ISE 540
Text Analytics

Mayank Kejriwal

Research Assistant Professor/Research Lead

Department of Industrial and Systems Engineering

Information Sciences Institute

USC Viterbi School of Engineering

kejriwal@isi.edu

Learning=Generalization

H. Simon -

"Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the task or tasks drawn from the same population more efficiently and more effectively the next time."

The ability to perform a task in a situation which has never been encountered before

Why study learning?

- Computer systems with new capabilities.
  - Develop systems that are too difficult to impossible to construct manually .
  - Develop systems that can automatically adapt and customize themselves to the needs of the individual users through experience.
  - Discover knowledge and patterns in databases, database mining, e.g. discovering purchasing patterns for marketing purposes.

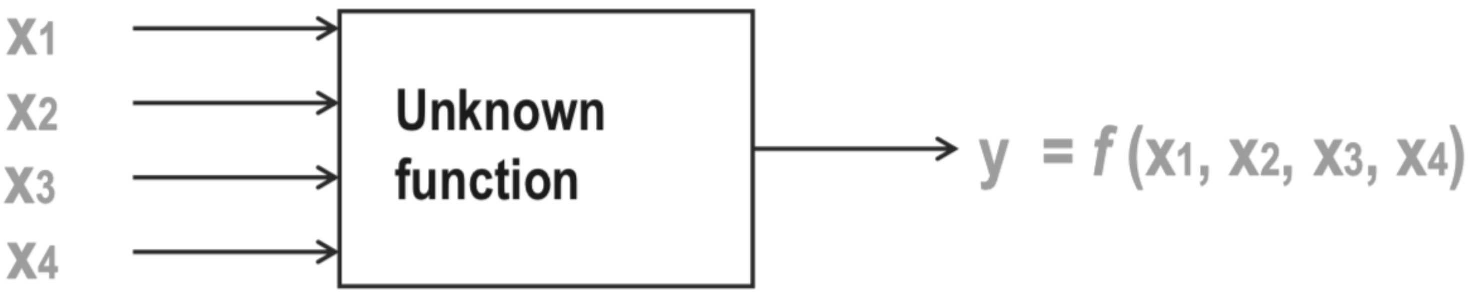
Why study learning?

- Time is right.
  - Initial algorithms and theory in place.
  - Growing amounts of on-line data
  - Computational power available.
  - Necessity: many things we want to do cannot be done by “programming”.

Supervised machine learning

- Given: Examples $(x,f(x))$ of some unknown function f
- Find: A good approximation of f

- x provides some representation of the input
  - The process of mapping a domain element into a representation is called Feature Extraction. (Hard; ill-understood; important)
  - $x \in \{0,1\}^n$ or $x \in \Re^n$
- The target function (label)
  - $f(x) \in \{-1,+1\}$ Binary Classification
  - $f(x) \in \{1,2,3,.,k-1\}$ Multi-class classification
  - $f(x) \in \Re$ Regression



Example	x_1	x_2	x_3	x_4	y
1	0	0	1	0	0
2	0	1	0	0	0
3	0	0	1	1	1
4	1	0	0	1	1
5	0	1	1	0	0
6	1	1	0	0	0
7	0	1	0	1	0

Can you learn this function?

What is it?

Complete Ignorance:

There are 2^{16} = 65536 possible functions over four input features.

We can't figure out which one is correct until we've seen every possible input-output pair.

After seven examples we still have 2^9 possibilities for f

Example	x_1	x_2	x_3	x_4	y
	0	0	0	0	?
	0	0	0	1	?
	0	0	1	0	0
	0	0	1	1	1
	0	1	0	0	0
	0	1	0	1	0
	0	1	1	0	0
	0	1	1	1	?
	1	0	0	0	?
	1	0	0	1	1
	1	0	1	0	?
	1	0	1	1	?
	1	1	0	0	0
	1	1	0	1	?
	1	1	1	0	?
	1	1	1	1	?

Is Learning Possible?

Yes, 2^9 is correct (the total number of possible functions is 2^{total number of 'empty' function values or 'question marks'}

General machine learning strategies

- Develop representation languages for expressing concepts
  - Serve to limit the expressivity of the target models
  - E.g., Functional representation (n-of-m); Grammars; stochastic models;
- Develop flexible hypothesis spaces:
  - Nested collections of hypotheses. Decision trees, neural networks
  - Hypothesis spaces of flexible size

In either case:

- Develop algorithms for finding a hypothesis in our hypothesis space, that fits the data
- And hope that they will generalize well

Terminology

- Target function (concept): The true function $f : X \rightarrow \{...\text{Labels}...\}$
- Concept: Boolean function. Example for which $f(x) = 1$ are positive examples; those for which $f(x) = 0$ are negative examples (instances)

- Hypothesis: A proposed function h, believed to be similar to f. The output of our learning algorithm.
- Hypothesis space: The space of all hypotheses that can, in principle, be output by the learning algorithm.

- Classifier: A discrete valued function produced by the learning algorithm. The possible value of f: $\{1,2,\ldots K\}$ are the classes or class labels. (In most algorithms the classifier will actually return a real valued function that we'll have to interpret).

- Training examples: A set of examples of the form $\{(x, f(x))\}$

Key issues

- Modeling
  - How to formulate application problems as machine learning problems ? How to represent the data?
  - Learning Protocols (where is the data & labels coming from?)

- Representation:
  - What are good hypothesis spaces ?
  - Any rigorous way to find these? Any general approach?

- Algorithms:
  - What are good algorithms? (The Bio Exam Model)
  - How do we define success?
  - Generalization Vs. over fitting
  - The computational problem

Generalization vs. overfitting

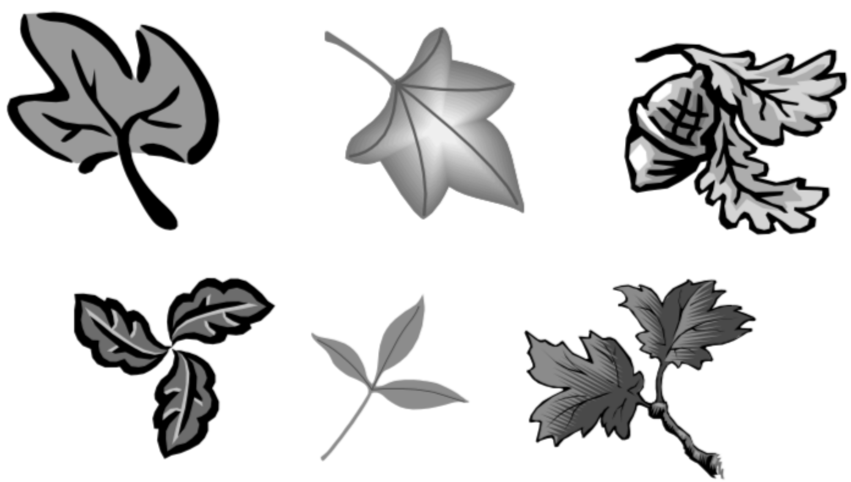
What is a Tree ?

A botanist

A tree is something with leaves I've seen before

Her brother

A tree is a green thing



Neither will generalize well