

# Machine Learning - Lecture 16

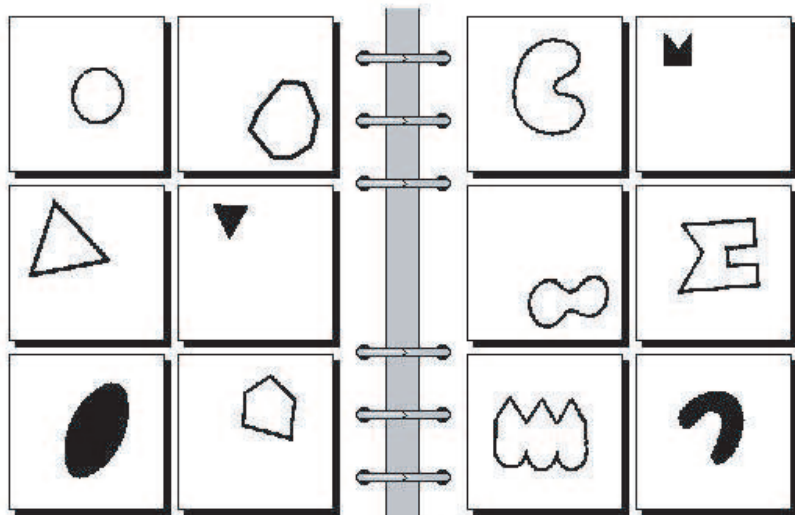
## Machine Discovery

*Chris Thornton*

November 25, 2011

# Illustrative modeling problem

How are the six images on the left different to the six on the right?



# Bongard problems

Six boxes on the left, and another six on the right. The ones on the left conform to a pattern, or rule, and the six on the right don't. The task of the problem-solver is to find this pattern or rule.

# Bongard problems

Six boxes on the left, and another six on the right. The ones on the left conform to a pattern, or rule, and the six on the right don't. The task of the problem-solver is to find this pattern or rule.

- ▶ <http://www.cs.indiana.edu/hfoundal/res/bps/bpidx.htm>

# Bongard problems

Six boxes on the left, and another six on the right. The ones on the left conform to a pattern, or rule, and the six on the right don't. The task of the problem-solver is to find this pattern or rule.

- ▶ <http://www.cs.indiana.edu/hfoundal/res/bps/bpidx.htm>

# Structure involving relationships

Up to this point, the methods we've looked at have all aimed to model patterns in terms of shapes or areas of the data space.

Not all patterns are of this form.

Where classifications are based on relationships between values, there is no dependency between classes and *absolute* values.

So, no reason for examples of a particular class to gather in any particular part of the space.

# Letter analogy problems

If 'abc' goes to 'abd', what does 'ijk' go to?

Popular answers:

# Letter analogy problems

If 'abc' goes to 'abd', what does 'ijk' go to?

Popular answers:

- ▶ ijl - replace the rightmost letter by its successor,



# Letter analogy problems

If 'abc' goes to 'abd', what does 'ijk' go to?

Popular answers:

- ▶ ijl - replace the rightmost letter by its successor,
- ▶ ijd - replace the rightmost letter by d,

# Letter analogy problems

If 'abc' goes to 'abd', what does 'ijk' go to?

Popular answers:

- ▶ ijl - replace the rightmost letter by its successor,
- ▶ ijd - replace the rightmost letter by d,
- ▶ ijk - replace c with d.

# Letter analogy problems

If 'abc' goes to 'abd', what does 'ijk' go to?

Popular answers:

- ▶ ijl - replace the rightmost letter by its successor,
- ▶ ijd - replace the rightmost letter by d,
- ▶ ijk - replace c with d.

Let's say we have data giving examples of such problems.

What are the significant patterns?

How could they be identified and modeled?

# Spot the rule?

13	2	2	3	8	2	2	1	2	4	-->	4
8	3	6	4	6	2	8	3	8	1	-->	5
12	1	5	2	3	3	3	2	3	1	-->	4
13	4	13	3	8	2	8	1	8	3	-->	5
9	3	10	1	11	2	12	1	13	4	-->	7
10	4	10	3	1	3	1	4	10	2	-->	5
13	4	11	4	11	3	13	4	13	4	-->	5
9	2	4	2	5	2	13	2	10	2	-->	6
7	4	12	4	12	2	4	2	12	1	-->	4
13	2	8	2	1	3	1	3	1	4	-->	4
10	3	10	1	5	2	13	2	10	2	-->	4
13	4	3	4	4	1	3	4	3	4	-->	4
11	2	8	4	4	4	4	2	4	4	-->	4
11	3	11	4	13	1	13	1	13	3	-->	5
2	3	2	1	2	1	2	2	1	4	-->	9
8	2	2	2	9	2	11	2	13	2	-->	6

# Hand rankings in poker

Input vectors represent a hand of five playing cards.

Input variables are in twos, where the first number is the card value and second number represents the suit.

The class variable is the rank of the hand in poker.

pair < threes < full house < run < etc.

13	2	2	3	8	2	2	1	2	4	-->	4
8	3	6	4	6	2	8	3	8	1	-->	5
9	2	4	2	5	2	13	2	10	2	-->	6
9	3	10	1	11	2	12	1	13	4	-->	7

Should we expect examples of a particular rank to clump together in the data space?

# How can relational structure be identified and modelled?

We need ways to identify and model relationships in the data.

# How can relational structure be identified and modelled?

We need ways to identify and model relationships in the data.

- ▶ What relationships should we look for?

# How can relational structure be identified and modelled?

We need ways to identify and model relationships in the data.

- ▶ What relationships should we look for?
- ▶ Where should we look for them?



# How can relational structure be identified and modelled?

We need ways to identify and model relationships in the data.

- ▶ What relationships should we look for?
- ▶ Where should we look for them?

An early example of a relational method called BACON was developed by Langley and co-workers in the 1970s.

BACON is provided with knowledge of mathematical relationships.

It then searches through the space of possible compositions of those relationships, testing to see how well each one predicts the data.

# BACON discovers Kepler's third law

Using this methodology, BACON achieved a number of successes, including the discovery of Kepler's third law of planetary motion.

This states that the squares of the periods of planets are proportional to the cubes of the mean radii of their orbits.

```
\begin{center} \mbox{\epsfig{file=copied-pics/two-planets-in-orbit.eps}}
```

(In other words, it states that the square of the year is proportional to the cube of the average distance from the sun.)

# Modeling the rule

If  $y$  represents the length of the planet's year and  $d$  represents the average distance from the sun, Kepler's third law states that

$$\frac{y^2}{d^3}$$

is constant.

# How BACON works

In discovering Kepler's third law, BACON starts out with just the raw values of  $y$  and  $d$ .

It then constructs increasingly complex formulae using division and multiplication operators:

Planet	$y$	$d$	$y/d$	$(y/d)/d$	$((y/d)/d)y$	$((y/d)/d)y$
Mercury	0.24	0.39	0.62	1.61	0.39	1.00
Venus	0.61	0.72	0.85	1.18	0.72	1.00
Earth	1.00	1.00	1.00	1.00	1.00	1.00
Mars	1.88	1.52	1.23	0.81	1.52	1.00
Ceres	4.60	2.77	1.66	0.60	2.76	1.00
Jupiter	11.86	5.20	2.28	0.44	5.20	1.00
Saturn	29.46	9.54	3.09	0.32	9.54	1.00
Uranus	84.01	19.19	4.38	0.23	19.17	1.00
Neptune	164.80	30.07	5.48	0.18	30.04	1.00
Pluto	248.40	39.52	6.29	0.16	39.51	1.00
T.Beta	680.00	77.22	8.81	0.11	77.55	1.00

Process stops once a constant value is found



## Other types of BACON

The team behind BACON have created other versions of the program (GLAUBER, STAHL and DALTON et al.) by varying the subset of mathematical relationships used.

Provided that the search space used is appropriately customised, the program is guaranteed to succeed, i.e., to 'discover' whatever law applies.

Hence, these methods are describes as doing **machine dicovery**.

# Problems with BACON

The BACON method is sensitive to noisy data and depending how the search is organised, it may also be sensitive to the instantiation and ordering of variables.

the big problem is that it requires relationships and variables to be configured so as to ensure that the search succeeds.

This is easy enough where the make-up of the target relationship is known.

However, where the aim is to discover regularities of an unknown form, it may be much more challenging.

People have also looked at ways of identifying and modeling analogical relationships.

A prominent approach here is the **structure-mapping** framework of Gentner and colleagues.

The key idea in this is that the *strength* of analogy between two concepts depends on similarities in their relational structure.



# The atom/solar-system analogy

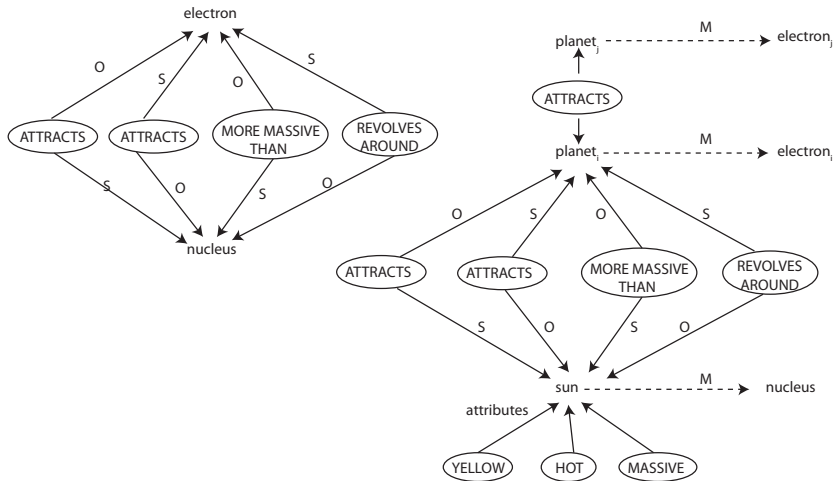
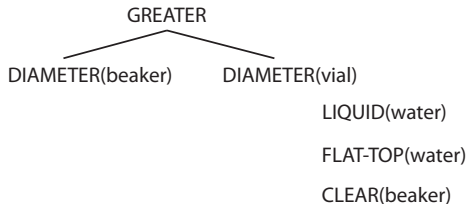
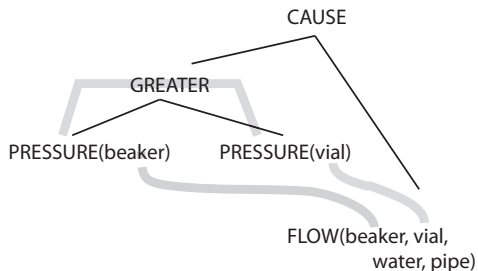


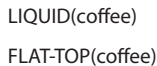
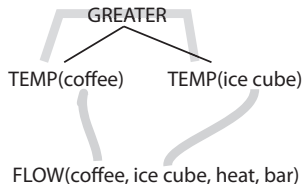
Figure: After Gentner, 1983, p. 160.

# Finding the structure mapping

## WATER FLOW



## HEAT FLOW



# The searcher's dilemma

All these methods search through some space of possible relationships, or relational structures, looking for one which works as a model.

This will only work if the space *contains* a satisfactory model.

So we need to know quite a bit about the solution in order to find it this way.

In simple cases, there may be no difficulty. But in realistic scenarios, it may be very difficult to identify appropriate domain knowledge.

# Summary

- ▶ Patterns may involve relationships

# Summary

- ▶ Patterns may involve relationships
- ▶ Examples of a class distributed, not clumped

# Summary

- ▶ Patterns may involve relationships
- ▶ Examples of a class distributed, not clumped
- ▶ BACON approach

# Summary

- ▶ Patterns may involve relationships
- ▶ Examples of a class distributed, not clumped
- ▶ BACON approach
- ▶ Structure-mapping approach



# Summary

- ▶ Patterns may involve relationships
- ▶ Examples of a class distributed, not clumped
- ▶ BACON approach
- ▶ Structure-mapping approach
- ▶ The searcher's dilemma

# Summary

- ▶ Patterns may involve relationships
- ▶ Examples of a class distributed, not clumped
- ▶ BACON approach
- ▶ Structure-mapping approach
- ▶ The searcher's dilemma

# Questions

- ▶ It took Kepler 10 years to discover his laws of planetary motion. How do we explain the fact that BACON was able to discover the same law in a matter of minutes?

- ▶ It took Kepler 10 years to discover his laws of planetary motion. How do we explain the fact that BACON was able to discover the same law in a matter of minutes?