



Machine Learning

Chapter 1 - Introduction

Lecturer: Duc Dung Nguyen, PhD.

Contact: nddung@hcmut.edu.vn

Faculty of Computer Science and Engineering
Hochiminh city University of Technology

What is Machine learning?

What is Machine learning?

- Arthur Samuel (1959): "Field of study that gives computers the ability to learn without being explicitly programmed"
- Tom Mitchell (1997): "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 ".

Machine Learning

- How to construct programs that automatically improve with experience.

Machine Learning

- How to construct programs that automatically improve with experience.
- The scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead.

Machine Learning

- How to construct programs that automatically improve with experience.
- The scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead.
- A subset of artificial intelligence.

Experience

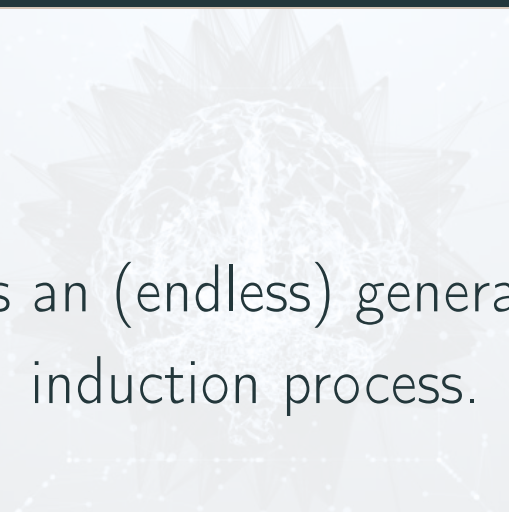
Example	Gray?	Mammal?	Large?	Vegetarian?	Wild?	Elephant
1	+	+	+	+	+	+
2	+	+	+	-	+	+
3	+	+	-	+	+	-
4	-	+	+	+	+	-
5	+	-	+	-	+	-
1	+	+	+	+	-	+

Prediction

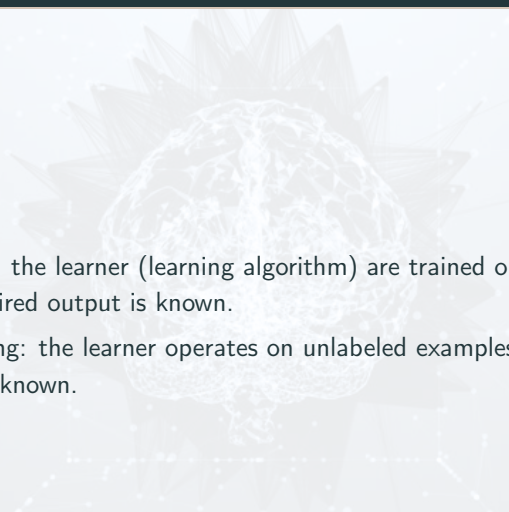
7	+	+	+	-	+	?
8	+	-	+	-	+	?
9	+	+	+	-	-	?

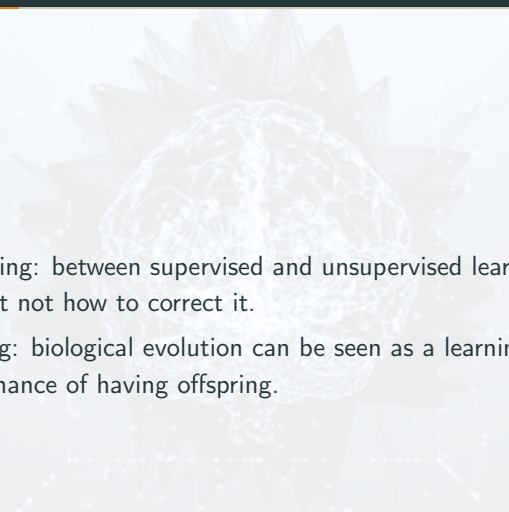
A large, faint, stylized graphic of a brain with neural connections is centered in the background. The brain is composed of a complex network of white lines and dots, giving it a digital or neural appearance. It is surrounded by a larger, more abstract shape that resembles a star or a flower with many points, also made of white lines and dots. The entire background is a light gray color.

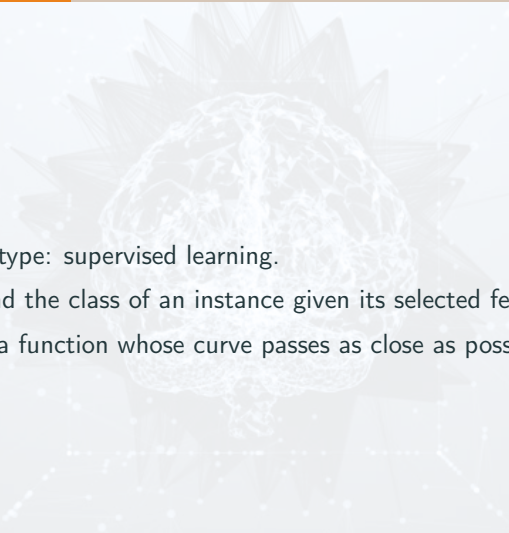
What is learning?

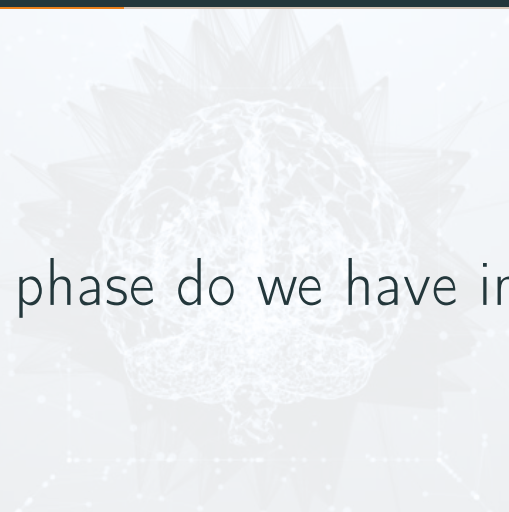
A faint, stylized graphic of a lotus flower is centered in the background. The lotus is composed of many small, interconnected nodes and lines, giving it a digital or network-like appearance. The petals are layered, and the overall color is a light, muted green or grey, blending into the light blue background.

Learning is an (endless) generalization or induction process.

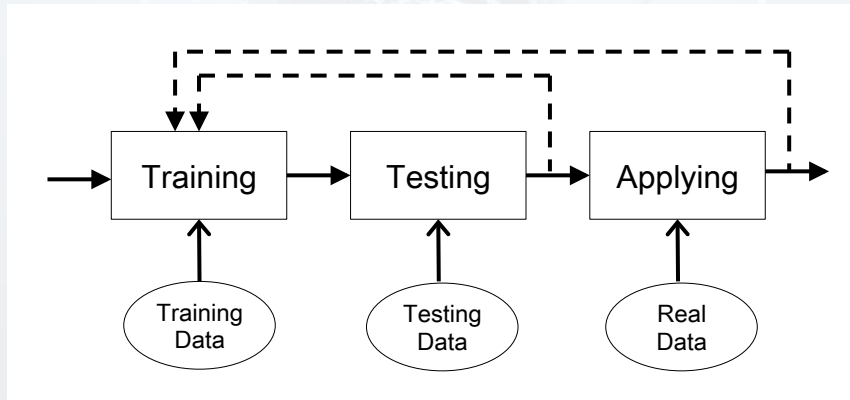
- 
- A faint, stylized background graphic of a human brain with a network of white lines representing neural connections or data flow, set against a light gray background.
- Supervised learning: the learner (learning algorithm) are trained on labeled examples, i.e., input where the desired output is known.
 - Unsupervised learning: the learner operates on unlabeled examples, i.e., input where the desired output is unknown.

- 
- A large, faint, stylized graphic of a lotus flower is centered in the background. The lotus is composed of many small, interconnected dots and lines, giving it a digital or network-like appearance. The petals are layered, and the overall color is a light gray or off-white, blending into the light blue background.
- Reinforcement learning: between supervised and unsupervised learning. It is told when an answer is wrong, but not how to correct it.
 - Evolutionary learning: biological evolution can be seen as a learning process, to improve survival rates and chance of having offspring.

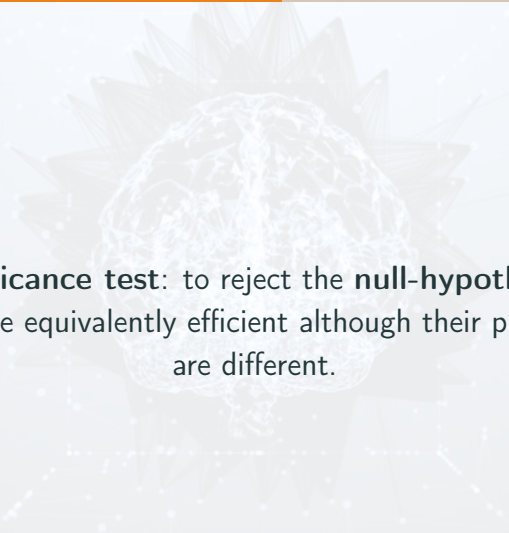
- 
- A faint, stylized background graphic of a human brain with a network of white lines representing neural connections or data flow, set against a light gray background.
- The most common type: supervised learning.
 - Classification: to find the class of an instance given its selected features.
 - Regression: to find a function whose curve passes as close as possible to all of the given data points.

A faint, stylized graphic of a brain with neural connections is centered in the background. The brain is composed of a network of white lines and dots, giving it a digital or neural appearance. It is surrounded by a larger, more abstract shape that resembles a sunburst or a starburst, also made of white lines and dots. The entire background is a light gray color.

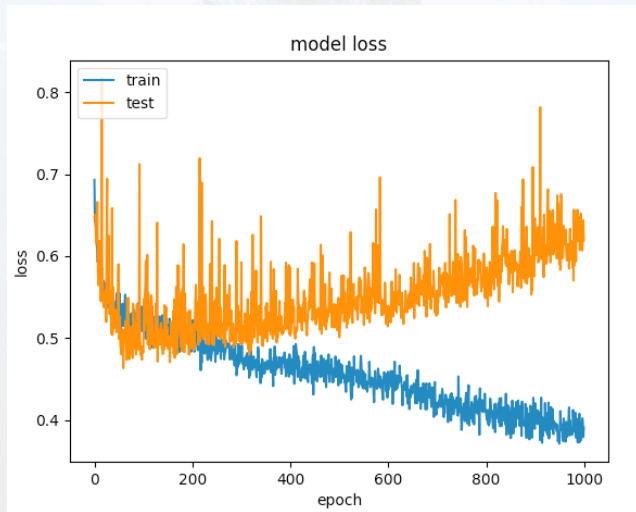
How many phase do we have in machine learning?



- K-fold cross validation:
 - Randomly partitioned k equal sized sub-samples.
 - $k - 1$ for training and 1 for testing.
 - k times (folds) of validation and taking the average.

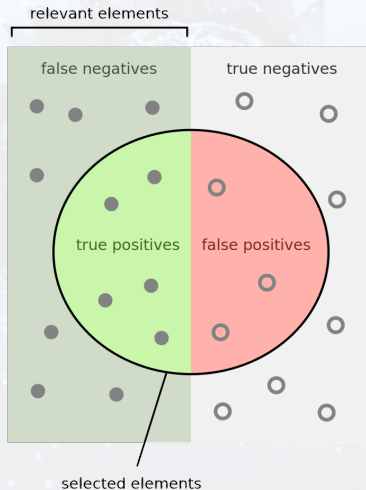
A faint, stylized background graphic of a human brain with a network of white lines representing neural connections or data flow, set against a light gray background.

Statistical significance test: to reject the **null-hypothesis** that the two compared systems are equivalently efficient although their performance measures are different.



Overfitting

- There is noise in the data
- The number of training examples is too small to produce a representative sample of the target concept.



- Precision:

$$P = \frac{\text{number of correct system answers}}{\text{number of system answers}}$$

- Recall:

$$R = \frac{\text{number of correct system answers}}{\text{number of correct problem answers}}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

F1 score: want to seek a balance between Precision and Recall
It is good when there is an uneven class distribution.

$$F1 = 2 \frac{P * R}{P + R}$$

Example	Quality	Price	Buy
1	Good	Low	Yes
2	Bad	High	No
3	Good	High	?
4	Bad	Low	?

- A learner that makes no **prior assumptions** regarding the **identity of the target concept** cannot classify any unseen instances.

- A learner that makes no **prior assumptions** regarding the **identity of the target concept** cannot classify any unseen instances.
- A learner that makes no **a priori** assumptions regarding the identity of the target concept has no rational basis for classifying any unseen instances.

- A learner that makes no **prior assumptions** regarding the **identity of the target concept** cannot classify any unseen instances.
- A learner that makes no **a priori** assumptions regarding the identity of the target concept has no rational basis for classifying any unseen instances.
- The **inductive bias** (learning bias): the set of assumptions that the learner uses to predict outputs given inputs that it has not encountered.

Common inductive bias in ML:

- **Maximum conditional independence:** if the hypothesis can be cast in a Bayesian framework, try to maximize conditional independence (Naive Bayes classifier).

Common inductive bias in ML:

- **Maximum conditional independence:** if the hypothesis can be cast in a Bayesian framework, try to maximize conditional independence (Naive Bayes classifier).
- **Minimum cross-validation error:** when trying to choose among hypotheses, select the hypothesis with the lowest cross-validation error.

Common inductive bias in ML:

- **Maximum conditional independence:** if the hypothesis can be cast in a Bayesian framework, try to maximize conditional independence (Naive Bayes classifier).
- **Minimum cross-validation error:** when trying to choose among hypotheses, select the hypothesis with the lowest cross-validation error.
- **Maximum margin:** when drawing a boundary between two classes, attempt to maximize the width of the boundary (SVM). The assumption is that distinct classes tend to be separated by wide boundaries.

Common inductive bias in ML:

- **Minimum description length:** when forming a hypothesis, attempt to minimize the length of the description of the hypothesis. The assumption is that simpler hypotheses are more likely to be true.

Common inductive bias in ML:

- **Minimum description length:** when forming a hypothesis, attempt to minimize the length of the description of the hypothesis. The assumption is that simpler hypotheses are more likely to be true.
- **Minimum features:** unless there is good evidence that a feature is useful, it should be deleted.

Common inductive bias in ML:

- **Minimum description length:** when forming a hypothesis, attempt to minimize the length of the description of the hypothesis. The assumption is that simpler hypotheses are more likely to be true.
- **Minimum features:** unless there is good evidence that a feature is useful, it should be deleted.
- **Nearest neighbors:** assume that most of the cases in a small neighborhood in feature space belong to the same class.