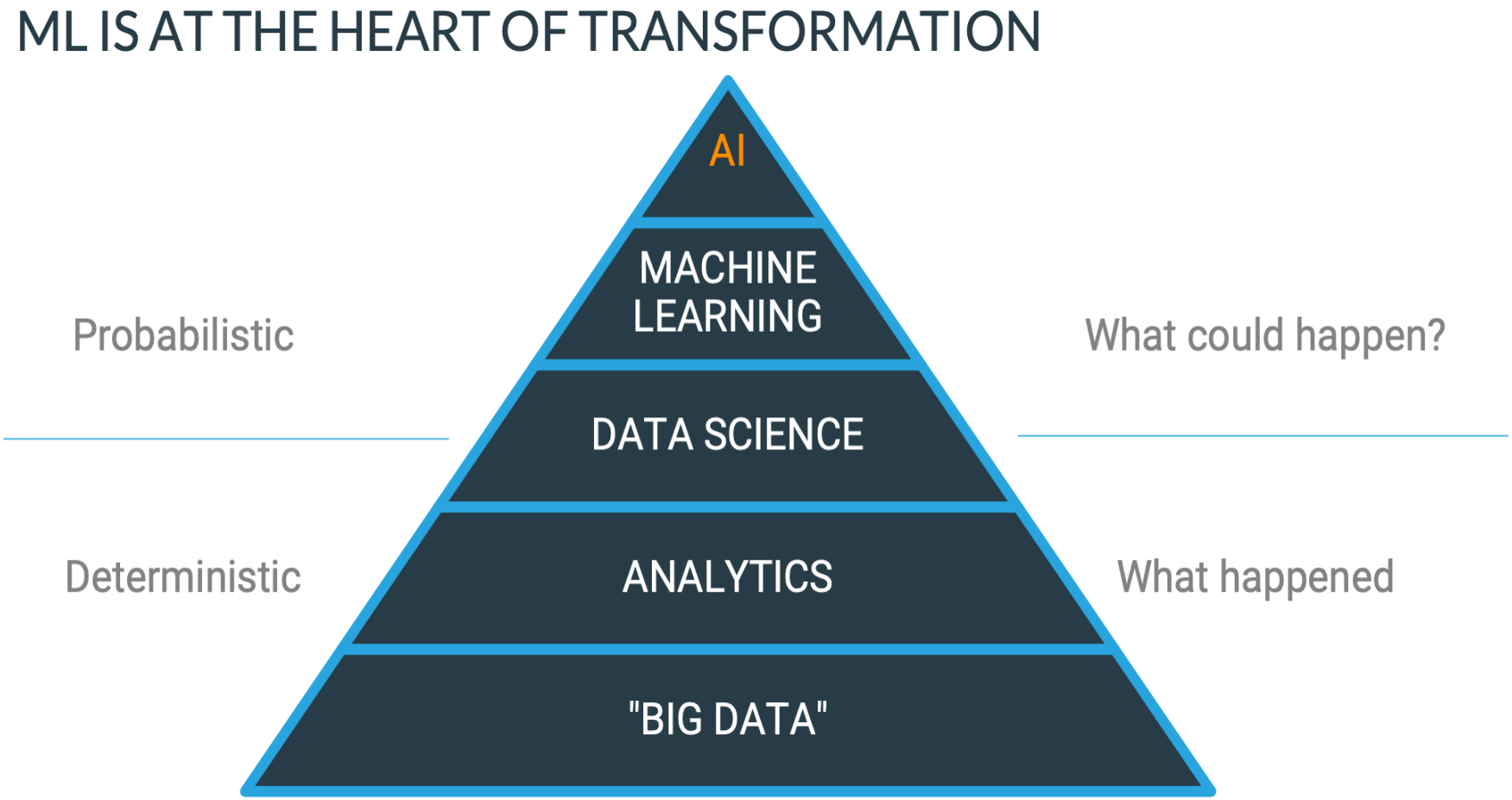
Machine Learning

is

the science (and art) of programming computers so they can *learn from data*.

Week 1: Introduction to ML

What is ML


ML IS AT THE HEART OF TRANSFORMATION
AI
MACHINE LEARNING
DATA SCIENCE
ANALYTICS
"BIG DATA"
Probabilistic
What could happen?
Deterministic
What happened

Machine Learning & AI

- **Artificial Intelligence:** Design an intelligent agent that perceives its environment and make decisions to maximize chances of achieving its goal.

  Subfields: Robotics, Machine Learning, Natural Language processing

- **Machine Learning:** Give "computer the ability to learn without being explicitly programmed."

  (Author Samuel, 1959)

Why Learning

Machine Learning is used when:

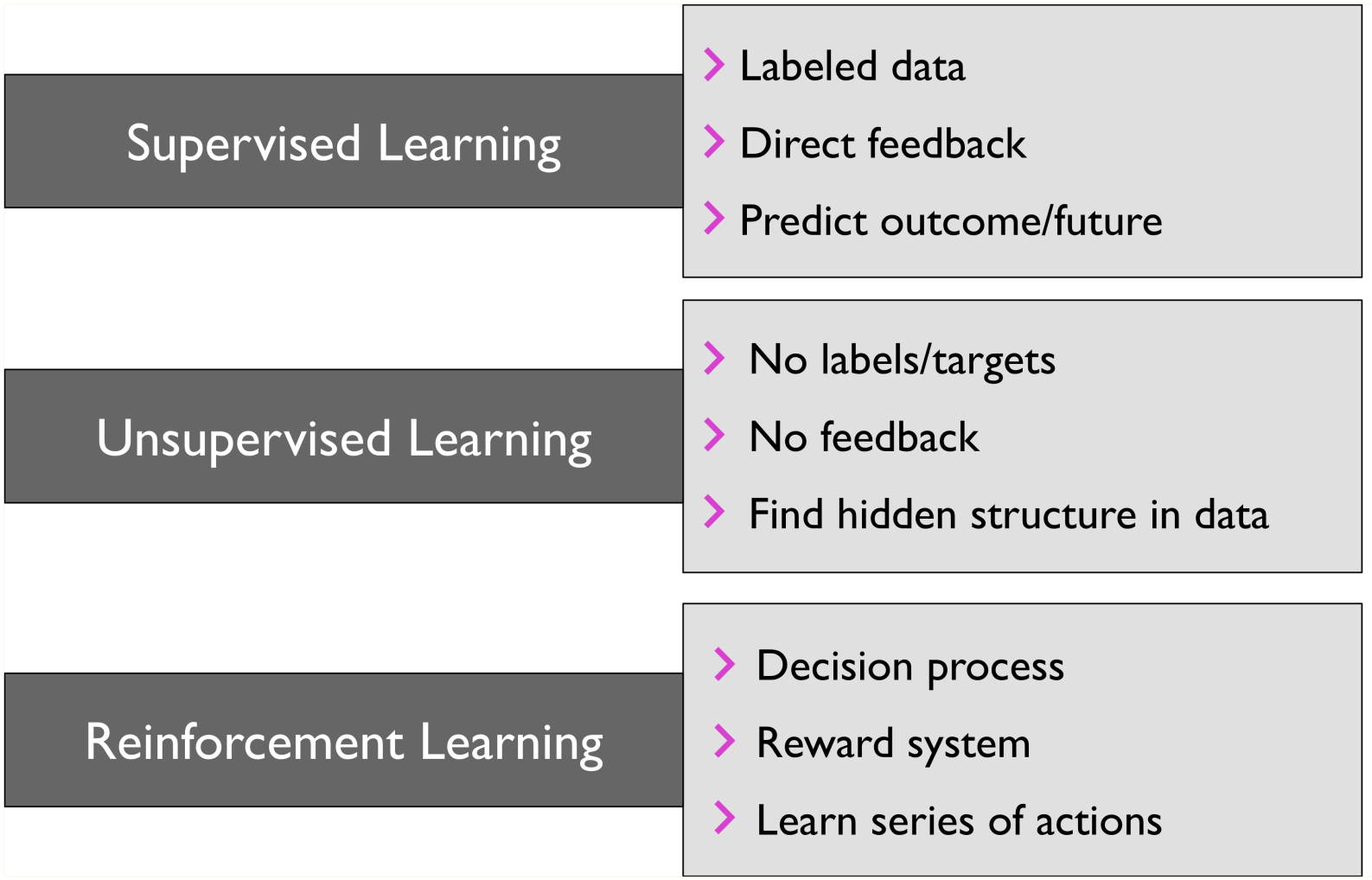
- Human expertise does not exist

  (e.g. navigating on Mars)
- Humans are unable to explain their expertise (e.g. speech recognition)
- Solution changes in time

  (e.g. routing on a computer network. self-driving cars)

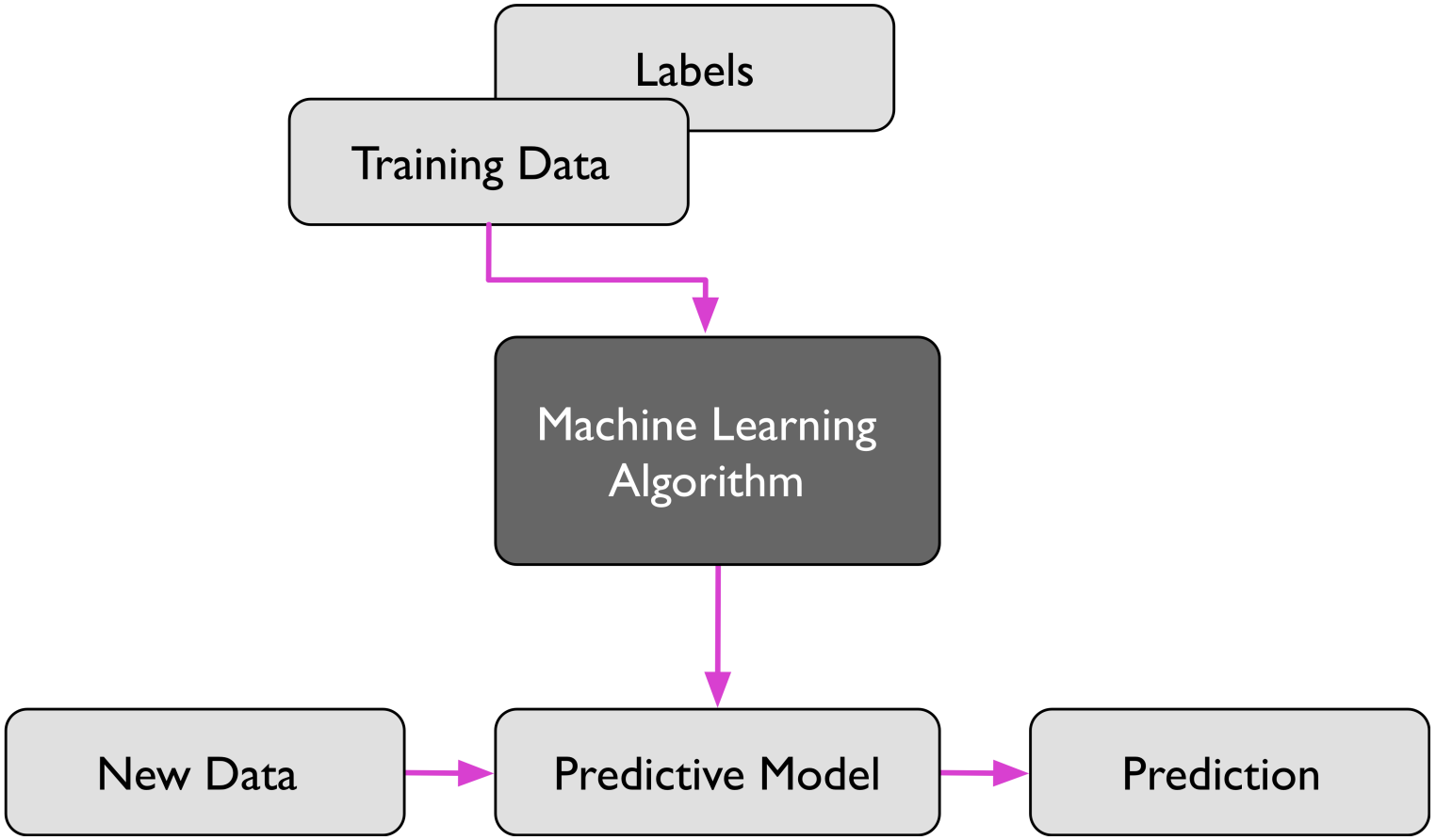
IBM Deep Blue defeated the world chess champion Garry Kasparov May 1997. Rule-based system.

Professional Go player Lee Sedol reviewing his math with AlphaGo after defeat March 2016. Atlantic


Supervised Learning
Labeled data
Direct feedback
Predict outcome/future
Unsupervised Learning
No labels/targets
No feedback
Find hidden structure in data
Reinforcement Learning
Decision process
Reward system
Learn series of actions

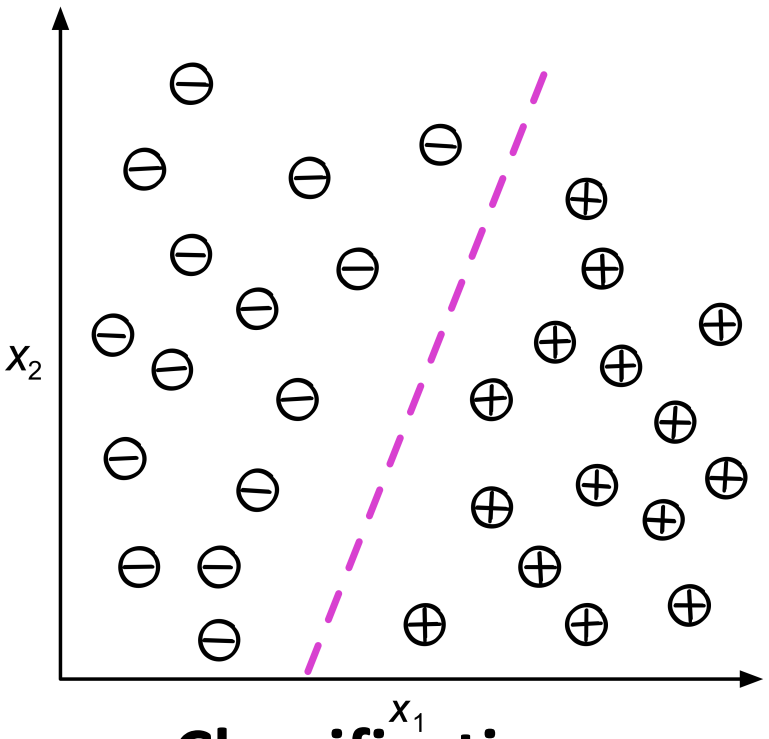
Supervised Learning

After sufficient training, a supervised technique provides target for new inputs

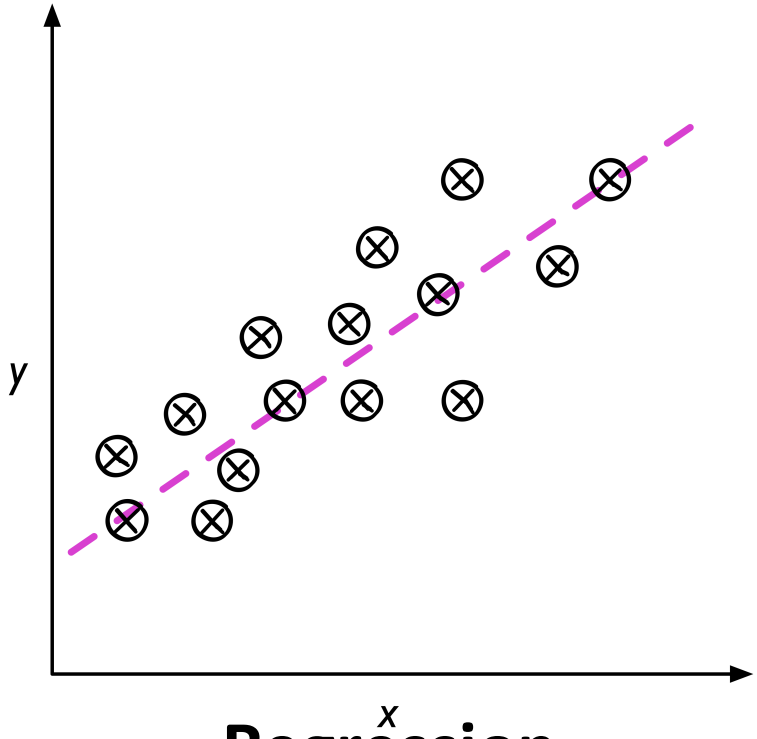


Supervised Learning

- Knowledge of outputs
- Labeled data



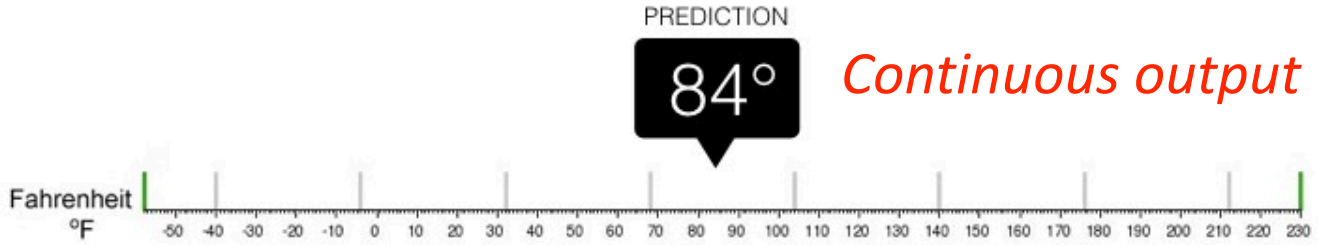
Classification



Regression

Regression

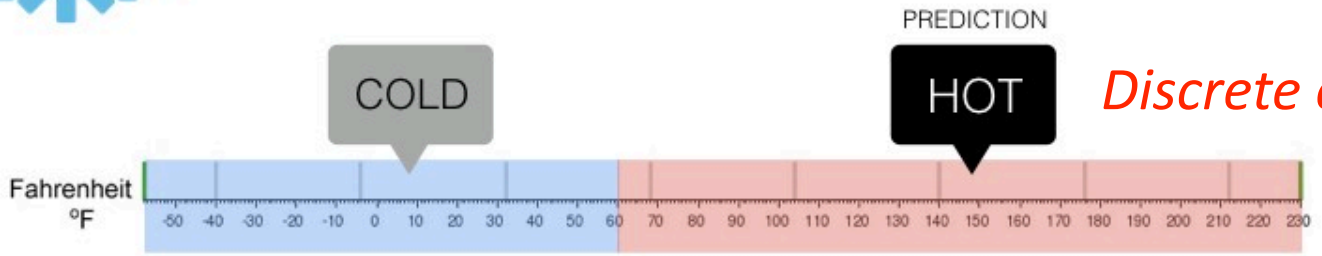
What is the temperature going to be tomorrow?



Continuous output

Classification

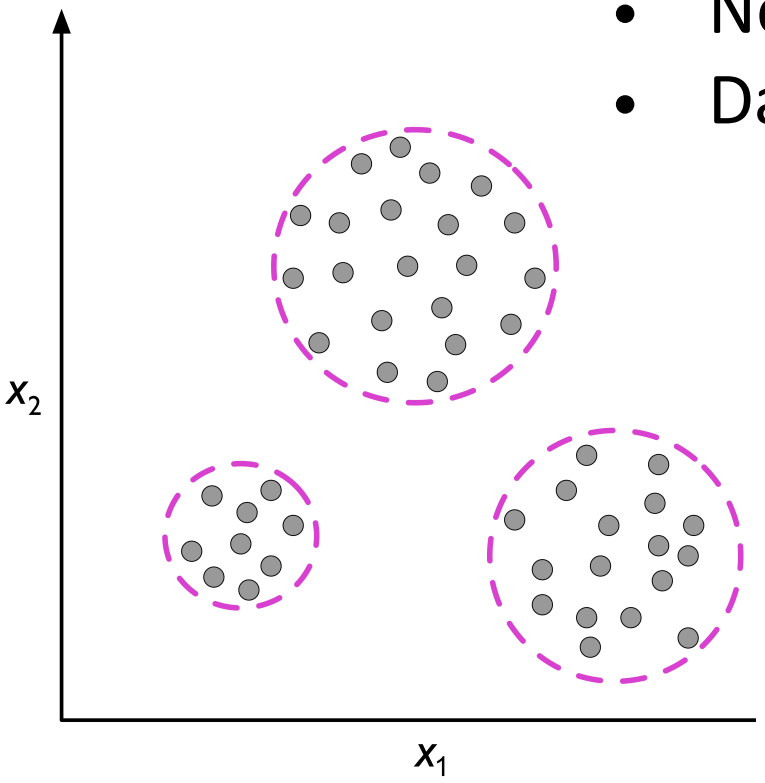
Will it be Cold or Hot tomorrow?



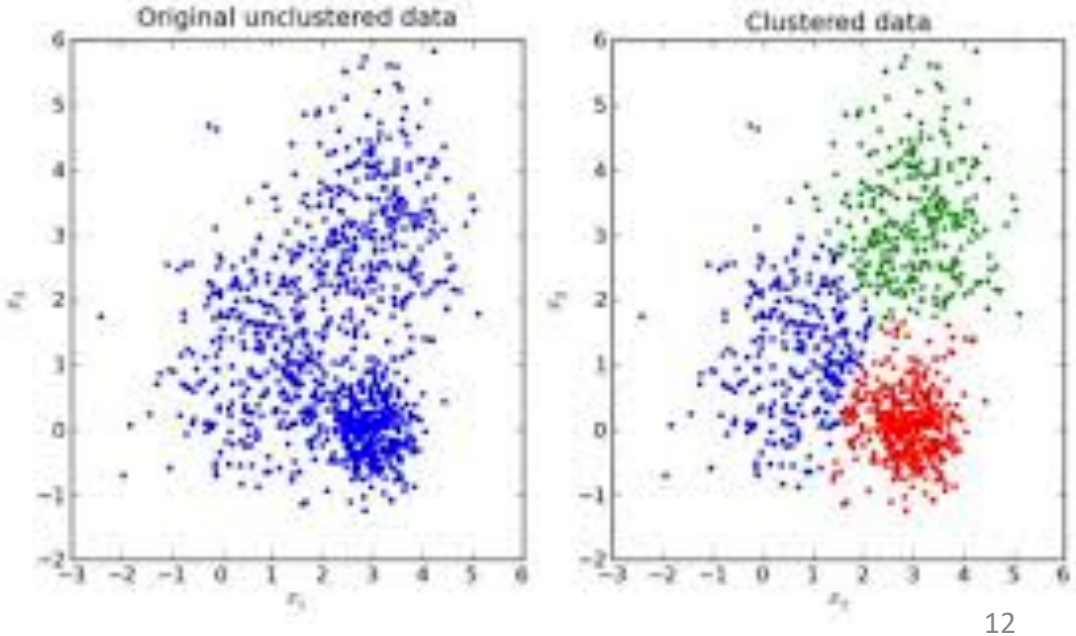
Discrete output

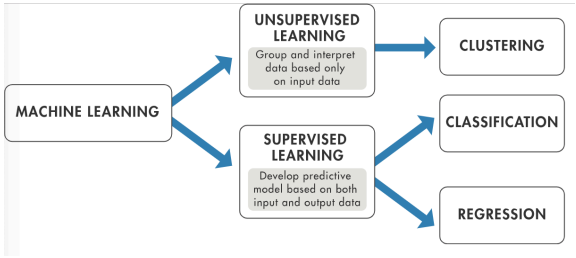
Unsupervised Learning

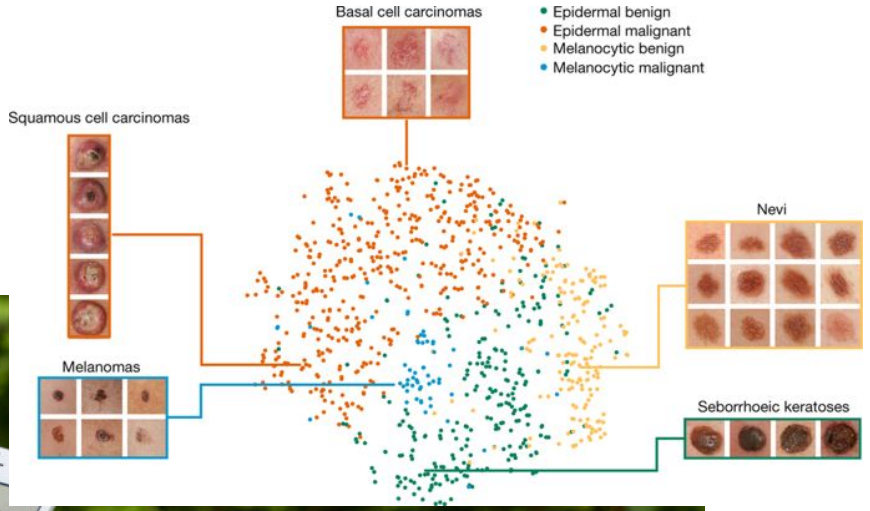
- No knowledge of outputs
- Data is unlabeled



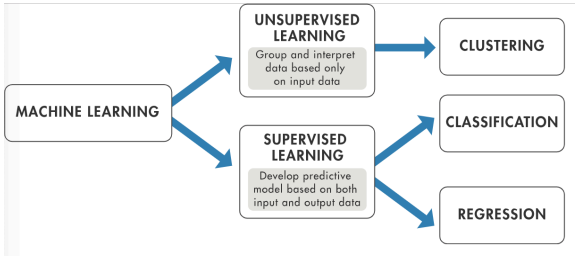
Clustering algorithm



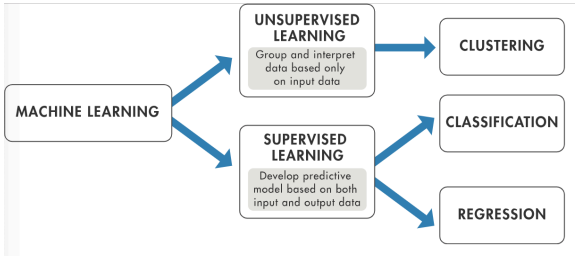




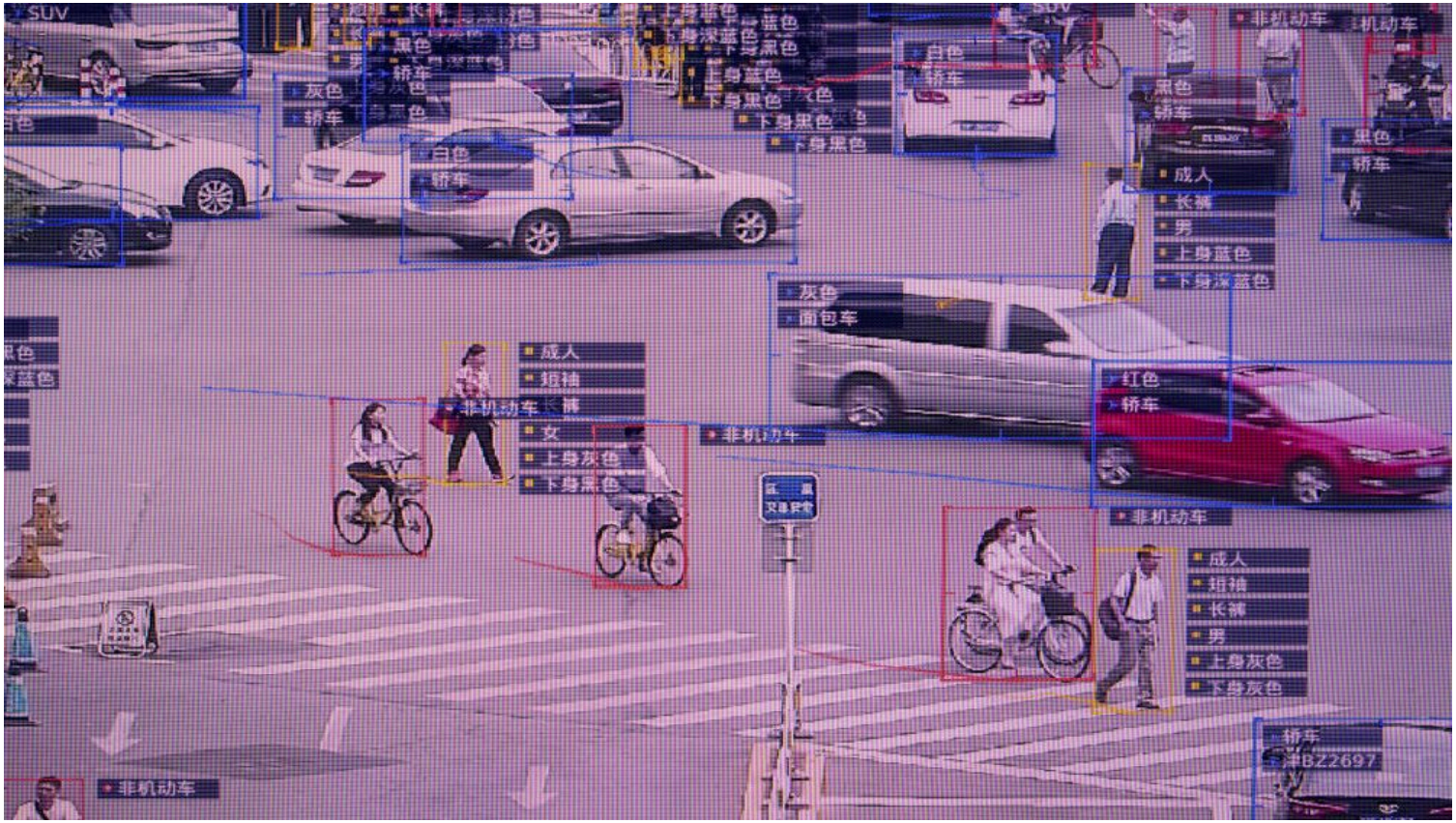
Machine learning makes drug discovery and patient diagnosis quicker, cheaper and more effective. E.g. Dermatologist-level detection of skin cancer with deep neural networks. Nature



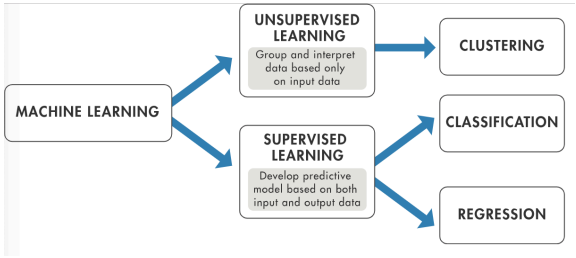
More than 80% of TV shows on Netflix are found through its recommendation engine based on machine learning to identify viewer's preferences.



Camera Surveillance

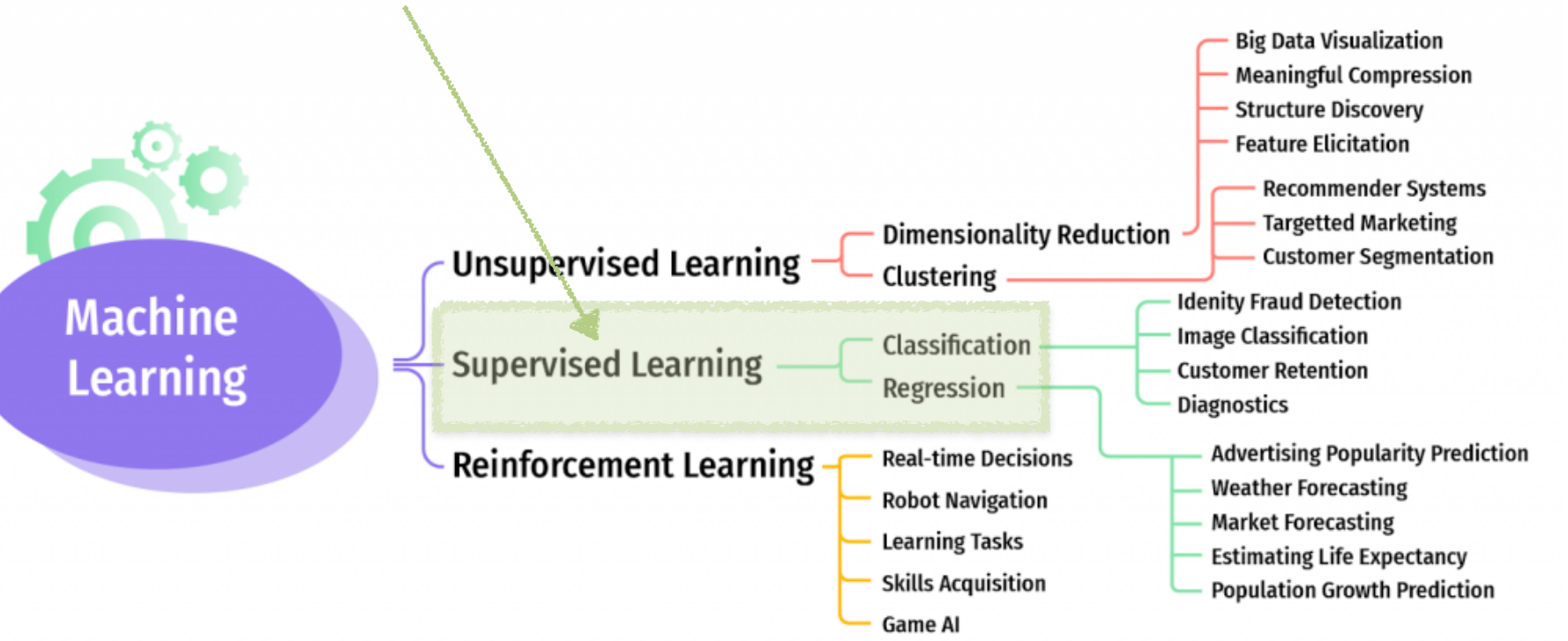


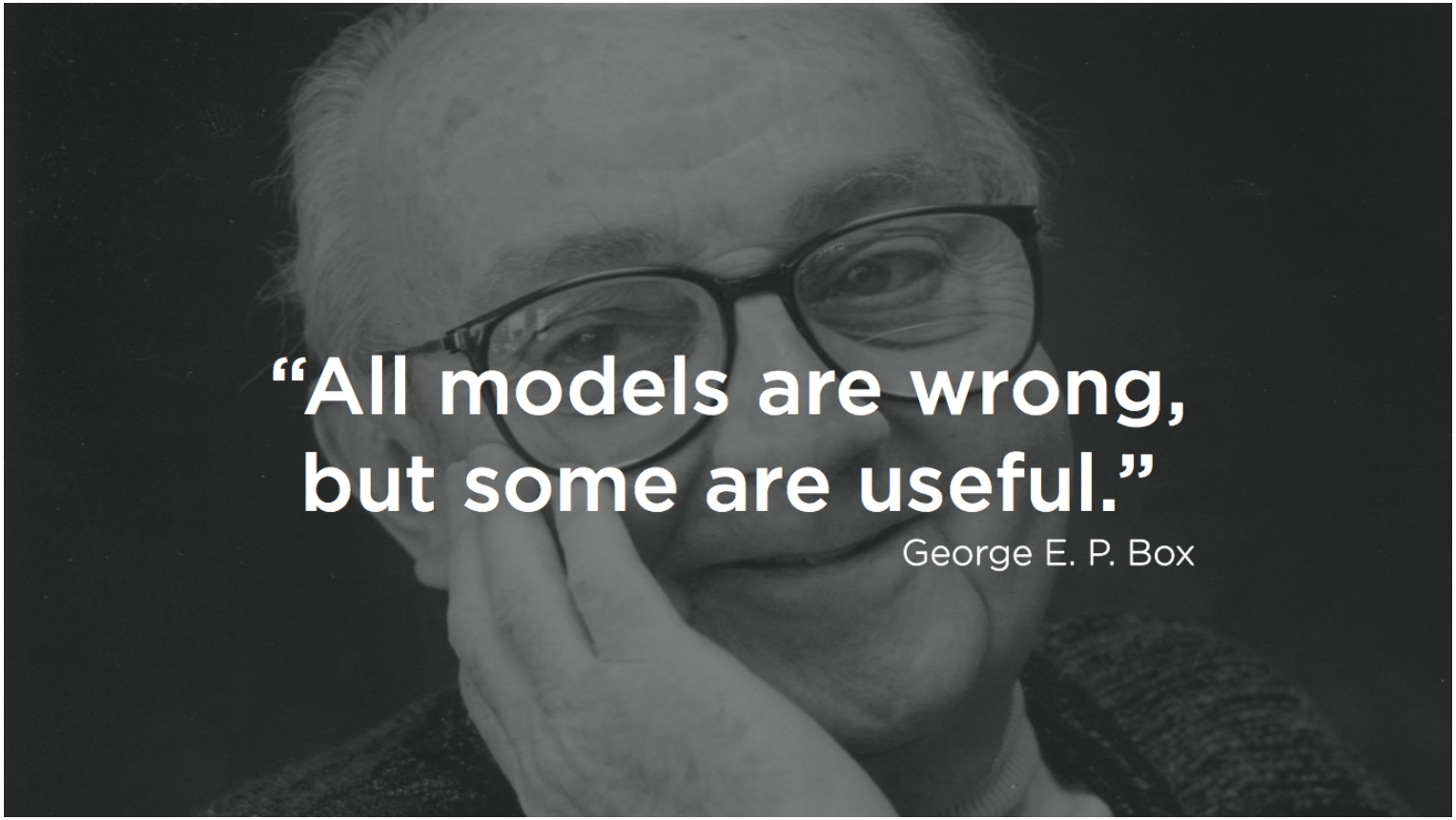
China's mass surveillance system, coupled with facial recognition technology to build a social credit system.



Predict house pricing based on previously house sale data.

Our course focus




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

This week's discussion question:

"It's not who has the best algorithm that wins; It's who has the most data."

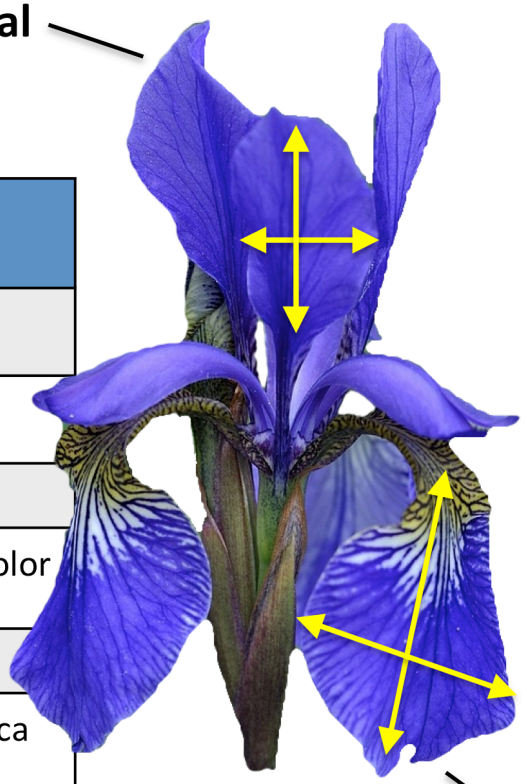
Fun ML Demo

- https://experiments.withgoogle.com/drum-machine
- https://experiments.withgoogle.com/quick-draw
- https://teachablemachine.withgoogle.com

Basic Terminology

Samples
(instances, observations)

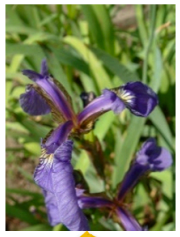
	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
			...		
50	6.4	3.5	4.5	1.2	Versicolor
			...		
150	5.9	3.0	5.0	1.8	Virginica



Features **(X)**
(attributes, measurements, dimensions)

Class labels ***(y)***
(targets)

Terminology & Notation



	A	B	C	D	E
1	Sepal Length	Sepal Width	Petal Length	Petal Width	Class
2	5.1	3.5	1.4	0.2	Iris-setosa
3	4.9	3	1.4	0.2	Iris-setosa
4	4.7	3.2	1.3	0.2	Iris-setosa
5	4.6	3.1	1.5	0.2	Iris-setosa
6	5	3.6	1.4	0.2	Iris-setosa
7	5.4	3.9	1.7	0.4	Iris-setosa
8	4.6	3.4	1.4	0.3	Iris-setosa
9	5	3.4	1.5	0.2	Iris-setosa
10	4.4	2.9	1.4	0.2	Iris-setosa
11	4.9	3.1	1.5	0.1	Iris-setosa
12	5.4	3.7	1.5	0.2	Iris-setosa
13	4.8	3.4	1.6	0.2	Iris-setosa
14	4.8	3	1.4	0.1	Iris-setosa
15	4.3	3	1.1	0.1	Iris-setosa
16	5.8	4	1.2	0.2	Iris-setosa
17	5.7	4.4	1.5	0.4	Iris-setosa
18	5.4	3.9	1.3	0.4	Iris-setosa
19	5.1	3.5	1.4	0.3	Iris-setosa
20	5.7	3.8	1.7	0.3	Iris-setosa
21	5.1	3.8	1.5	0.3	Iris-setosa
22	5.4	3.4	1.7	0.2	Iris-setosa
23	5.1	3.7	1.5	0.4	Iris-setosa
24	4.6	3.6	1	0.2	Iris-setosa
25	5.1	3.3	1.7	0.5	Iris-setosa

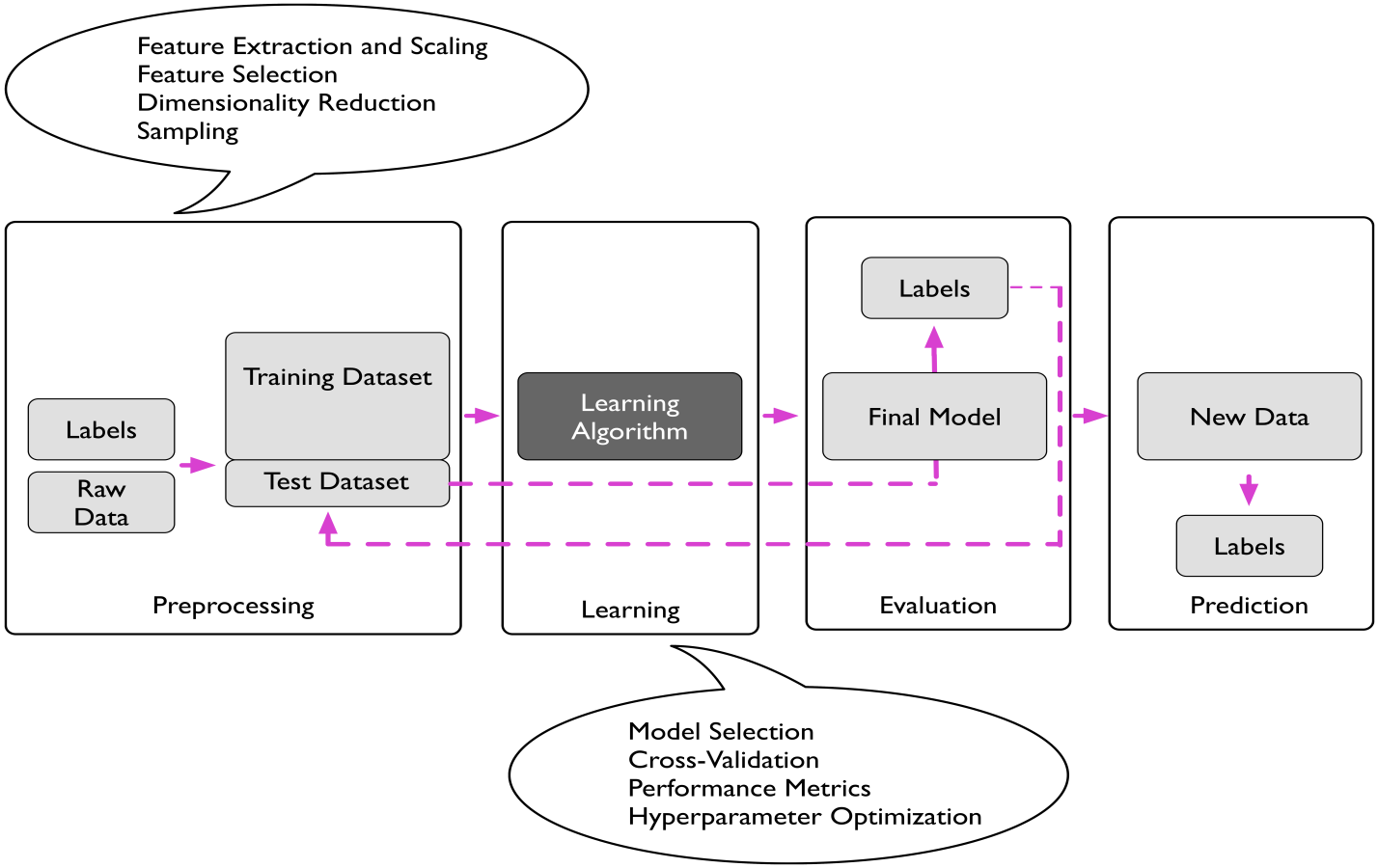
Feature Table	
Feature # 1	**Sepal Length**
Feature # 2	**Sepal Width**
Feature # 3	**Petal Length**
Feature # 4	**Petal Width**

INPUT: The features of the 150 samples can be written as a 150*4 matrix**.**

Target: [Setosa, Versicolor, Virginica]

TARGET: The targets/labels of the 150 samples can be written as a 150-dimensional column vector.

ML Roadmap


Feature Extraction and Scaling
Feature Selection
Dimensionality Reduction
Sampling
Labels
Raw Data
Training Dataset
Test Dataset
Preprocessing
Learning Algorithm
Learning
Labels
Final Model
Evaluation
New Data
Labels
Prediction
Model Selection
Cross-Validation
Performance Metrics
Hyperparameter Optimization

Exercise

- Identify the four ML steps in the first_ml_iris example.
    1. Pre-processing
    2. Learning
    3. Evaluation
    4. Prediction