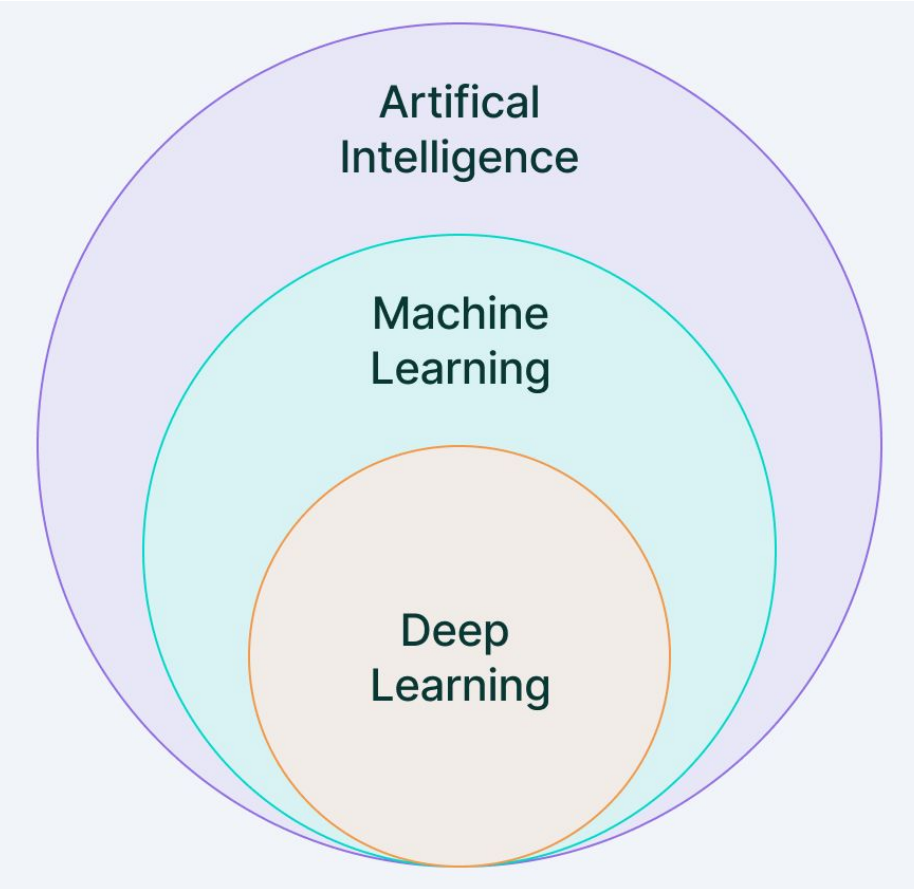
Introduction to machine learning

What is machine learning?

What is machine learning?

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** of **tasks** in **T**, as measured by **P**, improves with **experience E**.



Artificial intelligence (AI): aims at building systems that simulate intelligence behaviour

Machine learning (ML): a subset of AI that learns to make decisions by fitting mathematical models to data.

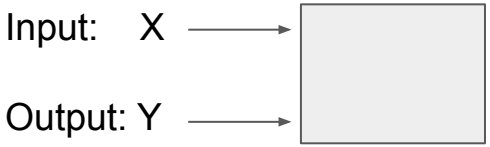
Deep Learning (DL): machine learning algorithms that make use of neural networks.

In regards to experience E machine learning is:

1. **Supervised learning**
   a. Classification
   b. Regression
2. **Unsupervised learning**
   a. Clustering
   b. Dimensionality reduction
3. **Reinforcement learning**

Supervised learning

Training



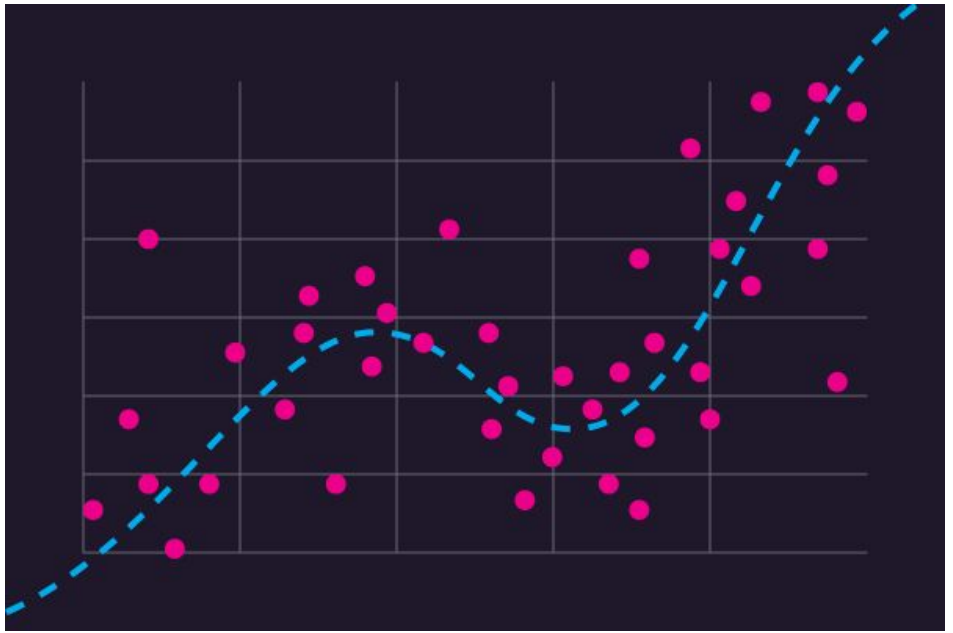
We need lots of pairs (X,Y)

Testing

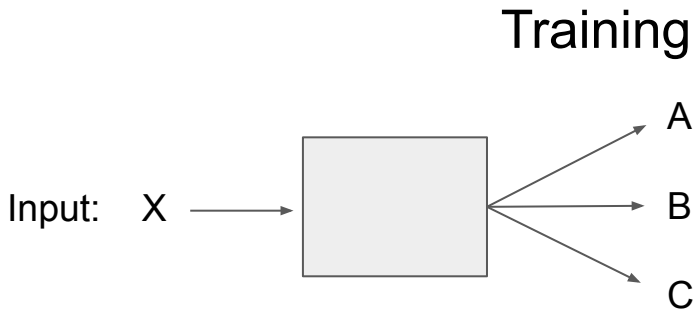
Supervised learning tasks

Classification

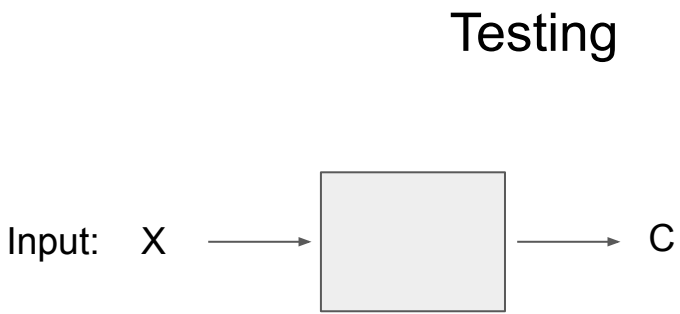
Regression



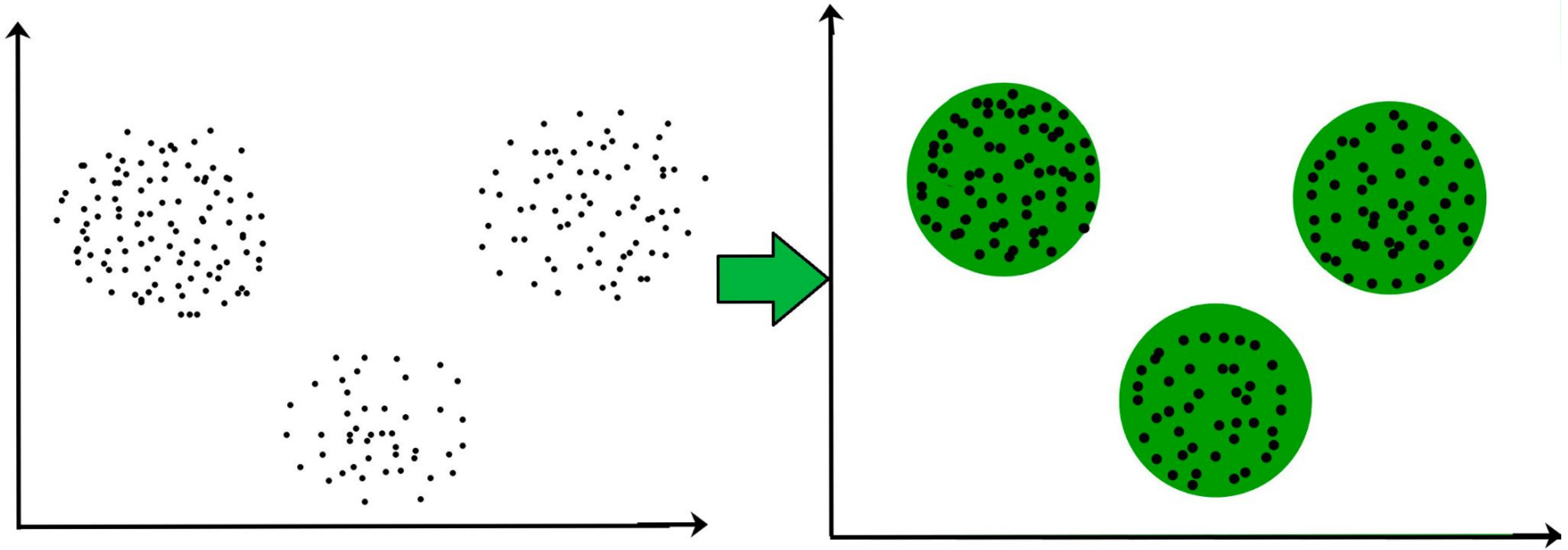
Unsupervised learning



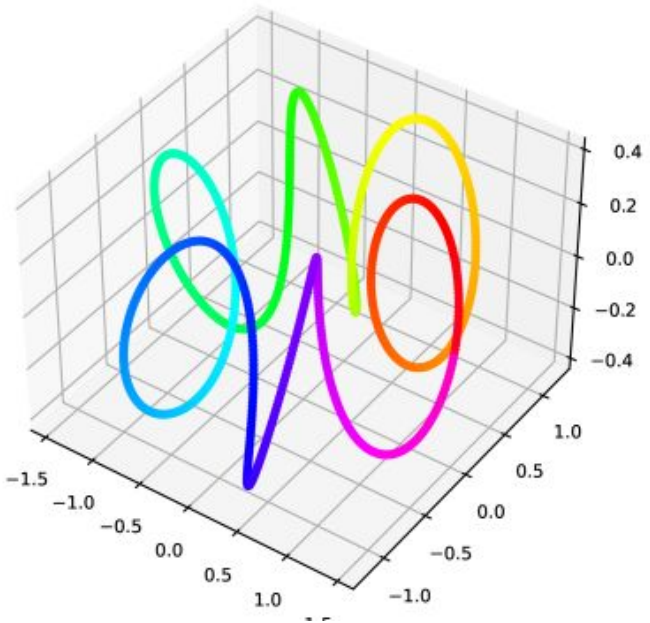
Goal is to find underlying patterns in the data



Unsupervised learning: clustering

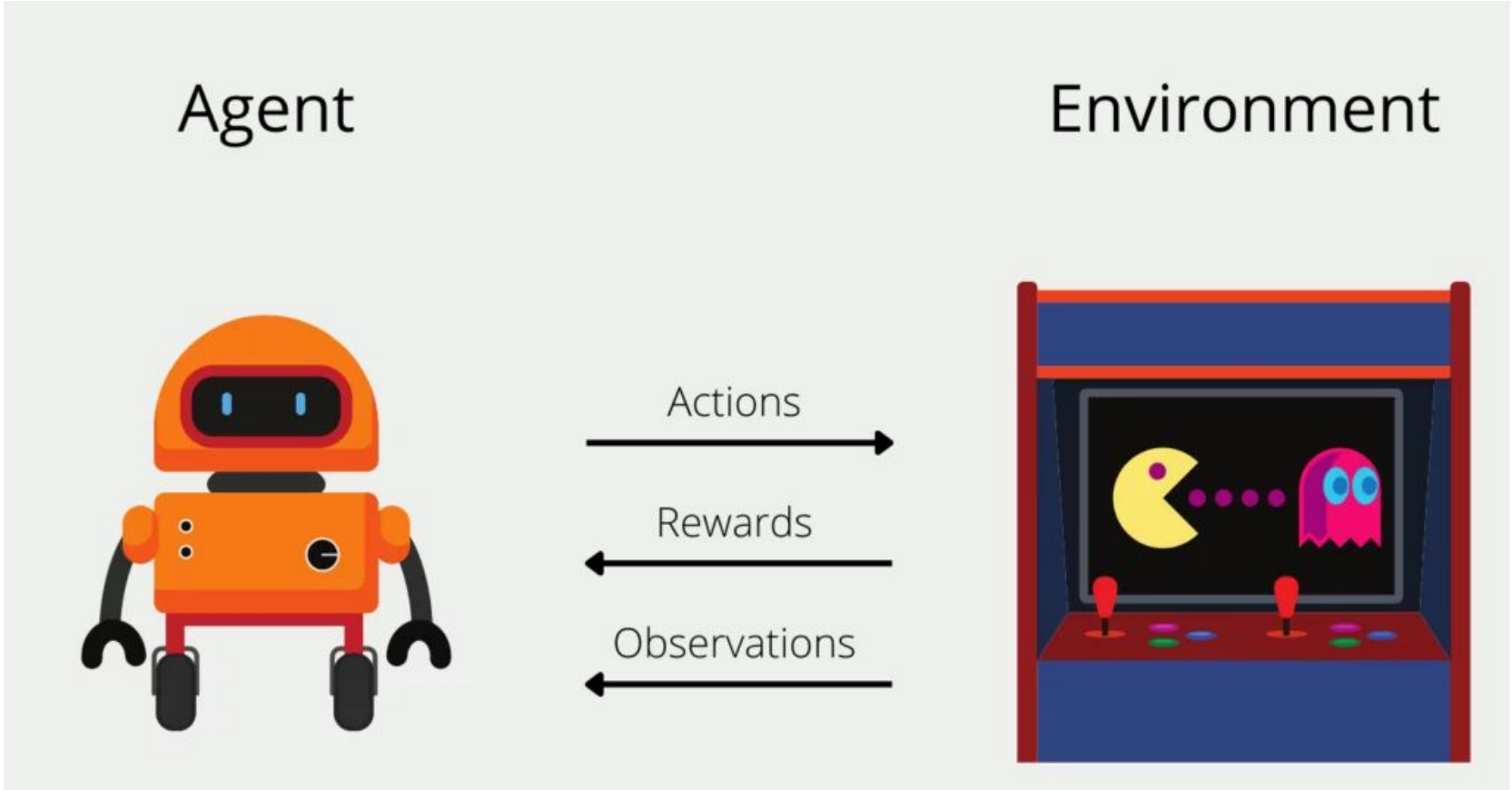


Unsupervised learning: dimensionality reduction



3D ⟶ 1D

Reinforcement learning

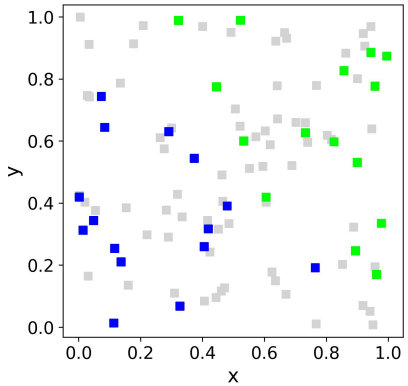


MACHINE LEARNING JARGON

Features are observable quantities, known for all objects (input)

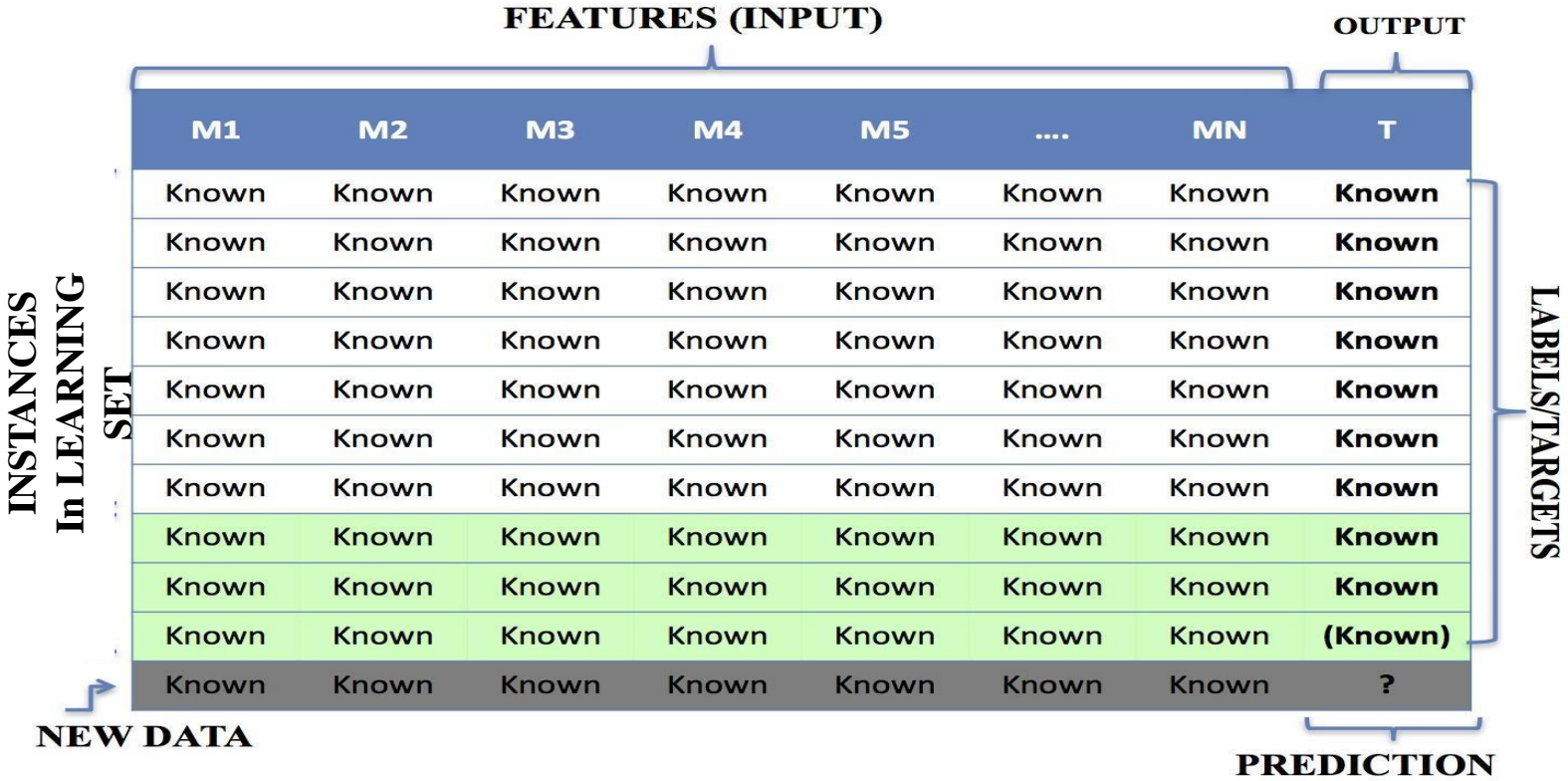
Label or Target is the property that we want to predict (if it exists)

Instances (or examples) are the objects in our data set

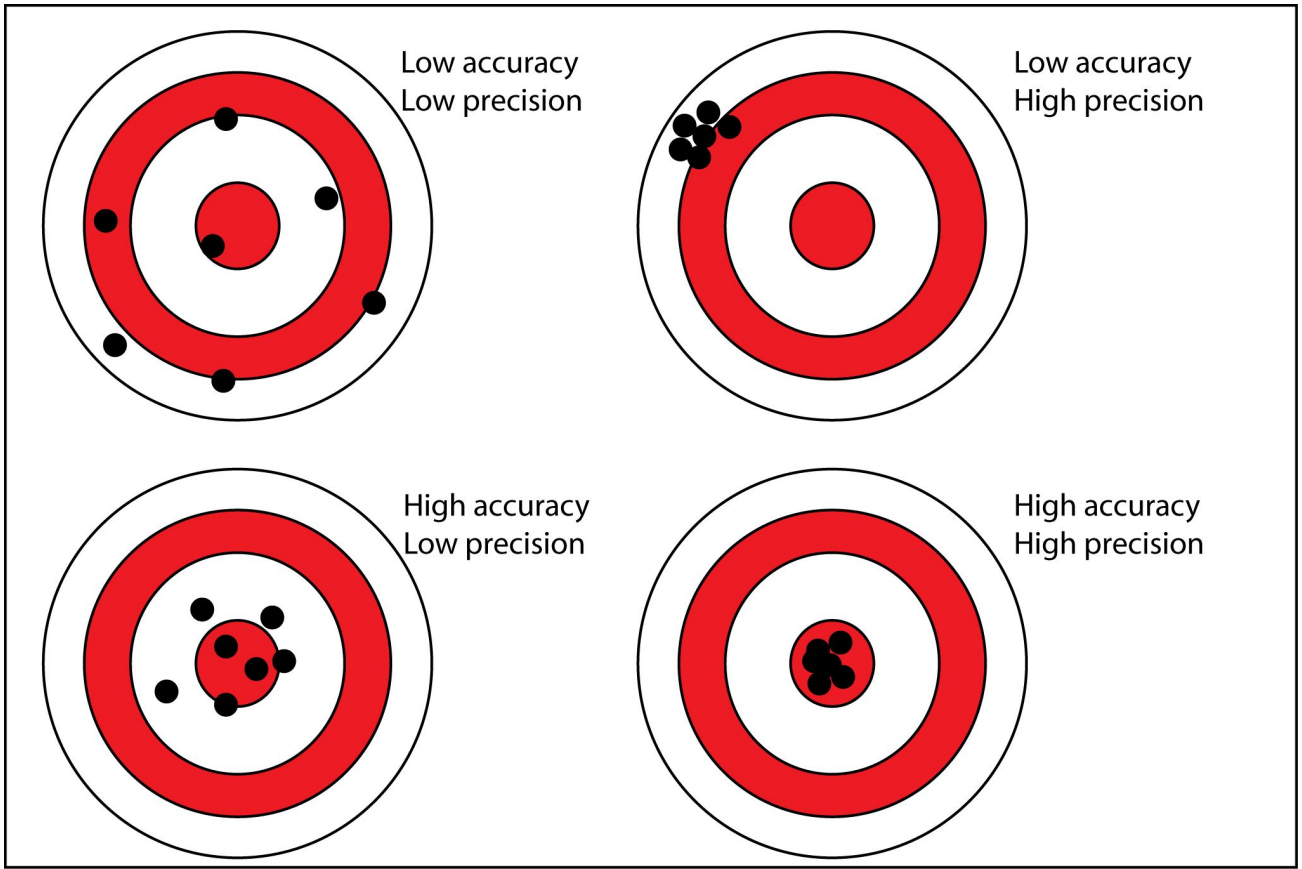


In this case…?

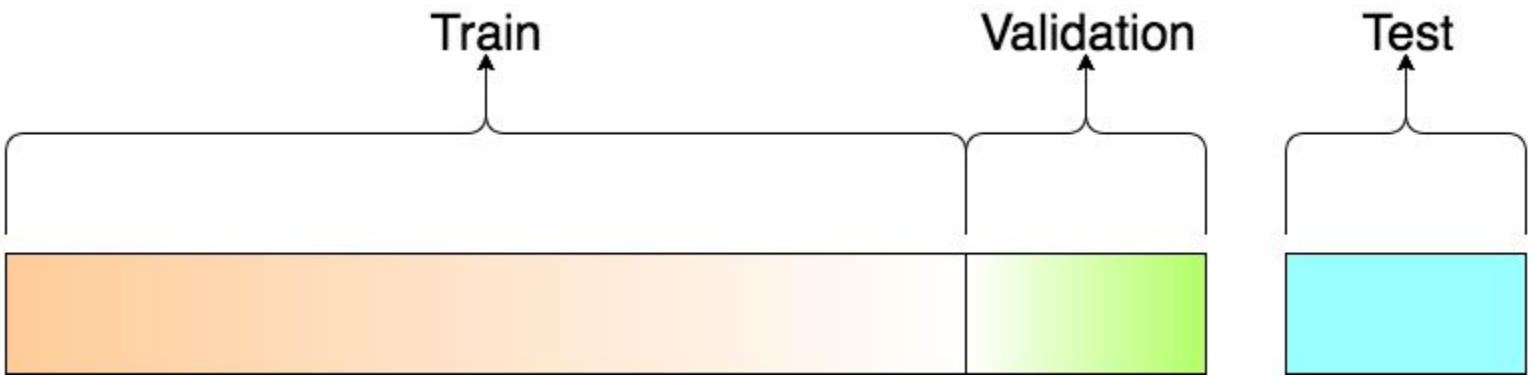
SUPERVISED LEARNING VISUAL SUMMARY



Accuracy and precision



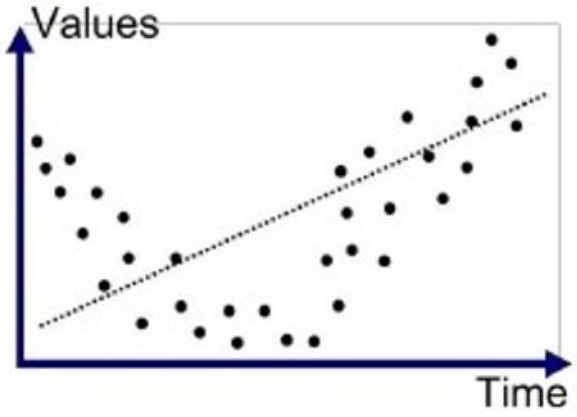
Validation vs testing sets



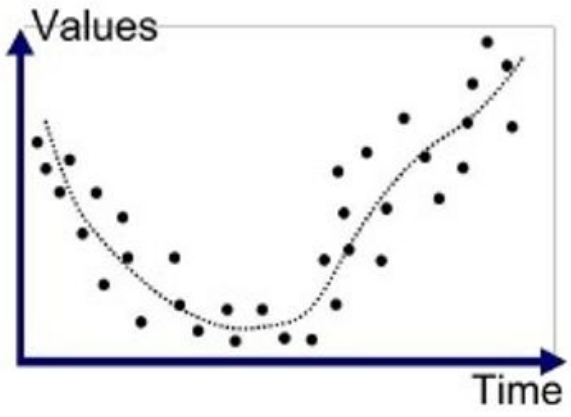
- Test set should never be seeing while training.
- Test set should be independent from training and validation set. Avoid leakage.
- Evaluate the model on the validation set often. Keep the model with the best validation loss.
- All datasets should be representative of the underlying data distribution

Training, validation, and testing sets

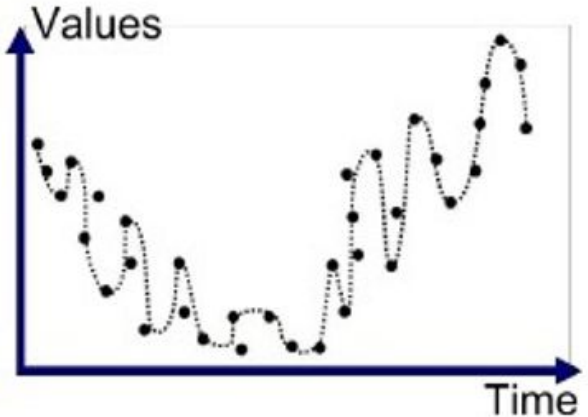
$$Y = f(X)$$



Underfitted



Good Fit/Robust



Overfitted

Training, validation, and testing sets