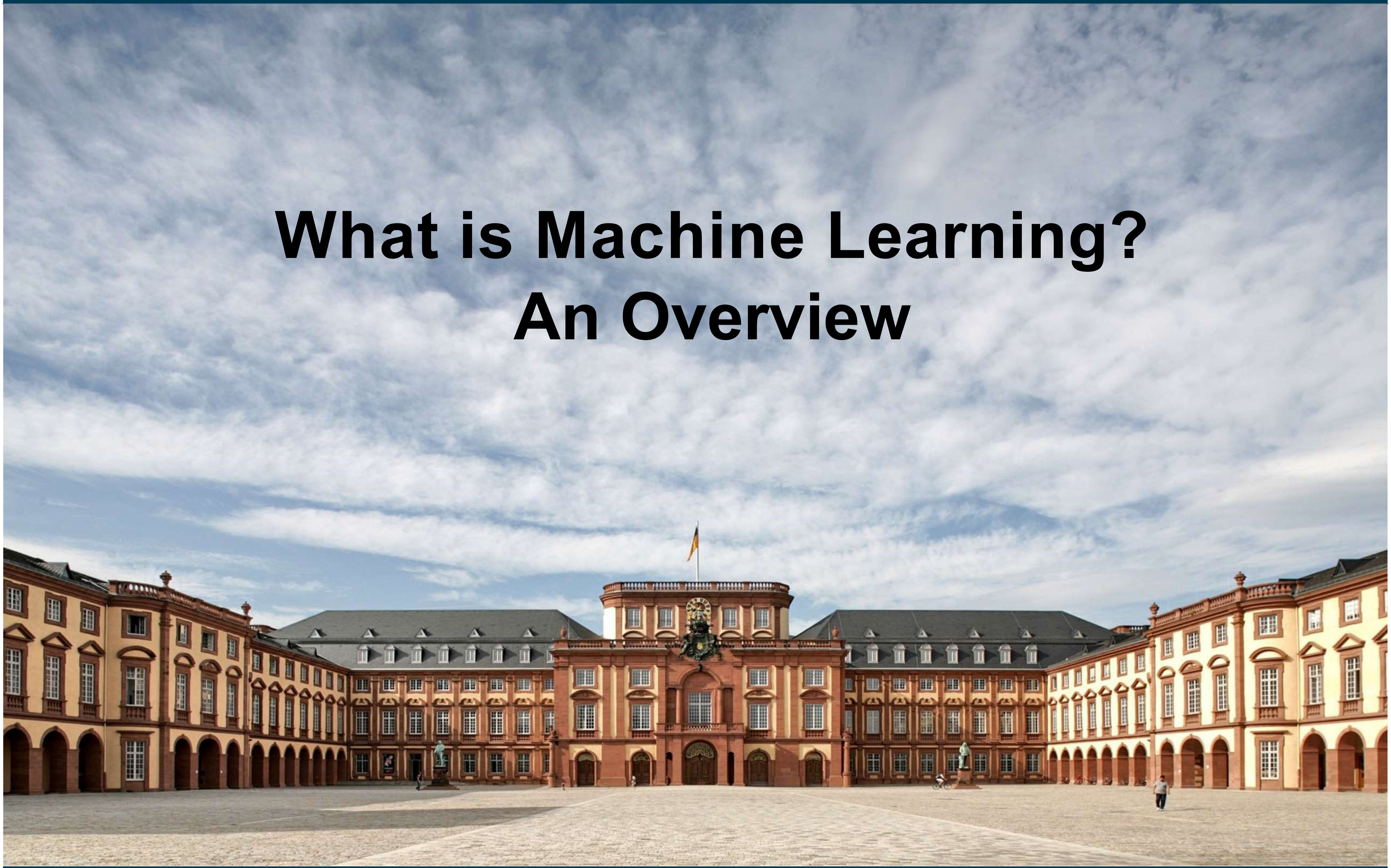
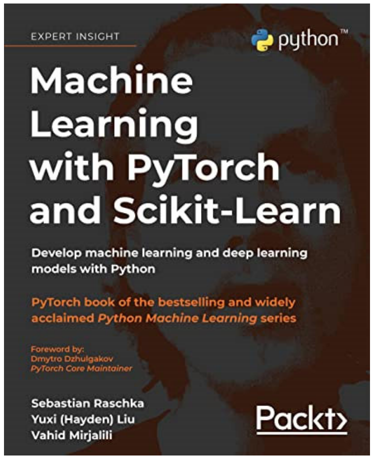
What is Machine Learning?
An Overview

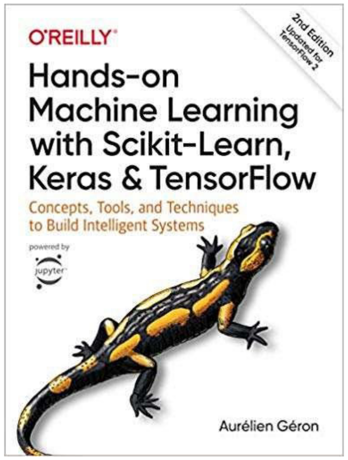


Recommended Books

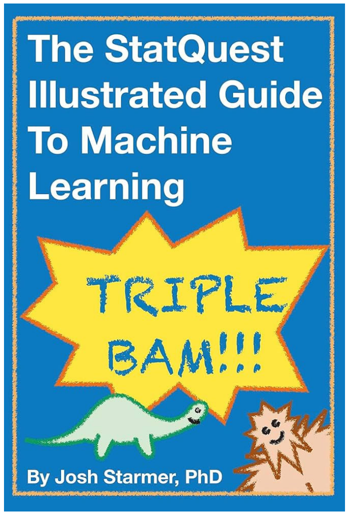
Sebastian Raschka , Yuxi (Hayden) Liu, Vahid Mirjalili:
Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python.
Packt.



Aurélien Géron:
Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow.
2nd or 3rd Edition, O'Reilly, 2019 or 2022



Josh Starmer:
The StatQuest Illustrated Guide To Machine Learning.
StatQuest Publications.



https://www.youtube.com/@statquest YouTube

What is Machine Learning?

"Machine learning is the hot new thing."

-- John L. Hennessy, President of Stanford (2000-2016)

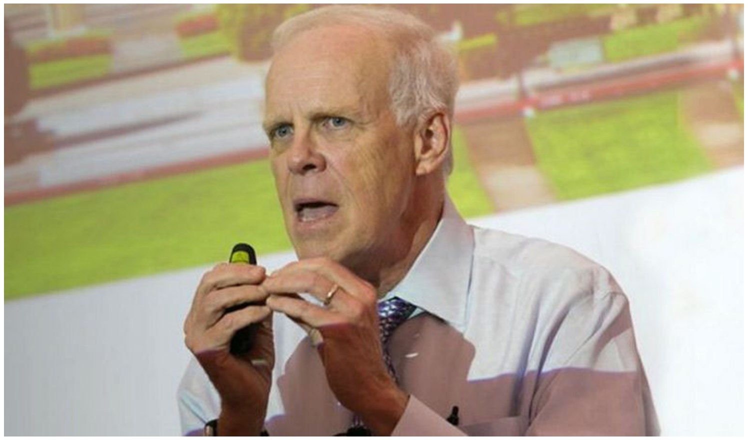


Image Source: https://www.innovateli.com/hennessy-grad-keeps-gifting/

"A breakthrough in machine learning would be worth ten Microsofts"

-- Bill Gates, Microsoft Co-founder

Image source: https://www.gatesnotes.com/Books

[…] machine learning is a subcategory within the field of computer science, which allows you to implement artificial intelligence. So it’s kind of a mechanism to get you to artificial intelligence.

-- Rana el Kaliouby, CEO at Affectiva

Image Source: https://fortune.com/2019/03/08/rana-el-kaliouby-ceo-affectiva/

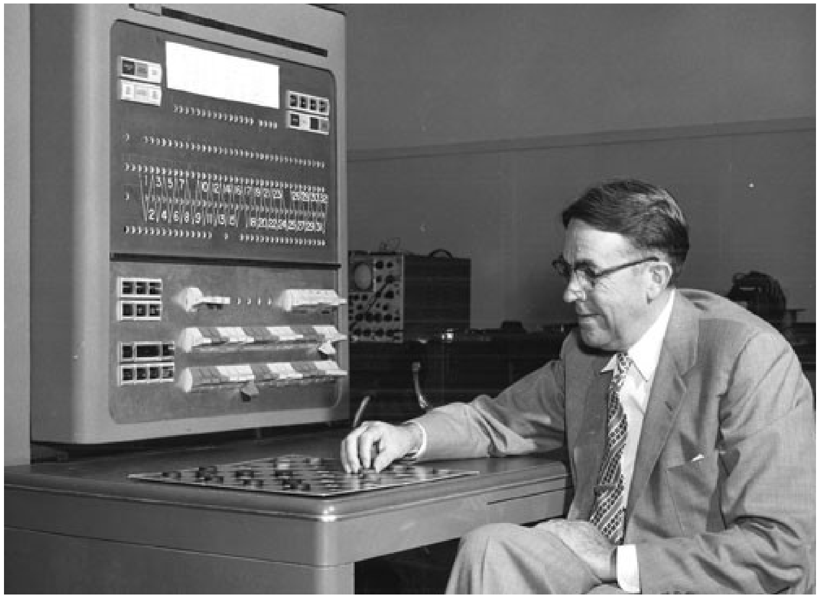


Image Source: https://history-computer.com/ModernComputer/thinkers/images/Arthur-Samuel1.jpg

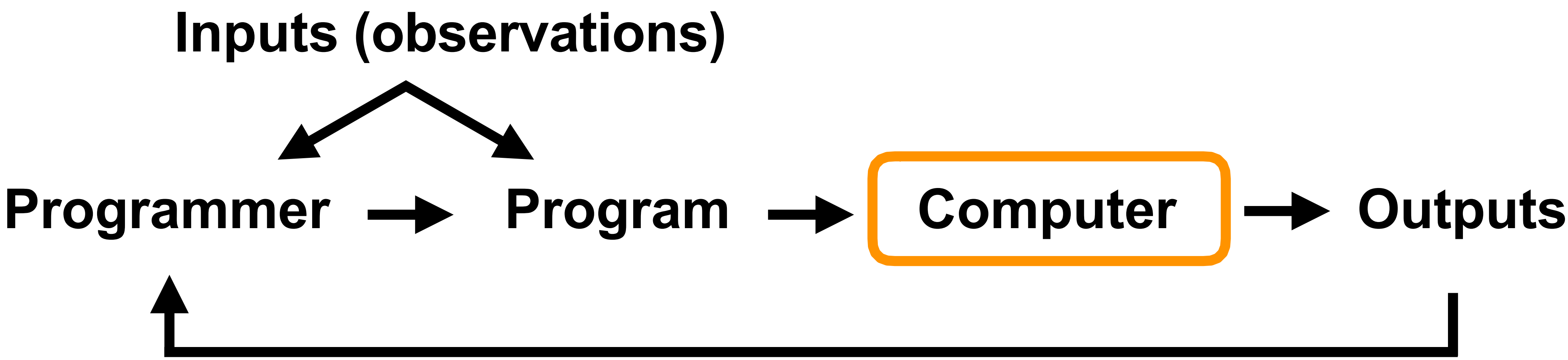
"Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed"

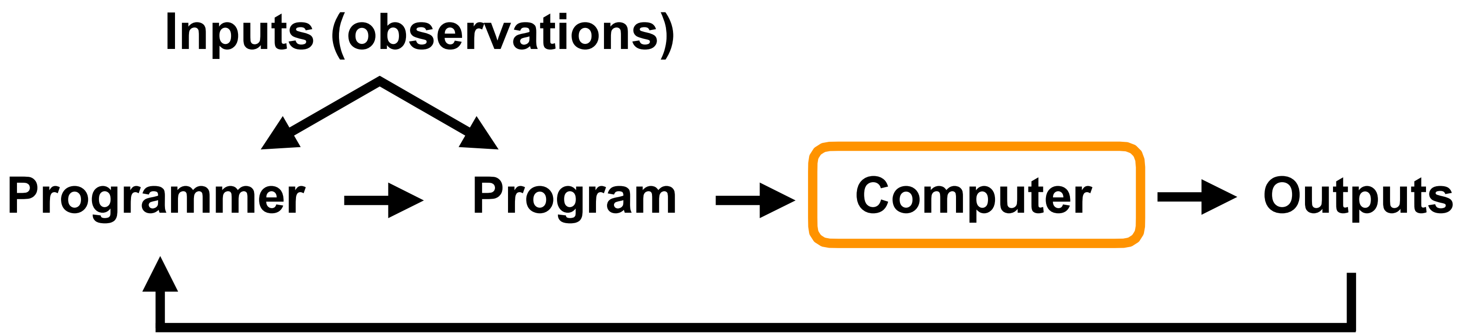
— Arthur L. Samuel, AI pioneer, 1959

(This is likely not an original quote but a paraphrased version of Samuel's sentence "Programming computers to learn from experience should eventually eliminate the need for much of this detailed programming e↵ort.")

Arthur L Samuel. "Some studies in machine learning using the game of checkers". In: *IBM Journal of research and development* 3.3 (1959), pp. 210–229.

The Traditional Programming Paradigm





Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed

— Arthur Samuel (1959)

We will not only use the machines for their intelligence, we will also collaborate with them in ways that we cannot even imagine.

-- Fei Fei Li, Director of Stanford's artificial intelligence lab

Image Source: https://en.wikipedia.org/wiki/Fei-Fei_Li#/media/File:Fei-Fei_Li_at_AI_for_Good_2017.jpg

"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."

— Tom Mitchell, Professor at Carnegie Mellon University

Tom M Mitchell et al. "Machine learning. 1997". In: *Burr Ridge, IL: McGraw Hill* 45.37 (1997), pp. 870–877.

"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."

— Tom Mitchell, Professor at Carnegie Mellon University

Handwriting Recognition Example:

- Task T: r?
- Performance measure P : ?
- Training experience E: ?

"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."

— Tom Mitchell, Professor at Carnegie Mellon University

Handwriting Recognition Example:

- Task T: ***classifying handwritten digits from images***
- Performance measure P *:* ***percentage of digits classified correctly***
- Training experience E**:** ***dataset of digits given classifications, i.e., MNIST***

Some Applications of Machine Learning

- Email spam detection
- Face detection and matching (e.g., iPhone X, Windows laptops, etc.)
- Web search (e.g., DuckDuckGo, Bing, Baidu, Google)
- Sports predictions
- ATMs (e.g., reading checks)
- Credit card fraud • Stock predictions
- Smart assistants (Apple Siri, Amazon Alexa, . . .)
- Product recommendations (e.g., Walmart, Netflix, Amazon)
- Self-driving cars (e.g., Uber, Tesla)
- Language translation (Google translate)
- Sentiment analysis • Drug design
- Medical diagnoses
- . . .

Things to Consider

It is a good exercise to think about how machine learning could be applied in these problem areas or tasks listed above:

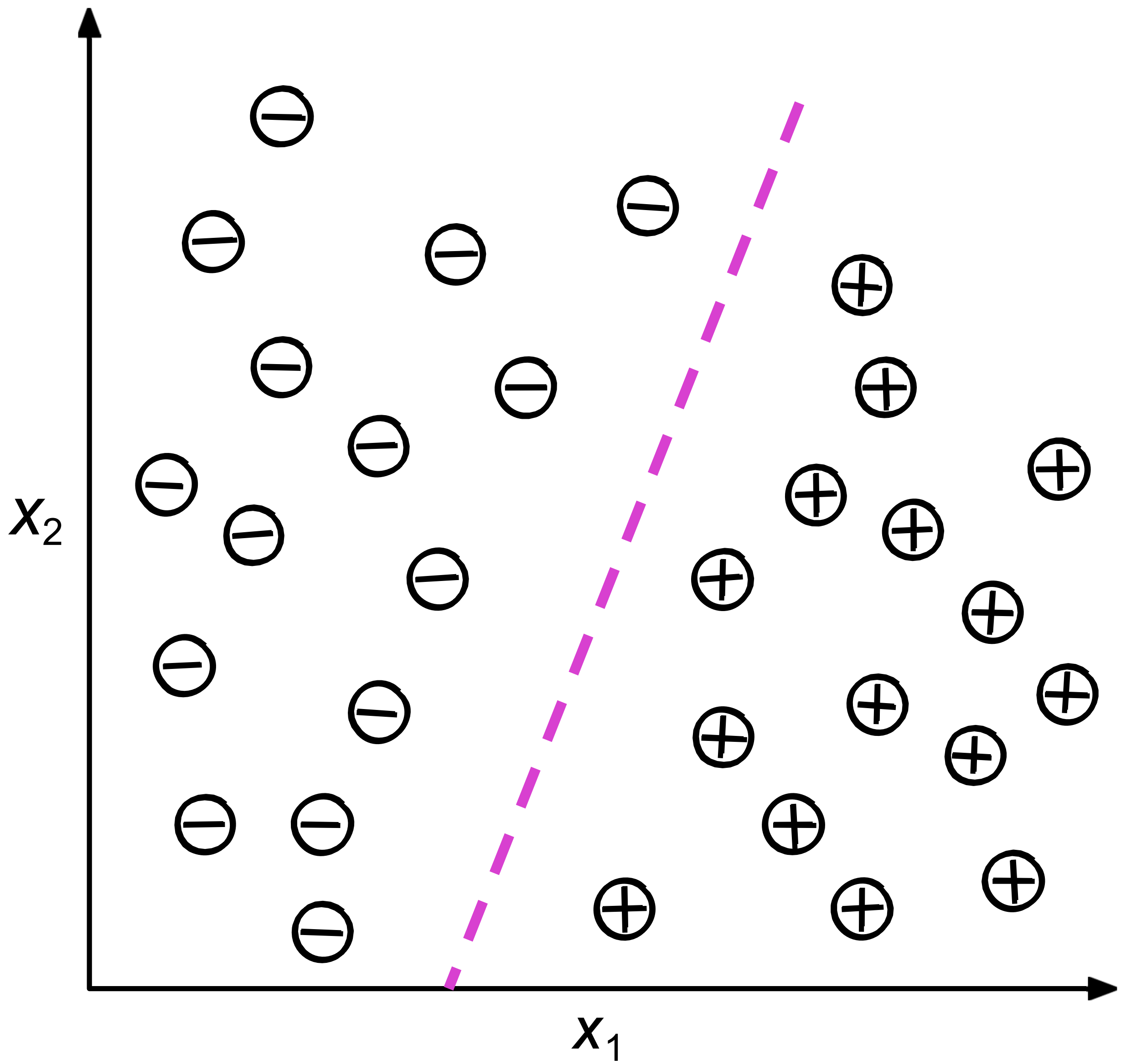
- What is the desired outcome?
- What could the dataset look like?
- Is this a supervised or unsupervised problem, and what algorithms would you use? (Supervised and unsupervised?!)
- How would you measure success?
- What are potential challenges or pitfalls?

Categories of Machine Learning

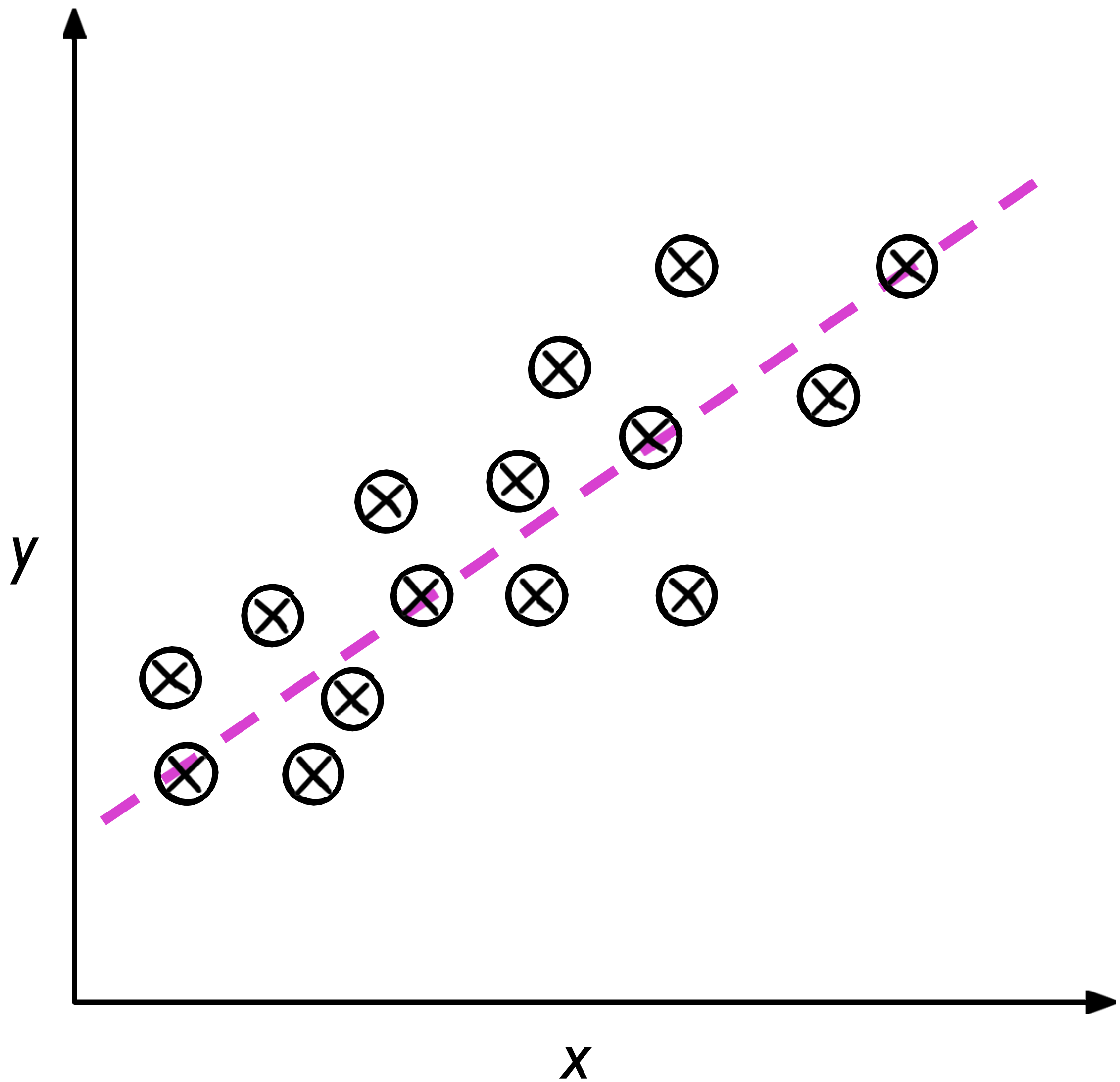
Supervised Learning

- Labeled data
- Direct feedback
- Predict outcome/future

Supervised Learning: Classification



Supervised Learning: Regression



Categories of Machine Learning

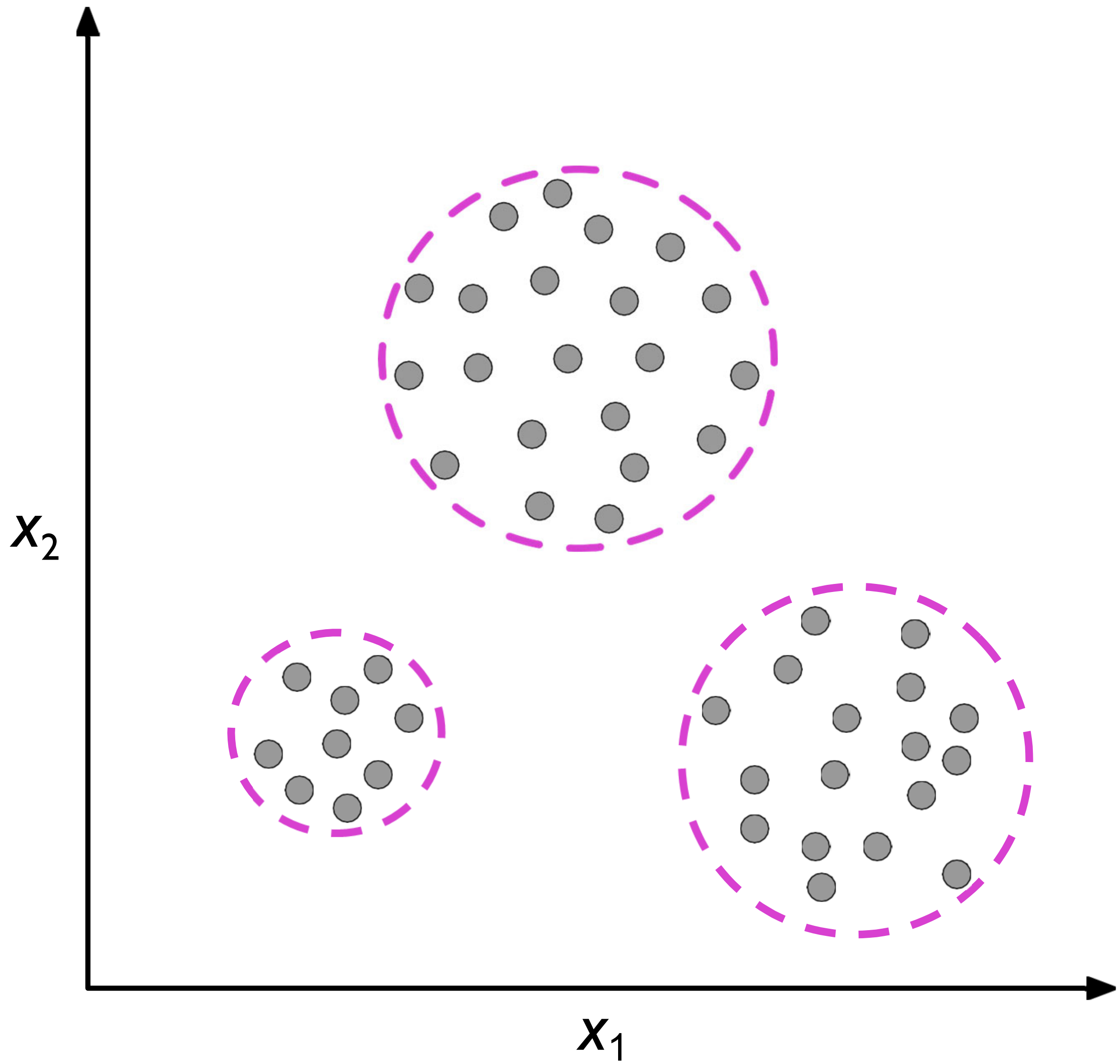
Supervised Learning

- Labeled data
- Direct feedback
- Predict outcome/future

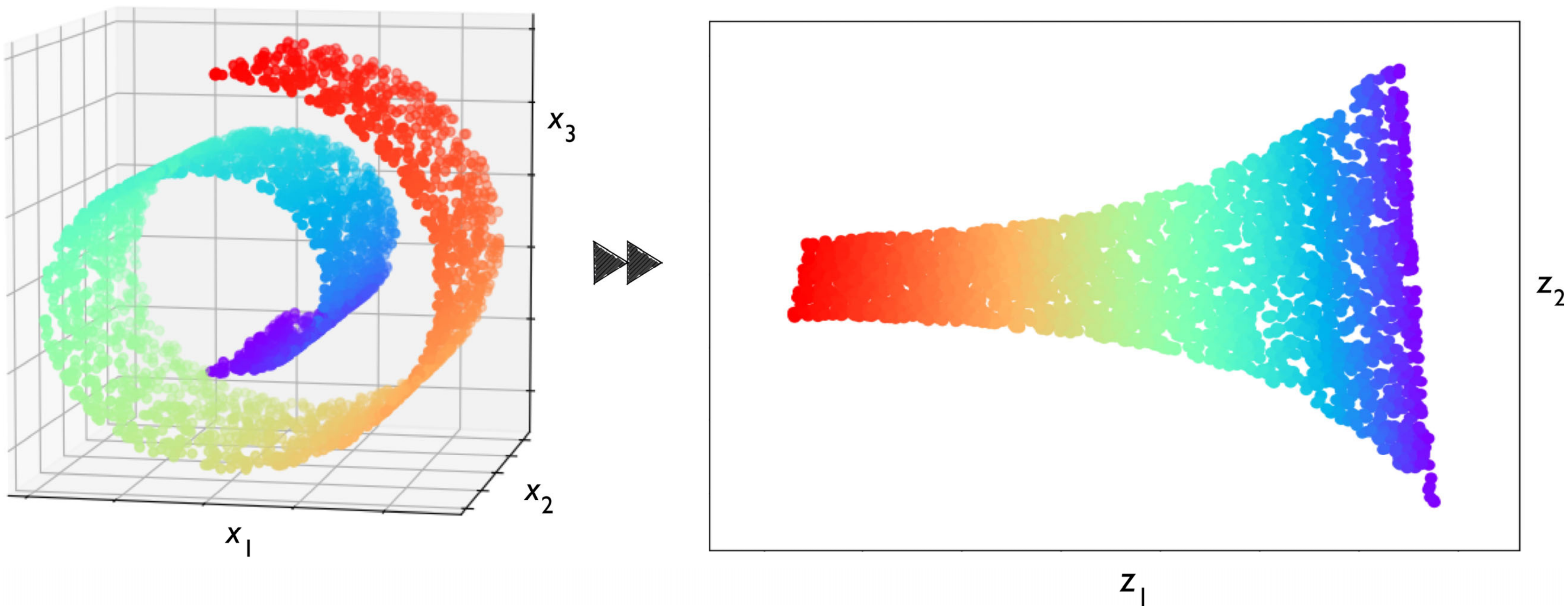
Unsupervised Learning

- No labels/targets
- No feedback
- Find hidden structure in data

Unsupervised Learning -- Clustering



Unsupervised Learning
-- Dimensionality Reduction



Categories of Machine Learning

Supervised Learning

- Labeled data
- Direct feedback
- Predict outcome/future

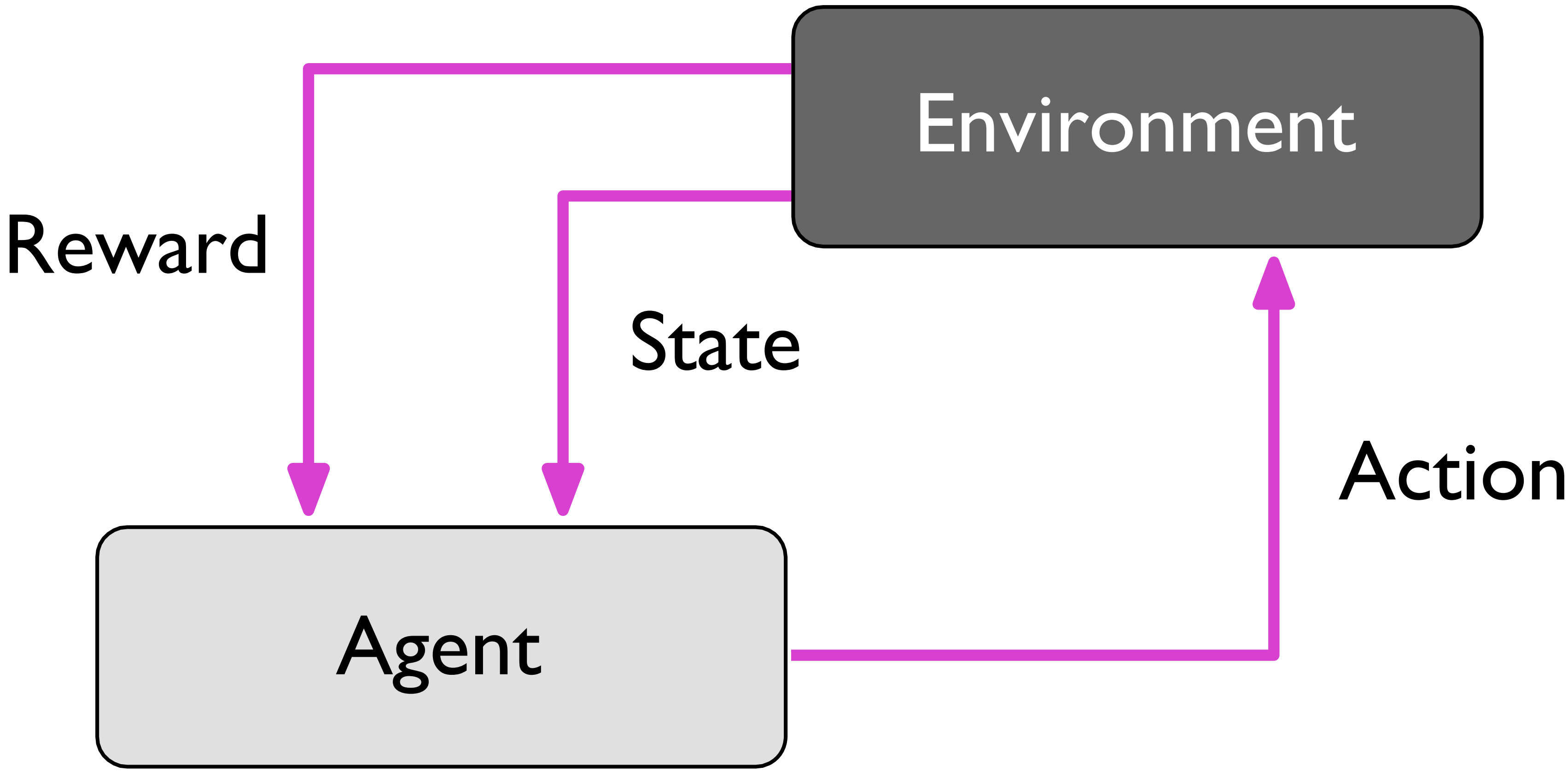
Unsupervised Learning

- No labels/targets
- No feedback
- Find hidden structure in data

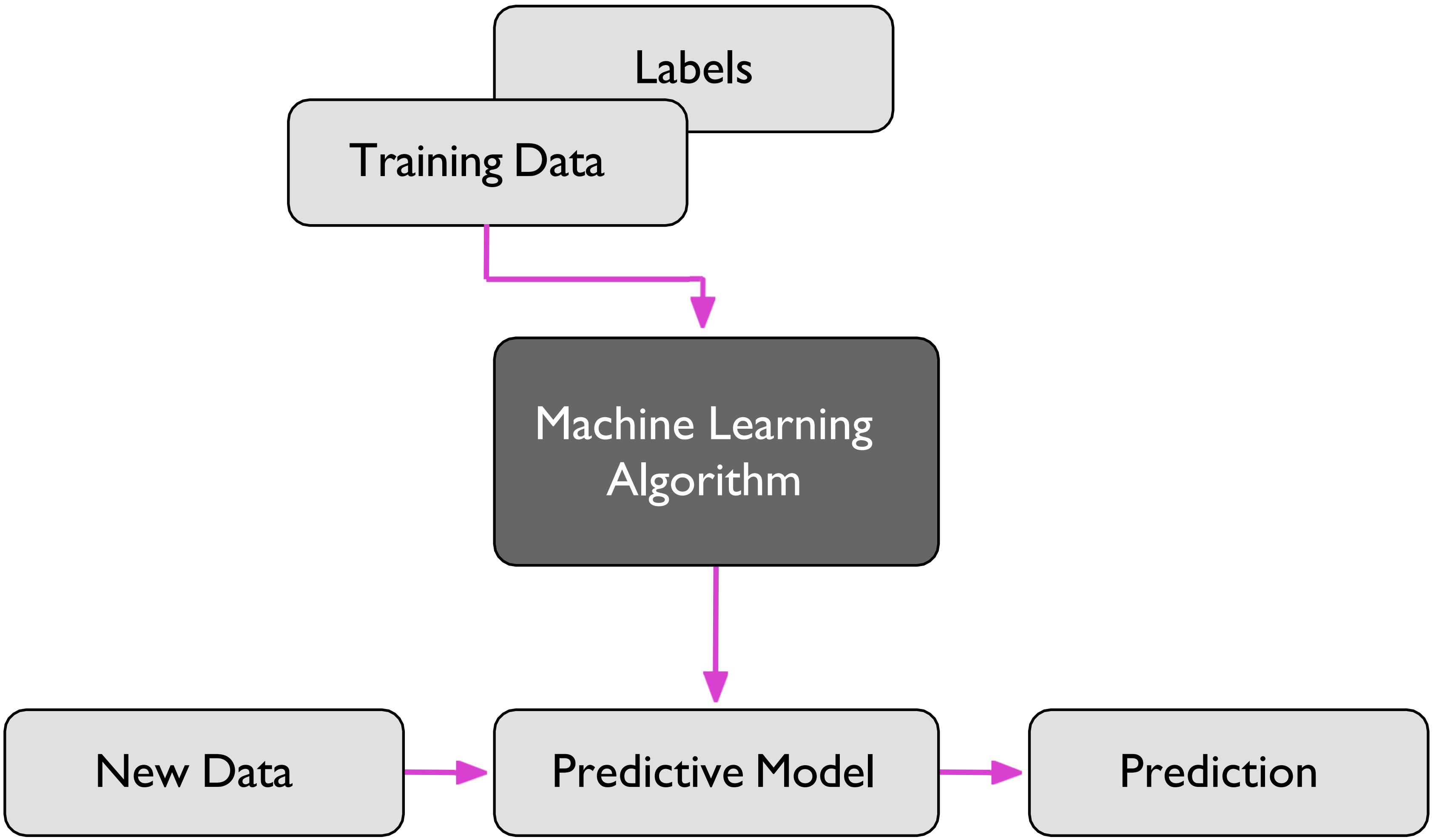
Reinforcement Learning

- Decision process
- Reward system
- Learn series of actions

Reinforcement Learning



Supervised Learning Workflow -- Overview



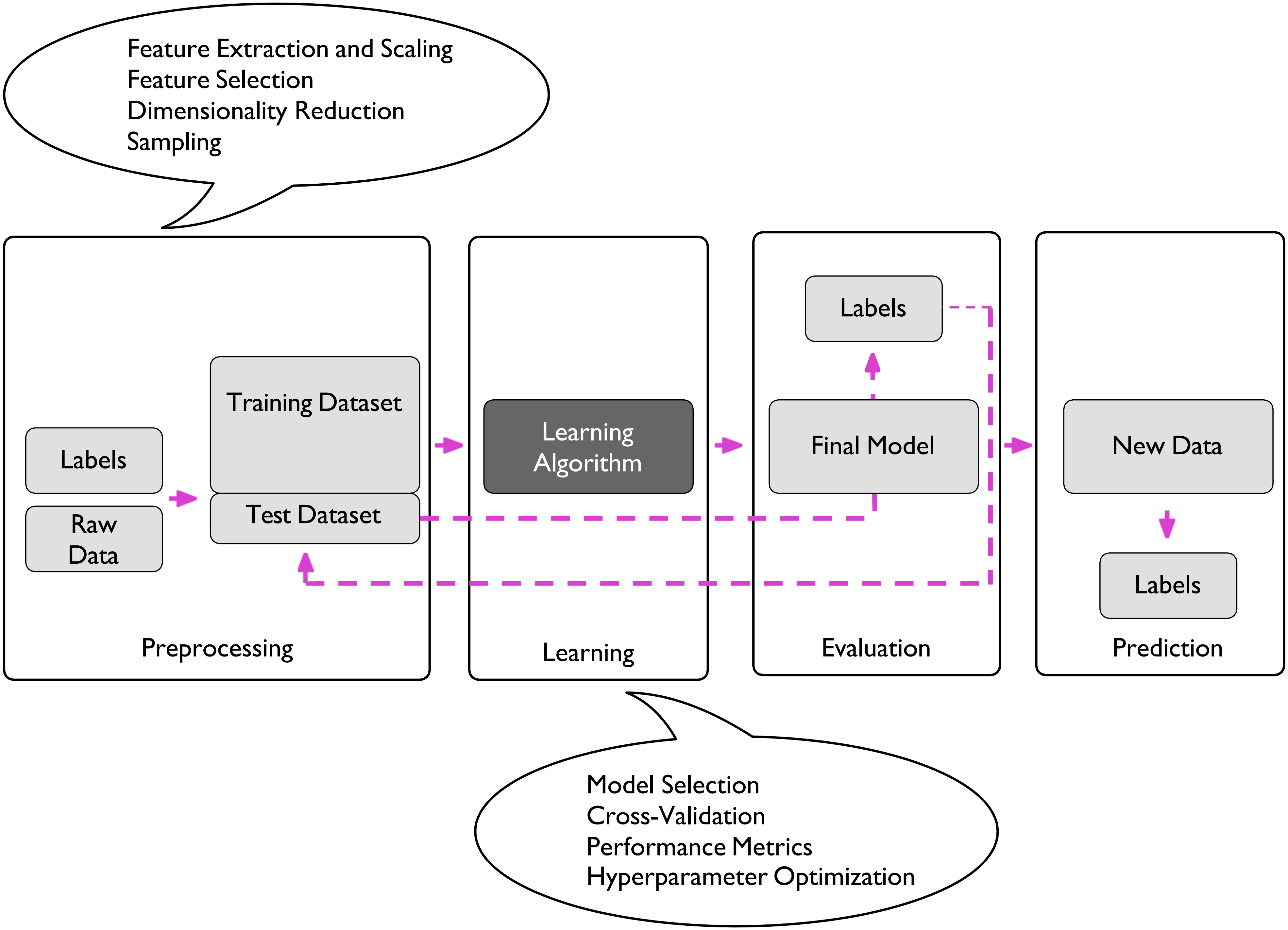
ML Terminology

- **Training example**: A row in the table representing the dataset. Synonymous to an observation, training record, training instance, training sample (in some contexts, sample refers to a collection of training examples)

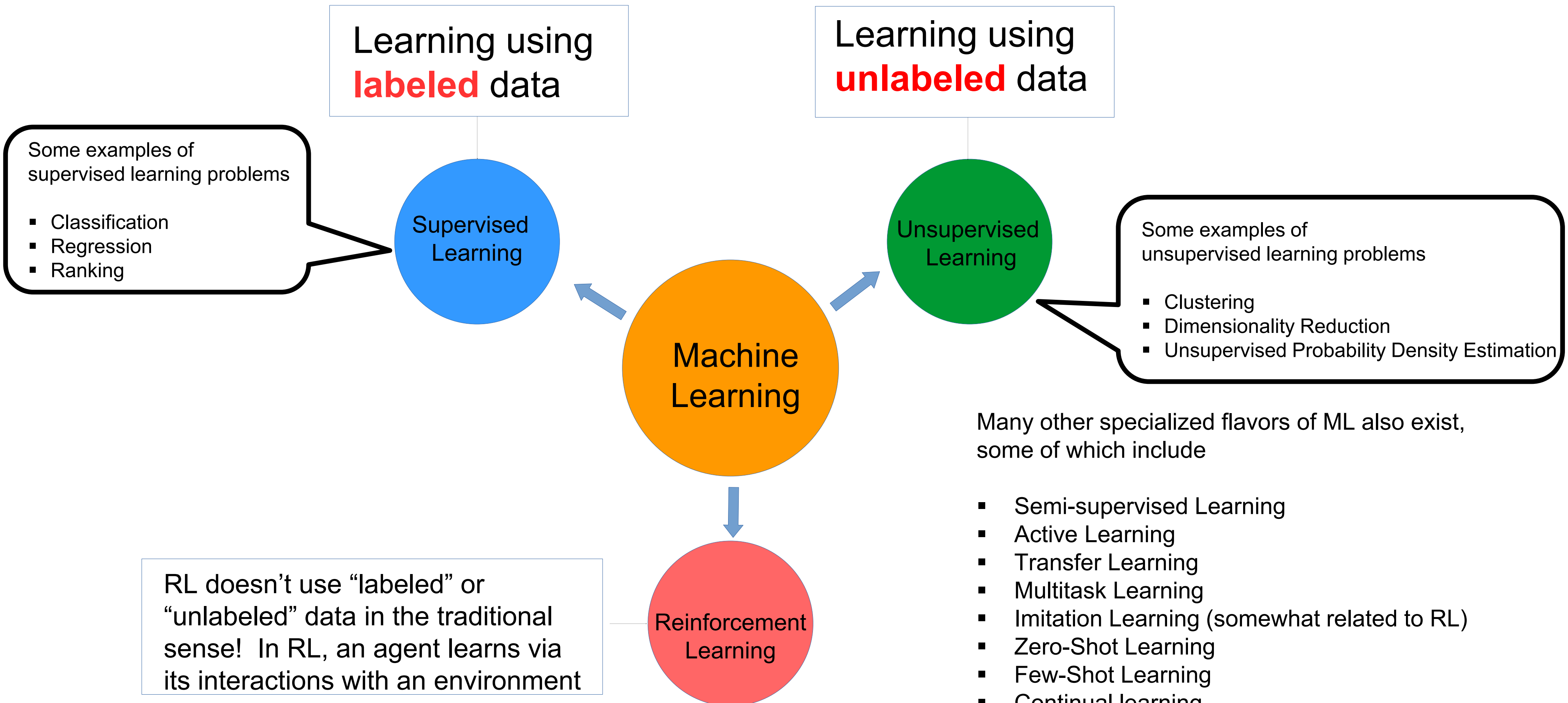
- **Feature**: a column in the table representing the dataset. Synonymous to predictor, variable, input, attribute, covariate.

- **Targets**: What we want to predict. Synonymous to outcome, output, ground truth, response variable, dependent variable, (class) label (in classification).

- **Output / prediction**: use this to distinguish from targets; here, means output from the model.


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

A Loose Taxonomy of ML

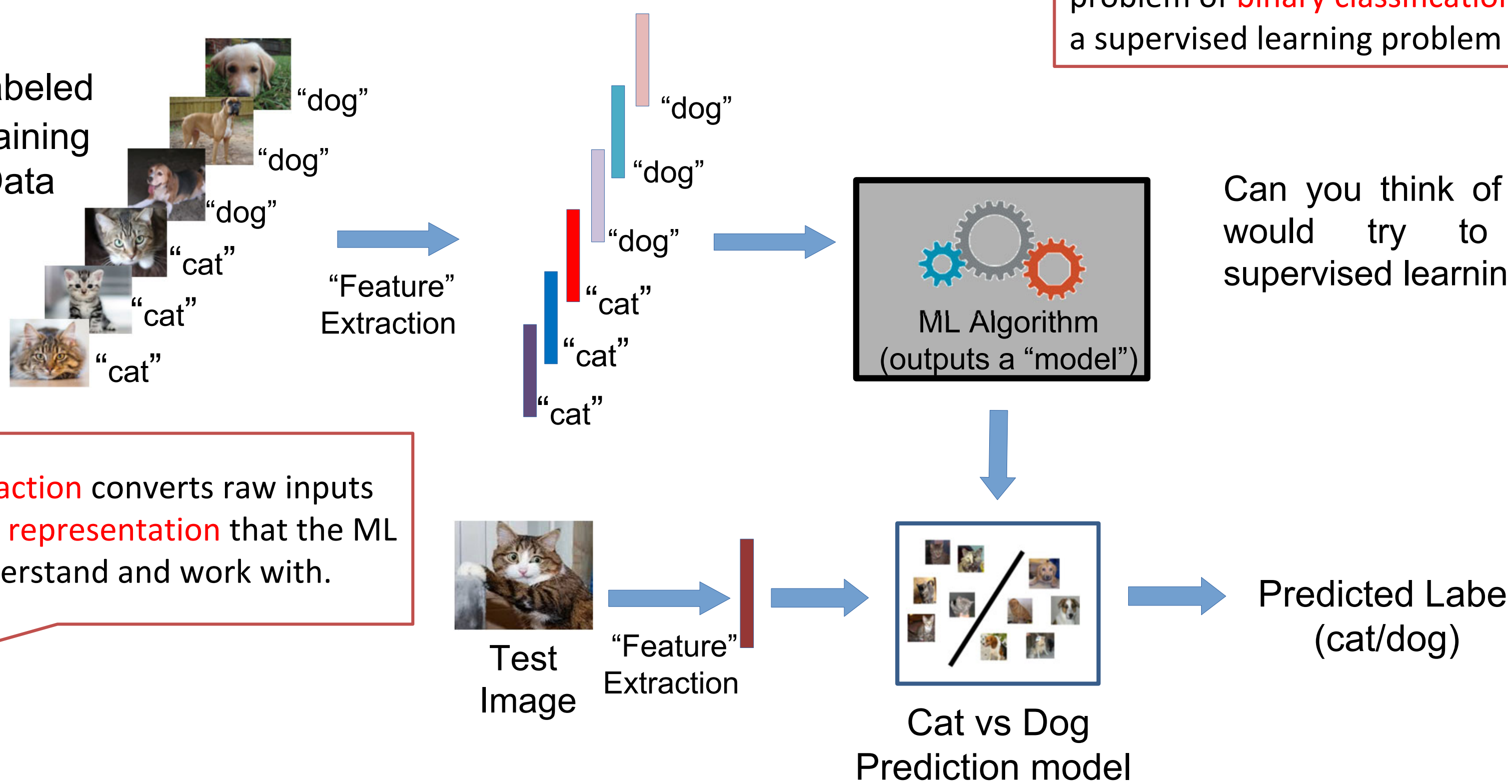


Many other specialized flavors of ML also exist, some of which include

- Semi-supervised Learning
- Active Learning
- Transfer Learning
- Multitask Learning
- Imitation Learning (somewhat related to RL)
- Zero-Shot Learning
- Few-Shot Learning
- Continual learning

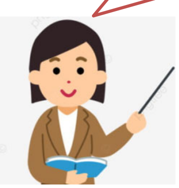
A Typical Supervised Learning Workflow

Note: This example is for the problem of binary classification, a supervised learning problem



Can you think of a problem you would try to solve using supervised learning?

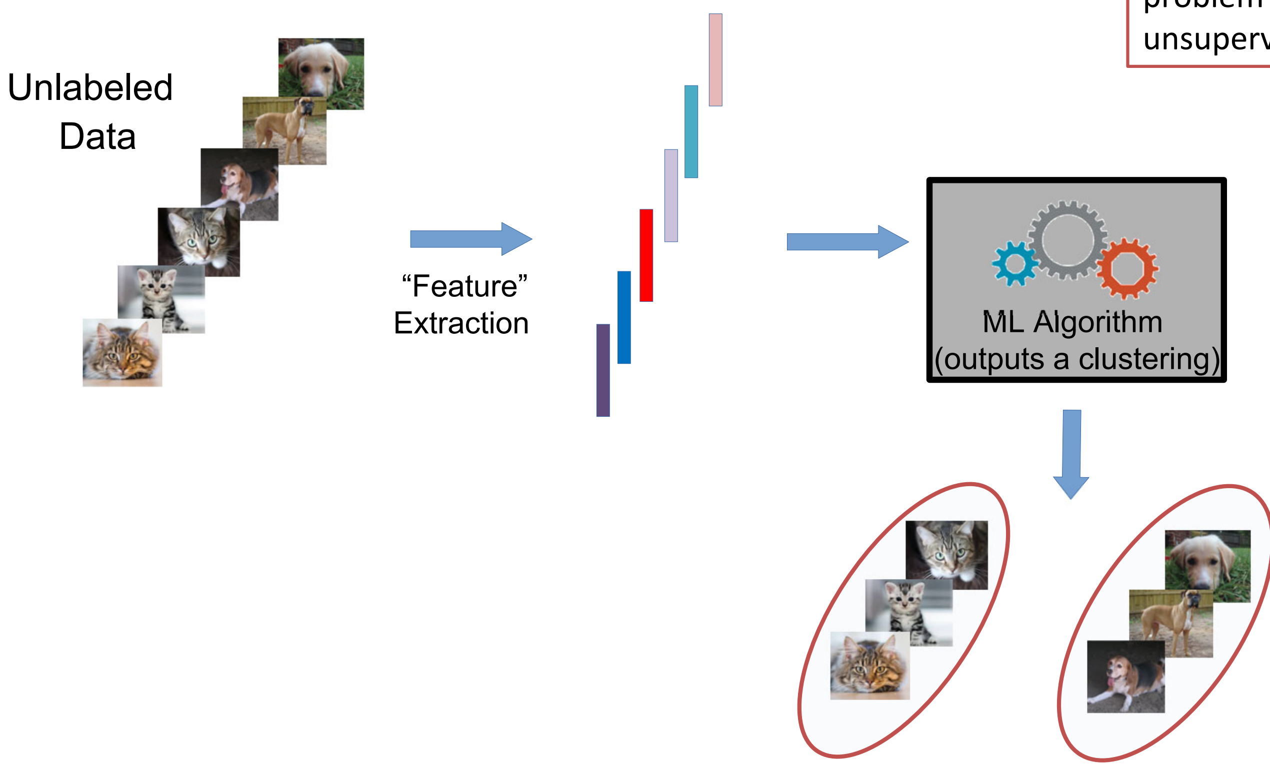
Feature extraction converts raw inputs to a numeric representation that the ML algo can understand and work with.



https://www.pinclipart.com/, http://www.pngtree.com

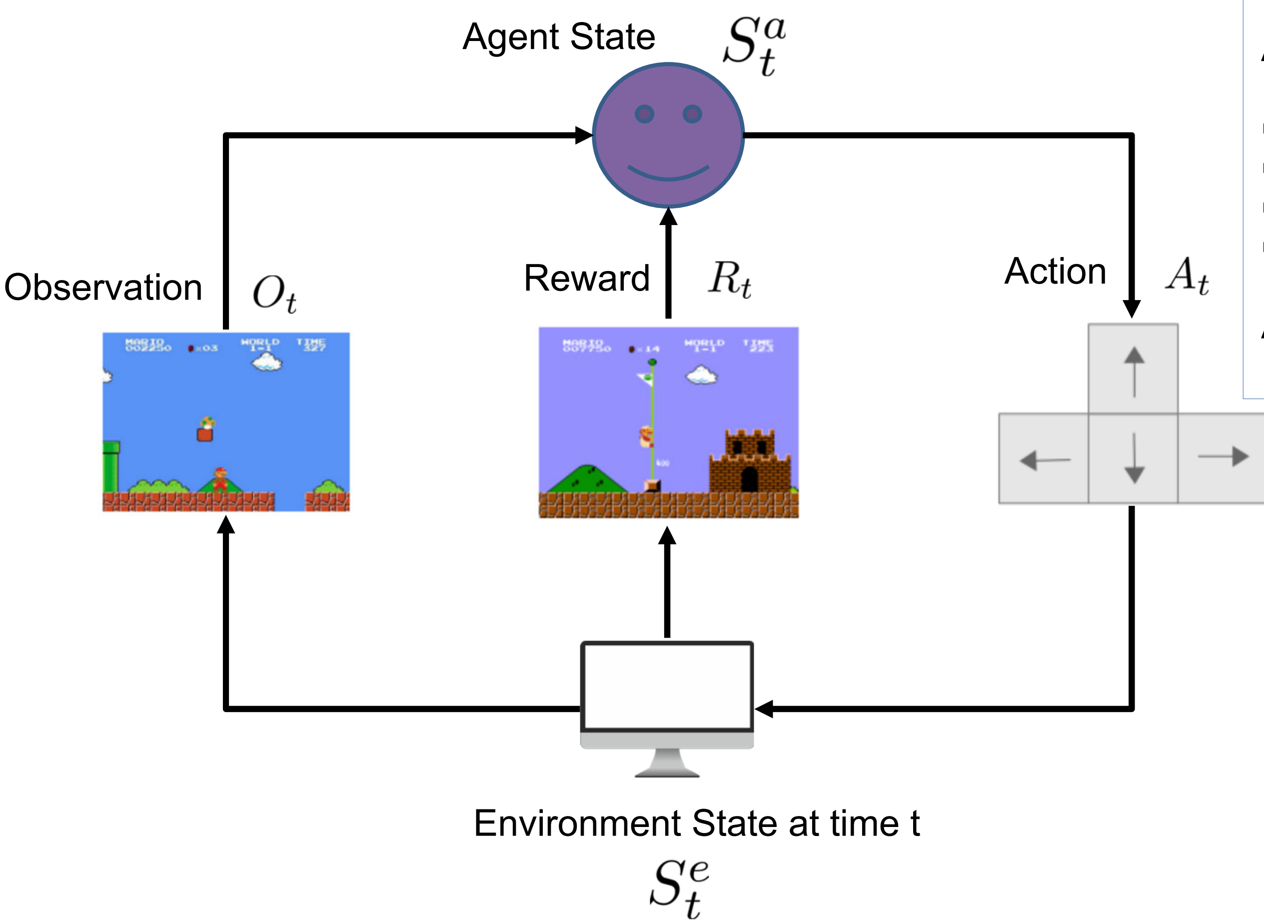
A Typical Unsupervised Learning Workflow

Note: This example is for the problem of data clustering, an unsupervised learning problem



https://www.pinclipart.com/, http://www.pngtree.com

A Typical Reinforcement Learning Workflow



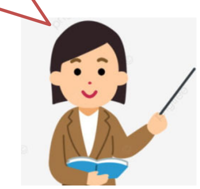
Wish to teach an agent optimal policy for some task

Agent does the following repeatedly

- Senses/observes the environment
- Takes an action based on its current policy
- Receives a reward for that action
- Updates its policy

Agent's goal is to maximize its overall reward

There IS supervision, not explicit (as in Supervised Learning) but rather implicit (feedback based)



Data and Features

- ML algos require a numeric feature representation of the inputs

- Features can be obtained using one of the two approaches
  - Approach 1: Extracting/constructing features <u>manually</u> from raw inputs
  - Approach 2: <u>Learning</u> the features from raw inputs

- Approach 1 is what we will focus on primarily for now

- Approach 2 is what is followed in Deep Learning algorithms (will see later)

- Approach 1 is not as powerful as Approach 2 but still used widely

Example: Feature Extraction for Text Data

- Consider some text data consisting of the following sentences:
  - John likes to watch movies
  - Mary likes movies too
  - John also likes football
- Want to construct a feature representation for these sentences
- Here is a "bag-of-words" (BoW) feature representation of these sentences

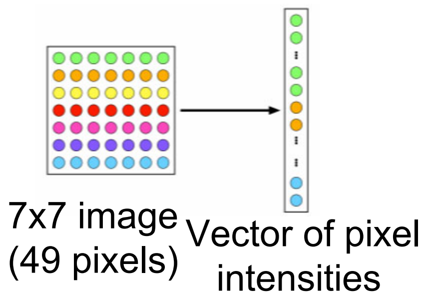
BoW is just one of the many ways of doing feature extraction for text data. Not the most optimal one, and has various flaws, but often works reasonably well

	John	likes	to	watch	movies	Mary	too	also	football
Sentence 1	1	1	1	1	1	0	0	0	0
Sentence 2	0	1	0	0	1	1	1	0	0
Sentence 3	1	1	0	0	0	0	0	1	1

- Each sentence is now represented as a binary vector (each feature is a binary value, denoting presence or absence of a word). BoW is also called "unigram" rep.

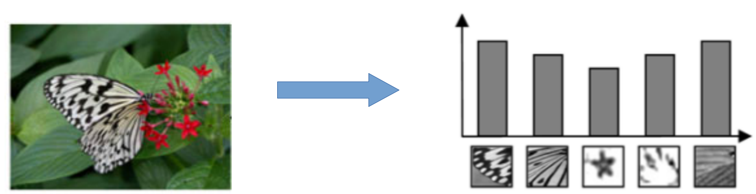
Example: Feature Extraction for Image Data

- A very simple feature extraction approach for image data is flattening



Flattening and histogram based methods destroy the spatial information in the image but often still work reasonably well

- Histogram of visual patterns is another popular feature extr. method for images



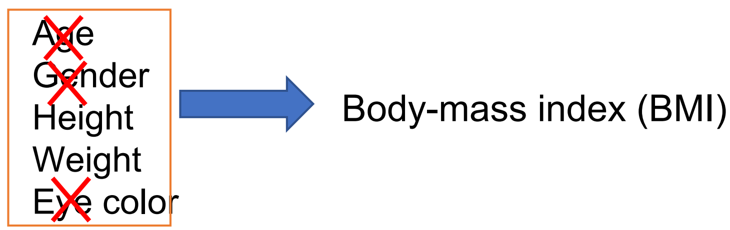
- Many other manual feature extraction techniques developed in computer vision and image processing communities (SIFT, HoG, and others)

Pic credit: cat.uab.cat/Research/object-recognition

Feature Selection

- Not all the extracted features may be relevant for learning the model (some may even confuse the learner)

- Feature selection (a step after feature extraction) can be used to identify the features that matter, and discard the others, for more effective learning

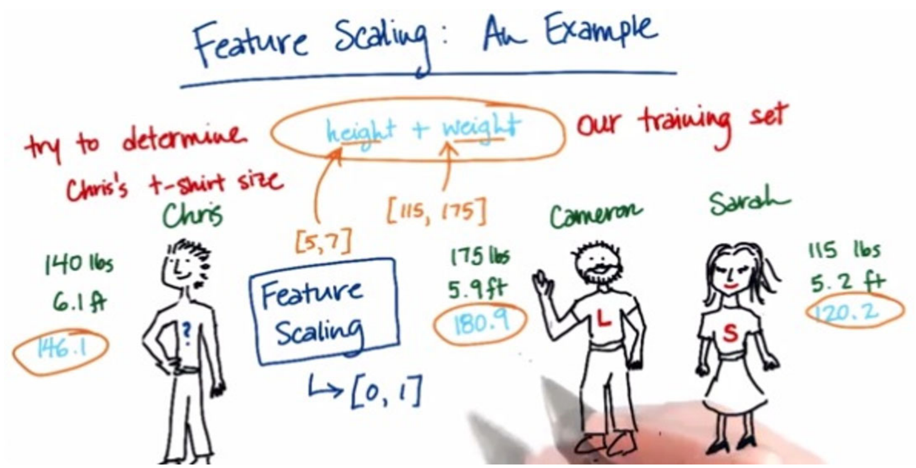


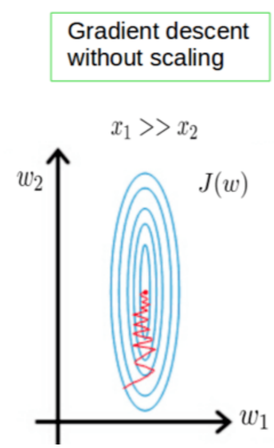
- Many techniques exist – some based on intuition, some based on algorithmic principles (will visit feature selection later)

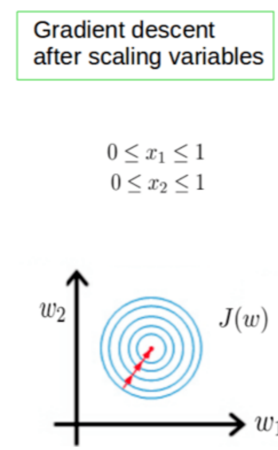
- More common in supervised learning but can also be done for unsup. learning

Some More Postprocessing: Feature Scaling

- Even after feature selection, the features may not be on the same scale
- This can be problematic when comparing two inputs – features that have larger scales may dominate the result of such comparisons
- Therefore helpful to standardize the features (e.g., by bringing all of them on the same scale such as between 0 to 1)







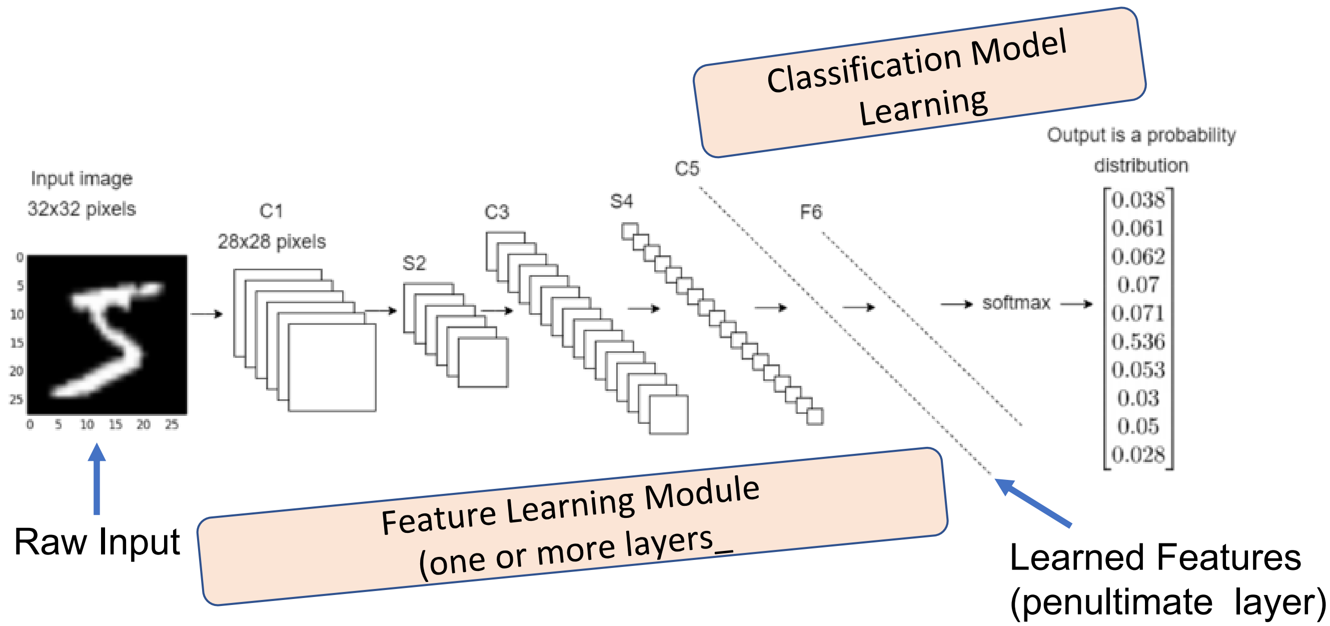
- Also helpful for stabilizing the optimization techniques used in ML algos

Pic credit: https://becominghuman.ai/demystifying-feature-scaling-baff53e9b3fd, https://stackoverflow.com/

Deep Learning: An End-to-End Approach to ML

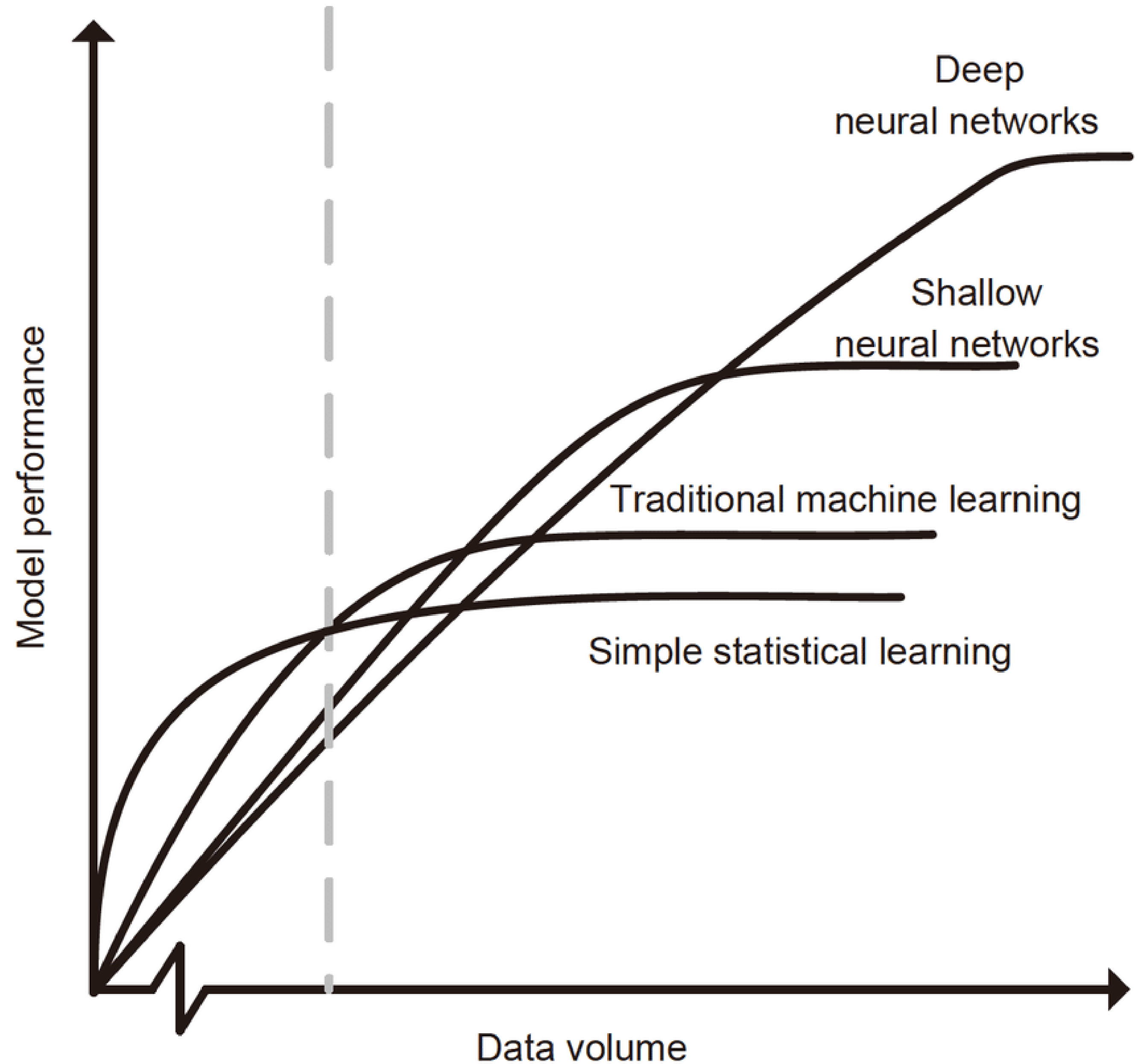
Deep Learning = ML with automated feature learning from the raw inputs

Feature extraction part is automated via the feature learning module



Pic an adaptation of the original from: https://deepai.org/

Machine Learning vs Deep Learning



Types of Features and Types of Outputs

- Features as well as outputs can be real-valued, binary, categorical, ordinal, etc.
- Real-valued: Pixel intensity, house area, house price, rainfall amount, temperature, etc
- Binary: Male/female, adult/non-adult, or any yes/no or present/absent type value
- Categorical/Discrete: Zipcode, blood-group, or any "one from a finite many choices" value
- Ordinal: Grade (A/B/C etc.) in a course, or any other type where relative values matter
- Often, the features can be of mixed types (some real, some categorical, some ordinal, etc.)

5 Steps for Approaching a Machine Learning Application

1. Define the problem to be solved.
2. Collect (labeled) data.
3. Choose an algorithm class.
4. Choose an optimization metric or measure for learning the model.
5. Choose a metric or measure for evaluating the model.

ML Terminology (Part 2)

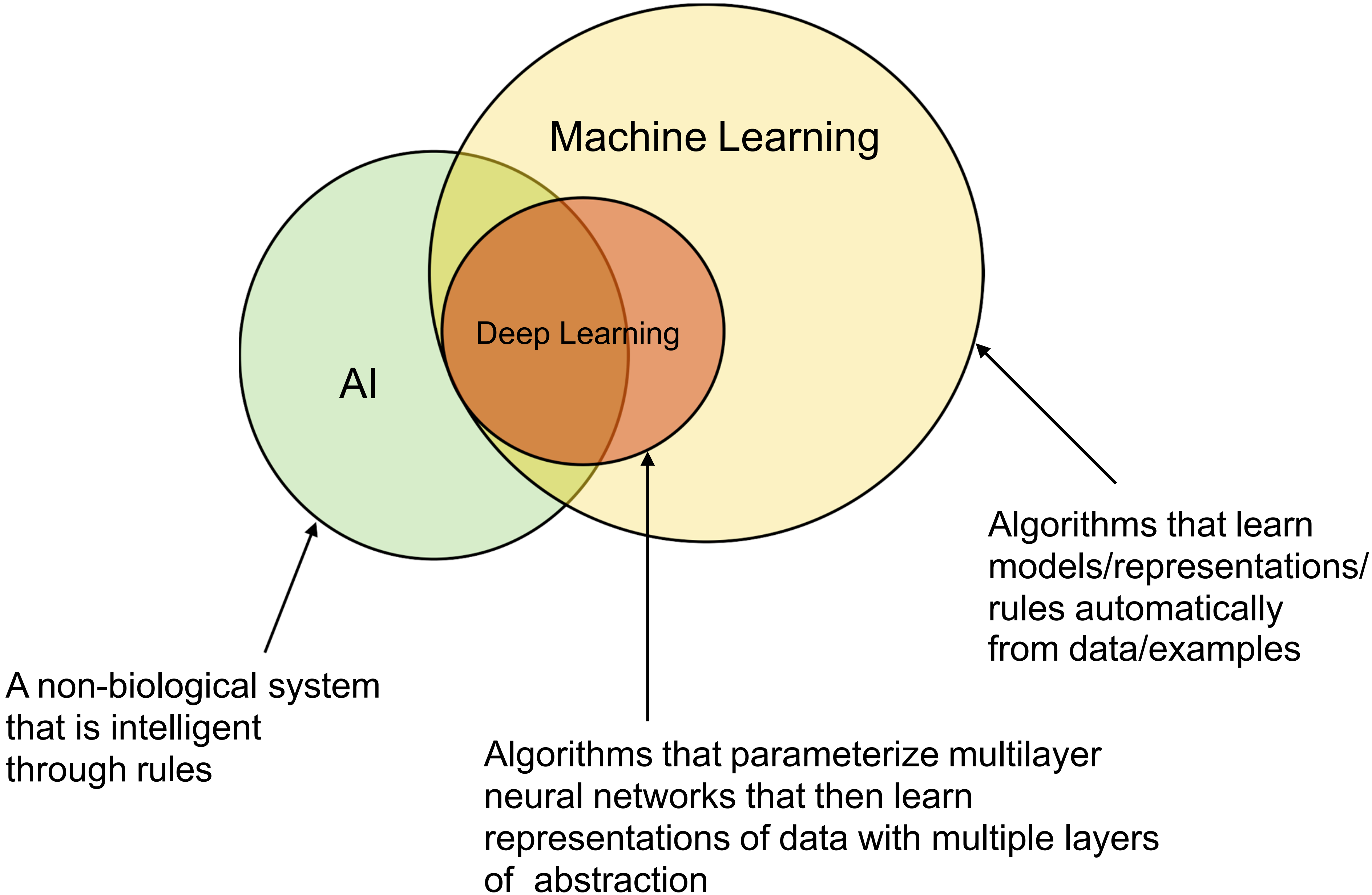
- **Loss function:** Often used synonymously with cost function; sometimes also called error function. In some contexts the loss for a single data point, whereas the cost function refers to the overall (average or summed) loss over the entire dataset. Sometimes also called empirical risk.

Other Metrics in Future Lectures

- Accuracy (1-Error)
- ROC AUC
- Precision
- Recall
- (Cross) Entropy
- Likelihood
- Squared Error/MSE
- L-norms
- Utility
- Fitness
- ...

But more on other metrics in future lectures.

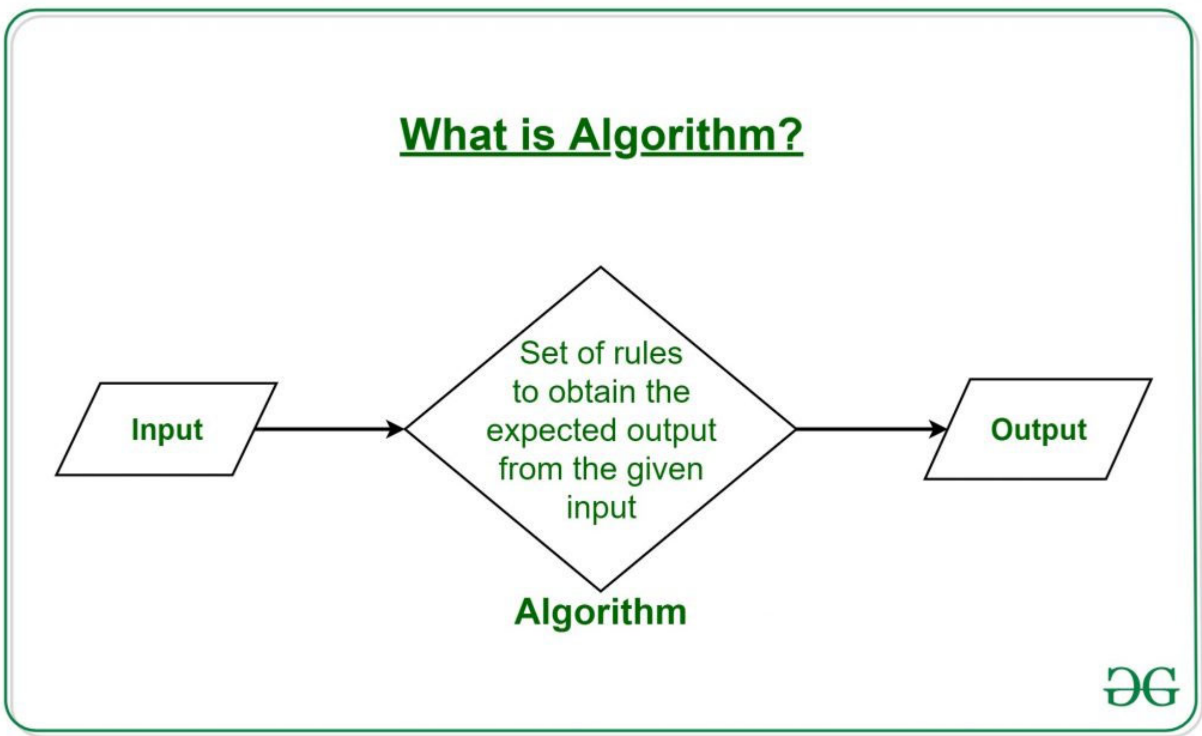
Machine Learning, AI, and Deep Learning

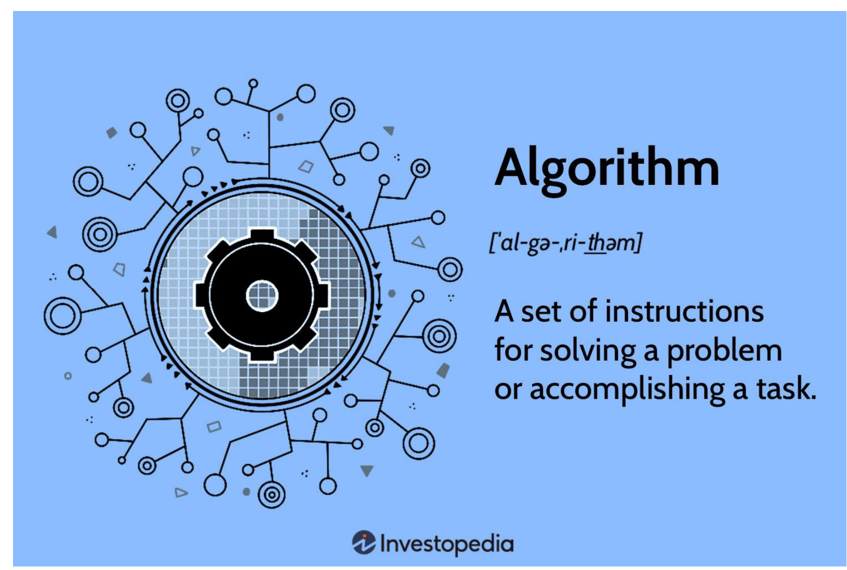


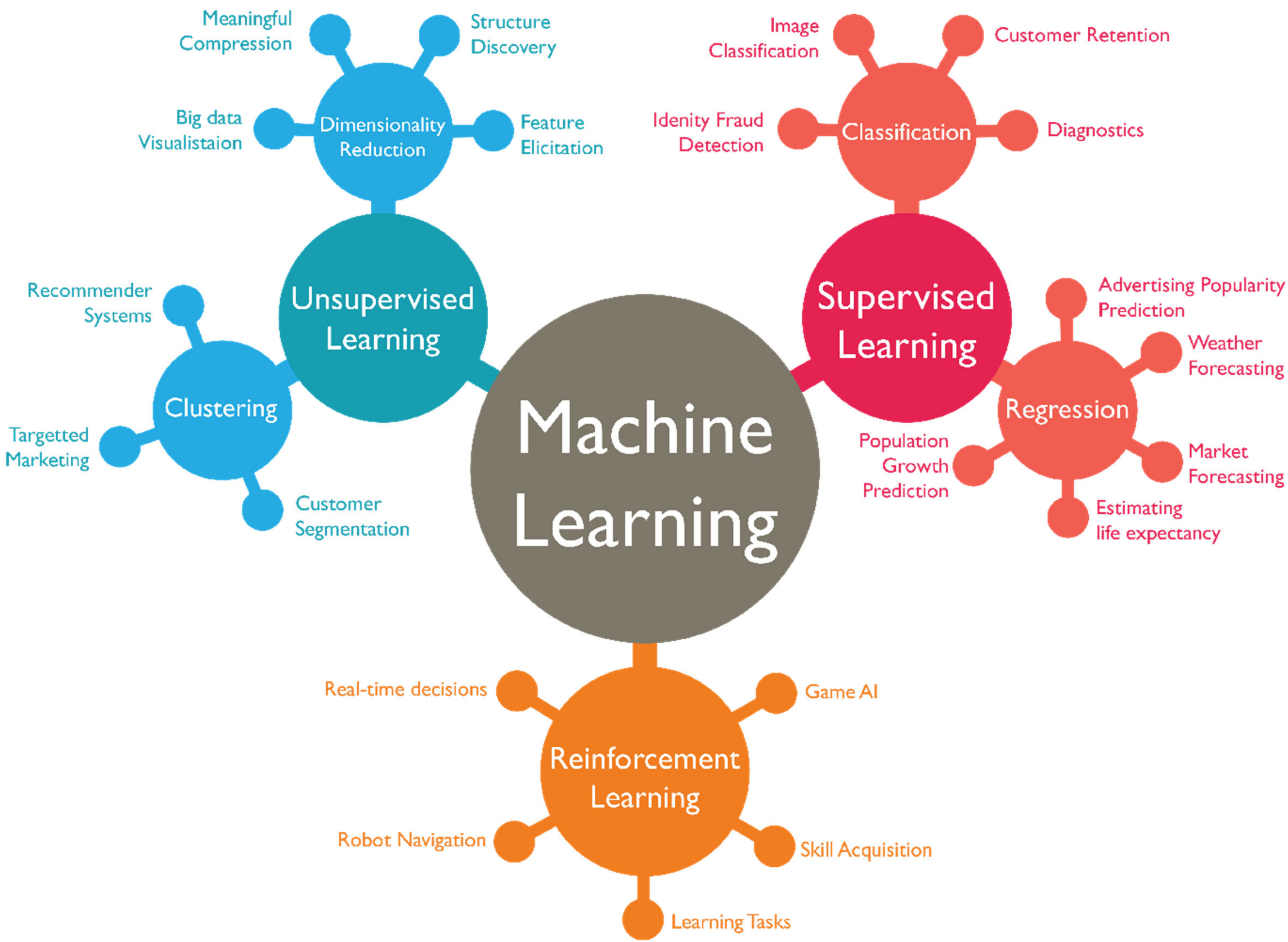
S5q#W .$:#/9#M L§

Machine learning (ML) is a branch of artificial intelligence (AI) and computer science that focuses on the using data and algorithms to enable AI to imitate the way that humans learn, gradually improving its accuracy.
-IBM

https://www.ibm.com/topics/machine-learning






Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks