being played and their frequencies, and on the right we see how many members are still active, and how many left the band.

Instrument

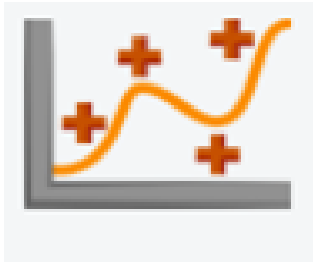


Current member





There are lots of different machine learning algorithms



Regression

- Boosting
- Decision tree
- K-nearest neighbors
- Linear
- Neural network
- Random forest
- Regularized linear
- Support vector machine



Classification

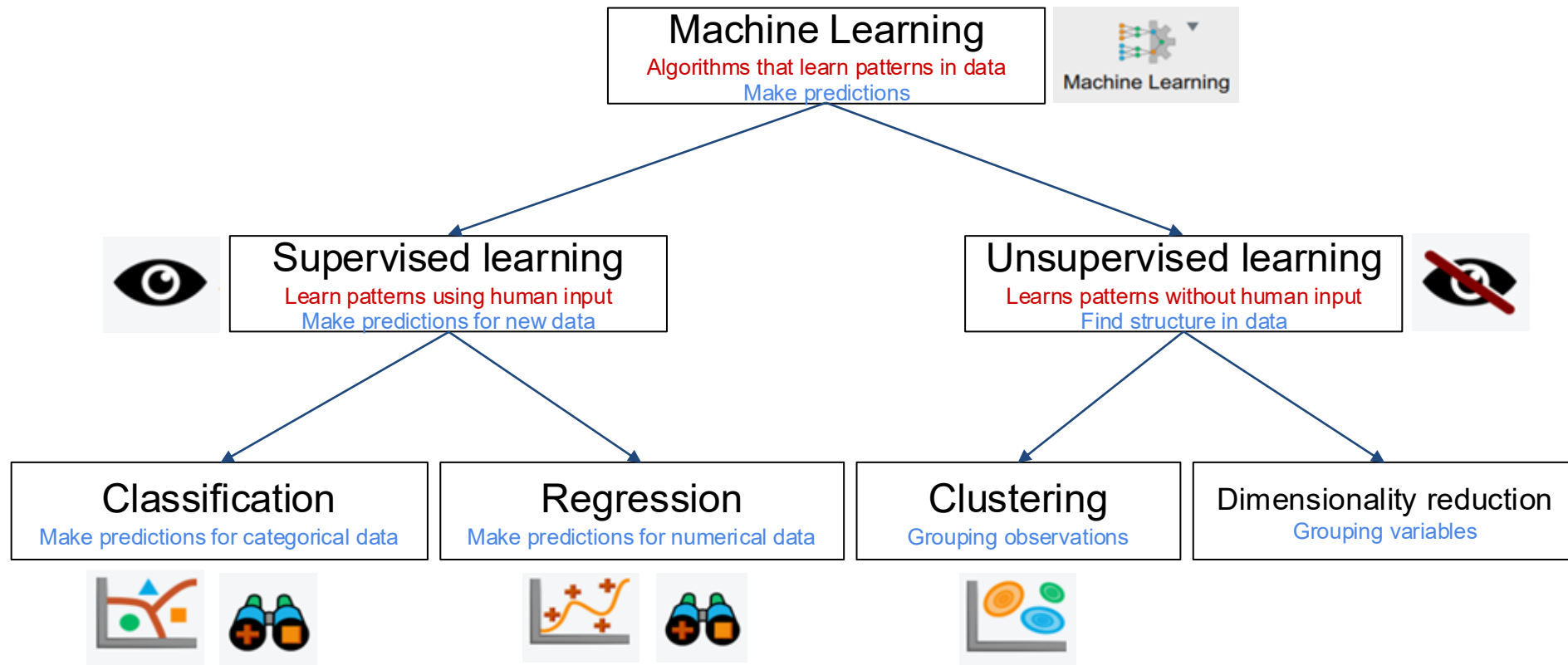
- Boosting
- Decision tree
- K-nearest neighbors
- Linear discriminant
- Logistic/Multinomial
- Naive Bayes
- Neural network
- Random forest
- Support vector machine



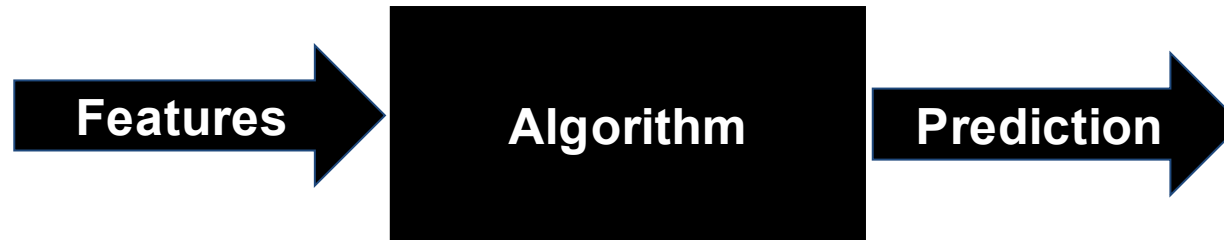
Clustering

- Density-based
- Fuzzy c-means
- Hierarchical
- Model-based
- Neighborhood-based
- Random forest

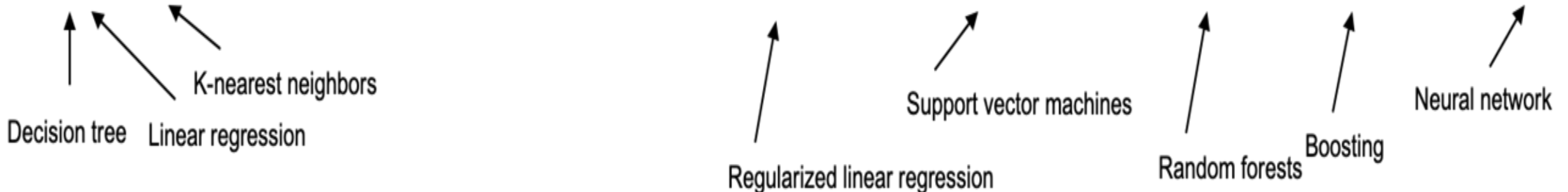
Machine learning is about finding patterns, the goal is making predictions



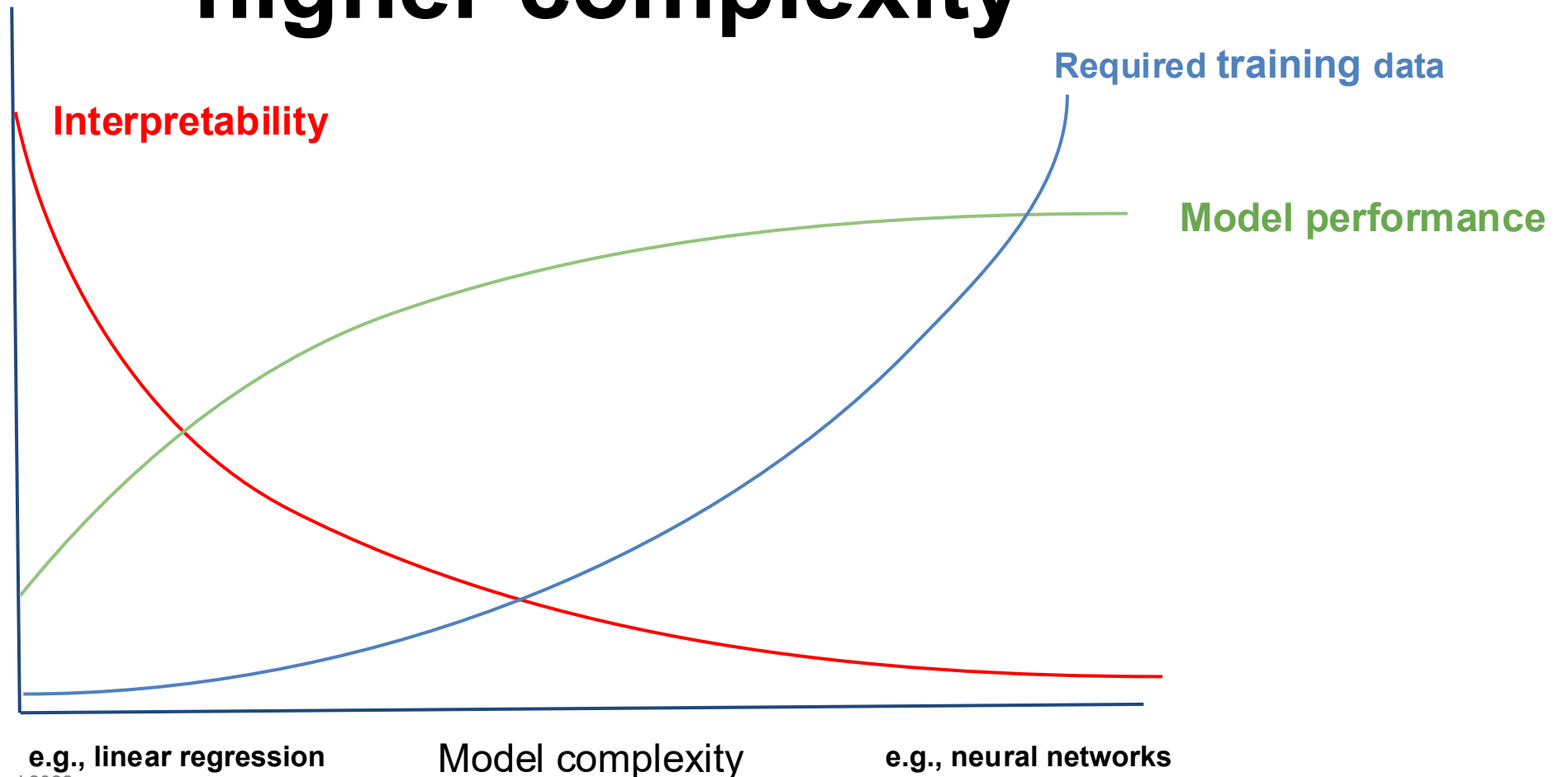
- Some algorithms are 'black boxes'



- Understanding all algorithms is not the most important, understanding how to evaluate their results is

Explainable**Not explainable**

Better performance (often) comes with higher complexity



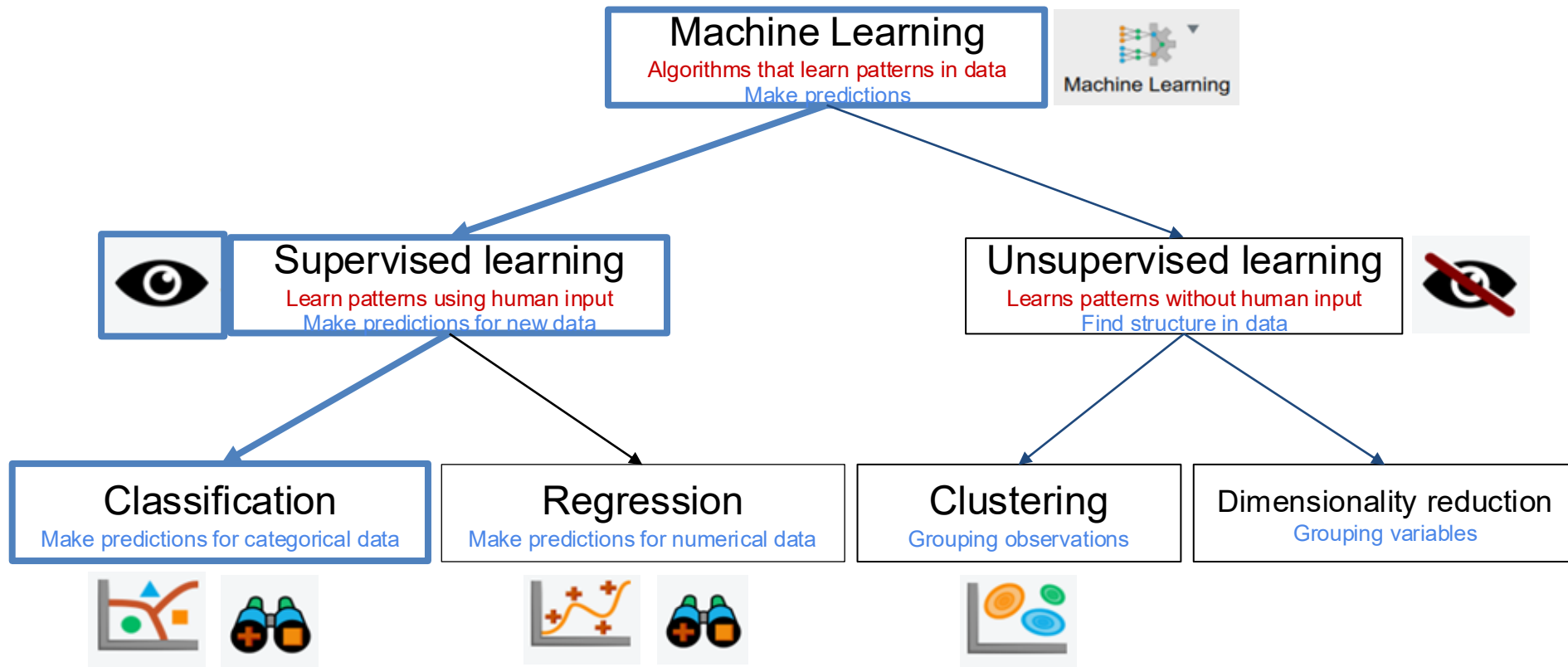
Learning objectives for today

- Train machine learning algorithms (regression, classification, clustering) using JASP
- Understand how to evaluate the quality of these algorithms
- Apply these algorithms to predict new data

Learning objectives for today

- Train machine learning algorithms (regression, **classification**, clustering) using JASP
- Understand how to evaluate the quality of these algorithms
- Apply these algorithms to predict new data

Machine learning is about finding patterns, the goal is making predictions



A typical classification workflow

1. Training the classification model
2. Optimizing the model (optional)
3. Evaluating the quality of the predictions of the model
4. Apply the model to new data

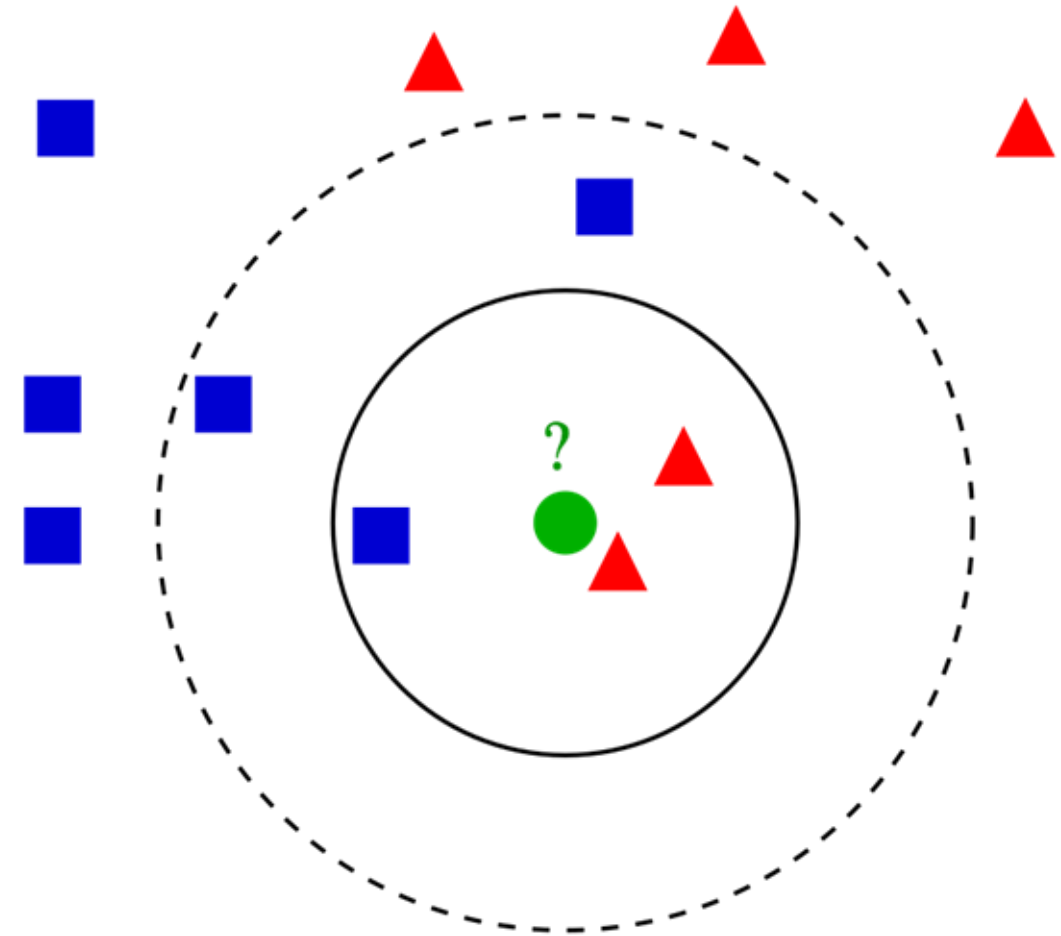
You typically split your dataset in 2 or 3 parts

1. Training set (typically 80%): Used for training the model
2. Validation set: Used for optimizing the model (optional)
3. Test set (typically 20%): Used for evaluating the model

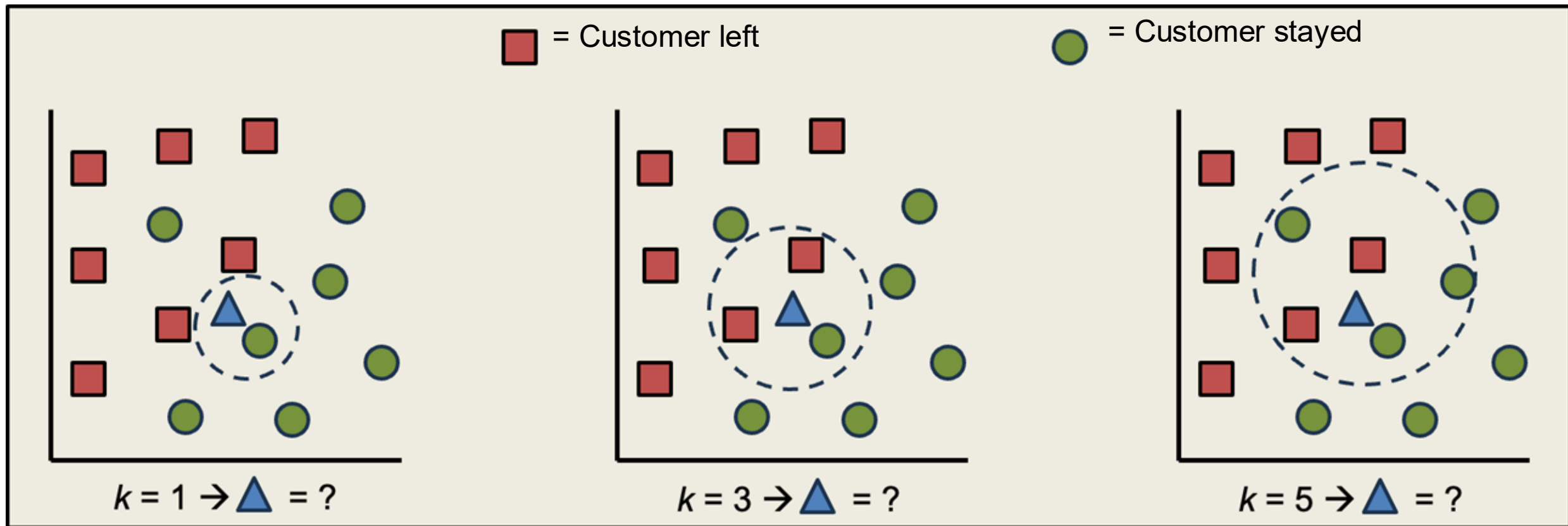


Algorithm: K-nearest neighbors

- Finds the k most similar observations in the training set and uses this as a basis for prediction.
- For a new observation you find the k most similar observations and **take the most occurring category**.



Algorithm: K-nearest neighbors



Exercise: K-nearest neighbors

Using pen and paper, draw up the data as follows:

- Number of products sold on the horizontal axis,
- Number of years as a customer on the vertical axis,
- Use 😊 for a customer who stayed,
- Use 😞 for a customer who left

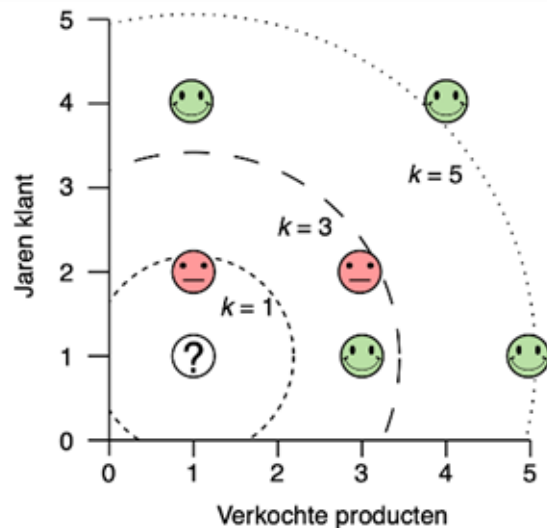
What is your prediction with $k = 1, 3$ and 5 when products sold = years as customer = 1? 😊 😞 😞?

What is your prediction with $k = 1, 3$ and 5 when products sold = 5 and years customer = 3 😊 😞 😞?

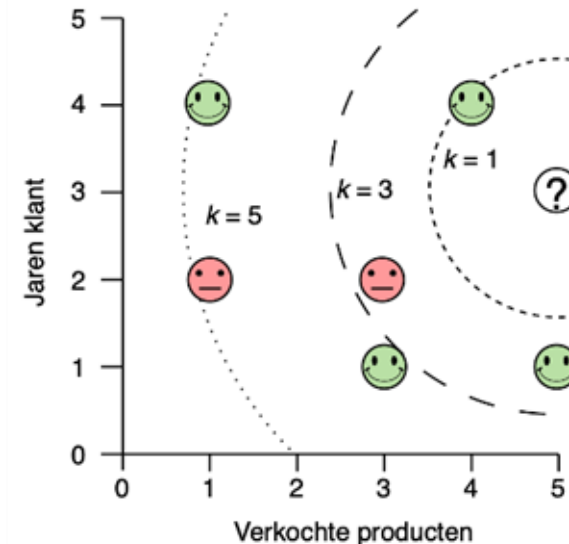
Sold products	Years customer	Outcome
1	4	Stayed
4	4	Stayed
1	2	Left
3	2	Left
3	1	Stayed
5	1	Stayed

Exercise: K-nearest neighbors

- What is your prediction with $k = 1, 3$ and 5 when products sold = years as customer = 1? 😊😐 or 😞? 😊 or 😞?
- What is your prediction with $k = 1, 3$ and 5 when products sold = 5 and years customer = 3 😊 or 😞?



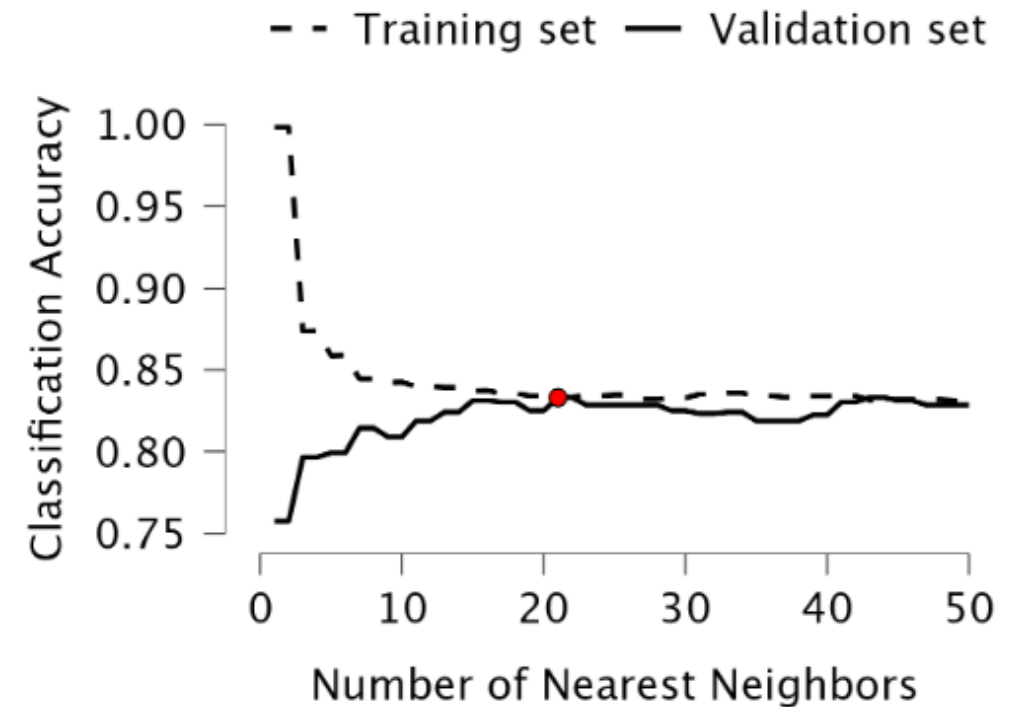
$k = 1 \rightarrow$ 😞
 $k = 3 \rightarrow$ 😞
 $k = 5 \rightarrow$ 😊



$k = 1 \rightarrow$ 😊
 $k = 3 \rightarrow$ 😊
 $k = 5 \rightarrow$ 😊

Validation: K-nearest neighbors

- $k = 1$ is a more flexible model, it gives perfect predictions on the training set, but does not generalize well to never before seen data
- $k = n$ is a more generalized model, it gives poor predictions on the training set but generalizes better to never before seen data.
- Optimal: somewhere in between.





Exercises!



edu.nl/8thy7



Thank you for your attention!

Library Academy 2026

