Induction & Decision Trees

CS / Philo 372
week 7

The problem of Induction

- D. Hume (1748)
  - two possible justifications for induction
    - the future must resemble the past
      - must it?
    - it has worked most often in the past, so it will probably continue to work most often in the future.
      - but this uses induction to justify inductuion
- B. Russell (1946)
  - "[induction is] an independent logical principle, incapable of being inferred either from experience or from other logical principles, and that without this principle, science is impossible"
  - So it is a postulate rather than a theorem

Escaping Hume

- K. Popper (1902-1994)
  - Science is not primarily induction but deduction
    - a hypothesis which does not allow for experimental tests of falsity is outside the bounds of science
      - Pauli "it is not even wrong"
- W. Whewell ~1847
  - Induction is a 2 step process
    - colligation: the mental operation of bringing together a number of empirical facts by “superinducing” upon them a conception which unites the facts and renders them capable of being expressed by a general law.
    - Explication: first clarify the colligation, then apply it. E.g., Kepler ellipse of Mars orbit, then apply it to other planets
  - "explication" is the step that justifies induction

Inductive learning
Characterizing the Problem

- Form of the element descriptions upon which learning is to take place
- Form of the feedback
- Representation of the learned Information
- Availability of "prior knowledge"
- Inductive Bias

Form of the available data

- The Physical Symbol System Hypothesis, Newell & Simon 1963
  - "a physical symbol system has the necessary and sufficient means of general intelligent action"
  - Interpreted to mean that "any system exhibiting intelligence must operate by manipulating data structures composed of symbols"
    - understand a picture by breaking it down to named things
    - understand a flower .....
  - 2 questions
    - Does the interpretation follow from the statement
    - Is the interpretation rational / correct?

What organizing principles are in these pictures?

taken from Abu-Mustafa 1995

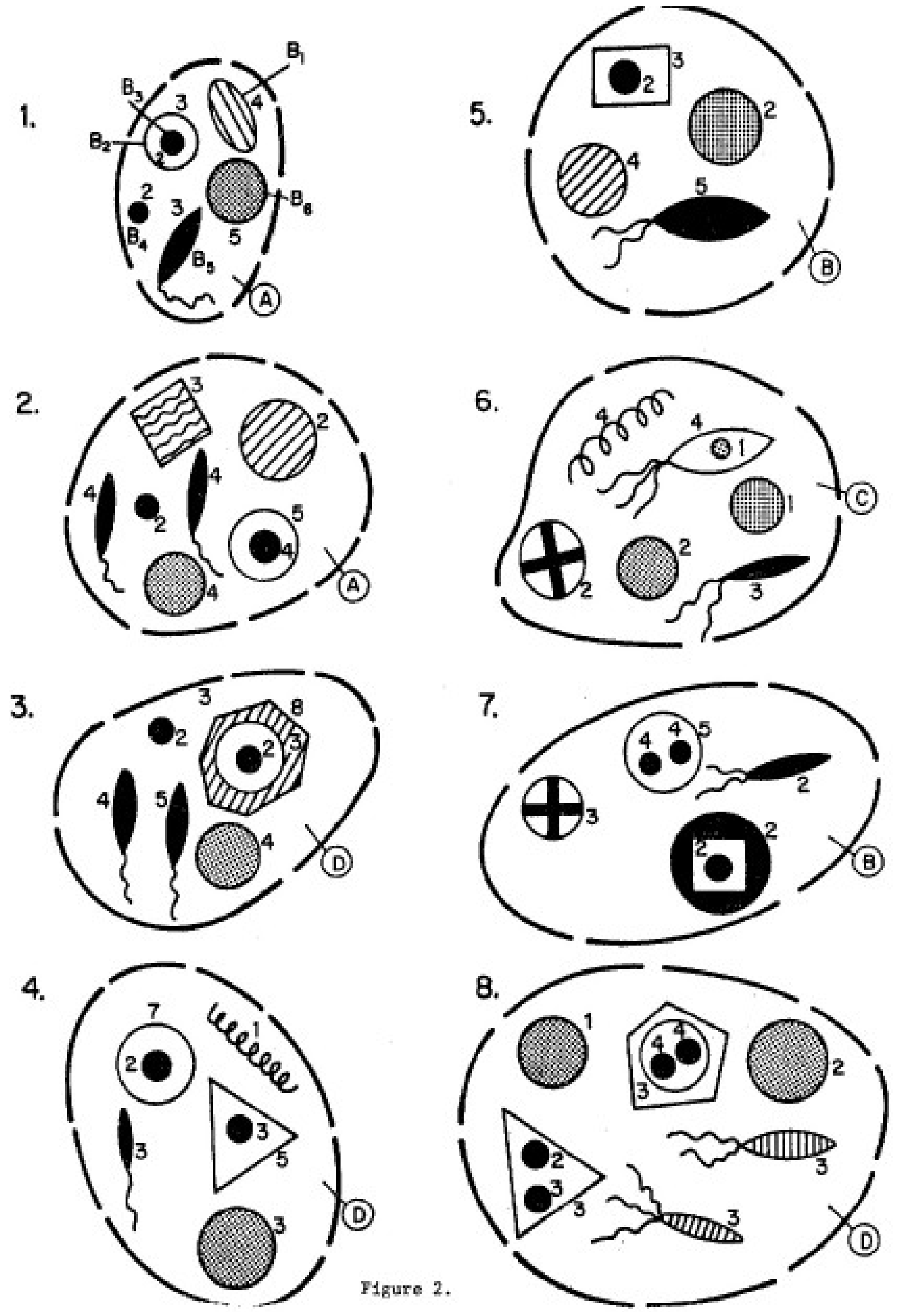


Figure 2.

What are the organizing principles here?

Taken from Michalski 1983

Forms of Feedback

- Supervised Learning
  - A teacher tells you what things are
    - that is poisonous, that is not
- Unsupervised Learning
  - No teacher
    - generally boils down to clustering
- Reinforcement Learning
  - teaching signal from environment at end of trial
    - Backgammon
    - foraging for food
      - I got sick after eating those mushrooms, that grass, ...

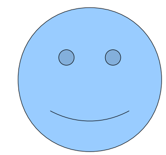
Representation of learning

- Decision Tree, Rules, Bayes nets, polynomials
- Representation matters
  - people want explanations not just instruction
    - explanations help when things go wrong
  - According to Whewell, explainability is vital, without it induction is unjustifiable and incomplete.

Prior Knowledge

- How do you use it?
  - In games often build it into the static evaluation function,
    - e.g. Samuels (1963) checker player
    - OTOH Tesauro's backgammon program
      - even here, the feature detectors represent built in prior knowledge (PSS)
- Learning with "hints"
  - Abu-Mustafa 1995

Induction on Computers

- Desirable Characteristics
  - Occam's razor
    - Directly measurable on a computer 
      - how many bits does it take to represent the hypothesis
      - should be fewer than required to represent the data
  - Generalization
    - accurately predicts examples not previously seen
      - predictability / prediction
      - smoothness
  - Speed
    - needs few examples to learn
    - CPU time for performance task is minimal
  - Explication

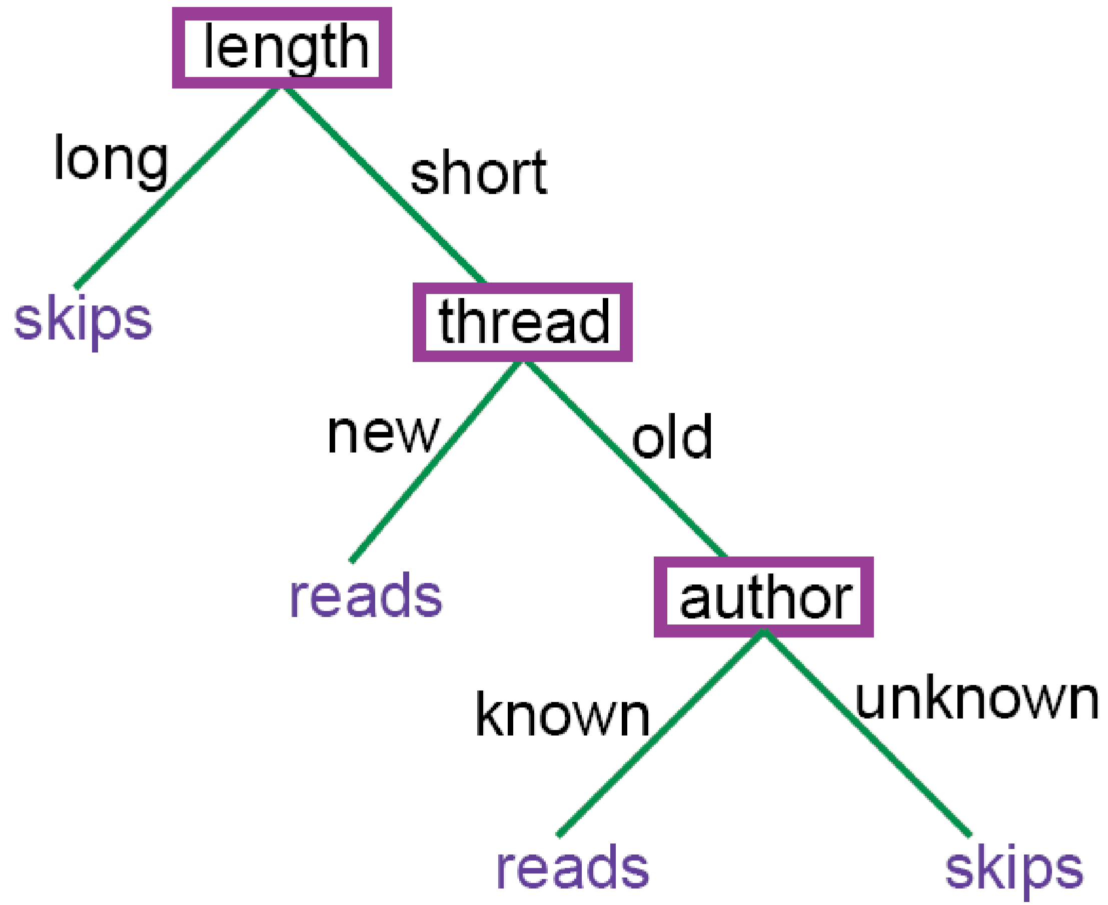
Decision Trees

- Bias: simple trees are preferred to complex ones (an Occam bias)
- Bias: Axis orthogonal splits of the data
- Approach: search through the space of decision trees starting with simple ones and getting more complex as needed
- A "decision tree" is a tree whose
  - non-leaf nodes are labeled with attribute/value pairs
  - leaf nodes are labeled with categories

Claims about Decision Trees

- are simple to understand and interpret.
- have value even with little hard data.
- use a white box model.
- can be combined with other decision techniques.

Sample Decision Tree
Choosing a book



Building a Decision Tree

- Pick a feature of the data
  - draw a branch & node for each value of that feature
  - in each node collect all examples that have the value of the attribute
  - If all examples are in the same category STOP
  - Call the examples you have the data and return to first step

http://www.cs.ualberta.ca/%7Eaixplore/learning/DecisionTrees/index.html
http://www.cs.ubc.ca/labs/lci/CIspace/Version4/dTree/index.html

Choosing the order of decisions

- Question – Goal:
  - Generalization accuracy
  - Minimal size tree
  - Cost of information
- Picking a good feature
  - the one that results in the cleanest subcategories?
  - the one that splits the dataset most evenly?
  - the lowest cost, safest, non-destructive?
    - Greedy?
- Build every possible tree – ?
  - which to use from decision "forest"
    - (Murphy & Pazzani, 1994)

A Synthetic Example

XYZ or AB

	A	B	X	Y	Z
YES	T	t	t	t	t
YES	T	t	t	t	F
yES	T	t	t	F	t
yES	T	t	t	F	F
yES	T	t	F	t	t
yES	T	t	F	t	F
yES	T	t	F	F	t
yES	T	t	F	F	F
yES	T	F	t	t	t
NO	T	F	t	t	F
nO	T	F	t	F	t
nO	T	F	t	F	F
nO	T	F	F	t	t
nO	T	F	F	t	F
nO	T	F	F	F	t
nO	T	F	F	F	F
yES	F	t	t	t	t
nO	F	t	t	t	F
nO	F	t	t	F	t
nO	F	t	t	F	F
nO	F	t	F	t	t
nO	F	t	F	t	F
nO	F	t	F	F	t
nO	F	t	F	F	F
yES	F	F	t	t	t
nO	F	F	t	t	F
nO	F	F	t	F	t
nO	F	F	t	F	F
nO	F	F	F	t	t
nO	F	F	F	t	F
nO	F	F	F	F	t
nO	F	F	F	F	F

- Given that we know the answer we can write the tree.
- It has 5 internal nodes and 6 leaves

Mushroom Data

http://www.ics.uci.edu/~mlearn/MLSummary.html

p,x,s,n,t,p,f,c,n,k,e,e,s,s,w,w,p,w,o,p,k,s,u
e,x,s,y,t,a,f,c,b,k,e,c,s,s,w,w,p,w,o,p,n,n,g
e,b,s,w,t,l,f,c,b,n,e,c,s,s,w,w,p,w,o,p,n,n,m
p,x,y,w,t,p,f,c,n,n,e,e,s,s,w,w,p,w,o,p,k,s,u
e,x,s,g,f,n,f,w,b,k,t,e,s,s,w,w,p,w,o,e,n,a,g
e,x,y,y,t,a,f,c,b,n,e,c,s,s,w,w,p,w,o,p,k,n,g
e,b,s,w,t,a,f,c,b,g,e,c,s,s,w,w,p,w,o,p,k,n,m
e,b,y,w,t,l,f,c,b,n,e,c,s,s,w,w,p,w,o,p,n,s,m
p,x,y,w,t,p,f,c,n,p,e,e,s,s,w,w,p,w,o,p,k,v,g
e,b,s,y,t,a,f,c,b,g,e,c,s,s,w,w,p,w,o,p,k,s,m
e,x,y,y,t,l,f,c,b,g,e,c,s,s,w,w,p,w,o,p,n,n,g
e,x,y,y,t,a,f,c,b,n,e,c,s,s,w,w,p,w,o,p,k,s,m
e,b,s,y,t,a,f,c,b,w,e,c,s,s,w,w,p,w,o,p,n,s,g
p,x,y,w,t,p,f,c,n,k,e,e,s,s,w,w,p,w,o,p,n,v,u
e,x,f,n,f,n,f,w,b,n,t,e,s,f,w,w,p,w,o,e,k,a,g
e,s,f,g,f,n,f,c,n,k,e,e,s,s,w,w,p,w,o,p,n,y,u
e,f,f,w,f,n,f,w,b,k,t,e,s,s,w,w,p,w,o,e,n,a,g
p,x,s,n,t,p,f,c,n,n,e,e,s,s,w,w,p,w,o,p,k,s,g
p,x,y,w,t,p,f,c,n,n,e,e,s,s,w,w,p,w,o,p,n,s,u
p,x,s,n,t,p,f,c,n,k,e,e,s,s,w,w,p,w,o,p,n,s,u
e,b,s,y,t,a,f,c,b,k,e,c,s,s,w,w,p,w,o,p,n,s,m
p,x,y,n,t,p,f,c,n,n,e,e,s,s,w,w,p,w,o,p,n,v,g
e,b,y,y,t,l,f,c,b,k,e,c,s,s,w,w,p,w,o,p,n,s,m
e,b,y,w,t,a,f,c,b,w,e,c,s,s,w,w,p,w,o,p,n,n,m
e,b,s,w,t,l,f,c,b,g,e,c,s,s,w,w,p,w,o,p,k,s,m
p,f,s,w,t,p,f,c,n,n,e,e,s,s,w,w,p,w,o,p,n,v,g
e,x,y,y,t,a,f,c,b,n,e,c,s,s,w,w,p,w,o,p,n,n,m
....

1. cap-shape: bell=b,conical=c,convex=x,flat=f, knobbed=k,sunken=s
2. cap-surface: fibrous=f,grooves=g,scaly=y,smooth=s
3. cap-color: brown=n,buff=b,cinnamon=c,gray=g,green=r, pink=p,purple=u,red=e,white=w,yellow=y
4. bruises?: bruises=t,no=f
5. odor: almond=a,anise=l,creosote=c,fishy=y,foul=f, musty=m,none=n,pungent=p,spicy=s
6. gill-attachment: attached=a,descending=d,free=f,notched=n
7. gill-spacing: close=c,crowded=w,distant=d
8. gill-size: broad=b,narrow=n
9. gill-color: black=k,brown=n,buff=b,chocolate=h,gray=g, green=r,orange=o,pink=p,purple=u,red=e, white=w,yellow=y
10. stalk-shape: enlarging=e,tapering=t
11. stalk-root: bulbous=b,club=c,cup=u,equal=e, rhizomorphs=z,rooted=r,missing=?
12. stalk-surface-above-ring: ibrous=f,scaly=y,silky=k,smooth=s
13. stalk-surface-below-ring: ibrous=f,scaly=y,silky=k,smooth=s
14. stalk-color-above-r: brown=n,buff=b,cinnamon=c,gray=g,orange=o, pink=p,red=e,white=w,yellow=y
15. stalk-clr-below-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o, pink=p,red=e,white=w,yellow=y
16. veil-type: partial=p,universal=u
17. veil-color: brown=n,orange=o,white=w,yellow=y
18. ring-number: none=n,one=o,two=t
19. ring-type: cobwebby=c,evanescent=e,flaring=f,large=l, none=n,pendant=p,sheathing=s,zone=z

.....

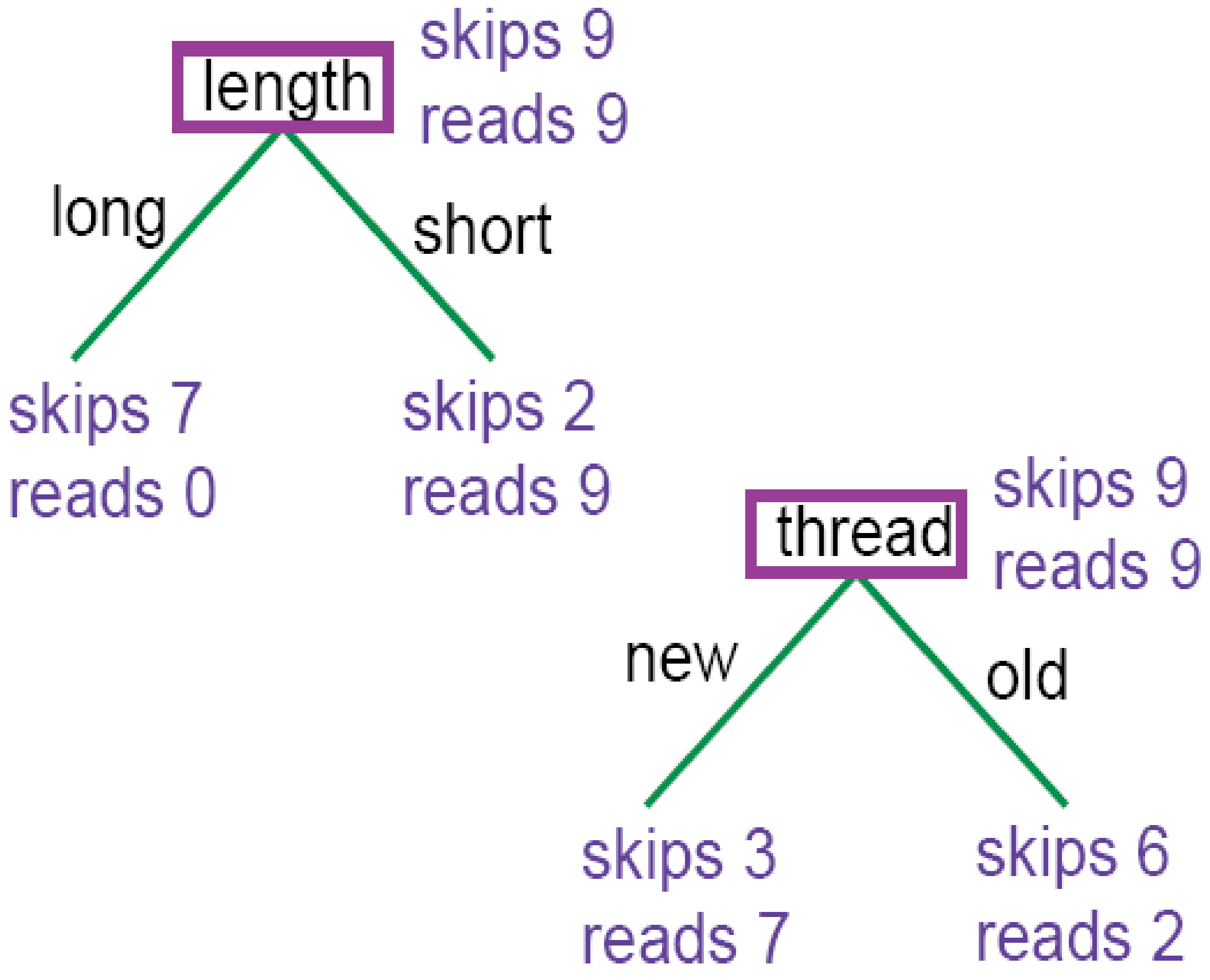
Problems

- Picking good features
- Features that are not categorical
  - e.g. numbers
- Missing data
- Noisy data
- Missing category labels

Picking Good Features

- The ID3 approach
  - Find a leaf that is not all the same catagory.
  - try all possible features to create new leaves
  - select the feature that maximizes the "information gained"
    - roughly this balances homogeneity of the leaves against the number of leaves added.

Picking Good Features (II)



Overfitting

- This algorithm gets into trouble overfitting the data. This occurs with noise and correlations in the training set that are not reflected in the data as a whole.
- To handle overfitting:
  - You can restrict the splitting, so that you split only when the split is useful.
  - You can allow unrestricted splitting and prune the resulting tree where it makes unwarranted distinctions.