

CSE446: Decision Trees

Winter 2016

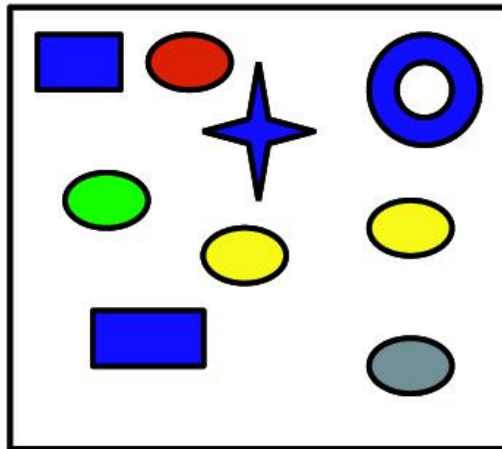
Ali Farhadi

Slides adapted from Carlos Guestrin, Andrew Moore, and Luke Zettlemoyer

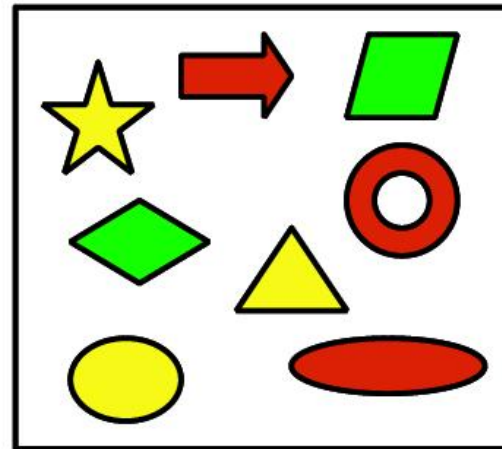
Administrative stuff

- Office hours
- Discussion board
- Anonymous feedback form
- Contact: **cse446-staff@cs.washington.edu**
- No Quiz sections
- Check the webpage regularly
- Will have sections on probability and linear algebra

yes



no



A learning problem: predict fuel efficiency

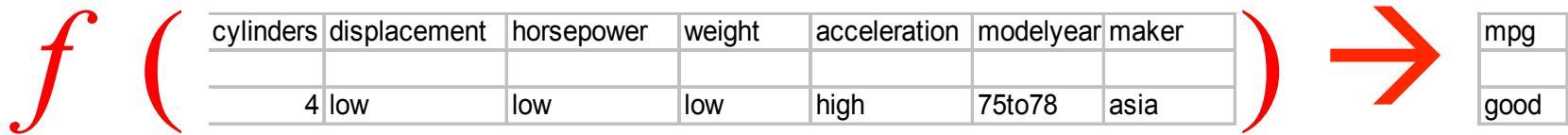
- 40 Records
- Discrete data (for now)
- Predict MPG
- Need to find:
 $f : X \rightarrow Y$

mpg	cylinders	displacement	horsepower	weight	acceleration	modelyear	maker
good	4	low	low	low	high	75to78	asia
bad	6	medium	medium	medium	medium	70to74	america
bad	4	medium	medium	medium	low	75to78	europa
bad	8	high	high	high	low	70to74	america
bad	6	medium	medium	medium	medium	70to74	america
bad	4	low	medium	low	medium	70to74	asia
bad	4	low	medium	low	low	70to74	asia
bad	8	high	high	high	low	75to78	america
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
bad	8	high	high	high	low	70to74	america
good	8	high	medium	high	high	79to83	america
bad	8	high	high	high	low	75to78	america
good	4	low	low	low	low	79to83	america
bad	6	medium	medium	medium	high	75to78	america
good	4	medium	low	low	low	79to83	america
good	4	low	low	medium	high	79to83	america
bad	8	high	high	high	low	70to74	america
good	4	low	medium	low	medium	75to78	europa
bad	5	medium	medium	medium	medium	75to78	europa

Y X

From the UCI repository (thanks to Ross Quinlan)

How to Represent our Function?



Conjunctions in Propositional Logic?

maker=asia \wedge weight=low

Need to find “Hypothesis”: $f : X \rightarrow Y$

Restricted Hypothesis Space

- Many possible representations
- Natural choice: *conjunction* of attribute constraints
- For each attribute:
 - Constrain to a specific value: eg **maker=asia**
 - Don't care: ?
- For example

<u>maker</u>	<u>cyl</u>	<u>displace</u>	<u>weight</u>	<u>accel ...</u>
asia	?	?	low	?

Represents **maker=asia** \wedge **weight=low**

Consistency

- Say an “example is consistent with a hypothesis” when the example *logically satisfies* the hypothesis

- Hypothesis: **maker=asia \wedge weight=low**

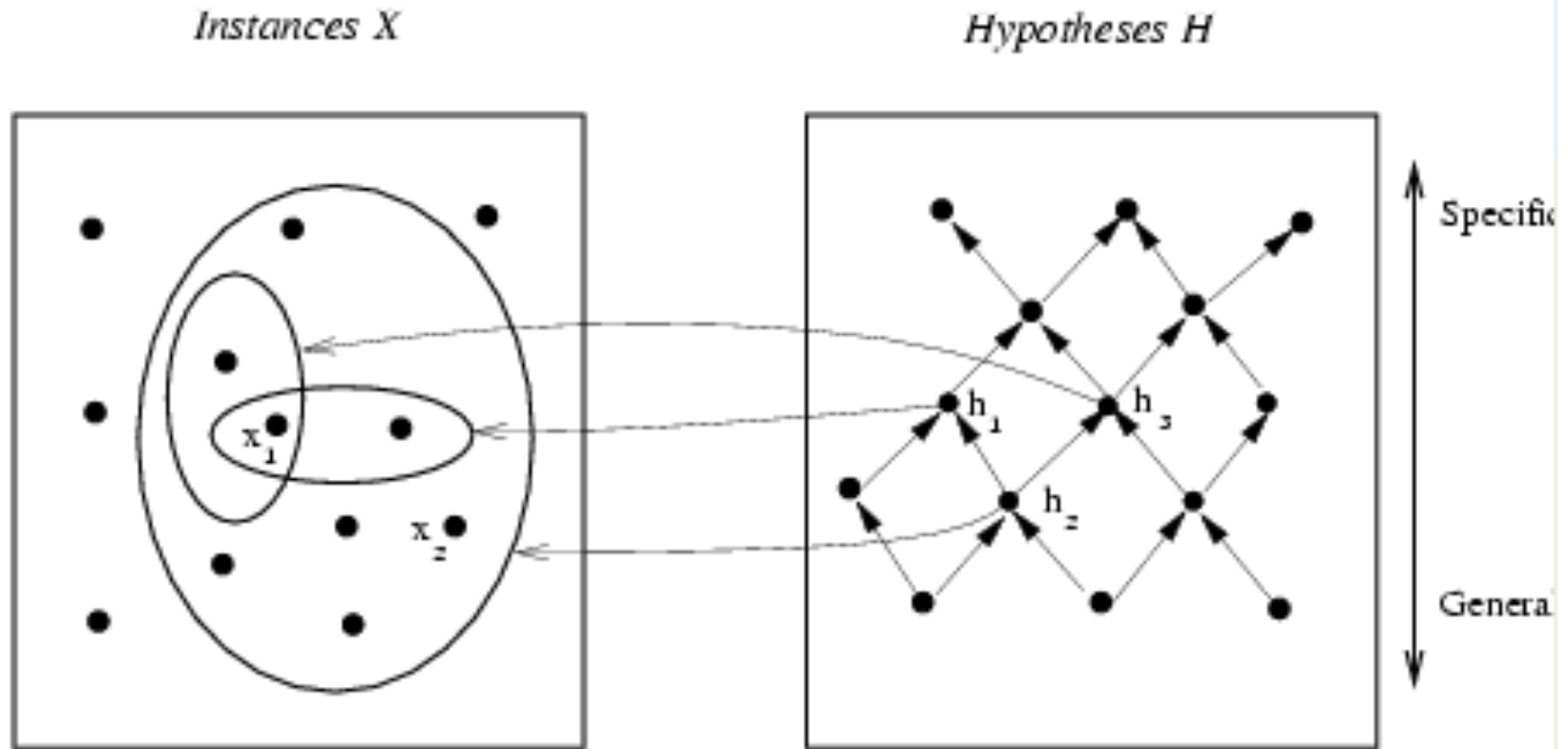
maker cyl displace weight accel

asia ? ? low ?

- Examples:

asia	5	low	low	low	...
usa	4	low	low	low	...

Ordering on Hypothesis Space



x_1	asia	5	low	low	low
x_2	usa	4	med	med	med

h_1 : maker=asia \wedge accel=low

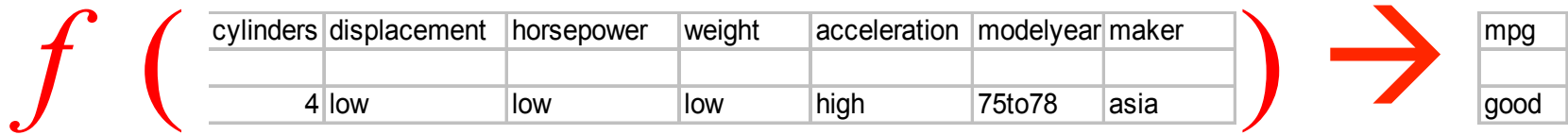
h_2 : maker=asia

h_3 : maker=asia \wedge weight=low

Version Space Algorithm

Ok, so how does it perform?

How to Represent our Function?



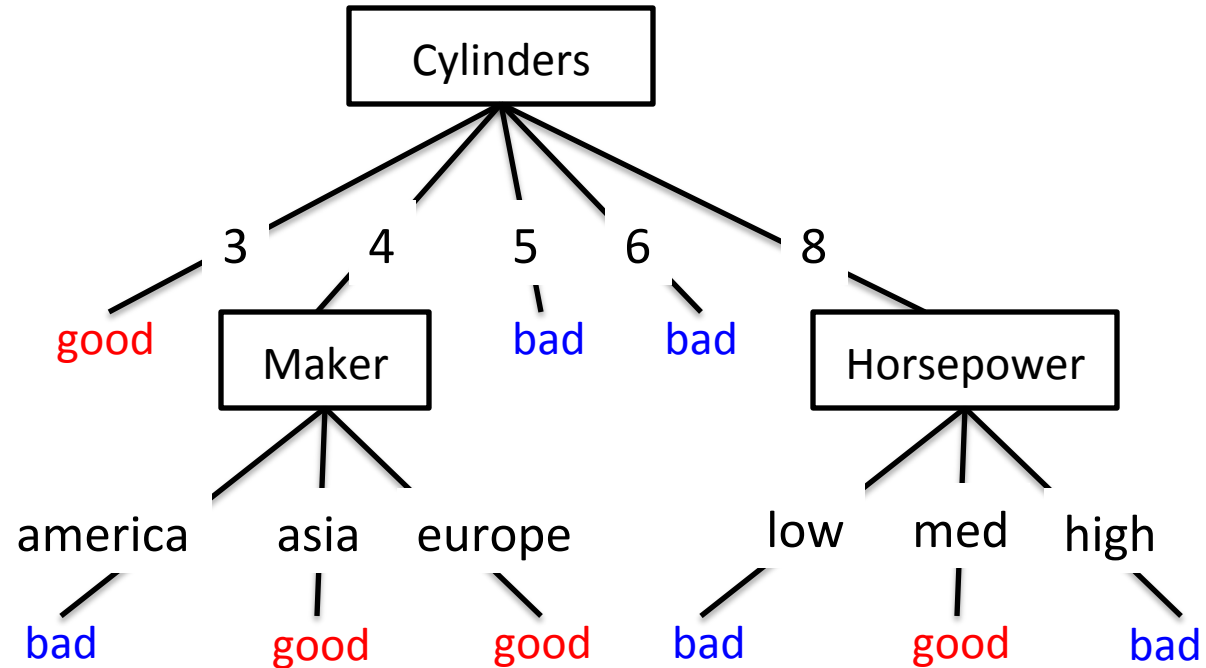
General Propositional Logic?

maker=asia \vee weight=low

Need to find “Hypothesis”: $f : X \rightarrow Y$

Hypotheses: decision trees $f : X \rightarrow Y$

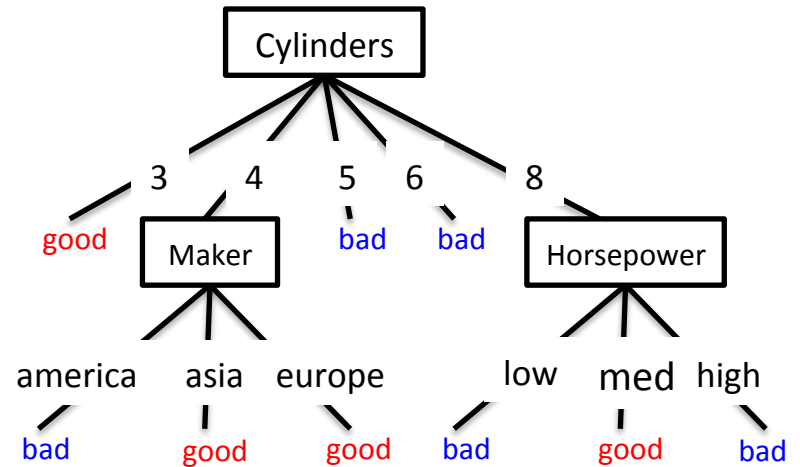
- Each internal node tests an attribute x_i
- Each branch assigns an attribute value $x_i = v$
- Each leaf assigns a class y
- To classify input x : traverse the tree from root to leaf, output the labeled y



Hypothesis space

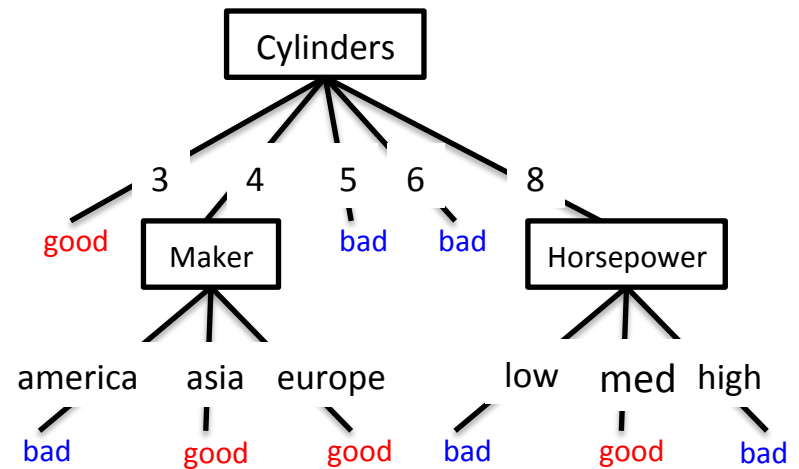
- How many possible hypotheses?
- What functions can be represented?

mpg	cylinders	displacement	horsepower	weight	acceleration	modelyear	maker
good	4	low	low	low	high	75to78	asia
bad	6	medium	medium	medium	medium	70to74	america
bad	4	medium	medium	medium	low	75to78	europa
bad	8	high	high	high	low	70to74	america
bad	6	medium	medium	medium	medium	70to74	america
bad	4	low	medium	low	medium	70to74	asia
bad	4	low	medium	low	low	70to74	asia
bad	8	high	high	high	low	75to78	america
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
bad	8	high	high	high	low	70to74	america
good	8	high	medium	high	high	79to83	america
bad	8	high	high	high	low	75to78	america
good	4	low	low	low	low	79to83	america
bad	6	medium	medium	medium	high	75to78	america
good	4	medium	low	low	low	79to83	america
good	4	low	low	medium	high	79to83	america
bad	8	high	high	high	low	70to74	america
good	4	low	medium	low	medium	75to78	europa
bad	5	medium	medium	medium	medium	75to78	europa



What functions can be represented?

- Decision trees can represent any boolean function!
- But, could require exponentially many nodes...

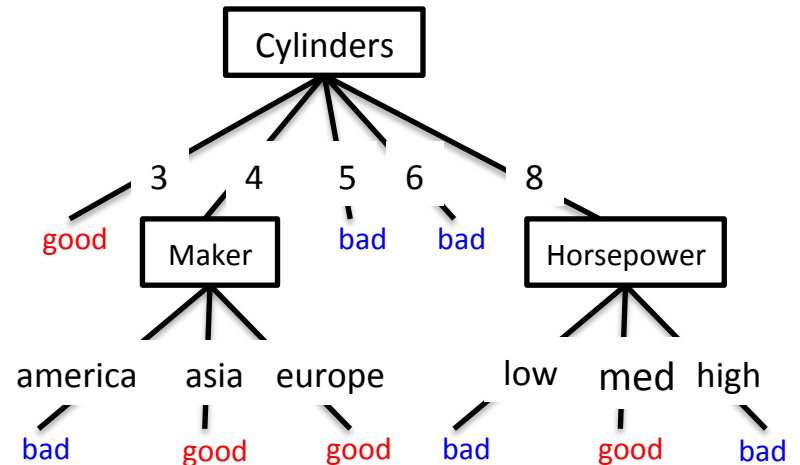


$cyl=3 \vee (cyl=4 \wedge (maker=asia \vee maker=europe)) \vee \dots$

Hypothesis space

- How many possible hypotheses?
- What functions can be represented?
- How many will be consistent with a given dataset?
- How will we choose the best one?
 - Lets first look at how to split nodes, then consider how to find the best tree

mpg	cylinders	displacement	horsepower	weight	acceleration	modelyear	maker
good	4	low	low	low	high	75to78	asia
bad	6	medium	medium	medium	medium	70to74	america
bad	4	medium	medium	medium	low	75to78	europa
bad	8	high	high	high	low	70to74	america
bad	6	medium	medium	medium	medium	70to74	america
bad	4	low	medium	low	medium	70to74	asia
bad	4	low	medium	low	low	70to74	asia
bad	8	high	high	high	low	75to78	america
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
bad	8	high	high	high	low	70to74	america
good	8	high	medium	high	high	79to83	america
bad	8	high	high	high	low	75to78	america
good	4	low	low	low	low	79to83	america
bad	6	medium	medium	medium	high	75to78	america
good	4	medium	low	low	low	79to83	america
good	4	low	low	medium	high	79to83	america
bad	8	high	high	high	low	70to74	america
good	4	low	medium	low	medium	75to78	europa
bad	5	medium	medium	medium	medium	75to78	europa



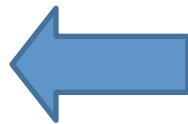
What is the Simplest Tree?

predict
mpg=bad

mpg	cylinders	displacement	horsepower	weight	acceleration	modelyear	maker
good	4	low	low	low	high	75to78	asia
bad	6	medium	medium	medium	medium	70to74	america
bad	4	medium	medium	medium	low	75to78	europa
bad	8	high	high	high	low	70to74	america
bad	6	medium	medium	medium	medium	70to74	america
bad	4	low	medium	low	medium	70to74	asia
bad	4	low	medium	low	low	70to74	asia
bad	8	high	high	high	low	75to78	america
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
bad	8	high	high	high	low	70to74	america
good	8	high	medium	high	high	79to83	america
bad	8	high	high	high	low	75to78	america
good	4	low	low	low	low	79to83	america
bad	6	medium	medium	medium	high	75to78	america
good	4	medium	low	low	low	79to83	america
good	4	low	low	medium	high	79to83	america
bad	8	high	high	high	low	70to74	america
good	4	low	medium	low	medium	75to78	europa
bad	5	medium	medium	medium	medium	75to78	europa

Is this a good tree?

[22+, 18-]

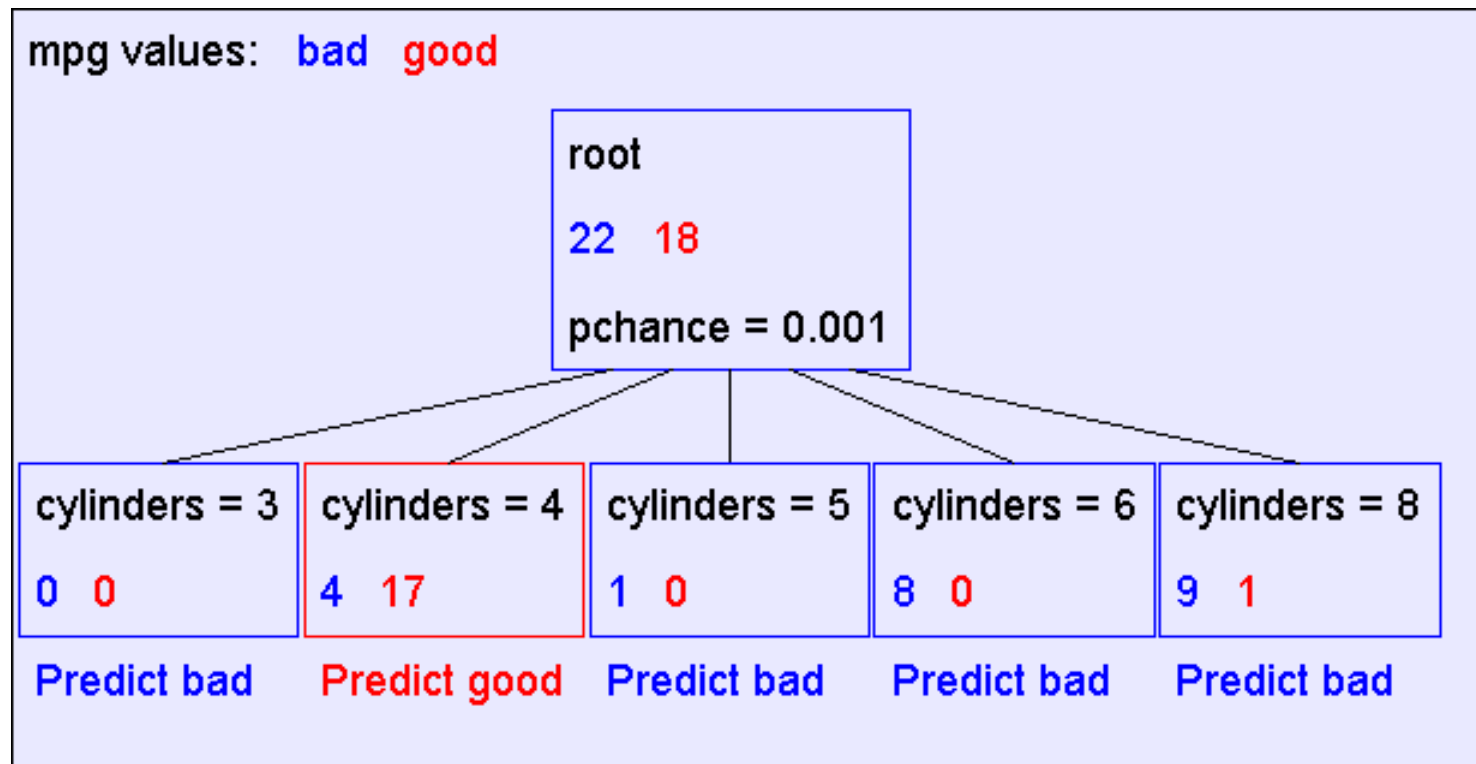


Means:

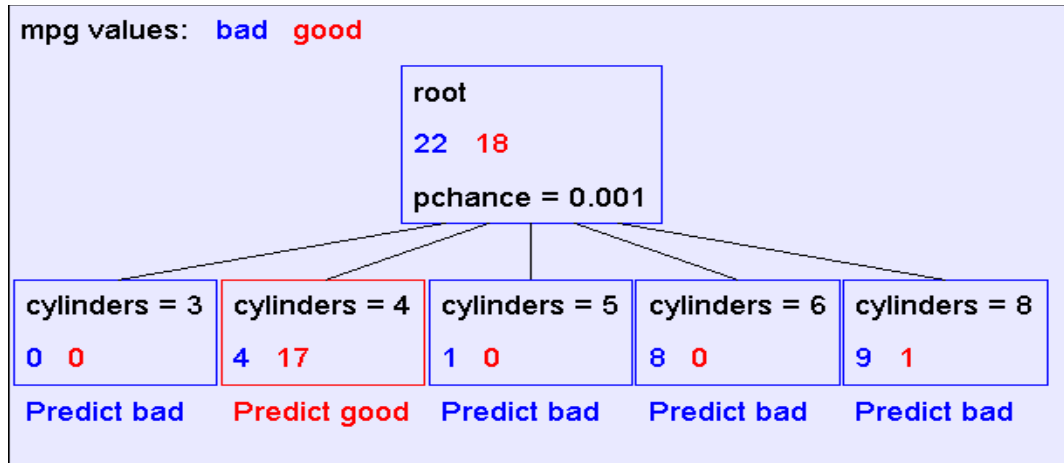
correct on 22 examples

incorrect on 18 examples

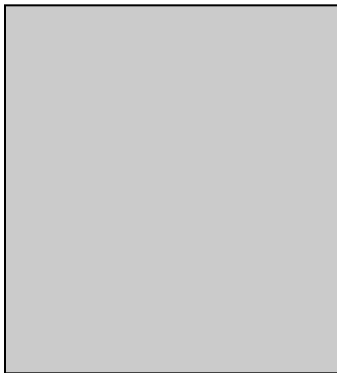
A Decision Stump



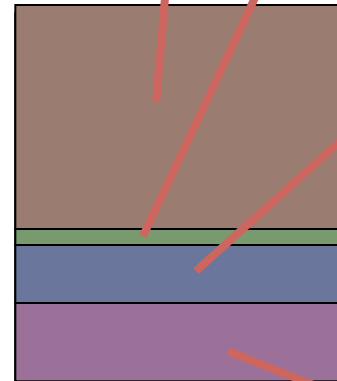
Recursive Step



Take the Original Dataset..



And partition it according to the value of the attribute we split on



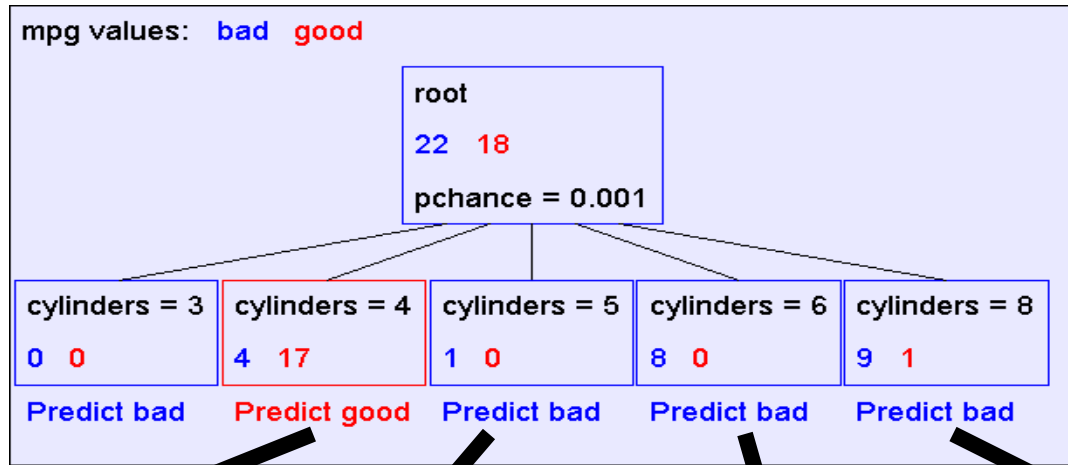
Records in which cylinders = 4

Records in which cylinders = 5

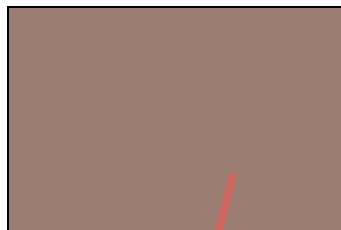
Records in which cylinders = 6

Records in which cylinders = 8

Recursive Step



Build tree from
These records..



Records in which
cylinders = 4

Build tree from
These records..



Records in which
cylinders = 5

Build tree from
These records..



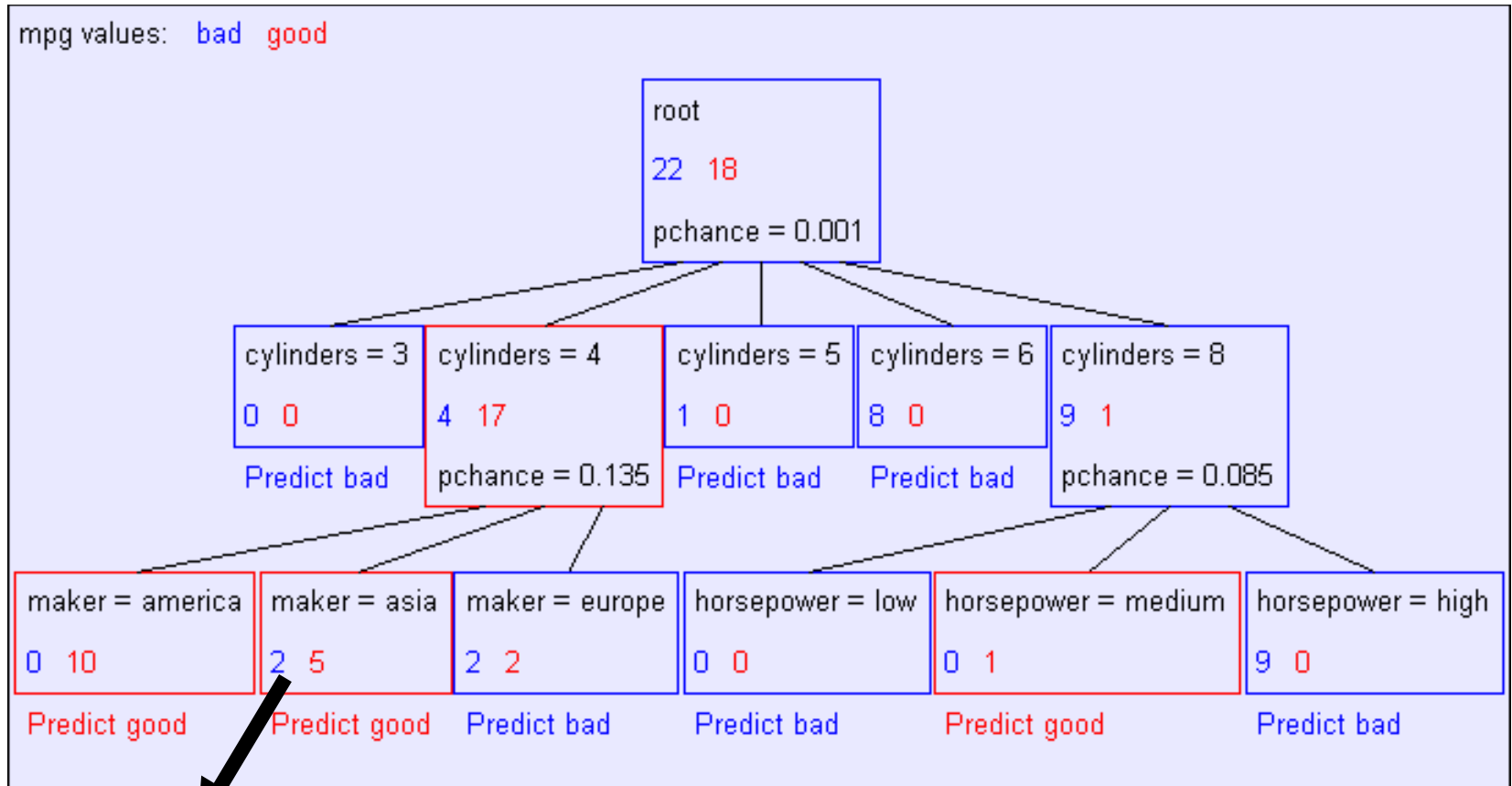
Records in which
cylinders = 6

Build tree from
These records..



Records in which
cylinders = 8

Second level of tree

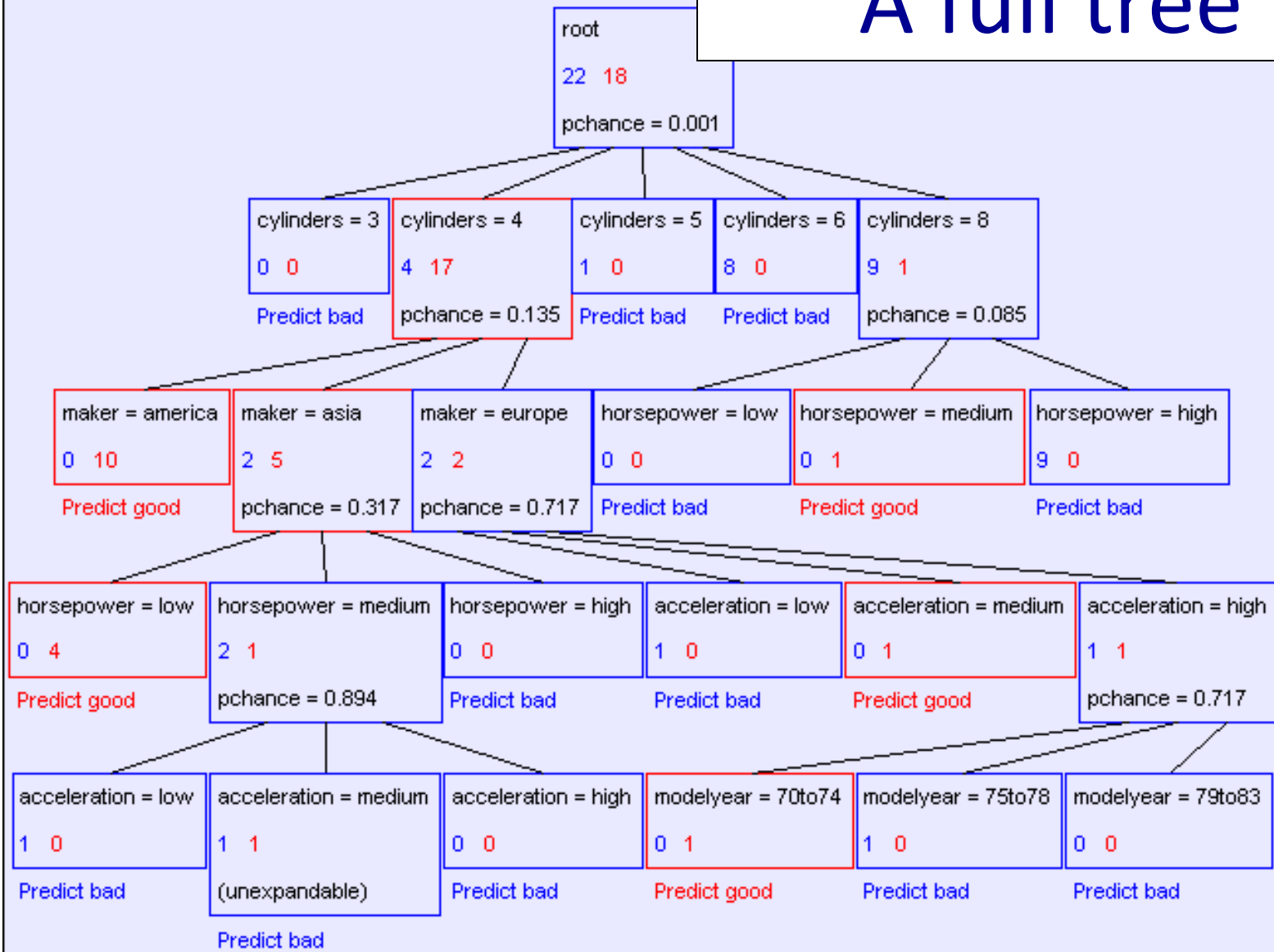


Recursively build a tree from the seven records in which there are four cylinders and the maker was based in Asia

(Similar recursion in the other cases)

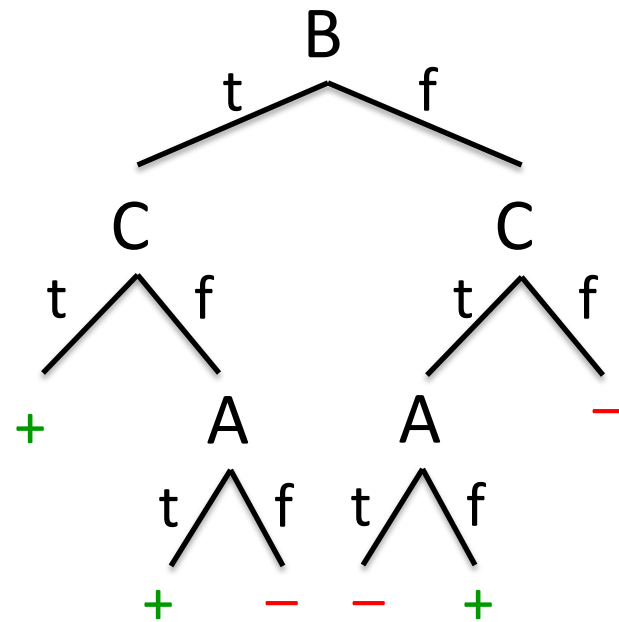
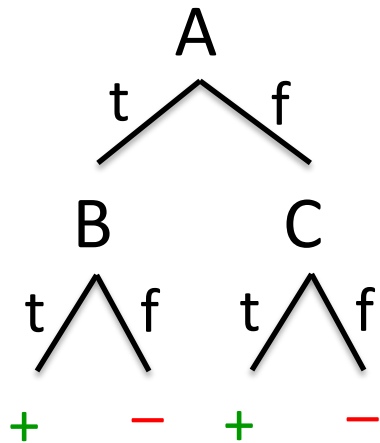
A full tree

mpg values: bad good



Are all decision trees equal?

- Many trees can represent the same concept
- But, not all trees will have the same size!
 - e.g., $\phi = (A \wedge B) \vee (\neg A \wedge C)$ -- ((A and B) or (not A and C))



- Which tree do we prefer?
 - Smaller tree has more examples at each leaf!

Learning decision trees is hard!!!

- Learning the simplest (smallest) decision tree is an NP-complete problem [Hyafil & Rivest '76]
- Resort to a greedy heuristic:
 - Start from empty decision tree
 - Split on **next best attribute (feature)**
 - Recurse

What defines a good attribute?

