Introduction

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.

Example

Experience

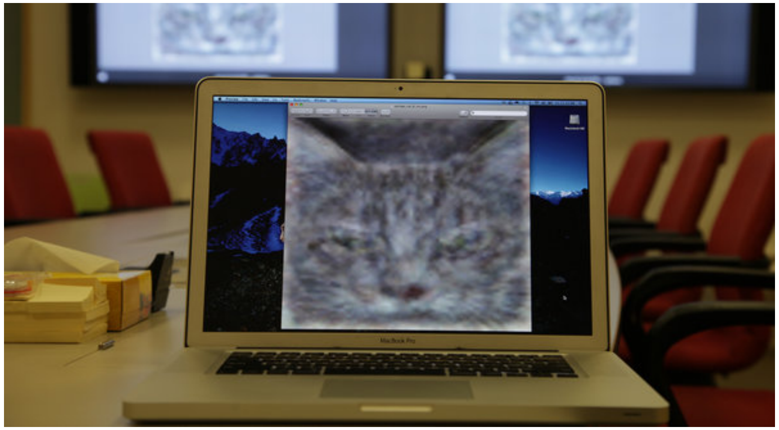
Example	GRAY?	MAMMAL?	LARGE?	VEGETARIAN?	WILD?	Elephant
1	+	+	+	+	+	+
2	+	+	+	-	+	+
3	+	+	-	+	+	- (*Mouse*)
4	-	+	+	+	+	- (*Giraffe*)
5	+	-	+	-	+	- (*Dinosaur*)
6	+	+	+	+	-	+

Prediction

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

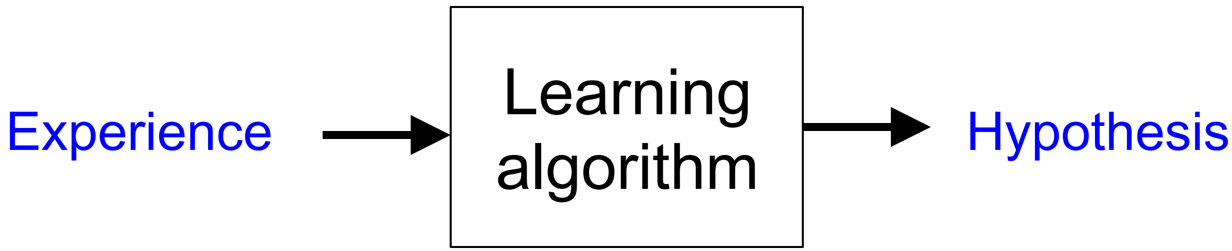
Example

- Deep learning: developed by a research group at Stanford and Google X.
- A system of 16,000 connected computer processors that can learn concepts without supervision.
- Featured in The New Youk Times in 2012.



Machine Learning

- What is learning?



Machine Learning

- Learning is an (endless) generalization or induction process.

Types of Machine Learning

- Supervised learning: the learner (learning algorithm) are trained on labelled examples, i.e., input where the desired output is known.

- Unsupervised learning: the learner operates on unlabelled examples, i.e., input where the desired output is unknown.

Types of Machine Learning

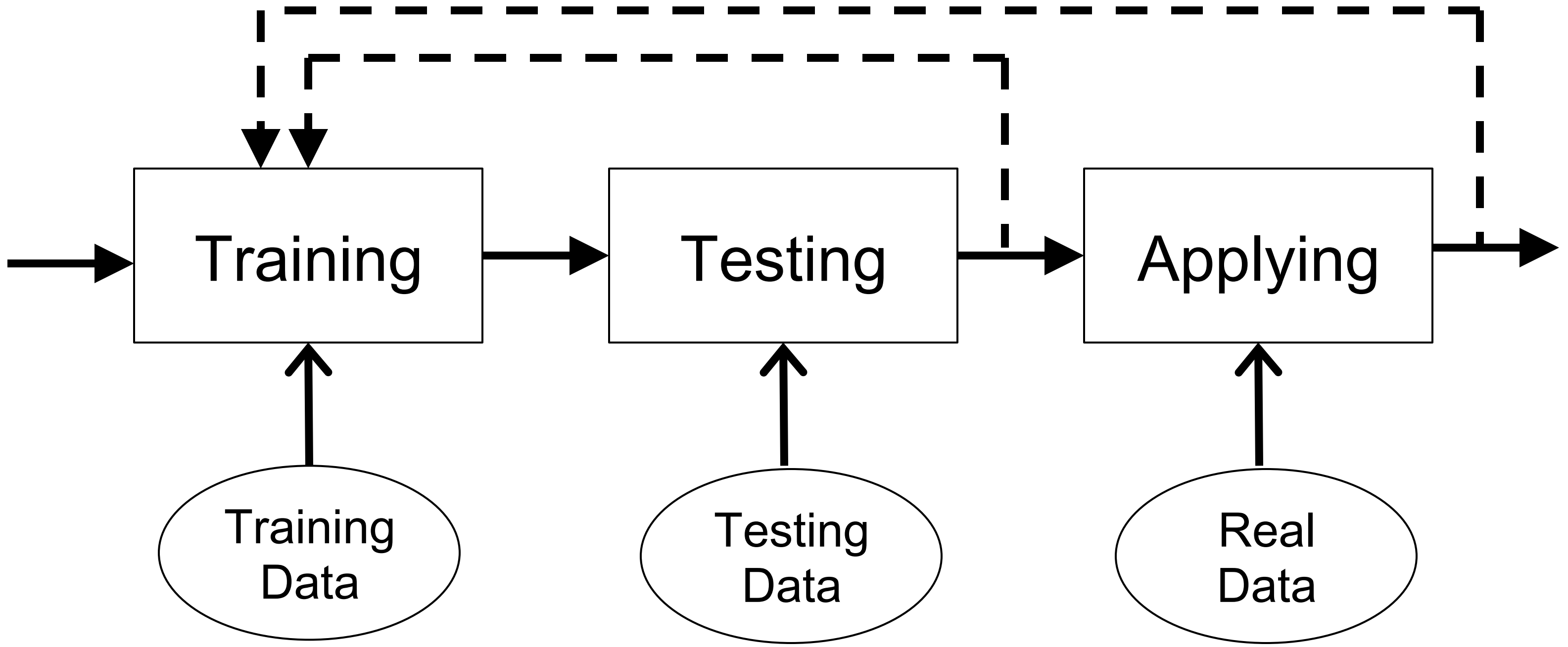
- Reinforcement learning: between supervised and unsupervised learning. It is told when an answer is wrong, but not how to correct it.

- Evolutionary learning: biological evolution can be seen as a learning process, to improve survival rates and chance of having offspring.

Types of Machine Learning

- The most common type: supervised learning.
  - Regression: to find a function whose curve passes as close as possible to all of the given data points.
  - Classification: to find the class of an instance given its selected features.

Phases of Machine Learning



Phases of Machine Learning

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

Phases of Machine Learning

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

Phases of Machine Learning

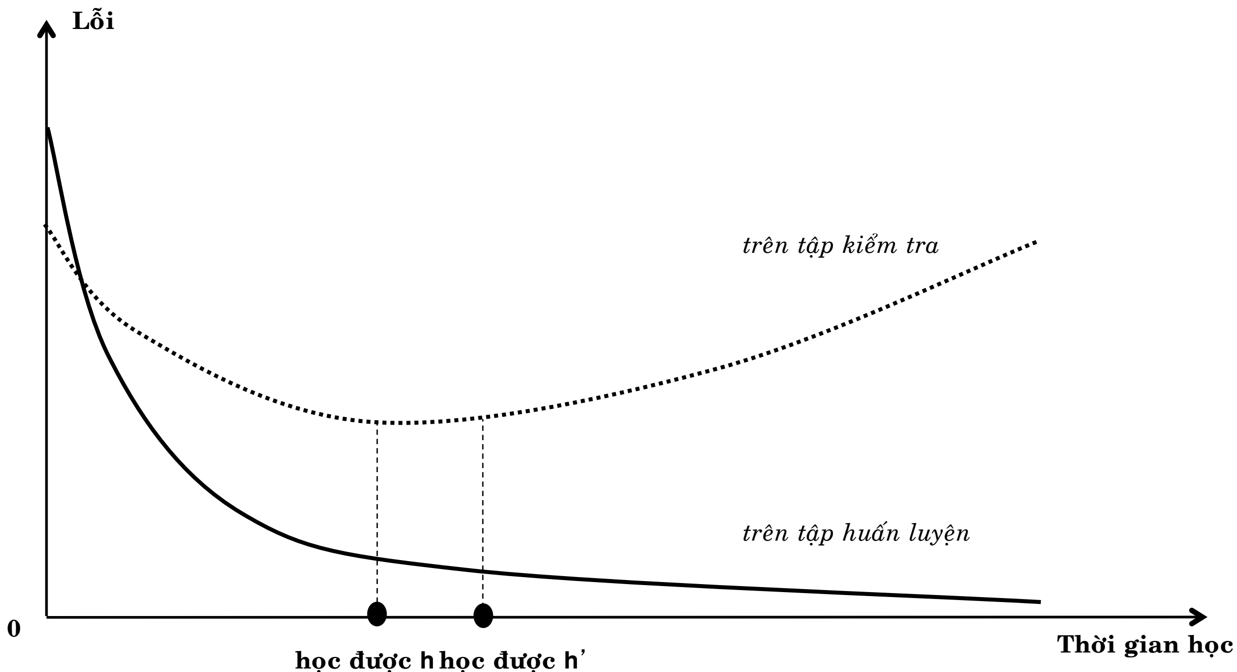
- Fisher's randomization:
  - Q testing cases.
  - $\delta = |m(A) - m(B)|$
  - $2^{|Q|}$ permutations of performances of A and B on Q cases.
  - N^+ = number of permutations whose A-B performance difference is greater than or equal to δ.
  - N^- = number of permutations whose A-B performance difference is smaller than or equal to $-\delta$.
  - two-sided p-value = $(N^+ + N^-)/2^{|Q|}$
  - $p \leq 0.05$ to reject the null-hypothesis

Phases of Machine Learning

- Overfitting: $h \in H$ is said to overfit the training data if there exists $h' \in H$, such that h has smaller error than h' over the training examples, but h' has a smaller error than h over the entire distribution of instances.

Phases of Machine Learning

- Overfitting:



Phases of Machine Learning

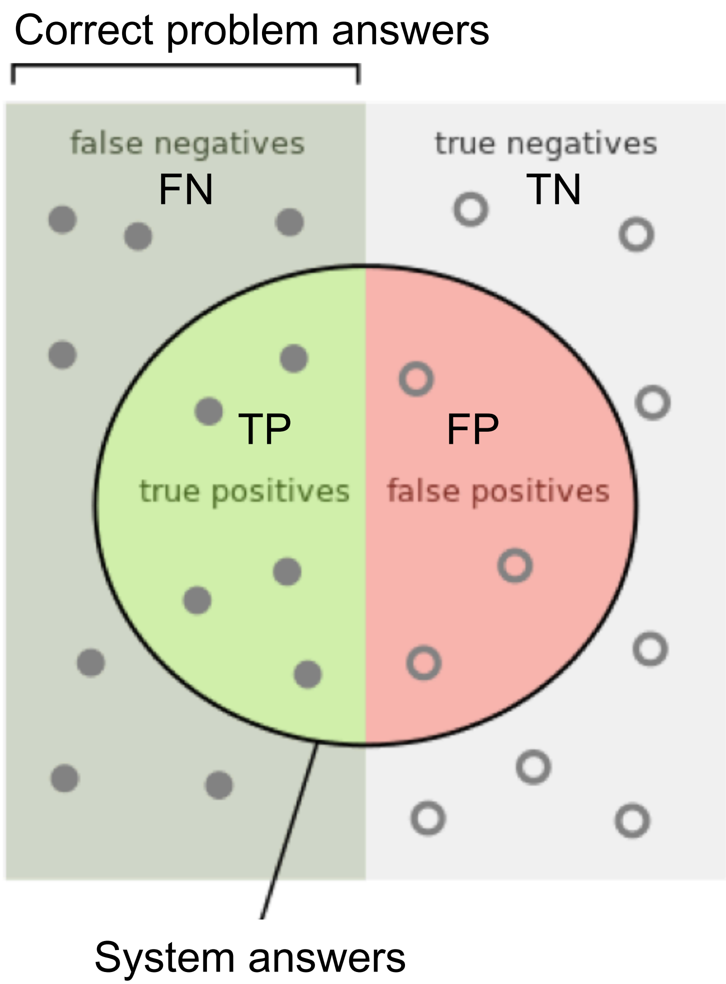
- Overfitting:
  - There is noise in the data
  - The number of training examples is too small to produce a representative sample of the target concept

Performance Measures

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

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

Performance Measures



Precision = TP/(TP + FP)

Recall = TP/(TP + FN)

Accuracy = (TP + TN)/(TP + TN + FP + FN)

Performance Measures

- **Precision** (P) = $\dfrac{\text{number of correct system answers}}{\text{number of system answers}}$

- **Recall** (R) = $\dfrac{\text{number of correct system answers}}{\text{number of correct problem answers}}$

- **F-measure** (F) = 2.P.R/(P + R)