# *Kinds of Learning*

PSY 322/ORF 322: Human Machine Interactions

Gilbert Harman
Department of Philosophy
Princeton University

Monday, March 1, 2004

## *Two Kinds of Learning*

- Declarative
  - Come up with a theory
  - Use scientific method
- Procedural
  - Acquire an ability
  - Make gradual adjustments, tuning

## *Handwriting Recognition*

- Possible purposes
  - Read addresses written on envelopes
  - Have a person sign something as a method of identification
- Possible approaches
  - Develop a theory of writing and use that theory
  - Use neural net learning
  - (Anecdote about Eric Ristad)

## *Science: Realism/Instrumentalism*

- Realism
  - The aim of science is to use data in order to learn about underlying reality
  - Plato's cave
  - Scientific theories purport to be true descriptions of the world.
  - Scientific theories should be internally consistent and consistent with each other
- Instrumentalism
  - The aim of science is to use data in order to anticipate further experience
  - Technology
  - Scientific theories are instruments of good prediction.
  - There is no requirement that scientific theories be consistent with each other. Incompatible theories can be useful in different ways and for different purposes.

## *East-West Differences in Science*

- History of Science in the East (India, China, Japan)
  - Mainly interested in technology and prediction
  - Heavy use of interpretation in order to square useful theoretical principles with a variety of background cosmologies
  - Attitude about conflicting cosmologies that there is some truth in all; no need to decide among them
  - Eastern technology in advance of Western technology
- History of Science in the West (Greece and Europe)
  - Conflicts in cosmologies and science have to be resolved
  - Interest in the underlying principles rather than in mere prediction or technological applications
  - Increasingly this attitude paid off, eventually leading to scientific advances bringing technological advances beyond what was possible in the East.

## *Approximations in Western Science*

- For many problems scientists use approximations where it would not be feasible to use more accurate principles
- Newton's laws rather than Relativity Theory.
- Other examples?
- The approximate theories are being used as instruments for predication and applications rather than as true representations of reality.
- So, an instrumentalist interpretation of theory is sometimes appropriate even where basic science has realist aims.

## *Strangeness of Contemporary Basic Science*

- Initially, atoms in science were conceived as indivisible particles.
- Bohr solar system model of an atom
- Are electrons really like planets?
- Quantum conceptions: particles or waves
- How many dimensions are there?
- What does it mean to have a Realist conception of current basic science?

## *William of Occam 1280-1349*

## *Occam's Razor*

- Basic principle: given two hypotheses that account equally well for the data, the simpler hypothesis is more likely to be correct.
- This does not imply that the world is simple, whatever that means
  - The data may rule out all fairly simple hypotheses
  - Occam's razor then favors a complicated hypotheses, as long as competing hypotheses are even more complicated
- Question: what justifies the use of Occam's Razor?
- Possible answer: without some such principle there is no way to choose among the infinitely many hypotheses that account equally well for the data

## *Karl Popper 1902-1994*

## *Falsifiability as a Methodological Principle*

- Karl Popper argues that a basic characteristic of scientific as opposed to metaphysical theories is that scientific theories are falsifiable by empirical data
- When theories are defended in a way that makes them unfalsifiable, they are no longer being treated as scientific theories
  - Astrology
  - Psychoanalysis

## *Complications: Auxiliary Assumptions and Noise*

- Auxiliary assumptions are needed to get from theory to predictions about data
  - Assumptions about instruments of observation
  - Assumptions about context of observation
  - One way to defend a theory when a prediction fails is to say the failure is due to a false auxiliary assumption
- Data are noisy
  - Measurements are always inexact
  - Conditions fail to be ideal
  - So even with OK auxiliary assumptions, theory will fit the data only approximately, never perfectly
  - If theory seems to fit the data too well, that can be a sign the data have been fudged—as in Leo Kamin's analysis of Cyril Burt's twin studies.

## *Using Simplicity and Data to Select a Theory*

- Balance empirical adequacy against simplicity. Minimize a hypothesis' empirical loss on the data + complexity of the hypothesis.
- How to measure simplicity?
  - Shortest theory formation, given a way to formulate theories
  - Minimize parameters, grouping theories into classes
  - Different types of simplicity ordering. Second version e.g. puts all linear hypotheses before all quadratic hypotheses

$$f(x) = ax + b$$

$$f(x) = ax^2 + bx + c$$

## *Problems with Parameters*

- The set of sine functions: $f(x) = sin(ax) + b$ almost always contains (infinitely many) functions that capture the data perfectly because sine functions of high frequency can make very fine distinctions.
- So, it is impossible in practice to falsify the hypothesis that the relevant function is a sine function.
- So, using parameters of a set to measure the simplicity of a set of hypotheses will always prefer a sine function over (e.g.) a quadratic function $f(x) = ax^2 + bx + c$.
- Also this violates Popper's falsifiability criterion.

## *Vladimir Vapnik*

## *Statistical Learning Theory and Instrumentalism*

- Traditional statistics
  - Assume the data results from a background probability function plus noise.
  - Use data to estimate the function.
  - Use that function to predict new data.
- Vapnik shows that this problem is usually "ill-posed" and cannot be solved
- Alternative apporach
  - Start with a set of hypotheses $H$
  - Goal: use data to find a hypothesis in $H$ that has close to the least error on new data.
  - Given an appropriate set of hypotheses $H$, this task is not "ill posed" and can be solved
  - Example: let $H$ be the set of linear hypotheses.
- This alternative approach gives up the goal of capturing reality in favor of the goal of finding a hypothesis that yields good predictions.

## *Shattering and VC Theory*

- A set $H$ of hypotheses *shatters* $N$ measurements iff, for every possible result of the measurements, there is a hypothesis in $H$ consistent with that result. So, no possible result of those $N$ measurements could refute the claim that some hypothesis in $H$ is correct.
- Vapnik-Chervonenkis Dimension (VC-Dimension) of a set $H$ of hypotheses is the smallest number $N$ such that no $N$ measurements are shattered by $H$.
  - Example: A linear hypothesis has a VC-dim of 3.

## *Results*

- Vapnik and Chervonenkis prove that it is possible to use empirical evidence to determine the best hypothesis in a set $H$ iff the set has finite VC-dimension.
- More generally, given a set of hypotheses of infinite VC-dimension, one can use VC-dimension as a measure of complexity and choose a hypothesis that minimizes empirical loss against that measure

END