Certificate I: Understanding AI and Machine Learning in Africa

Course AIML01: Artificial Intelligence – Past, Present, and Future

Module 2: The Nature of AI

Lecture 3: Statistical machine learning

Carnegie Mellon University
Africa

Learning Objectives

1. Explore the connection between machine learning and AI

2. Identify and discuss different forms of inference

3. Identify and discuss different forms of learning

Lecture Contents

1. Statistical machine learning and AI
2. Inference strategies: analogical, domain-specific, structural
3. Structural inference: regression, neural networks, Bayesian networks
4. Types of learning: supervised, unsupervised, reinforcement, self-supervised
5. Bayesian learning
6. Lecture summary
7. Recommended reading & references

Statistical Machine Learning and AI

- A parallel development in AI over the last 20 years: machine learning
  - Partial overlap with the AI connectionist approach
- Primarily based on a variety of statistical inference methods
  - Using large data sets
  - Estimate (i.e. learn) the parameters of a model that can
    - Classify data
    - Predict data

Statistical Machine Learning and AI

- Developed in conjunction with AI research in
  - Computer vision and speech (i.e., pattern recognition)
  - Robotics (e.g., reinforcement learning)
  - Neural networks (MLPs and deep neural networks)
- Today, some people use the terms AI and Machine Learning (ML) interchangeably
  - Mainly because of the current emphasis on deep learning and its success
- However, ML focusses on methods for data-driven statistical inference

Inference Strategies

There are several inference strategies in statistics (Danks 2014)

- Analogical inference
- Domain-specific inference
- Structural inference

Maps from a current situation or problem to one that is known from previous experience (either well known or personal)

Inferences are then made by the analogical mapping to translate previous outcomes to the current problem.

Inference Strategies

There are several inference strategies in statistics (Danks 2014)

- Analogical inference
- Domain-specific inference
- Structural inference

Maps from a current situation or problem to one that is known from previous experience (either well known or personal)

Inferences are then made by the analogical mapping to translate previous outcomes to the current problem.

James Clerk Maxwell famously used an analogy from fluid dynamics (how fluid flows) to infer a model of electromagnetism.

Inference Strategies

There are several inference strategies in statistics (Danks 2014)

- Analogical inference
- Domain-specific inference
- Structural inference

Maps from a current situation or problem to one that is known from previous experience (either well known or personal)

Inferences are then made by the analogical mapping to translate previous outcomes to the current problem.

James Clerk Maxwell famously used an analogy from fluid dynamics (how fluid flows) to infer a model of electromagnetism.

Analogical inference can support inferences for rare situations or situations for which there is limited data.

Analogical inference is very commonly used by humans; computational analogical inference (or analogical reasoning) is less common.

Inference Strategies

There are several inference strategies in statistics (Danks 2014)

- Analogical inference
- Domain-specific inference ← Uses techniques and knowledge that are tailored to a specific problem or environment

  Exploit specialized algorithms, constraints, domain knowledge
- Structural inference

Inference Strategies

There are several inference strategies in statistics (Danks 2014)

- Analogical inference
- Domain-specific inference
- Structural inference

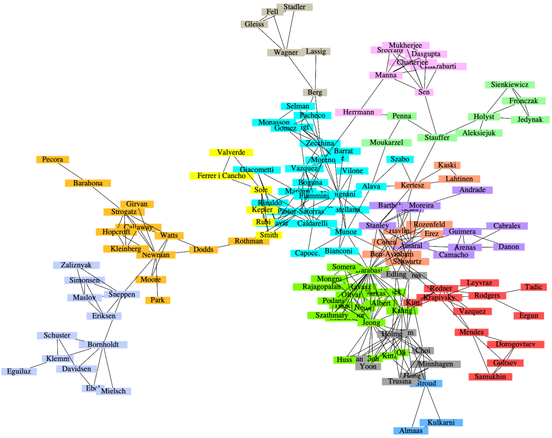
Uses domain-general algorithms: focus on the structure of the data, rather than features of the semantic content of the data: the exact meaning of the data is irrelevant to the learning algorithm

Inference Strategies

There are several inference strategies in statistics (Danks 2014)

- Analogical inference
- Domain-specific inference
- Structural inference

For example, detection of communities in complex networks often exploits patterns of connectivity rather than detailed knowledge of each entity



Communities of physicists who have co-authored papers on networks (Girvan and Newman, 2004)

Inference Strategies

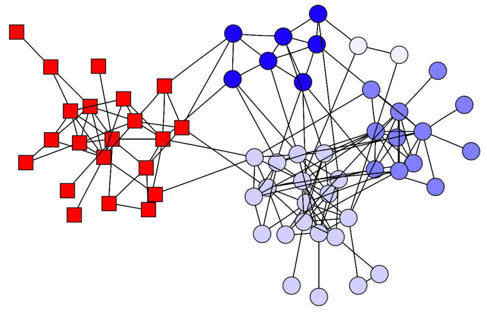
There are several inferentia strategies in statistics (Danks 2014)

– Analogical inference

– Domain-specific inference

– Structural inference

For example, detection of communities in complex networks often exploits patterns of connectivity rather than detailed knowledge of each entity



Communities structure in bottlenose dopphins (Girvan and Newman, 2004)

Inference Strategies

There are several inference strategies in statistics (Danks 2014)

- Analogical inference
- Domain-specific inference
- Structural inference

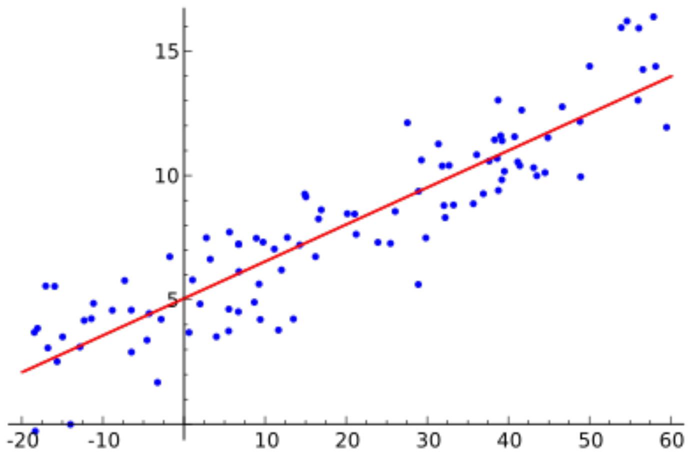
Two classes: logical methods and statistical methods

Depends on the existence of statistical regularities in the dataset

Structural Inference

Structural inference is the basis of many – perhaps most – machine learning methods; for example

- Regression
- Neural networks
- Bayesian networks

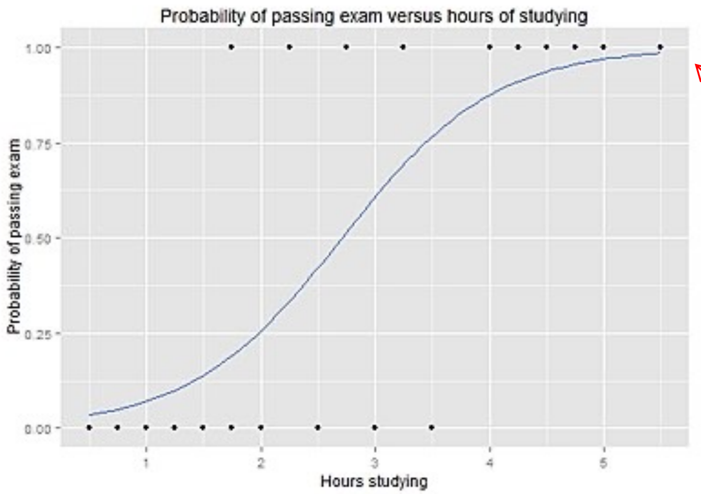


Linear regression: learning the relationship between a dependent variable (vertical axis) and an independent variable (horizontal axis)
https://en.wikipedia.org/wiki/Linear_regression

Structural Inference

Structural inference is the basis of many – perhaps most – machine learning methods; for example

- Regression
- Neural networks
- Bayesian networks



Two possible outcomes
(in this case, pass and fail)

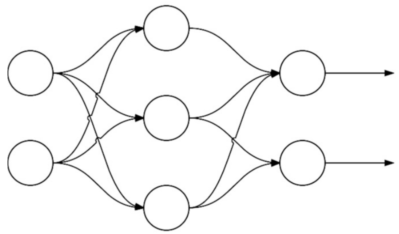
Logistic regression: learning a model of a binary dependent variable. The probability of the outcome (dependent variable) is shown on the vertical axis. The values of the independent variable are shown on the horizontal axis. Here the logistic curve is fitted to the data and shows the probability of passing an exam versus hours studying.
https://en.wikipedia.org/wiki/Logistic_regression

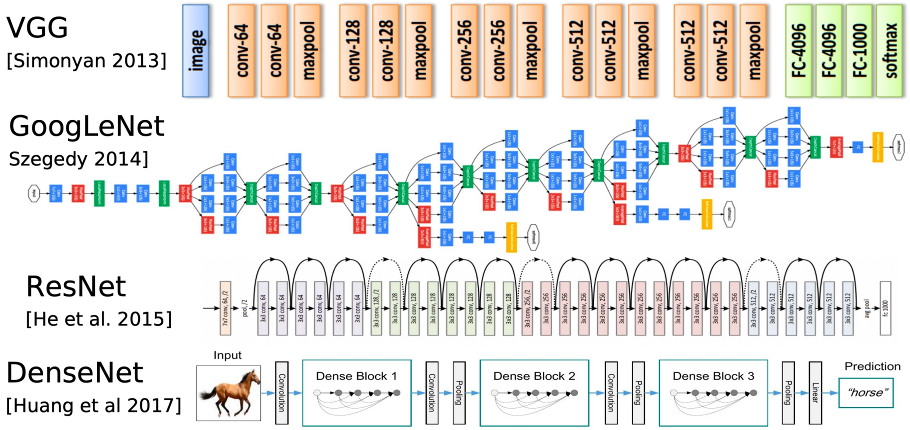
Structural Inference

Structural inference is the basis of many – perhaps most – machine learning methods; for example

- Regression
- Neural networks
- Bayesian networks

Multilayer Perceptron





Y. LeCun, Deep Learning Hardware: Past, Present, & Future, ISSCC 2019
https://drive.google.com/file/d/17w443t_5Atnwnu-iOrHKUPFik1pThyhx/view

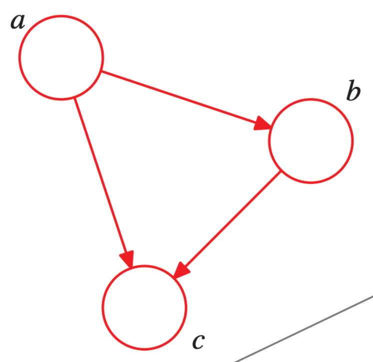
Structural Inference

Structural inference is the basis of many – perhaps most – machine learning methods; for example

- Regression
- Neural networks
- Bayesian networks

Bayesian networks are also known as a Bayes networks, Bayes nets, belief networks, decision networks, and probabilistic graphical models.

They were invented by Judea Pearl in 1985



$$p(a,b,c) = p(c|a,b)p(b|a)p(a)$$

A directed graphical model representing the joint probability distribution over three variables a, b and c, corresponding to the decomposition on the right-hand side.

They provide a diagrammatic way of expressing how a joint distribution can be factored into the product of distributions over smaller subsets of variables

https://www.microsoft.com/en-us/research/wp-content/uploads/2016/05/Bishop-MBML-2012.pdf

Structural Inference

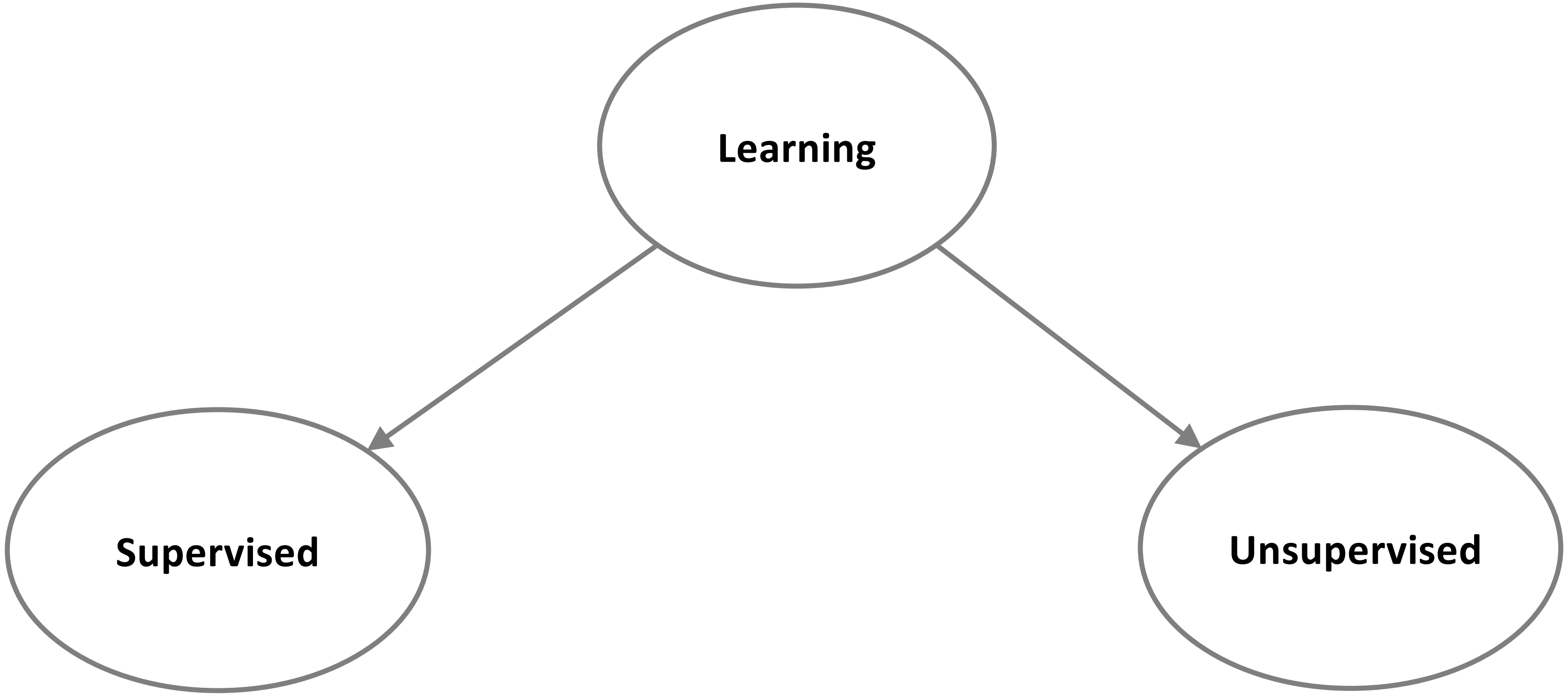
Structural inference techniques are data-centric ("data hungry")

Success of machine learning in making a major contribution to AI is due to

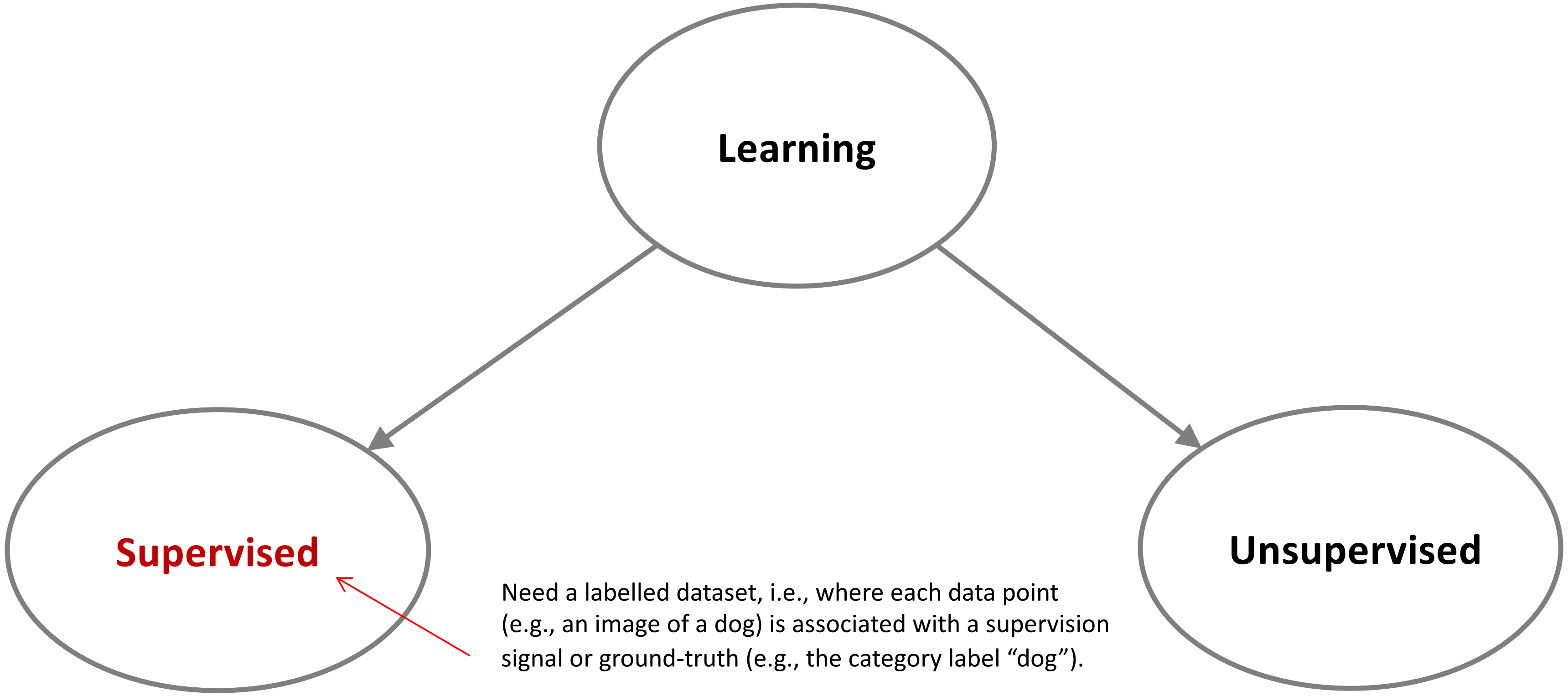
- Social media and the web for almost unlimited data
- Wider access to cloud-based parallel GPU computing systems for the computationally-intensive statistical computations on these large datasets

Facilitated the bootstrapping of neural network technology from the shallow MLP networks only trainable with small datasets in the 1980s, 1990s, and early 2000s, to the deep CNN trained on huge datasets in the last ten years or so.

Types of Learning



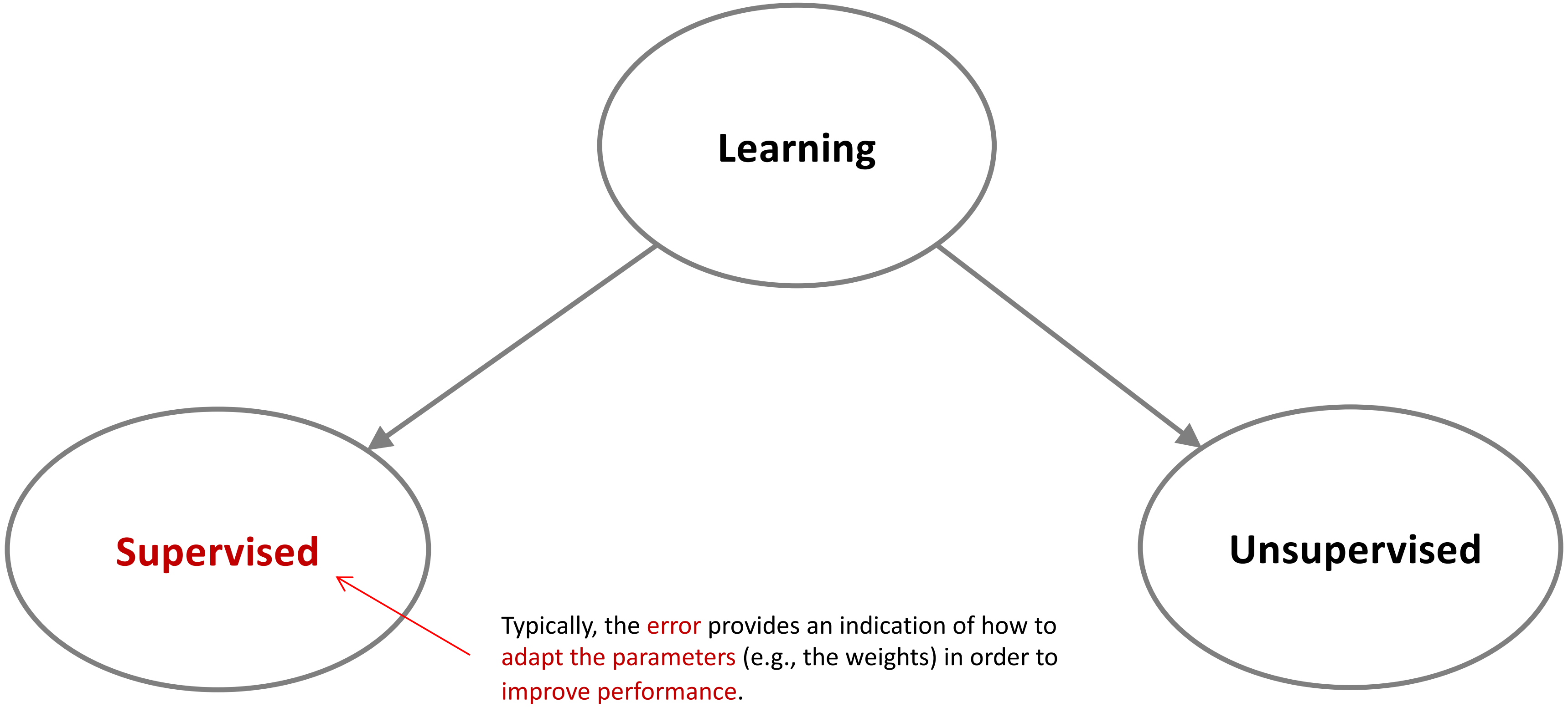
Types of Learning



Need a labelled dataset, i.e., where each data point (e.g., an image of a dog) is associated with a supervision signal or ground-truth (e.g., the category label “dog”).

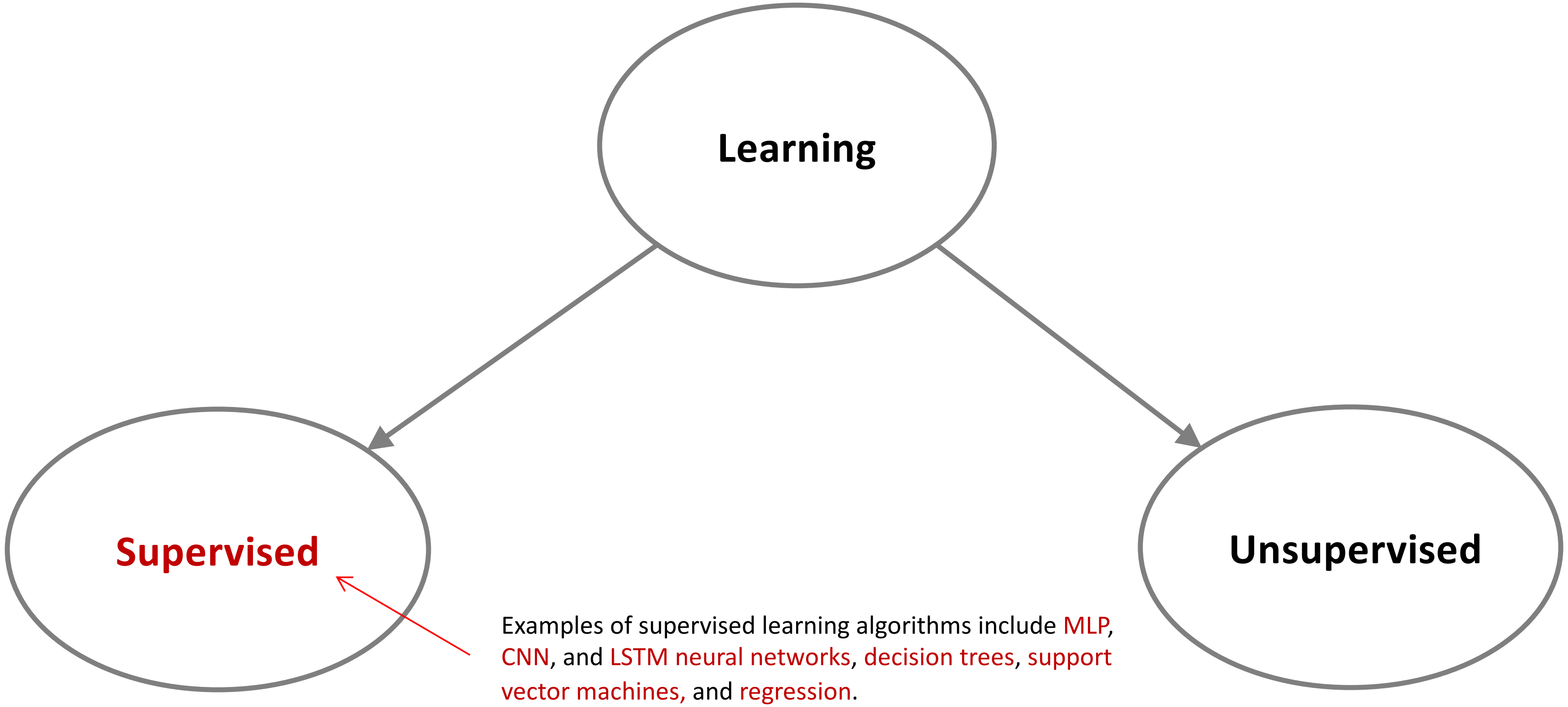
The learning algorithm has to find the parameters of the model (e.g., weights of a neural network) using the error between the model’s own estimate and the supervision label.

Types of Learning



Typically, the error provides an indication of how to adapt the parameters (e.g., the weights) in order to improve performance.

Types of Learning



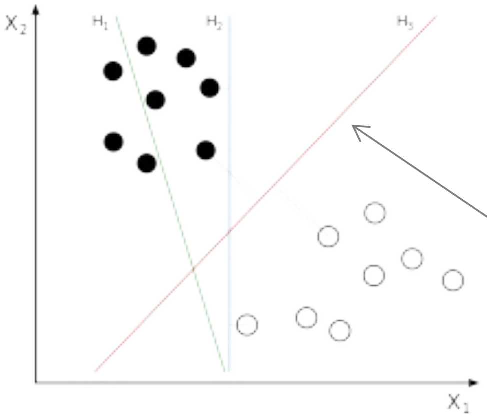
Examples of supervised learning algorithms include MLP, CNN, and LSTM neural networks, decision trees, support vector machines, and regression.

Types of Learning

Learning

Supervised

Unsupervised



A Support Vector Machine (SVM) learns the linear boundary that best separates the two classes.

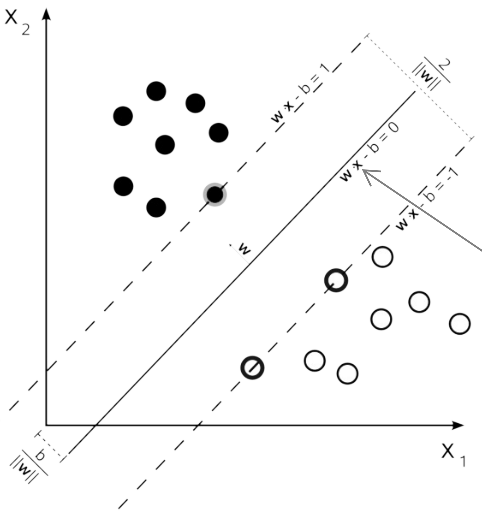
H3 separates classes “better” than H2

Types of Learning

Learning

Supervised

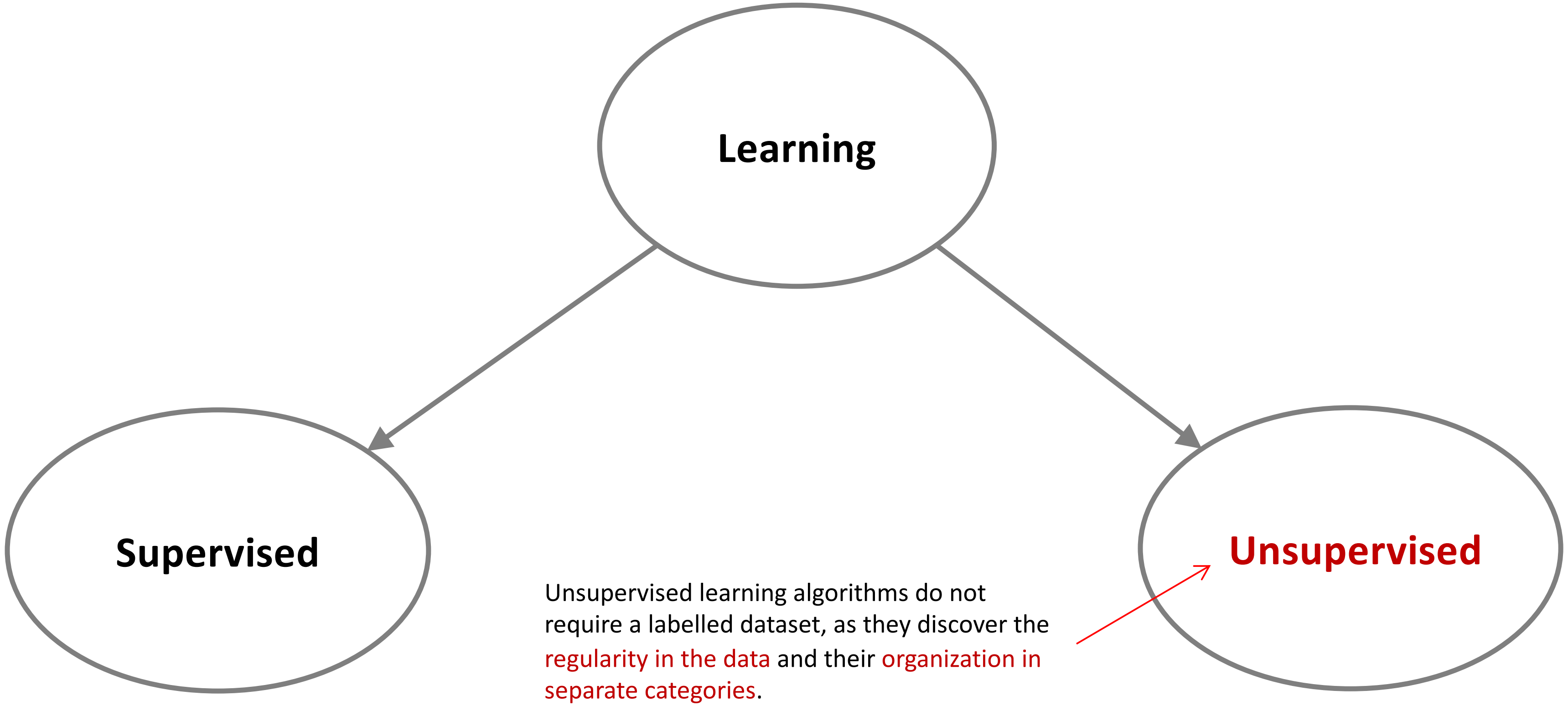
Unsupervised



A Support Vector Machine (SVM) learns the linear boundary that best separates the two classes.

H3 separates classes “better” than H2

Types of Learning



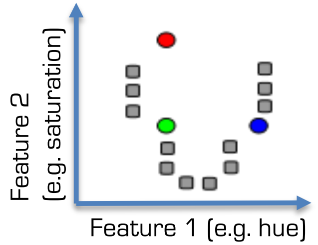
Unsupervised learning algorithms do not require a labelled dataset, as they discover the regularity in the data and their organization in separate categories.

Types of Learning

Learning

Supervised

Unsupervised



K-means clustering:

1. Randomly select k start points (centroids, nodes, centers)

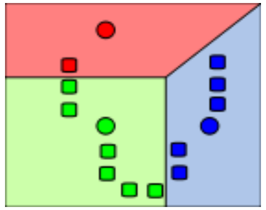
Illustrations from: http://en.wikipedia.org/wiki/K-means_clustering

Types of Learning

Learning

Supervised

Unsupervised



K-means clustering:

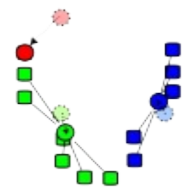
2. Assign all points to the closest centroid (in the feature space)

Types of Learning

Learning

Supervised

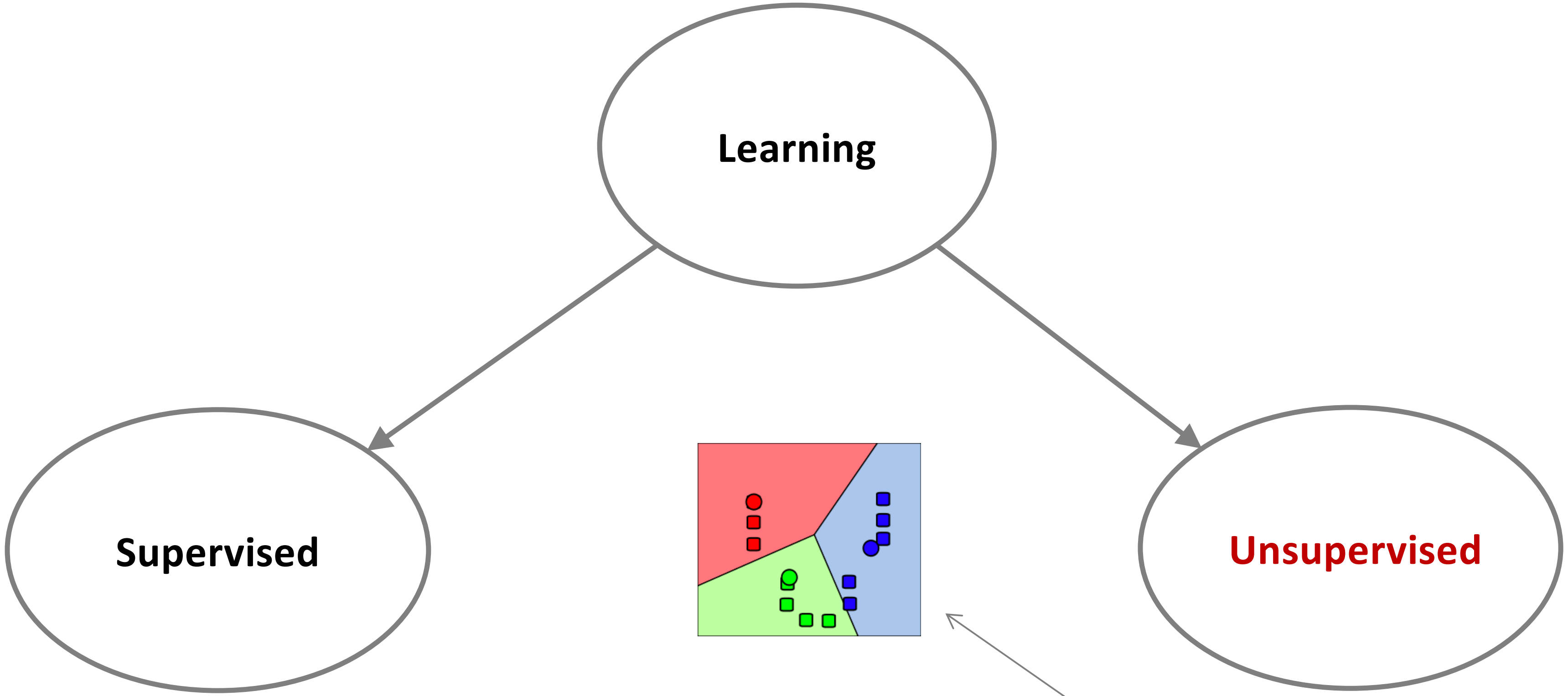
Unsupervised



K-means clustering:

3. Calculate new centroid (mean values) for each cluster

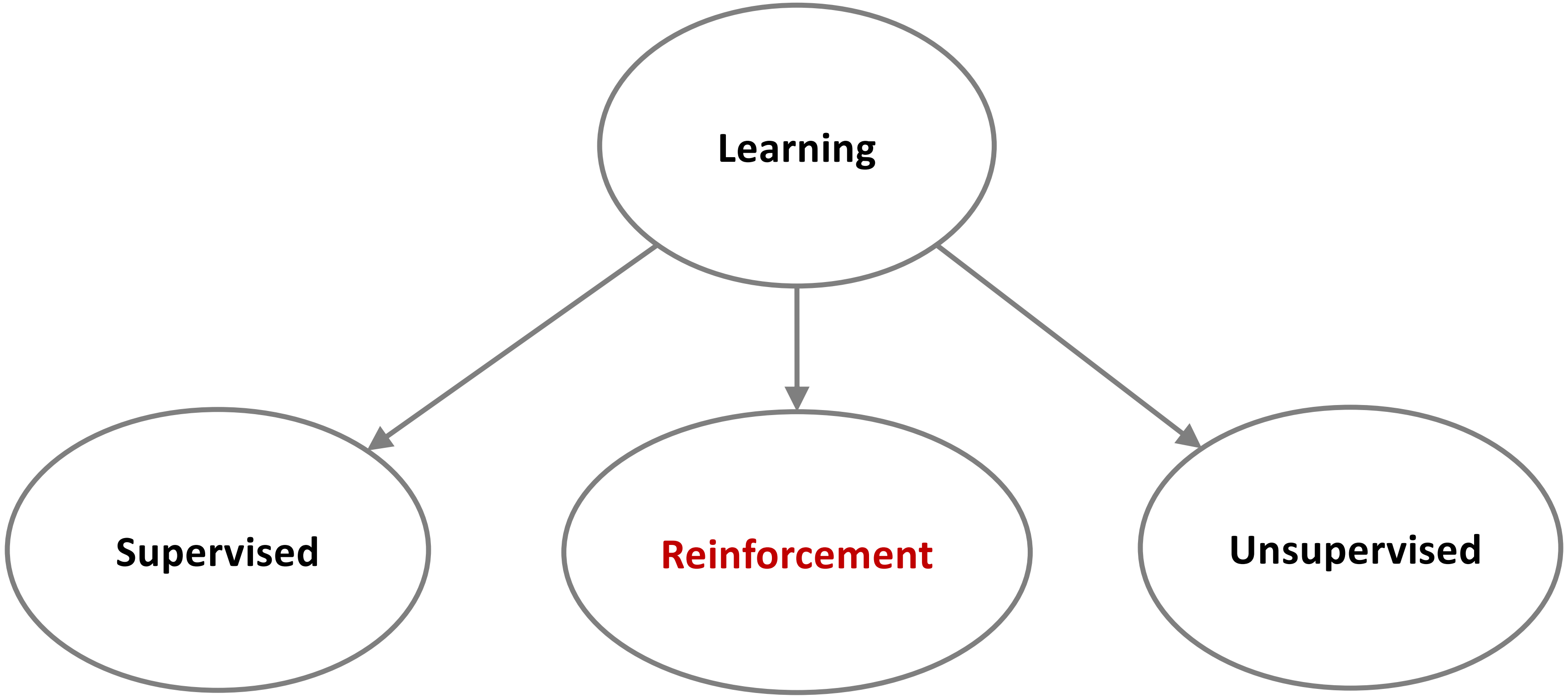
Types of Learning



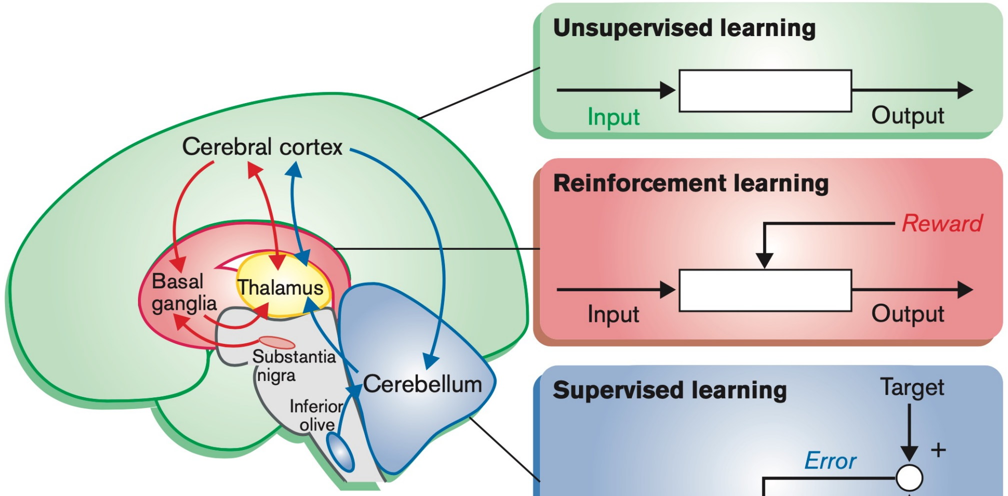
K-means clustering:

Iterate until all centroids remain stable

Types of Learning

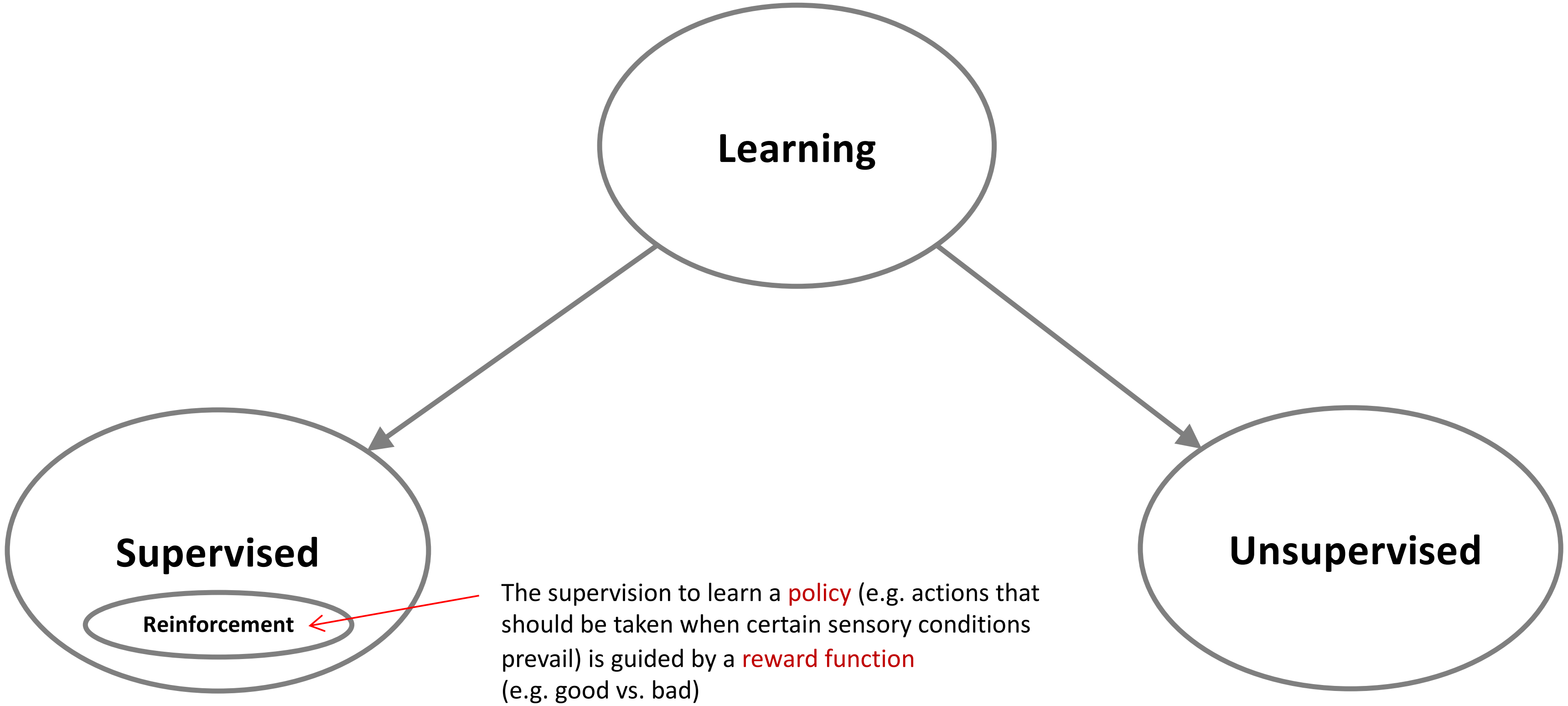


Types of Learning



K. Doya, 2000. Complementary roles of basal ganglia and cerebellum in learning and motor control, Current Opinion in Neurobiology, Vol. 10, pp. 732-739.

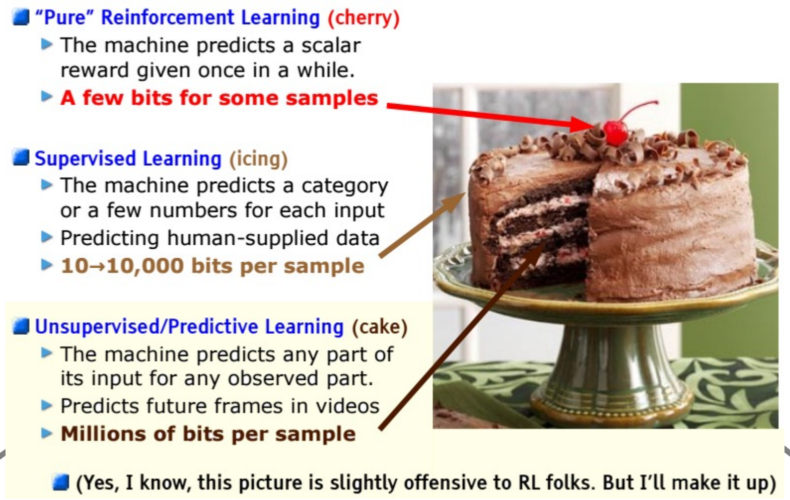
Types of Learning



The supervision to learn a policy (e.g. actions that should be taken when certain sensory conditions prevail) is guided by a reward function
(e.g. good vs. bad)

The cost function (or objective function) that governs the learning is based on maximizing the cumulative sum of rewards over time.

Types of Learning



Supervised

Reinforcement

Yann LeCun Cake Analogy
(Chief AI Scientist, Meta)
NIPS, 2016.

Unsupervised

"If intelligence is a cake, the bulk of the cake is unsupervised learning, the icing on the cake is supervised learning, and the cherry on the cake is reinforcement learning" (LeCun, NIPS 2016).

Types of Learning

Sample-efficient learning technique that uses any experience to provide a reward signal by assuming the goal equals whatever happened.

Supervised

Reinforcement

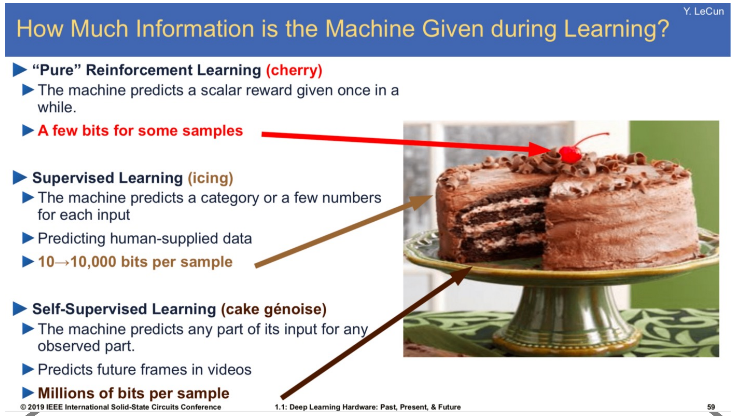
Unsupervised

Pieter Abbeel, NIPS 2017

"I prefer to eat a cake with a lot of cherries because I like reinforcement learning."

https://medium.com/syncedreview/yann-lecun-cake-analogy-2-0-a361da560dae

Types of Learning



Supervised

Reinforcement

Yann LeCun Cake Analogy 2.0

2019 International Solid-State Circuits Conference (ISSCC)

https://medium.com/syncedreview/yann-lecun-cake-analogy-2-0-a361da560dae

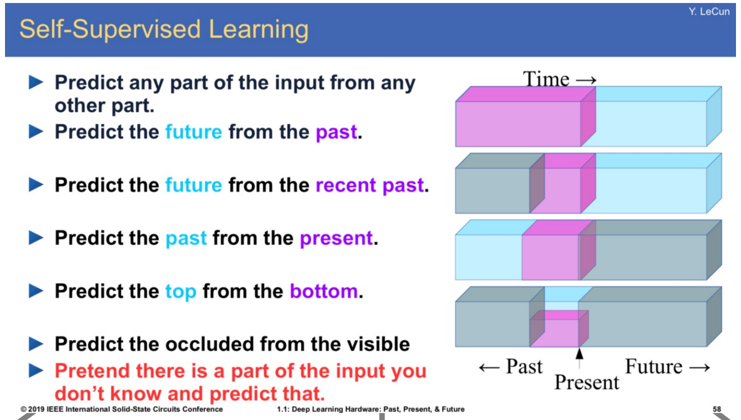
Also referred to as predictive learning

Unsupervised

Self-supervised

A variant of unsupervised learning where the data provides the supervision

Types of Learning



Supervised

Reinforcement

Yann LeCun Cake Analogy 2.0

2019 International Solid-State Circuits Conference (ISSCC)

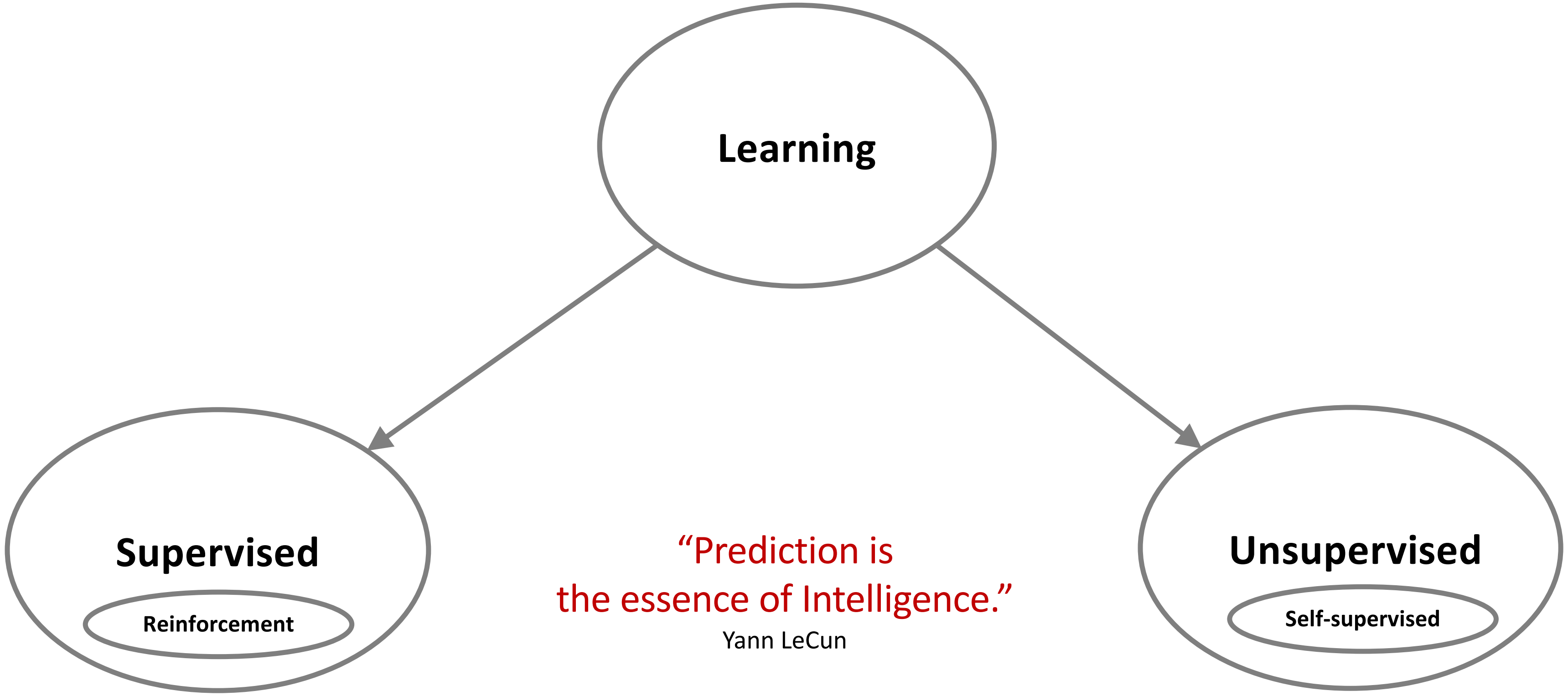
https://medium.com/syncedreview/yann-lecun-cake-analogy-2-0-a361da560dae

Unsupervised

Self-supervised

Self-supervised – predictive – learning leverages the power of unsupervised methods, such as autoencoders and word embeddings, for automatically extracting partial information from noisy or incomplete input data to predict the rest of the data.

Types of Learning



"Prediction is
the essence of Intelligence."
Yann LeCun

Types of Learning

1. Supervised
   Teaching data, in the form of required behavior, provide error signals

2. Reinforcement
   Teaching signals are scalar rewards or reinforcement signals
   (maximize the cumulative sum of rewards over time)

3. Unsupervised
   No teaching signals
   (uncover statistical regularities)

Bayesian Learning

Bayesian learning has re-emerged as a dominant theme in learning (Danks, 2014)

- A Bayesian learning algorithm requires
  - The specification of a (possibly very large) set of possible hypotheses or models
  - A probability distribution over those hypotheses: the prior probability distribution
- When provided with data
  - the learning algorithm then uses Bayes' Rule
  - to determine the probability distribution over those hypotheses, given that data

Bayesian Learning

It learns the probability that a hypothesis is true, given data

$$P(H_j|D)$$

This is computed from what we know

- the prior probabilities that each hypothesis is true:

$$P(H_j)$$

- the conditional probabilities (likelihoods) of that data occurring for each hypothesis

$$P(D|H_j)$$

Bayesian Learning

Using Bayes' Rule

$$P(H_j|D) = \frac{P(D|H_j)P(H_j)}{P(D)}$$

Bayes' Rule allows us to compute what we want to know ...

...from what we do know

$$= \frac{P(D|H_j)P(H_j)}{\sum_{j=1}^{n} P(D|H_j)P(H_j)}$$

Lecture Summary

1. Machine learning is a sub-set of AI and is one of the main reasons for the success of AI over the past 20 years

2. There are three main forms of inference: analogical, domain-specific, and structural

3. Structural inference is the most common and exploits the statistical patterns in the structure of the data, rather than the semantics of each data point

4. The three main structural inference techniques are regression, neural networks, and Bayesian networks

5. There are two main types of learning – supervised and unsupervised – with reinforcement learning and self-supervised learning forming respective subsets

Recommended Reading

Cangelosi, A. and Vernon, D. (2022). "Artificial Intelligence: Powering the Fourth Industrial Revolution", in *EPS Grand Challenges: Physics for Society at the Horizon 2050,* coordinated by the European Physical Society.
http://vernon.eu/publications/2022_Cangelosi_Vernon.pdf

Shanthamallu, U. S., Spanias, A., Tepedelenlioglu, C., and Stanley, M. (2017). "A brief survey of machine learning methods and their sensor and IoT applications," 8th International Conference on Information, Intelligence, Systems & Applications (IISA), 2017, pp. 1-8, doi: 10.1109/IISA.2017.8316459.
https://sensip.engineering.asu.edu/wp-content/uploads/2020/03/1_Machine-Learning-Survey-Paper.pdf

References

Danks D. (2014). Learning. In The Cambridge Handbook of Artificial Intelligence. Frankish, Keith, editor.; Ramsey, William M., 1960 - editor.
https://doi.org/10.1609/aimag.v27i4.1904.

Klix, F. (2001). Problem Solving: Deduction, Induction, and Analogical Reasoning. International Encyclopedia of Social & Behavioral Sciences. Elsevier Ltd. 12123–12130.
https://www.researchgate.net/publication/288151639_Problem_Solving_Deduction_Induction_and_Analogical_Reasoning