

Huawei HCIA-IA

Machine Learning - Basics Concepts

Lucas Sousa - Feb 2021

Agenda

Main topics

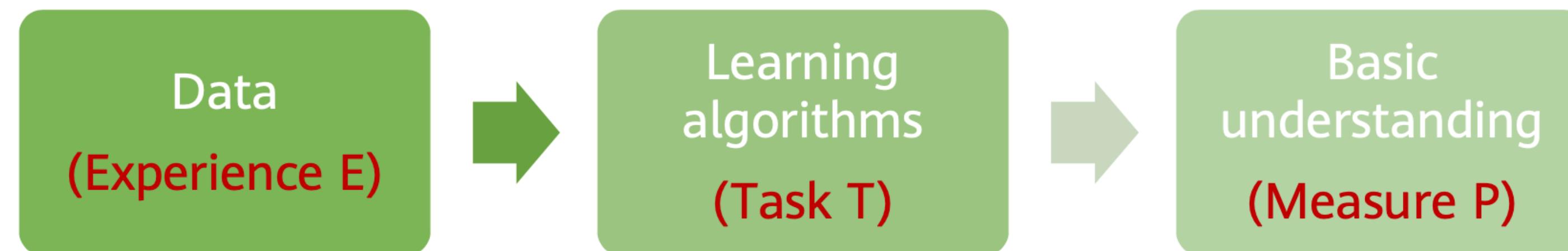
- ML Definition
- Datasets
- Mental problems
- ML Problems

Machine Learning Definition

Machine Learning

Definition

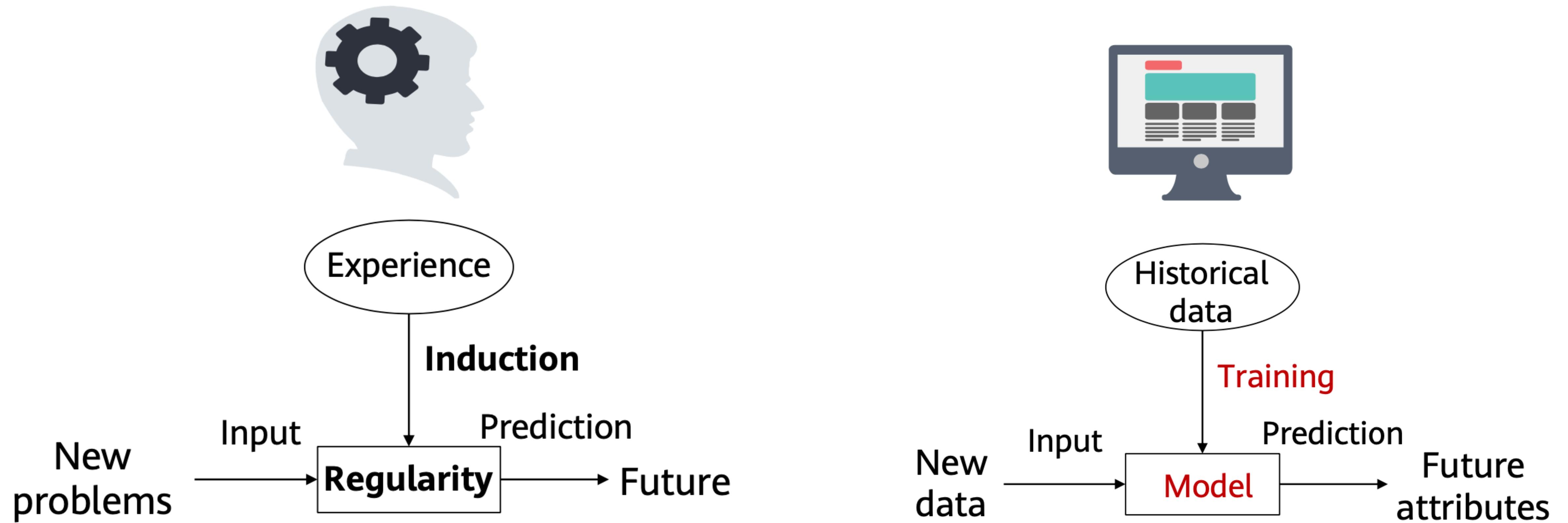
- Machine Learning (including deep learning) is the study of learning algorithms. A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.





Machine Learning

Human x Machine

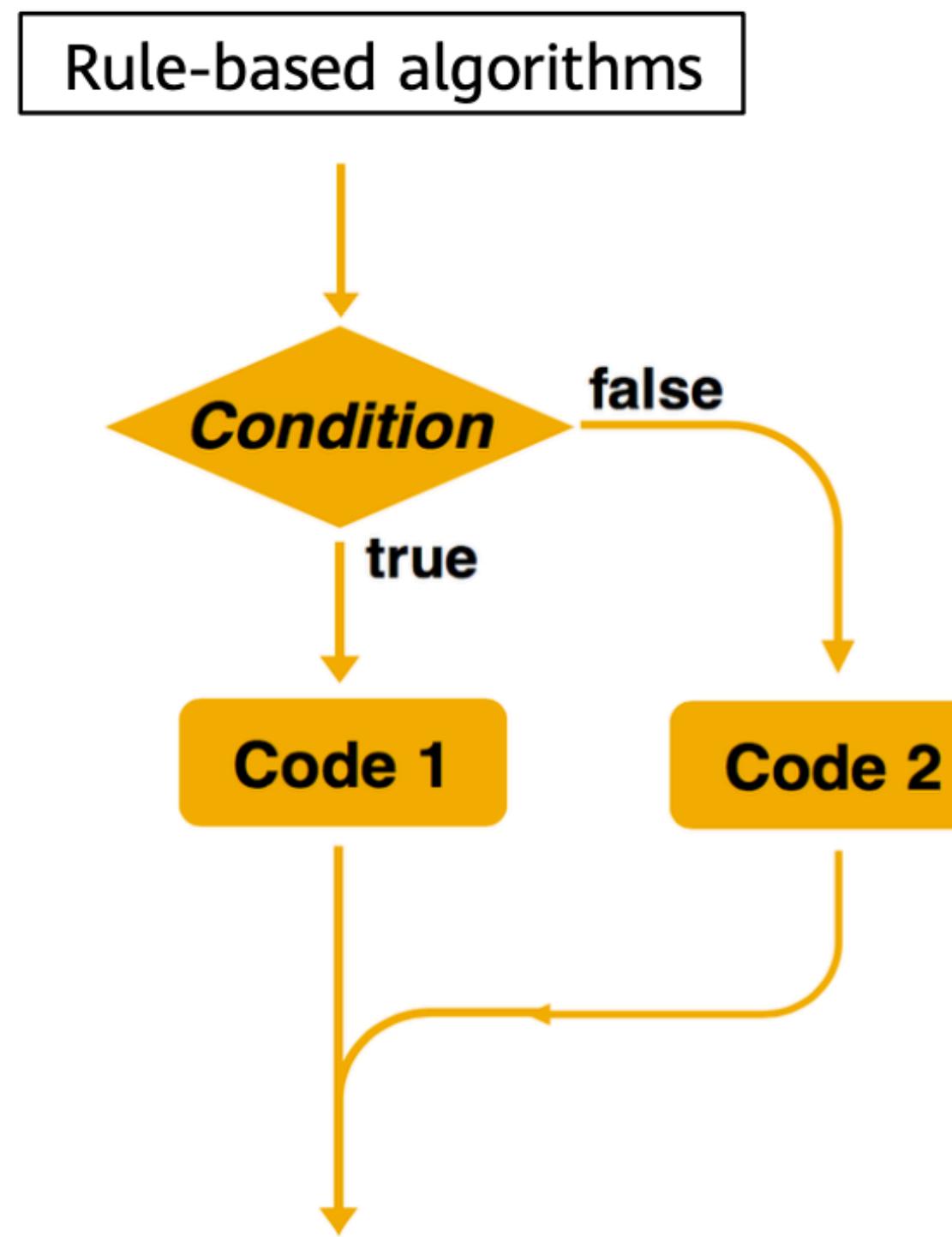


Face Recognition

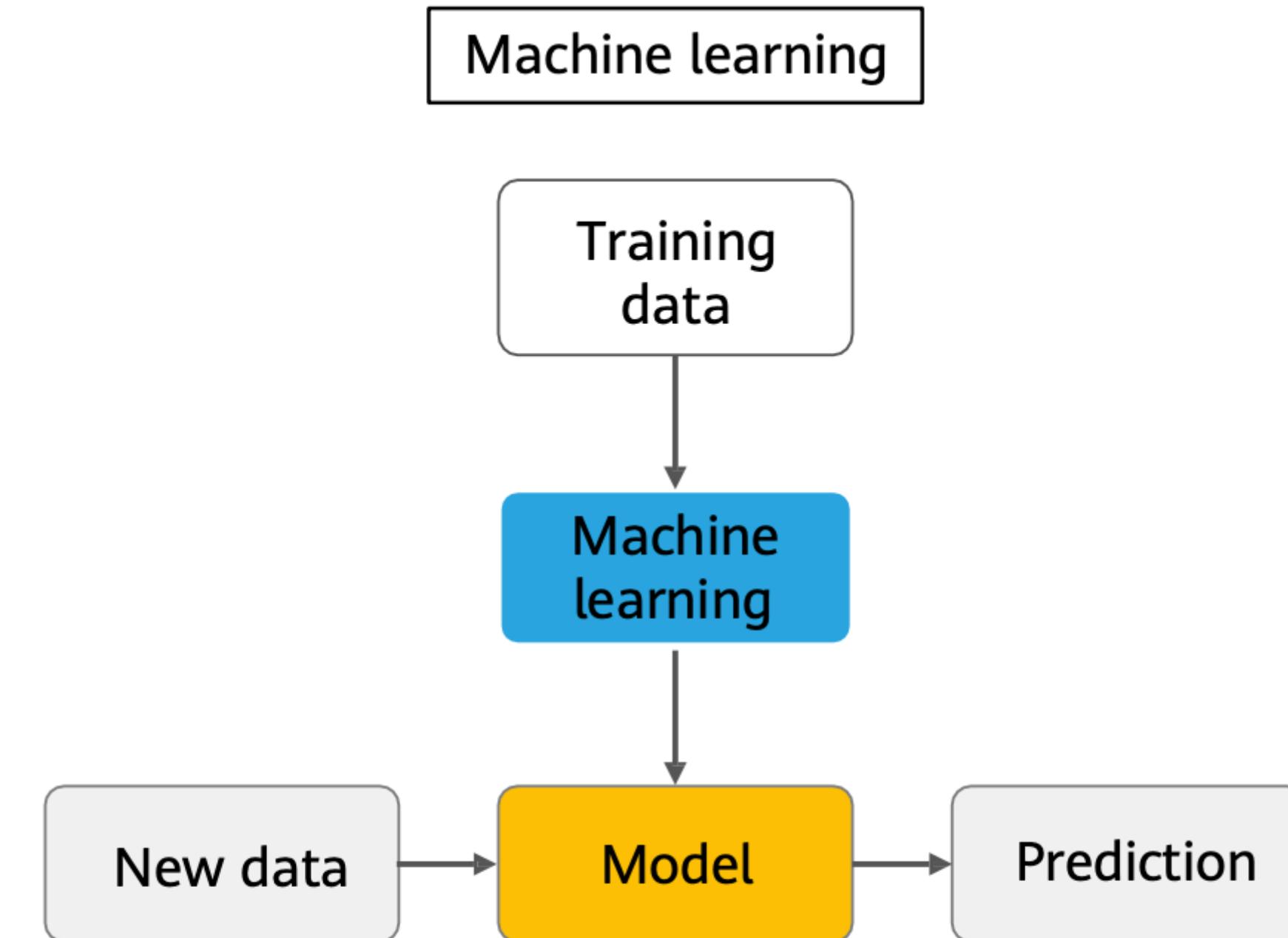
How to identify faces using only procedural language?



Differences between ML and Traditional Algorithms



- Explicit programming is used to solve problems.
- Rules can be manually specified.

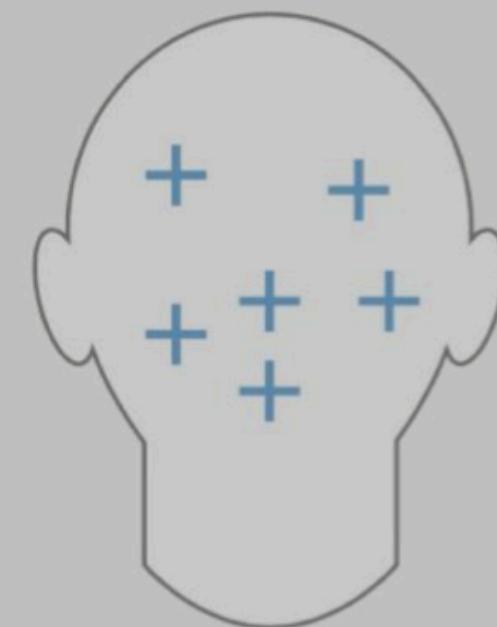


- Samples are used for training.
- The decision-making rules are complex or difficult to describe.
- Rules are automatically learned by machines.

Application Scenarios of Machine Learning

Machine Learning can be used in the following scenarios

Rules are complex or cannot be described, such as facial recognition and voice recognition.



Task rules change over time. For example, in the part-of-speech tagging task, new words or meanings are generated at any time.

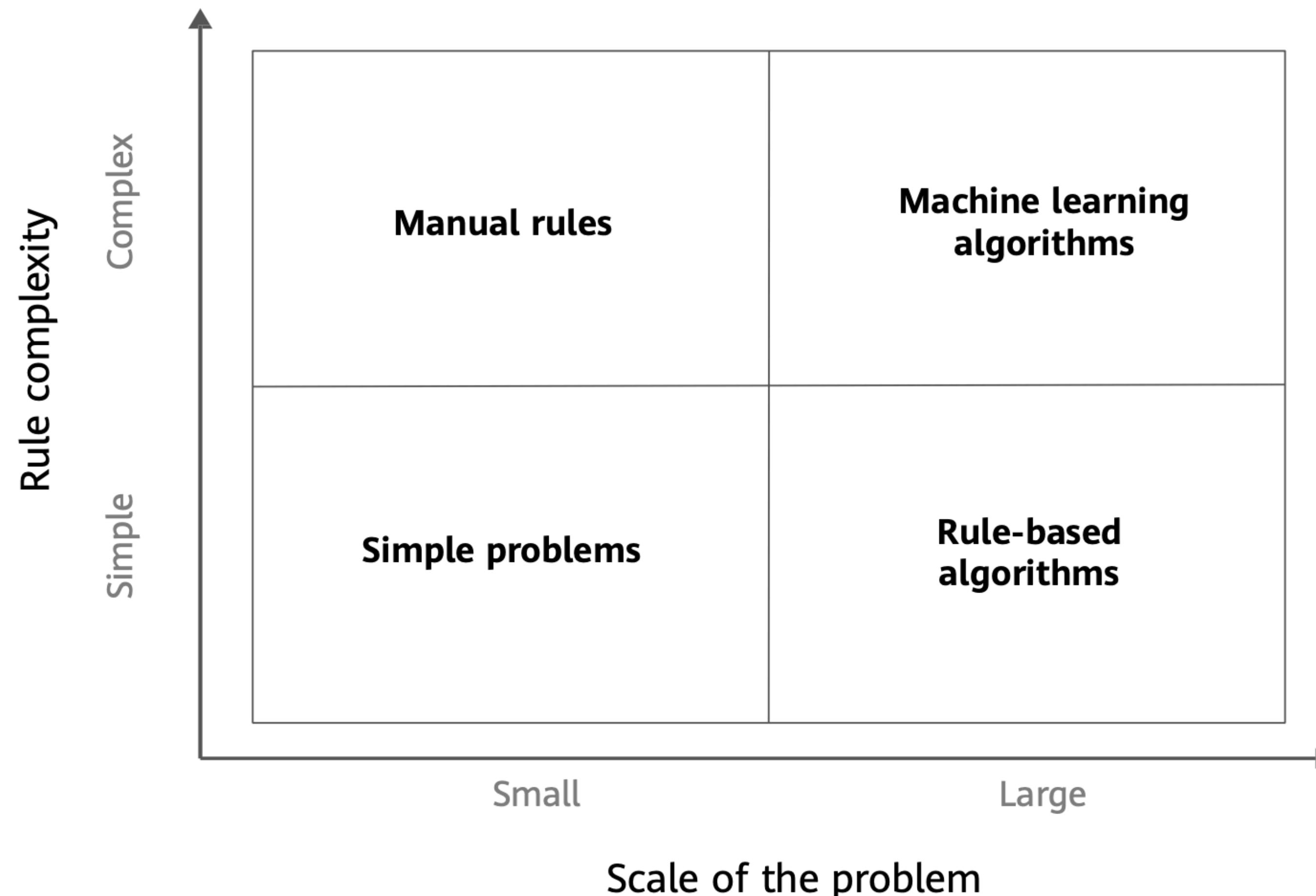


Data distribution changes over time, requiring constant readaptation of programs, such as predicting the trend of commodity sales.



Application Scenarios of Machine Learning

Complexity x Scale of the problem



Samples and Datasets

Samples and Datasets

The source of the knowledge

- Samples are \mathbb{R}^D vectors such that each dimension represents measures obtained from only object.
- In supervised learning, it is necessary to persist a label dimension.
- A dataset is a set of samples such that, in supervised learning tasks, each sample is associated with a label.

Iris dataset

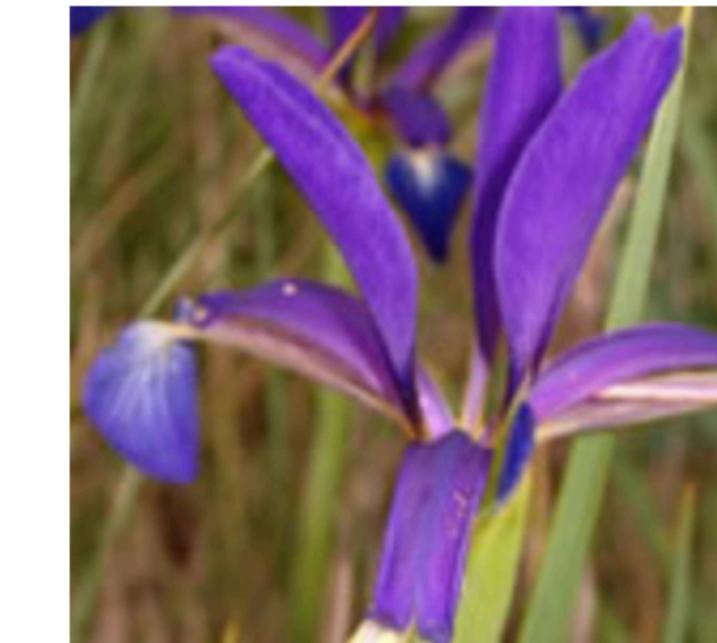
Hello World of ML



Iris Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Famous database; from Fisher, 1936

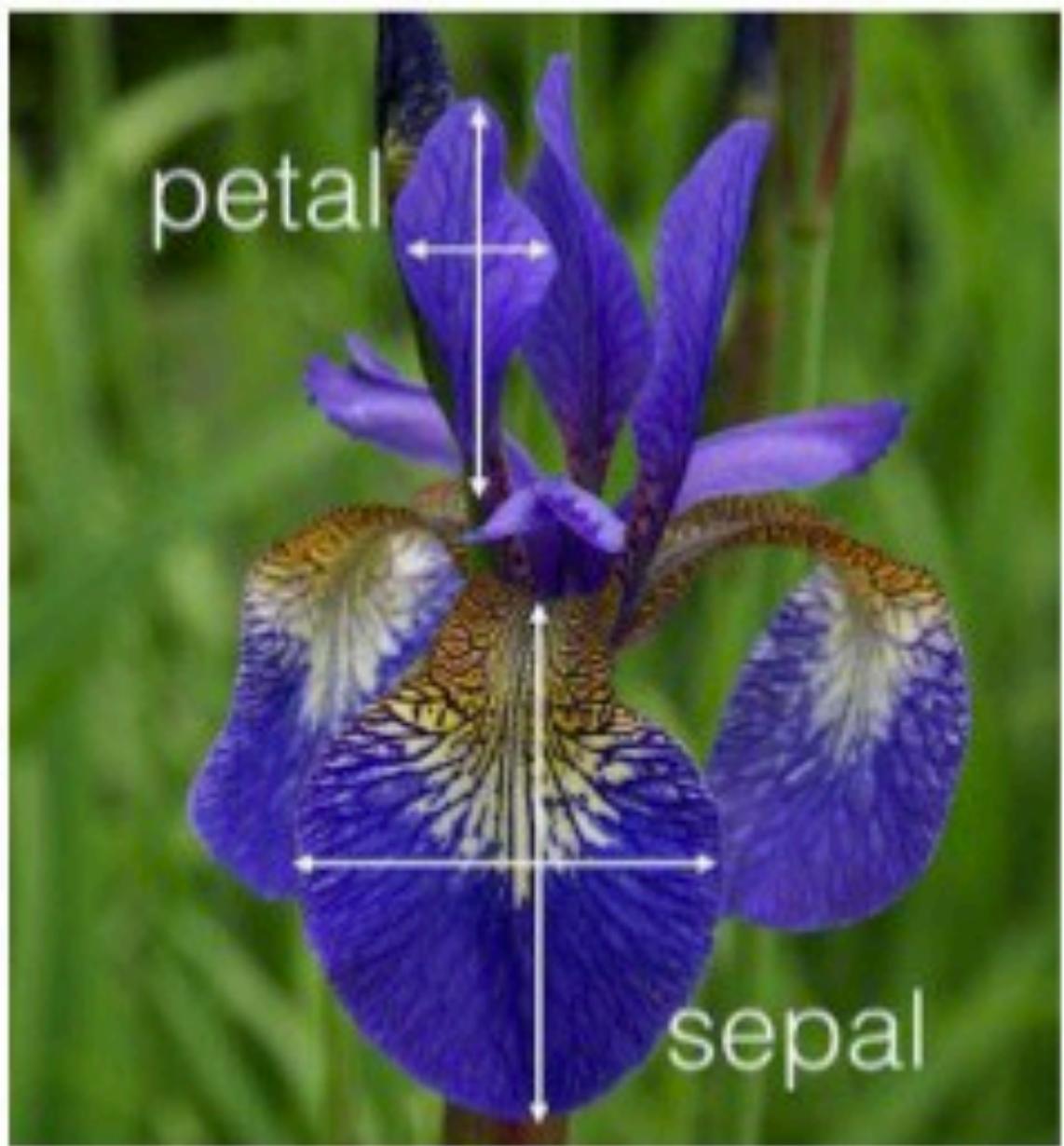


Data Set Characteristics:	Multivariate	Number of Instances:	150	Area:	Life
Attribute Characteristics:	Real	Number of Attributes:	4	Date Donated	1988-07-01
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	3827314

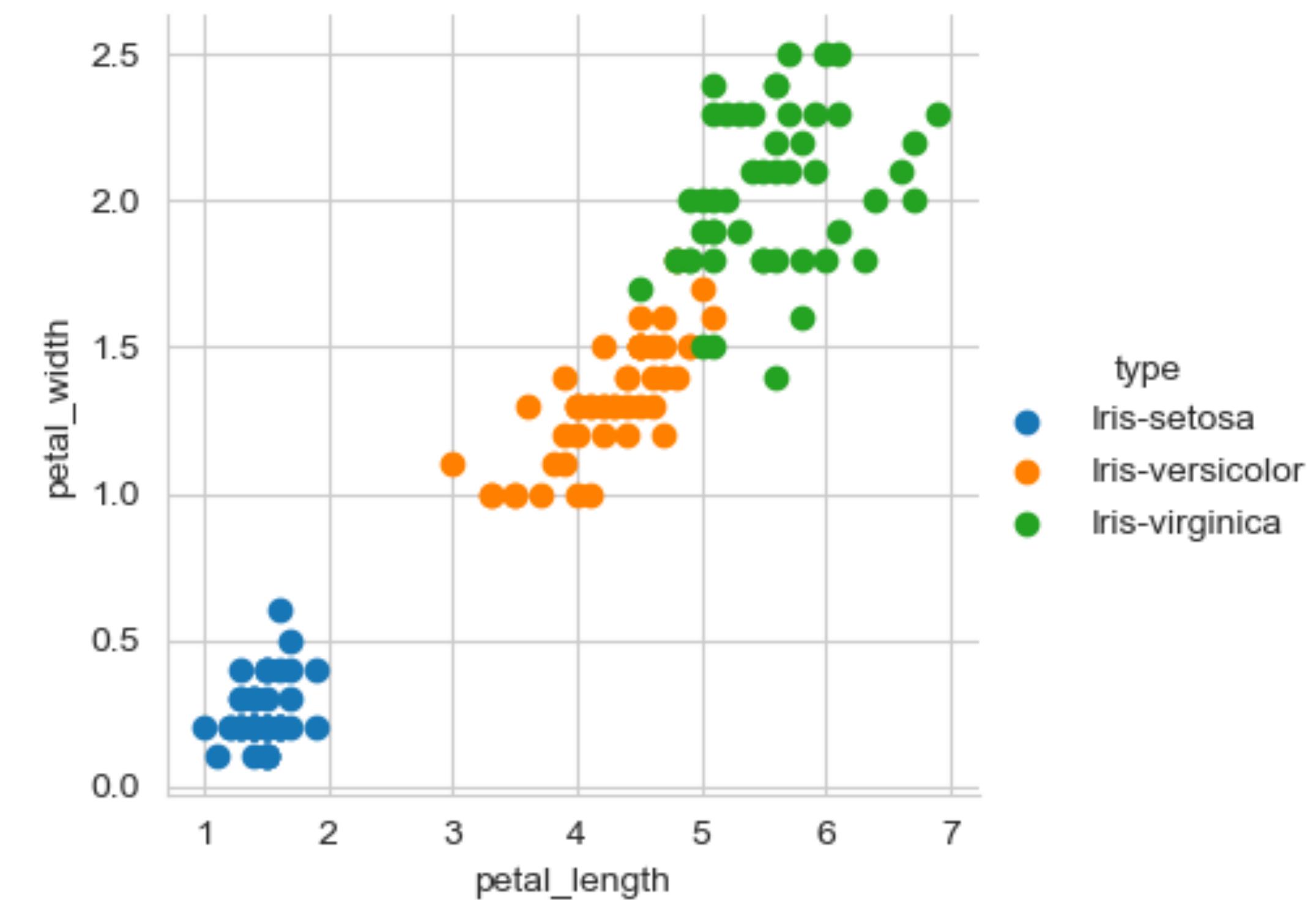
Iris dataset

Hello World of ML

Supervised learning **classification** problem
(using the [Iris flower data set](#))

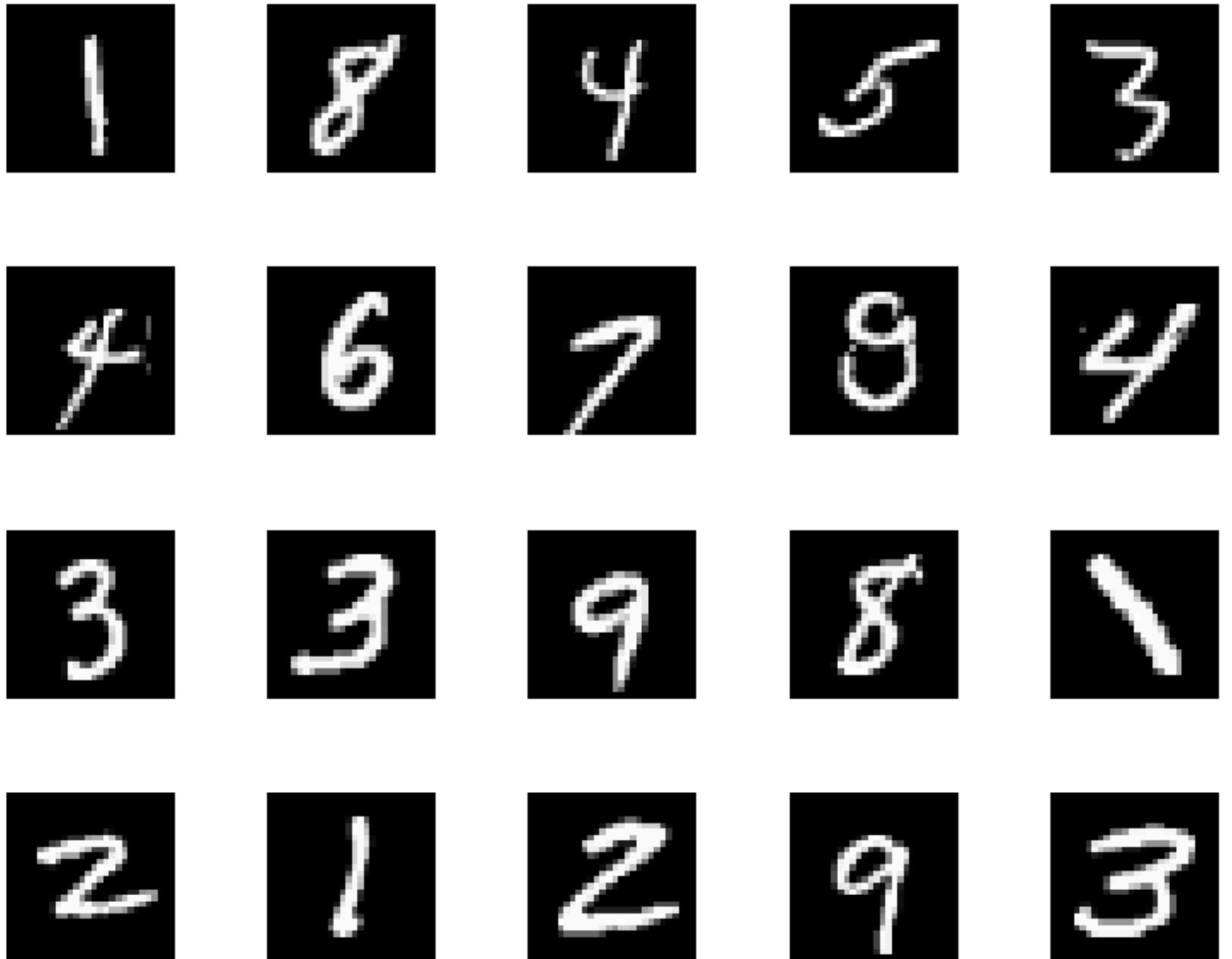


Training / test data				
Features			Labels	
Sepal length	Sepal width	Petal length	Petal width	Species
5.1	3.5	1.4	0.2	Iris setosa
4.9	3.0	1.4	0.2	Iris setosa
7.0	3.2	4.7	1.4	Iris versicolor
6.4	3.2	4.5	1.5	Iris versicolor
6.3	3.3	6.0	2.5	Iris virginica
5.8	3.3	6.0	2.5	Iris virginica

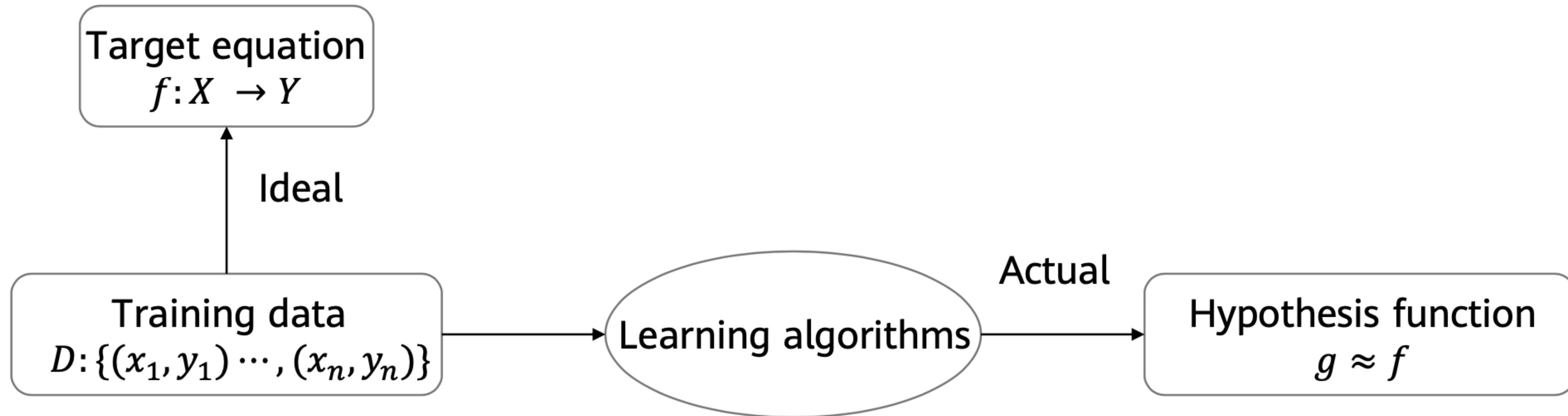


MNIST Dataset

- Handwritten digits
- Contains 60.000 training samples
- Contains 10.000 test samples



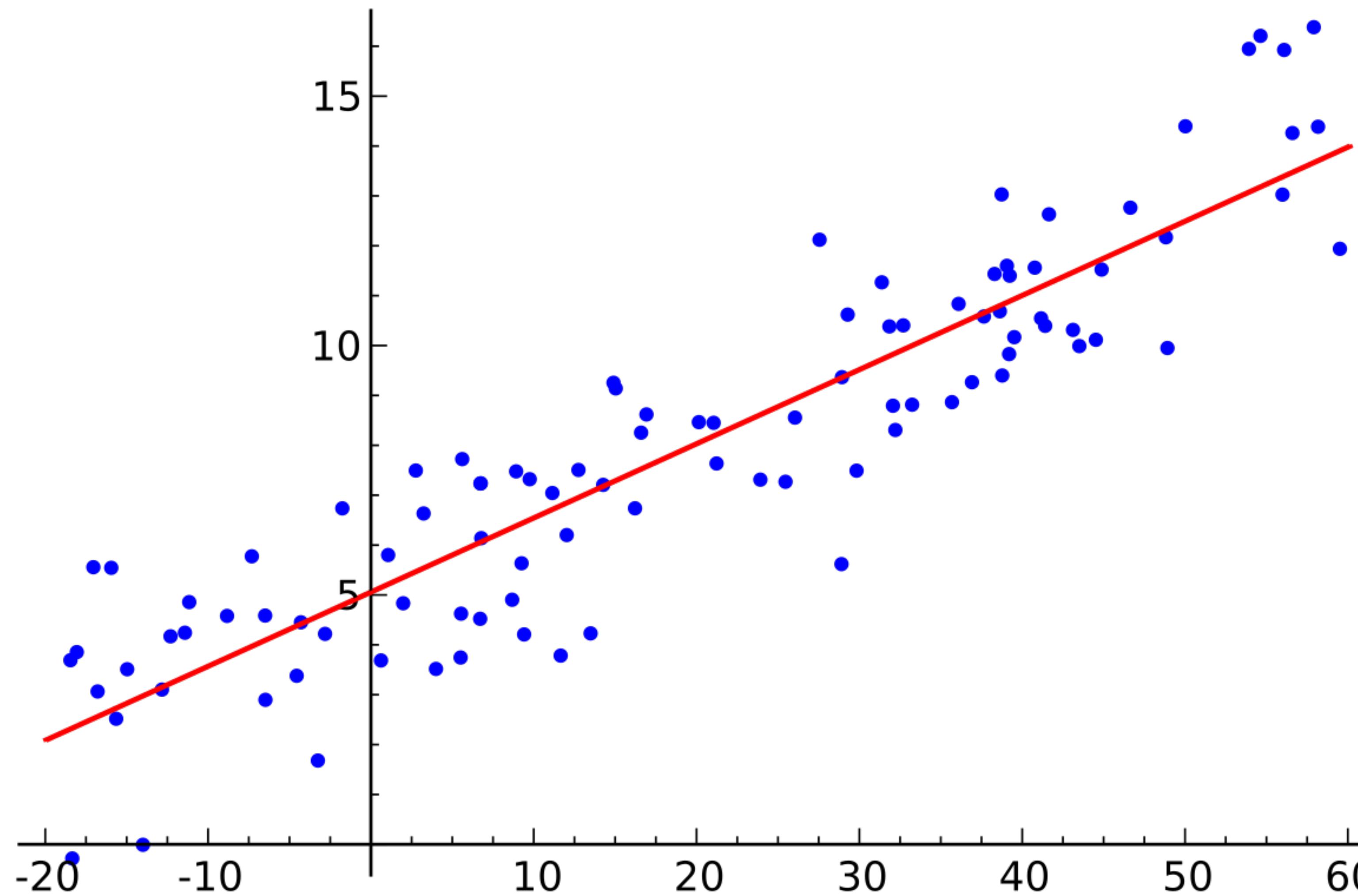
Rational Understanding of Machine Learning Algorithms



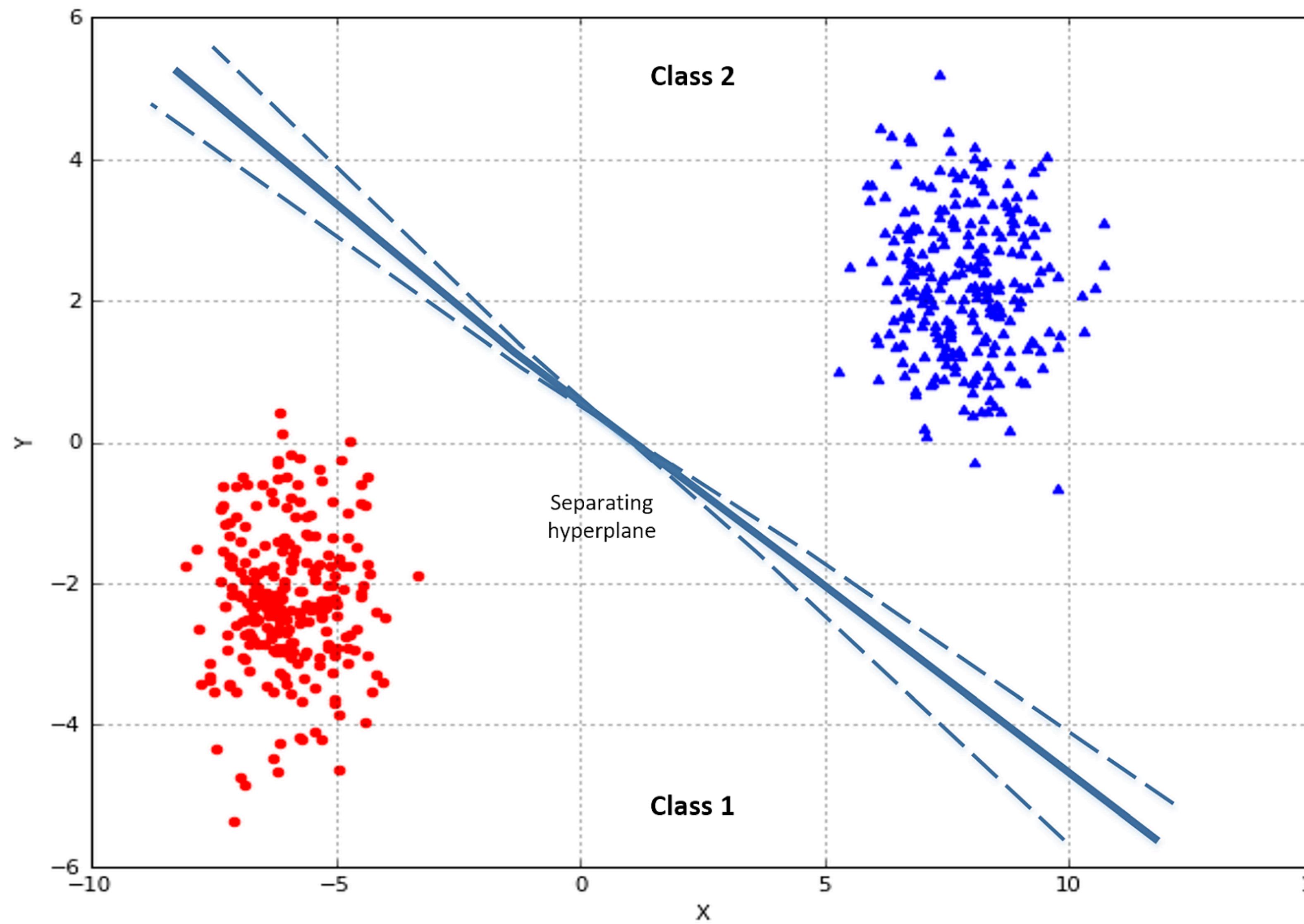
- Target function f is unknown. Learning algorithms cannot obtain a perfect function f .
- Assume that hypothesis function g **approximates** function f , but may be different from function f .

Machine Learning Problems

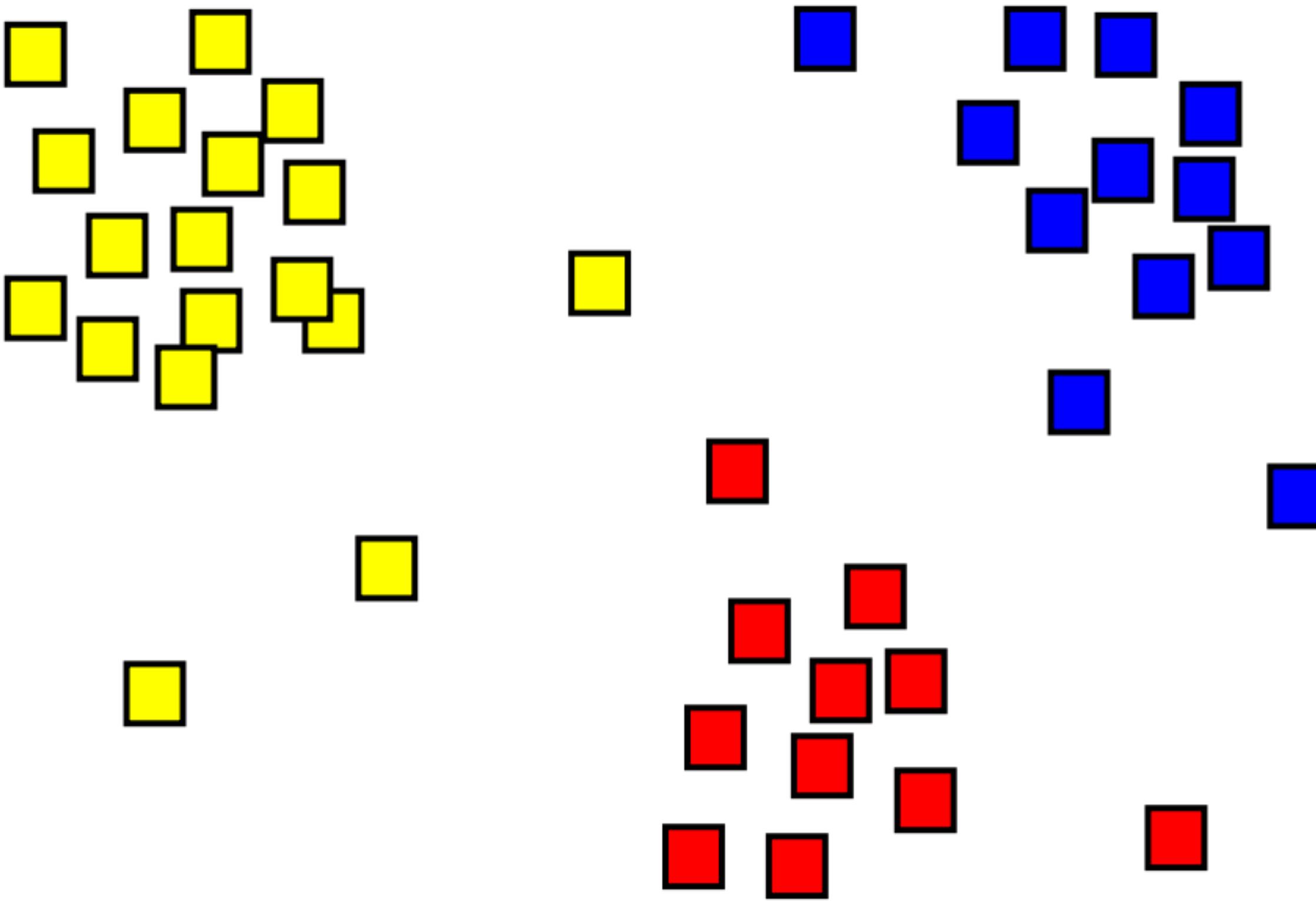
Regression



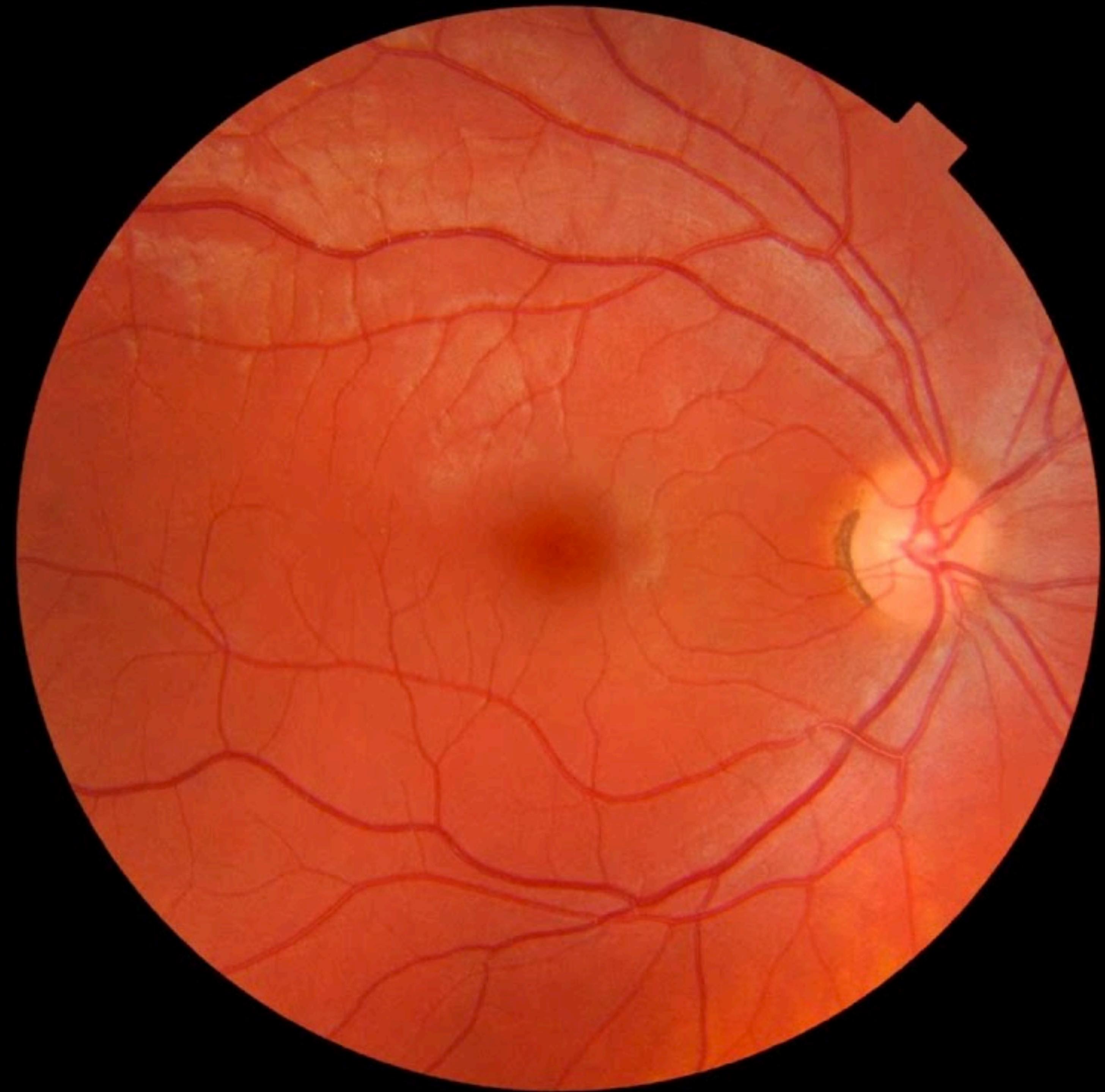
Classification



Clustering



Mental Exercises





Thanks for your attention

