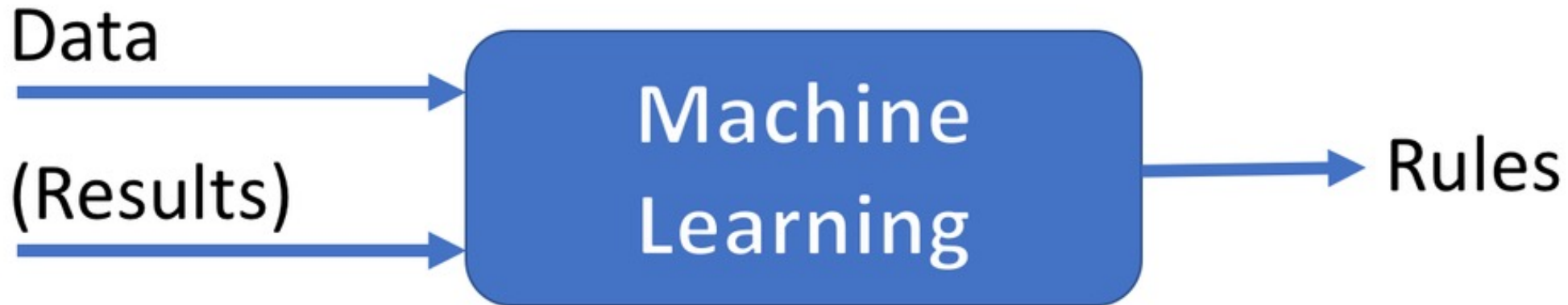


# Introduction to Machine Learning for everyone

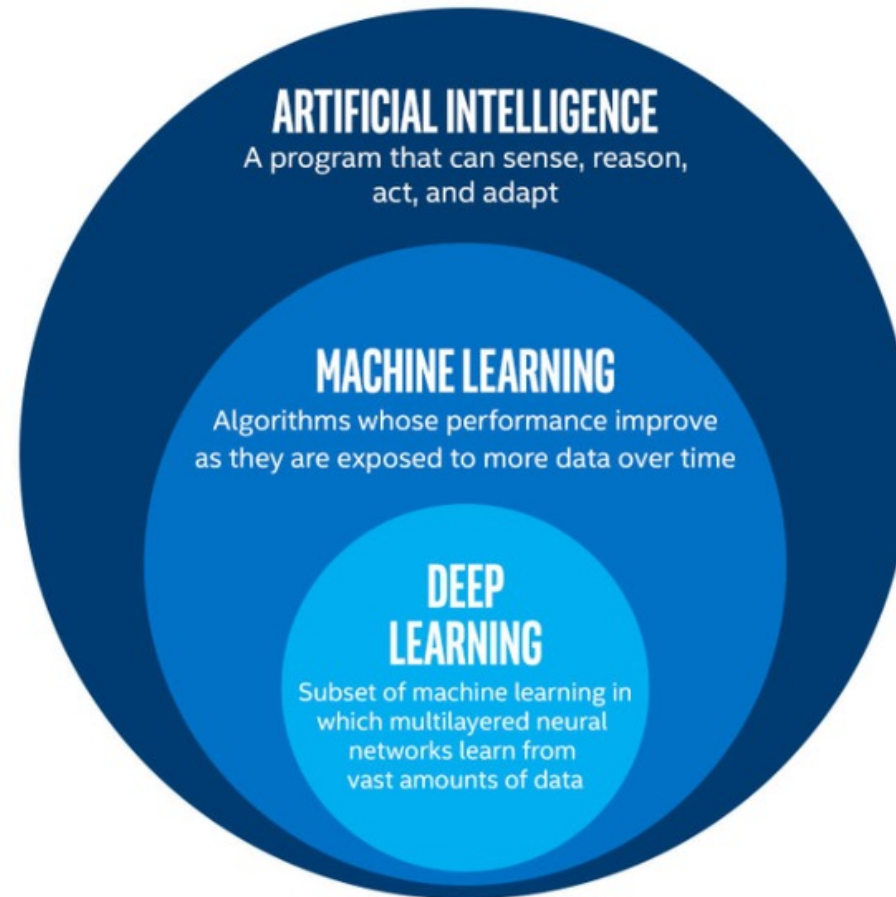
Thanh Tran, PhD candidate

<https://thanhvotran.github.io/>

# A new paradigm...



# AI, Machine Learning and Deep Learning



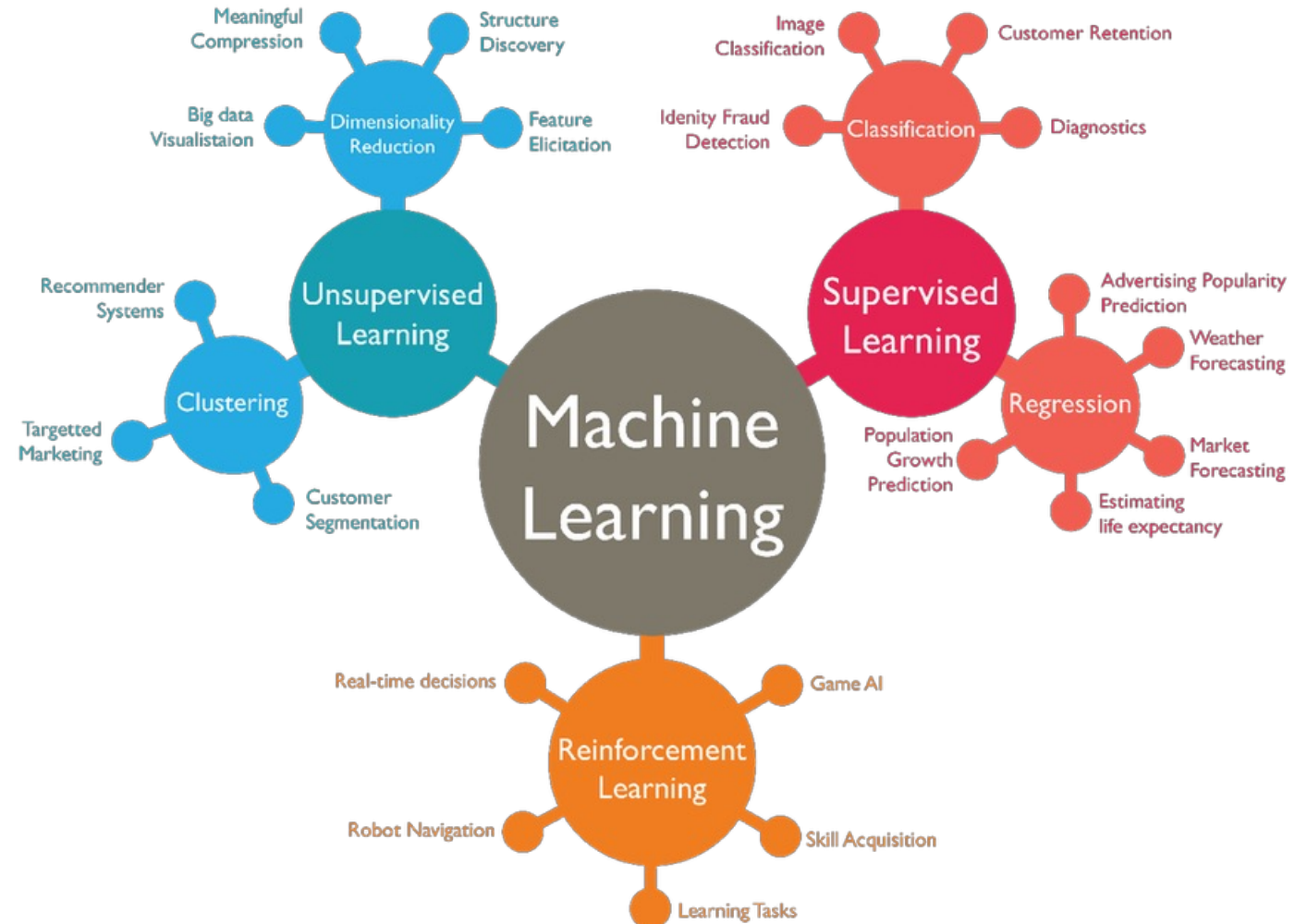
# Typology of ML systems

ML systems are traditionally classified in three categories, according to the amount and type of human supervision during training. Hybrid approaches exist.

**Supervised Learning:** expected results (called labels or tags) are given to the system along with training data.

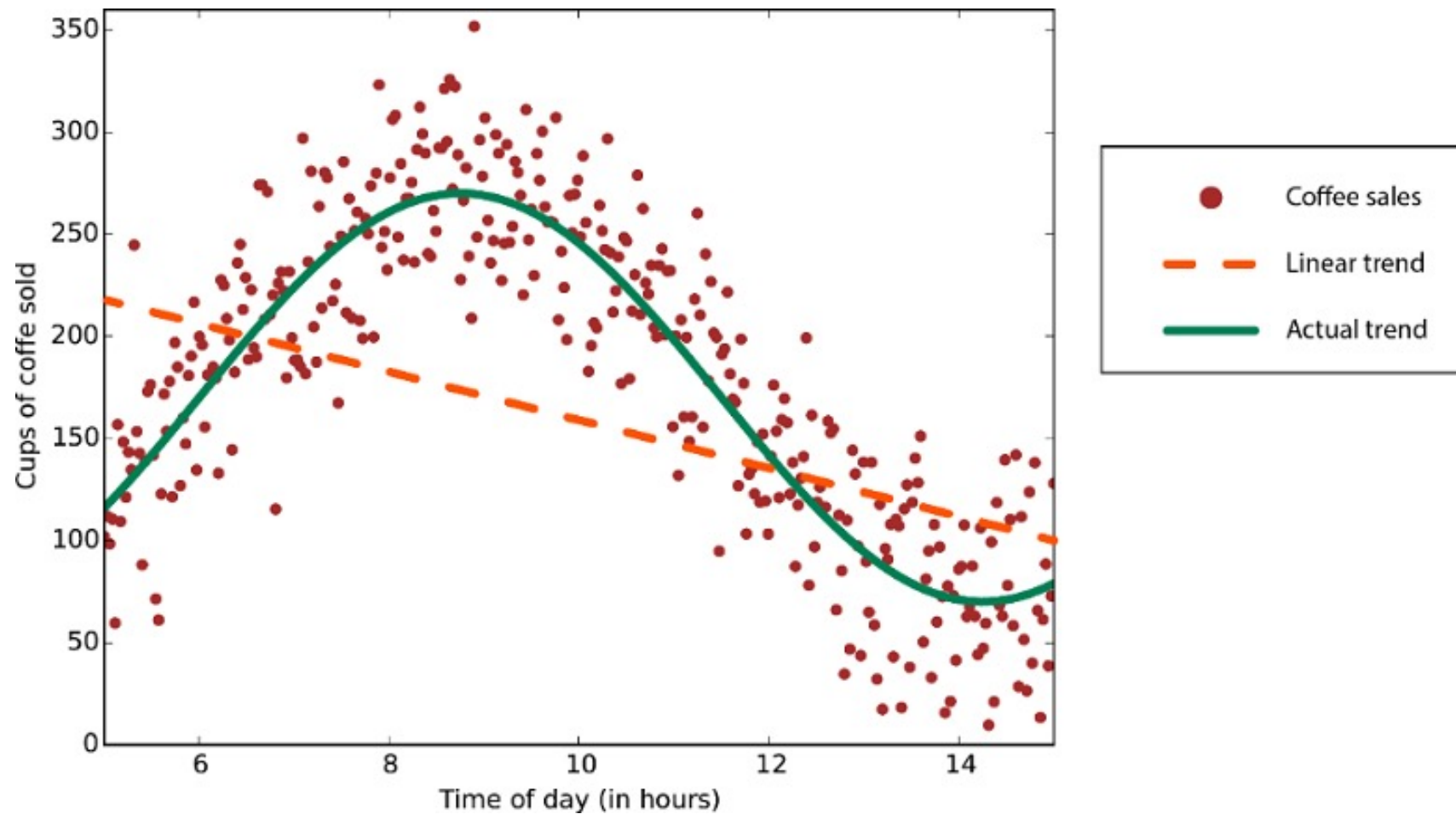
**Unsupervised Learning:** training data comes without the expected results. The system must discover some structure in the data by itself.

**Reinforcement Learning:** without being given an explicit goal, the system's decisions produce a reward it tries to maximize.



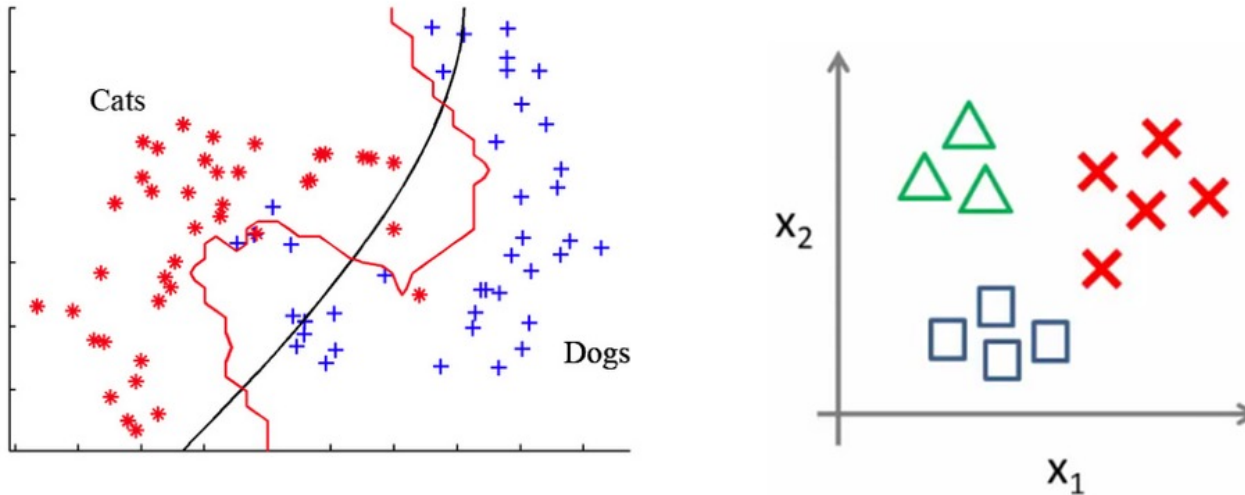
## Regression

The system predicts continuous values. Examples: temperature forecasting, asset price prediction...



## Classification

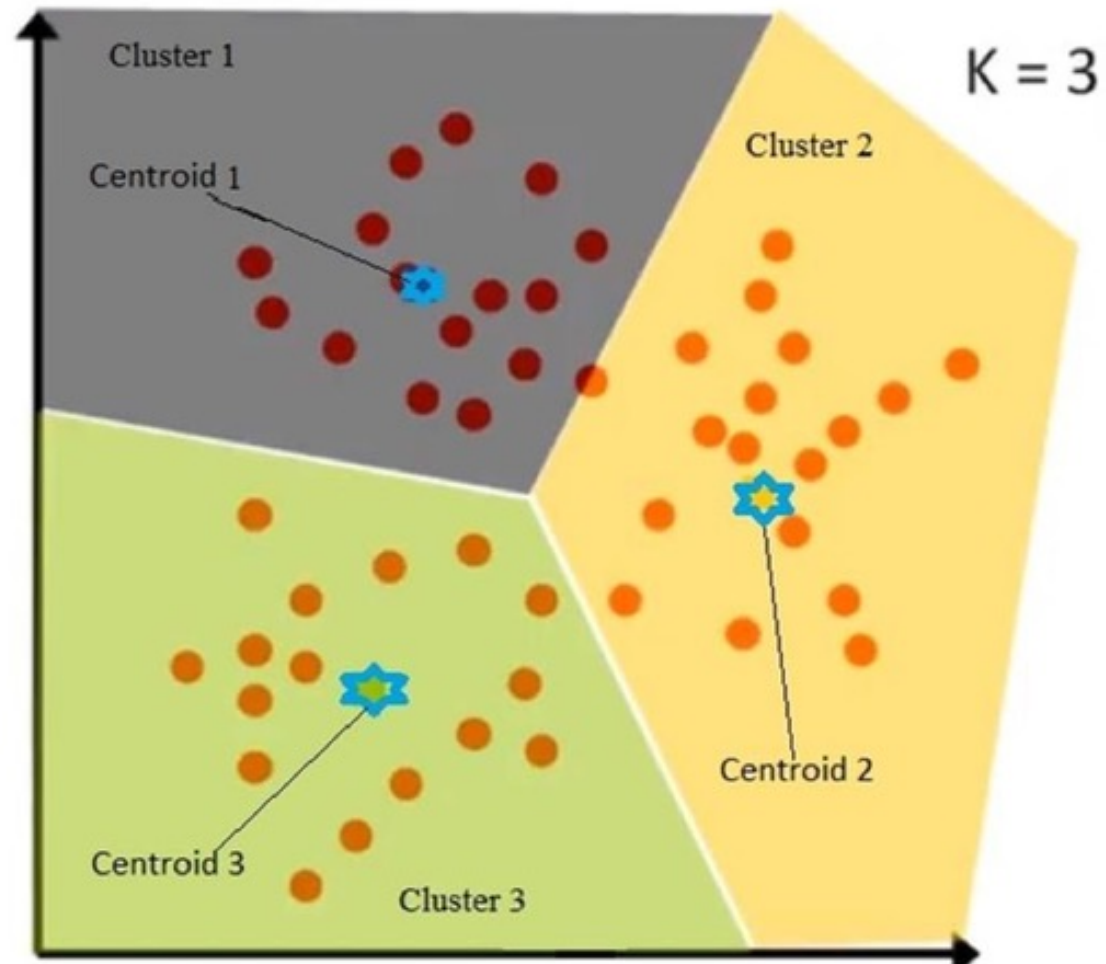
The system predicts **discrete** values: input is **categorized**.



- Binary: only two possible classes. Examples: cat/not a cat, spam/legit mail, benign/malignant tumor.
- Multiclass: several mutually exclusive classes. Example: handwritten digit recognition.
- Multilabel: several non-mutually exclusive classes. Example: face recognition.

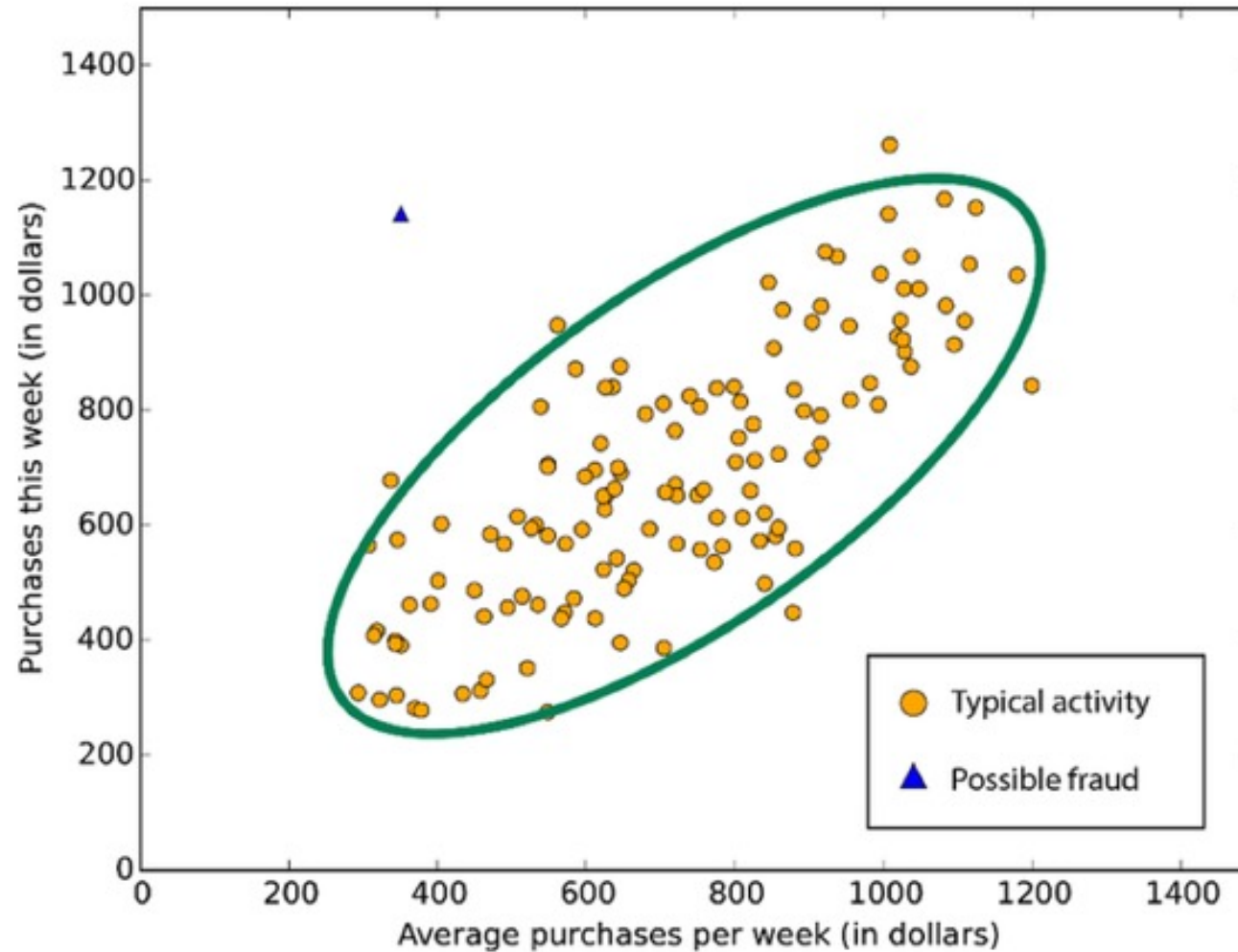
# Clustering

Data is partitioned into groups.



## Anomaly Detection

The system is able to detect abnormal samples (outliers).





# What do you want the machine learning system to do?

I want to see if there are natural clusters or dimensions in the data I have about different situations.

I want to learn what actions to take in different situations.

Do you want the ML system to be active or passive?

**ACTIVE**

The system's own actions will affect the situations it sees in the future.

**PASSIVE**

The system will learn from data I give it.

Do you have access to data that describes a lot of examples of situations and appropriate actions for each situation?

Will the system be able to gather a lot of data by trying sequences of actions in many different situations and seeing the results?

Could there be patterns in these situations that humans haven't recognized before?

Could a knowledgeable human decide what actions to take based on the data you have about the situation?

**UNSUPERVISED LEARNING MAY BE APPROPRIATE**

*clustering  
anomaly detection*

**SUPERVISED LEARNING MAY BE APPROPRIATE**

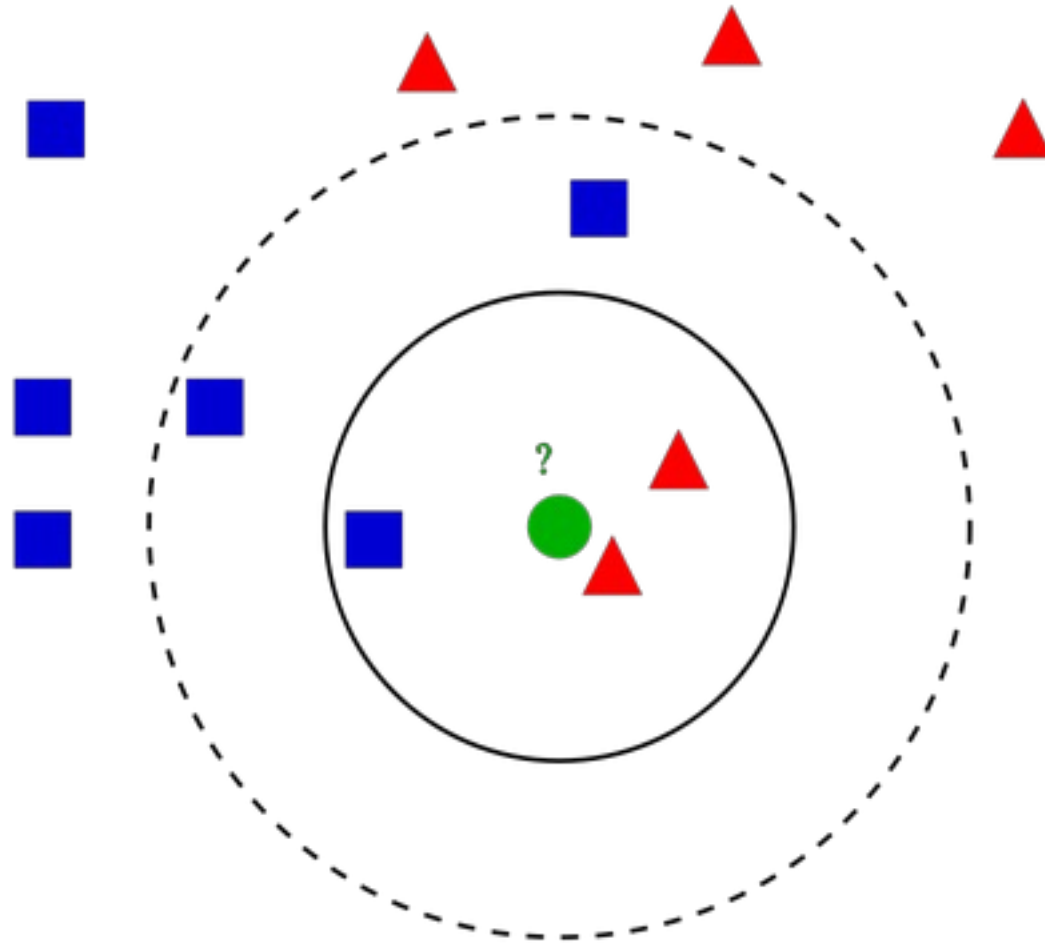
*neural nets  
support vector machines  
regression  
recommender systems*

**MACHINE LEARNING IS NOT USEFUL**

**REINFORCEMENT LEARNING MAY BE APPROPRIATE**

# How do “machines” learn actually?

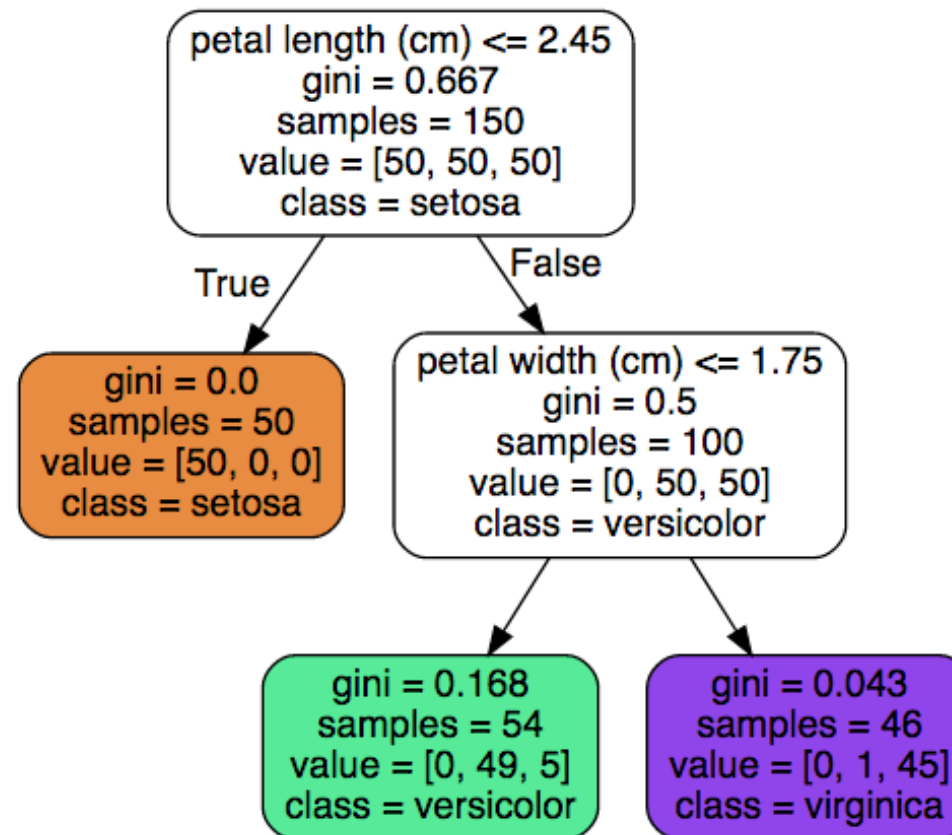
# Algorithm 1: K Nearest Neighbors



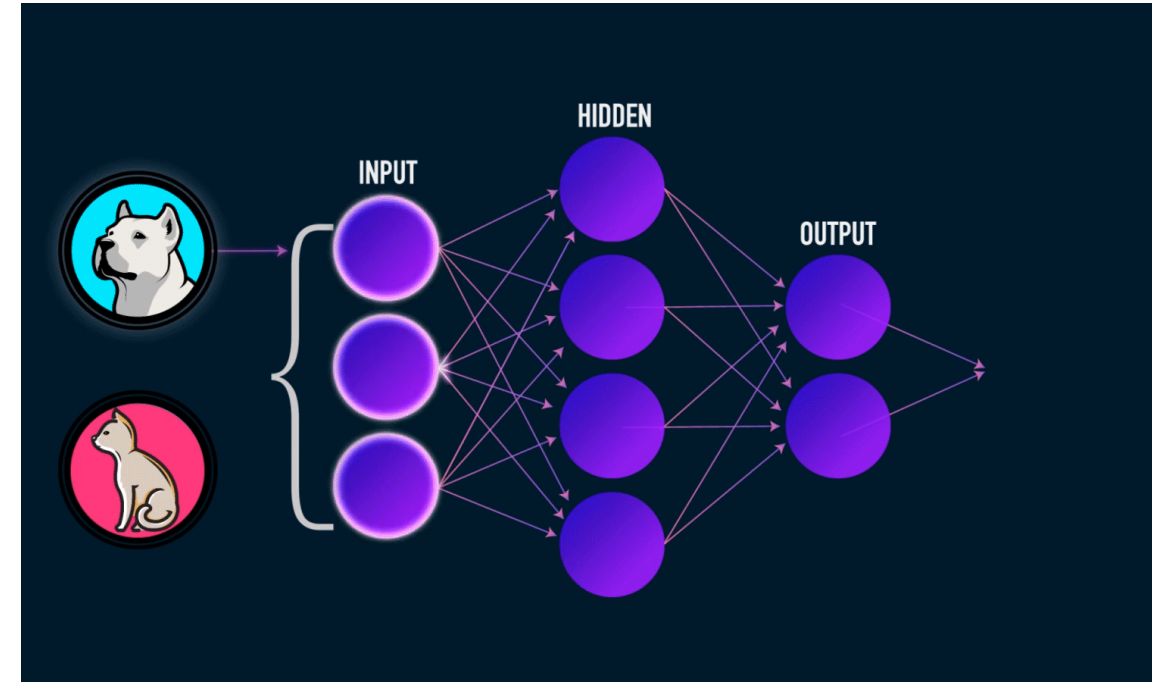
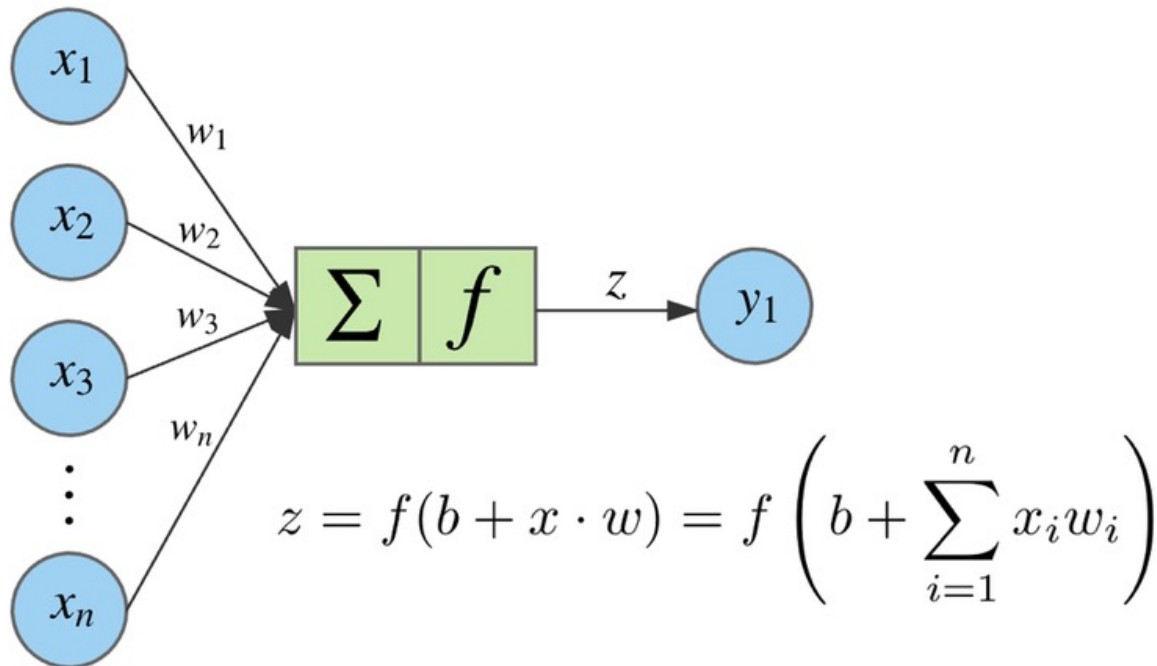
Prediction is based on the k nearest neighbors of a data sample.

# Algorithm 2: Decision Trees

Build a tree-like structure based on a series of discovered questions on the data.

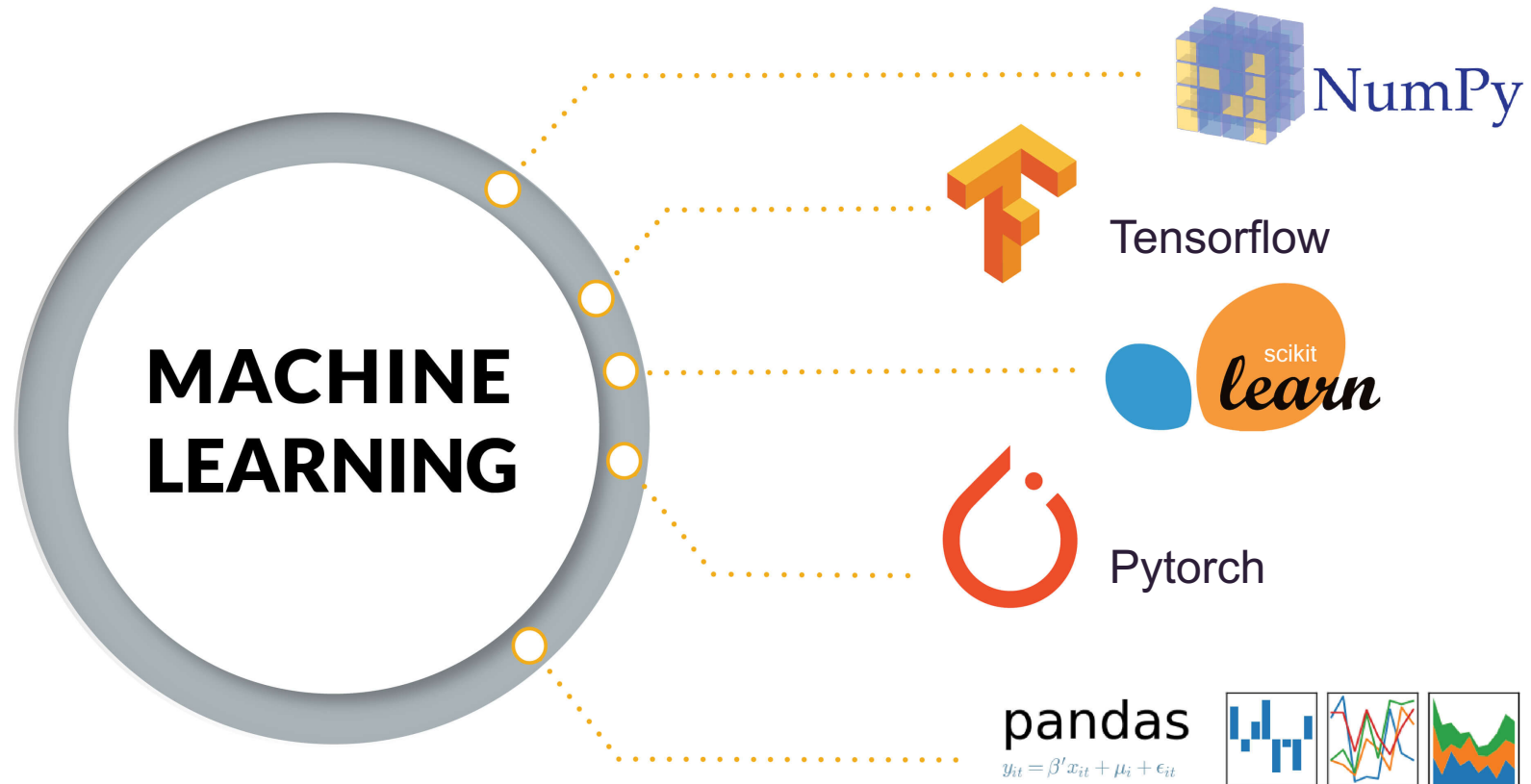


# Algorithm 3: Neural Networks (Deep Learning)



Layers of loosely neuron-inspired computation units that can approximate any continuous function.

# Python-based Machine Learning Libraries



# Linear Regression with Python

Problem: Predicting sales based on an advertisement on TV, Radio and Newspaper.

1		TV	Radio	Newspaper	Sales
2	1	230.1	37.8	69.2	22.1
3	2	44.5	39.3	45.1	10.4
4	3	17.2	45.9	69.3	9.3
5	4	151.5	41.3	58.5	18.5
6	5	180.8	10.8	58.4	12.9
7	6	8.7	48.9	75	7.2
8	7	57.5	32.8	23.5	11.8
9	8	120.2	19.6	11.6	13.2
10	9	8.6	2.1	1	4.8
11	10	199.8	2.6	21.2	10.6
12	11	66.4	5.9	84.5	8.6

# Reference

Machine Learning landscape's plots, <https://colab.research.google.com/github/bpesquet/mlhandbook/>

Example of LR, <https://www.mishrark.com/machine-learning/3-regression/simple-linear-regression-in-python>