Computational Learning and Discovery

GEORGE MASON UNIVERSITY

CSI 873 / MATH 689

Instructor: I. Griva

Wednesday 7:20 – 10:00 pm

Computational Learning = Machine Learning

Machine learning grew out of the more general area of Artificial Intelligence.

What is ML?

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

Tom Mitchell (1998): ML study algorithms that automatically improve their performance with experience.

Formally: 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.

Informally:

learning is generalization ability, or induction!

Learning can be

- **supervised**
- **unsupervised**
- **reinforced**

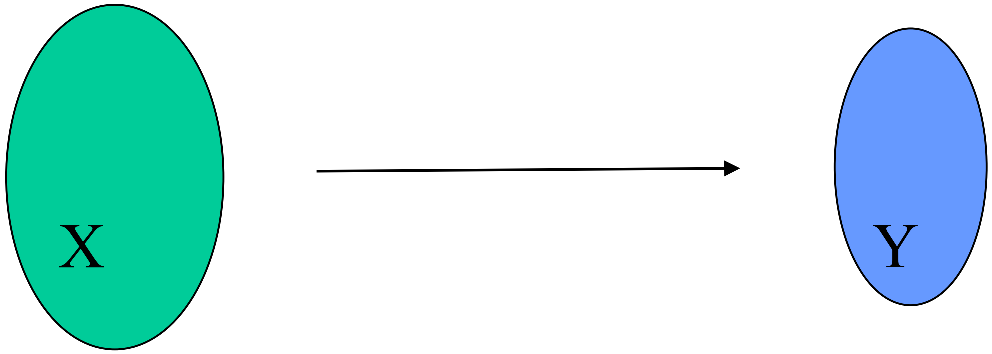
This class will focus mostly on supervised learning.

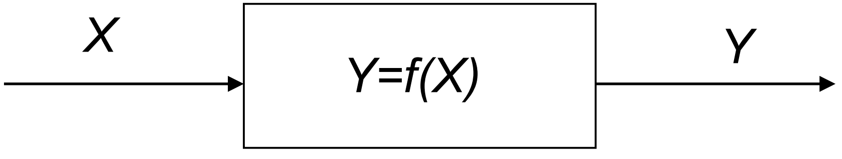
Supervised Learning

Given a set of training data:

$$(x_1, y_1), (x_2, y_2) \ldots, (x_l, y_l)$$

Find a function *Y=f(X)* with good generalization ability.





How can computers learn?

To learn? What does it mean?

Is it to acquire available information?

How can computers learn?

To learn? What does it mean?

Is it to acquire available information?

Is it to create unknown information?

How can computers learn?

To learn? What does it mean?

Is it to acquire available information? Yes!

Is it to create unknown information? Yes!

How can computers learn?

To learn? What does it mean?

Is it to acquire available information? Yes!

Is it to create unknown information? Yes!

What is the difference?

How can computers learn?

To learn? What does it mean?

Is it to acquire available information? Yes!

Is it to create unknown information? Yes!

What is the difference?

The latter is generalization!

The latter requires information processing!

What do we need for information processing?

Information! Processor!

Processing algorithms! Learning assumptions!

Measure of performance, or some criterion!

How the price of information processing can be characterized? Algorithmic complexity.

So the goal of a learning is ...

To create unknown information =

To generalize well with respect to some selected criterion

Looks challenging but important!

We may not know how to generalize ourselves.

But we want to construct algorithms of information processing that will be able to generalize well.

Looks challenging but important!

We may not know best conclusion ourselves.

But we want to construct algorithms of information processing that will be able to generalize well.

We want a computer to think and create information for us!

The class is about information processing algorithms that

Broadly:

...allow a computer to generalize!!!

What specifically do we want from a computer?

What specifically do we want from a computer?

What is the result of the information processing?

What specifically do we want from a computer?

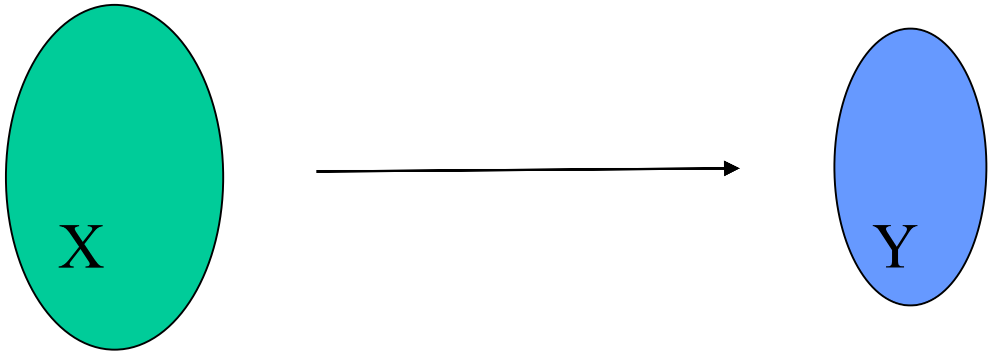
What is the result of the information processing?

It will be a function defined in a broad sense.

Given a set of training data:

$$(x_1, y_1), (x_2, y_2) \ldots, (x_l, y_l)$$

Find a function $Y = f(X)$ **with good generalization ability.**



Example of the function *Y=f(X)*

Y={Cold, Warm} *X={date}* *Fairfax, VA*

X	*f(X)*
Oct 15 ≤ X ≤ Apr 14	*Cold*
Apr 15 ≤ X ≤ Oct 14	*Warm*

Completely defined function =
Completely defined set of hypotheses

Example of the function $Y=f(X_1,X_2)$

$Y=\{$"I swim", "I do not swim"$\}$

$X_1=\{date\}$ $X_2=\{$"rain", "no rain"$\}$

X_1	X_2	$Y=f(X_1,X_2)$
May 27 ≤ X_1 ≤ Sep 4	rain	I do not swim
May 27 ≤ X_1 ≤ Sep 4	no rain	I swim
Sep 4 ≤ X_1 ≤ May 27	rain	I do not swim
Sep 4 ≤ X_1 ≤ May 27	no rain	I do not swim

Example of the function $Y=f(X)$

$Y=\{Cold, Warm\}$ $X=\{date\}$ Fairfax, VA

X	$f(X)$
Oct 15 $\leq X \leq$ Apr 14	Cold
Apr 15 $\leq X \leq$ Oct 14	Warm

Example of the function $Y=f(X)$

$Y=\{Cold, Warm\}$ $X=\{date\}$ *Fairfax, VA*

X	$f(X)$
Oct 15 $\leq X \leq$ Apr 14	*Cold*
Apr 15 $\leq X \leq$ Oct 14	*Warm*

X	$f(X)$
Dec 15 $\leq X \leq$ Jun 14	*Cold*
Jun 15 $\leq X \leq$ Dec 14	*Warm*

Example of the function $Y=f(X, A, B)$

$Y=\{Cold, Warm\}$ $X=\{date\}$ *Fairfax, VA*

X	$f(X)$
$A \leq X \leq B$	*Cold*
$B \leq X \leq A + 1$ *year*	*Warm*

Wider class of functions = less restrictive set of hypotheses

A and B are the parameters

Example of the function *Y=f(X, A, B)*

Y={Cold, Warm} *X={date}* *Fairfax, VA*

X	*f(X)*
$A \leq X \leq B$	*Cold*
$B \leq X \leq A + 1$ *year*	*Warm*

Wider class of functions = less restrictive set of hypotheses

A and B are the parameters

To find A and B using the data could be a goal of a particular learning algorithm!

The more we know about the target function, the better!

The more we know about the target function, the better!

Example: Analytical learning

Analytical learning uses prior knowledge and deductive reasoning to augment the information provided by training examples.

The more we know about the target function, the better!

Example: Bayesian learning

If we know that Y is governed by a probability distribution, then we can construct the function based on Bayesian principle.

The more we know about the target function, the better!

What if we do not know anything about the target function?

The more we know about the target function, the better!

What if we do not know anything about the target function?

We have to chose one from some class.

Finding the target function seems to be a very challenging task!

The more we know about the target function, the better!

What if we do not know anything about the target function?

We have to chose one from some class.

Finding the target function seems to be a very challenging task!

Is it possible to avoid finding the target function altogether?

How about imitating the target function instead of finding?!

$$X \longrightarrow \boxed{Y=f(X)} \longrightarrow Y$$

Examples: decision tree learning, artificial neural networks, support vector machines.

Given a set of training data:

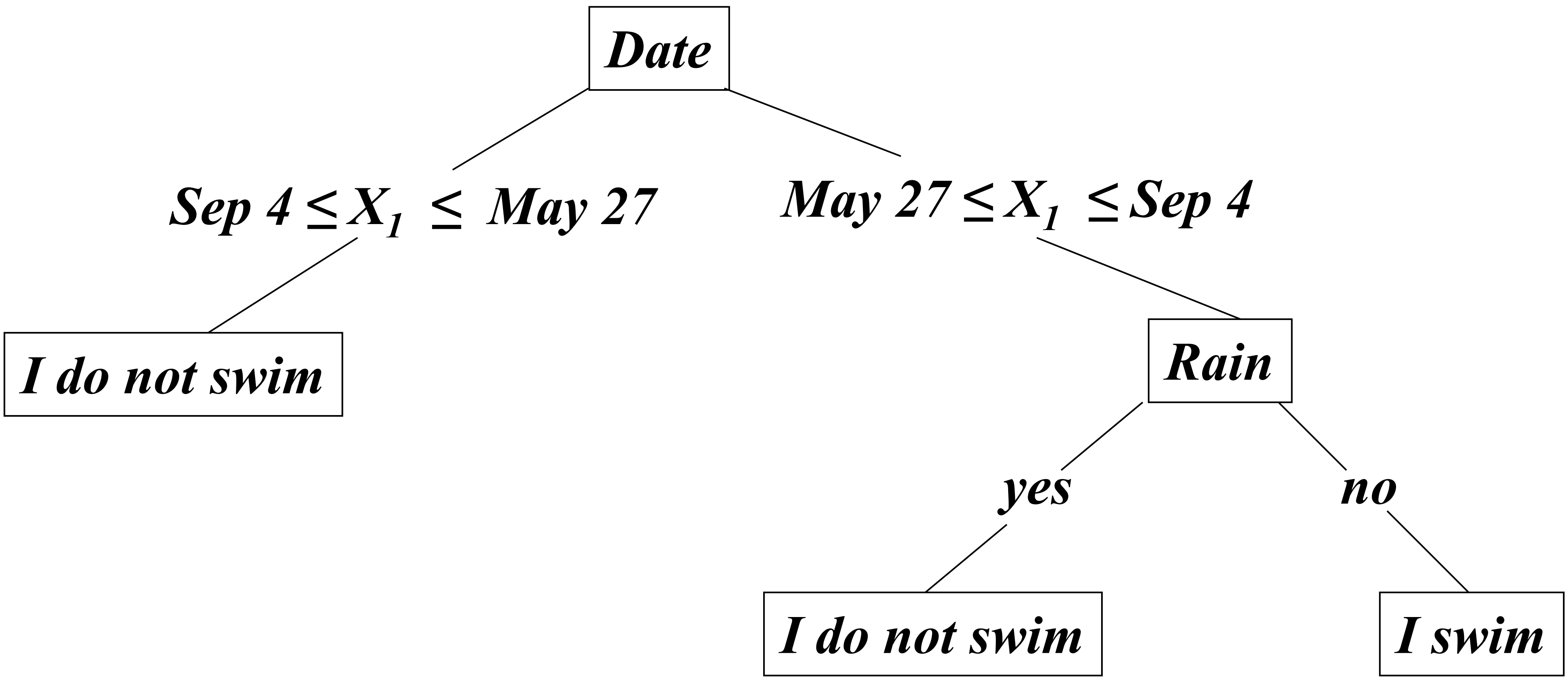
$$(x_1, y_1), (x_2, y_2) \ldots, (x_l, y_l)$$

Find a "black box" with good generalization ability.

Decision tree learning $Y=f(X_1,X_2)$

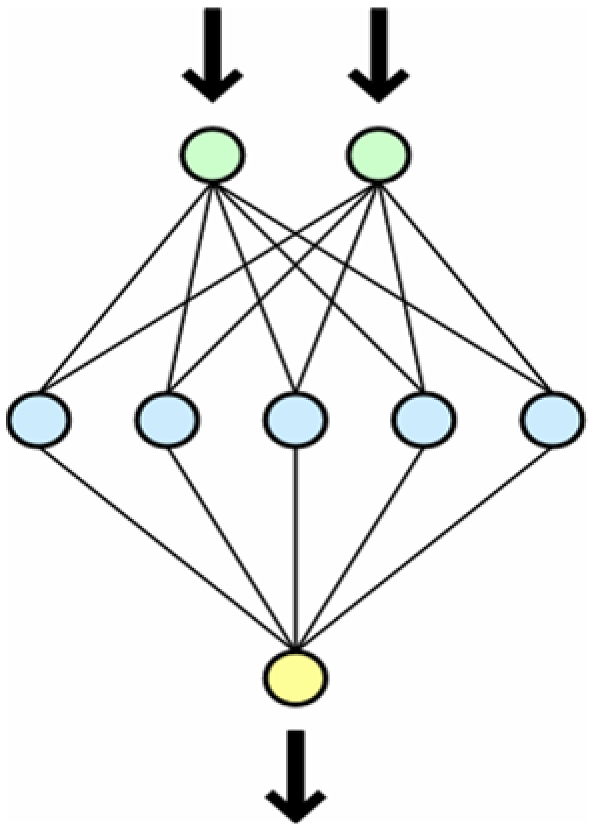
Y={"I swim", "I do not swim"}

X_1 ={date} X_2 ={"rain", "no rain"}



Artificial neural networks

Inspired by the fact that biological learning systems are built of very complex webs of interconnected neurons.

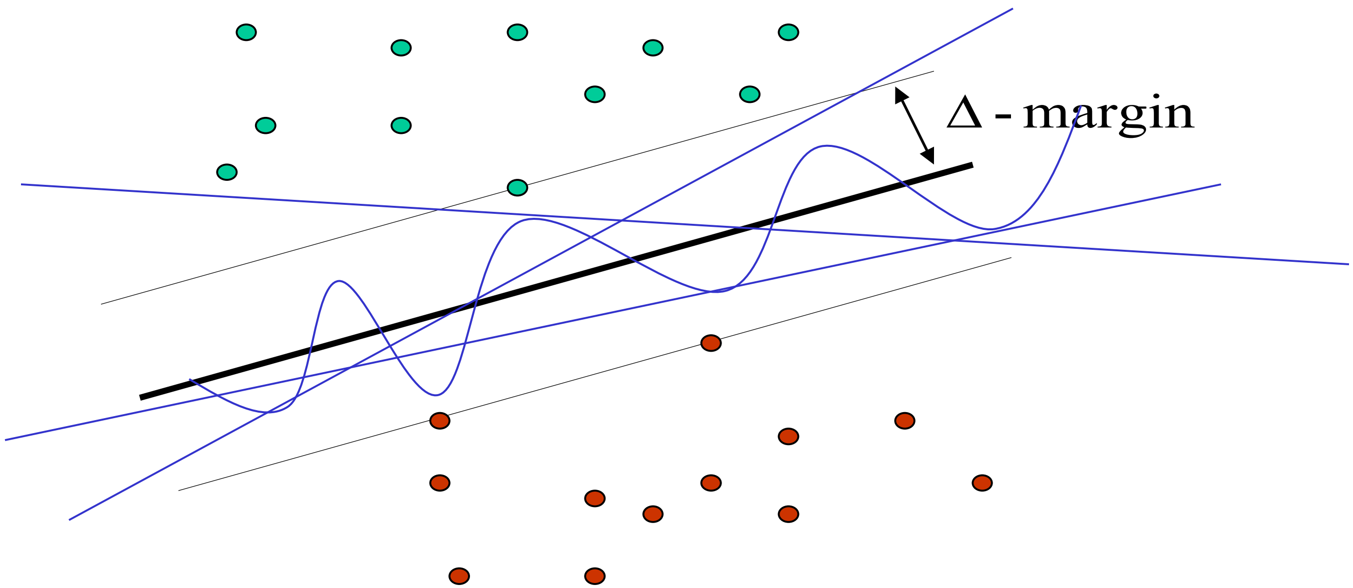


Bayesian learning

Provides probabilistic approach to learning. It is based on the assumption that the quantities of interest are governed by probability distributions and that optimal decisions can be made by reasoning about these probabilities.

Support vector machines

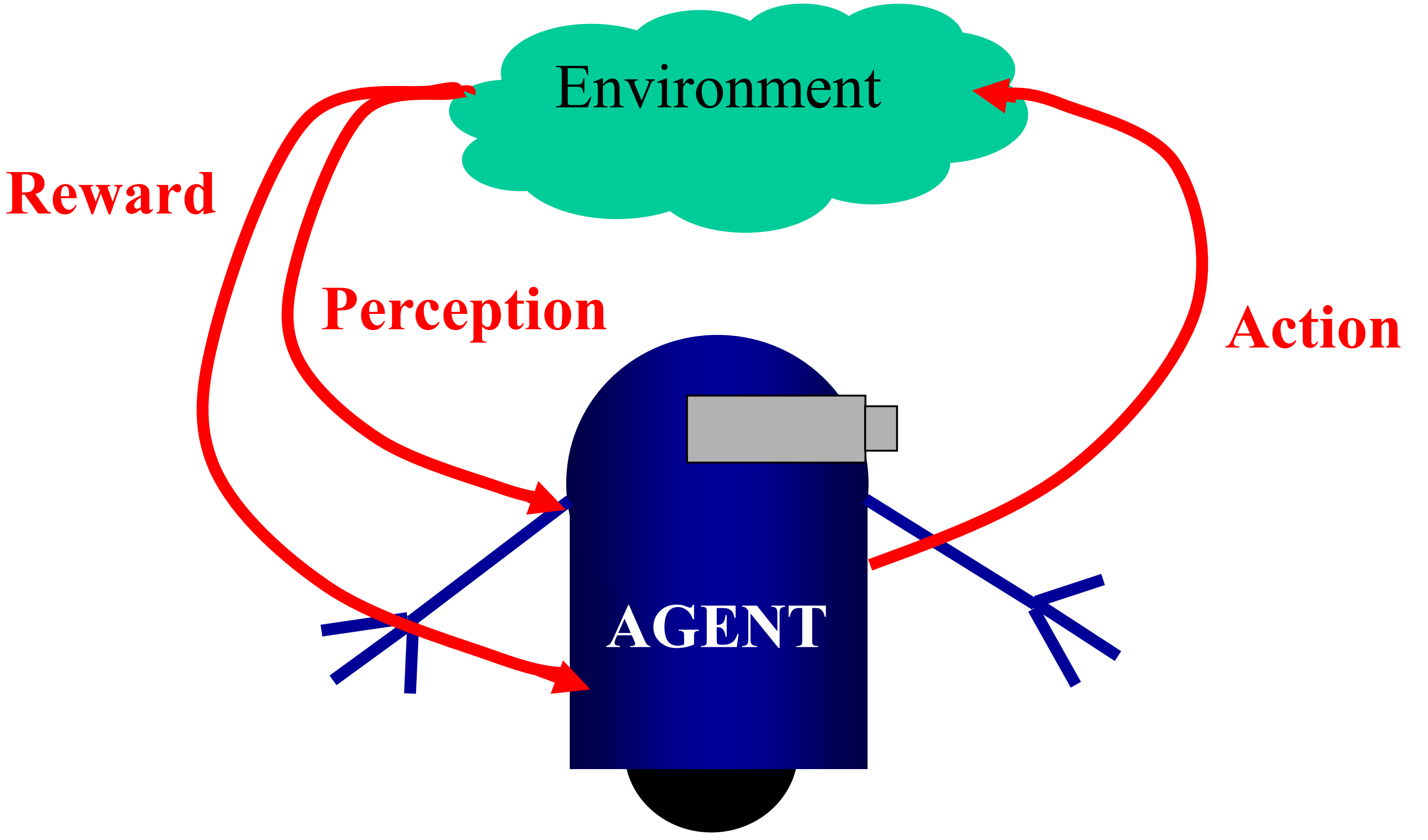
Based on fundamentals of statistical learning theory (Vapnik-Chervonenkis theory)



Instance – based learning.

Genetic algorithms.

Reinforcement learning.



By the end of this class you should be able to

- Understand the basic ideas of ML algorithms
- Identify problems that can be addressed with ML
- Apply ML algorithms for some problems
- Start doing research in ML

Some applications areas for machine learning

- Medical diagnosis
- Credit card applications or transactions
- Fraud detection in e-commerce
- Worm detection in network packets
- Spam filtering in email
- Recommended articles in a newspaper
- Recommended books, movies, music, or jokes
- Financial investments
- DNA sequences
- Spoken words
- Handwritten letters
- Astronomical images
- Playing checkers, chess, or backgammon
- Driving a car
- Learning driving patterns
- Flying a plane, helicopter, or rocket
- Controlling a mobile robot
- Drug discovery

Related research areas:

- Artificial Intelligence
- Data Mining
- Probability and Statistics
- Information theory
- Optimization
- Computational complexity theory
- Control theory (adaptive)
- Psychology (developmental, cognitive)
- Neurobiology
- Linguistics
- Philosophy