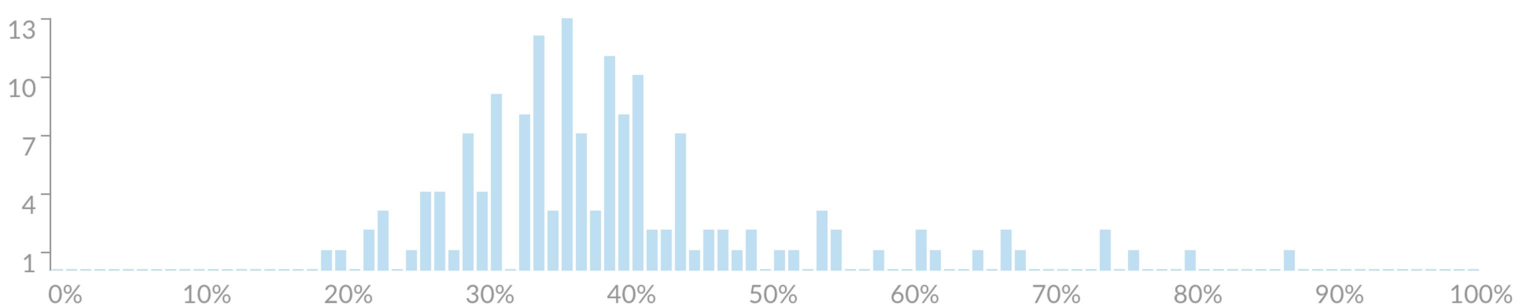
Learning & Decision Trees

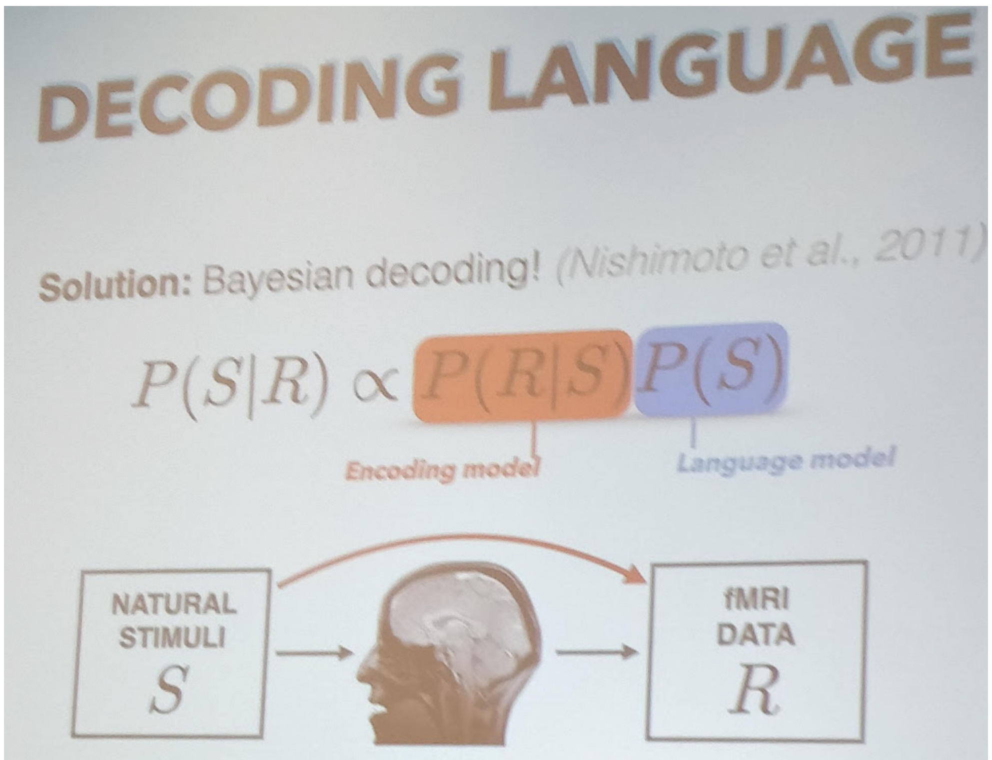
Announcements

- A2 due November 7
- A3 and optional A4 to follow (A4 will have a deadline during the last week of classes)
- Midterm grade distribution:

Average Score	High Score	Low Score	Standard Deviation	Average Time
39%	87%	19%	8.28	01:14:88



Fun example of Bayes' theorem application in decoding of brain recordings



Actual stimulus

i got up from the air mattress and pressed my face against the glass of the bedroom window expecting to see eyes staring back at me but instead finding only darkness

Decoded stimulus

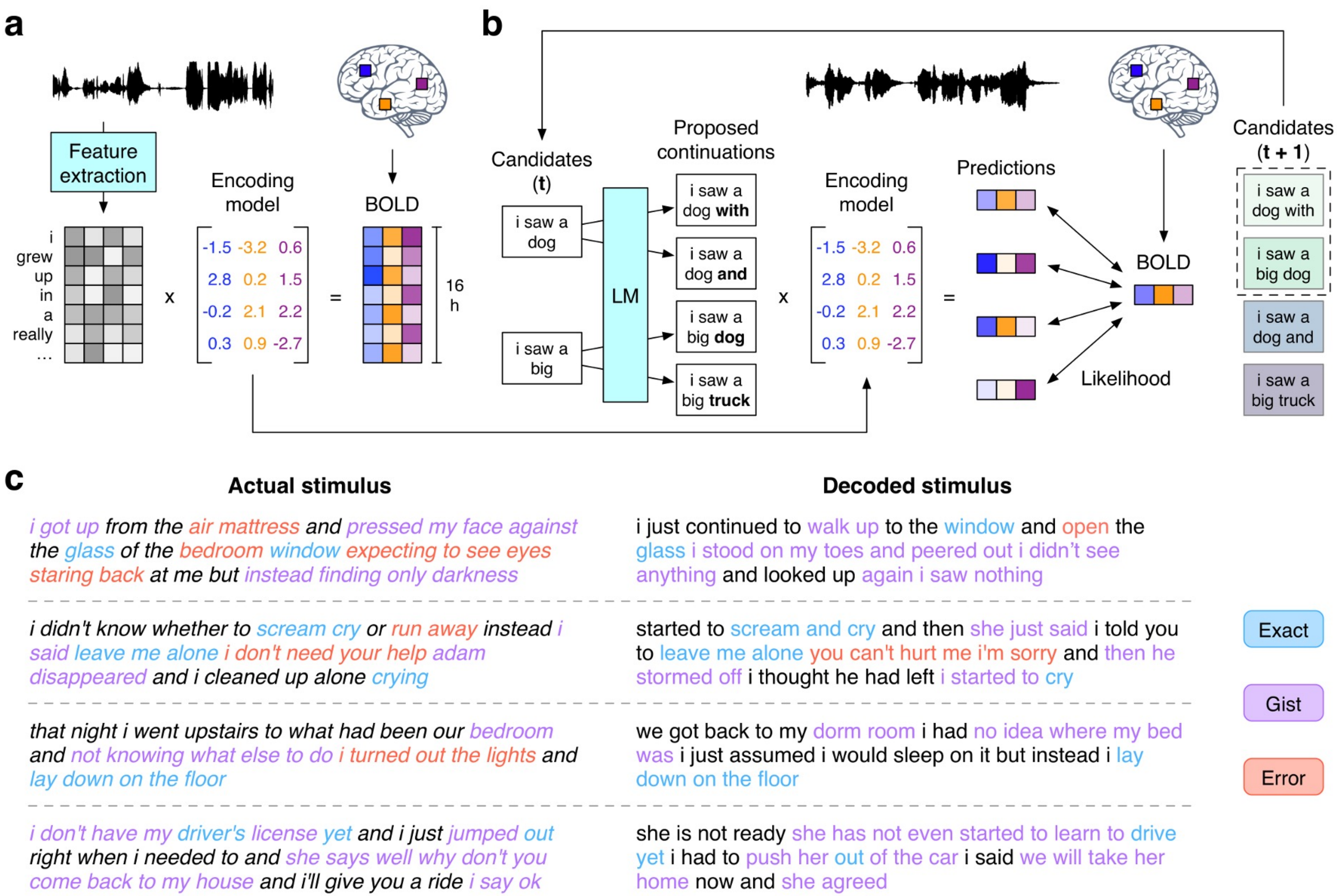
i just continued to walk up to the window and open the glass i stood on my toes and peered out i didn't see anything and looked up again i saw nothing

Self-supervised models of audio effectively explain human cortical responses to speech

Aditya R Vaidya, Shailee Jain, Alexander G Huth

https://arxiv.org/pdf/2205.14252

Fun example of Bayes' theorem application in decoding of brain recordings



https://arxiv.org/pdf/2205.14252

Learning

- Generalization from experience

"Our experience of the world is specific, yet we are able to formulate general theories that account for the past and predict the future."

--M.R. Genesereth and N.J. Nilsson, *Logical Foundations of AI*, 1987

- Agent has made observations (data)
- Now must make sense of it (hypotheses)

Tasks & settings
Classification
Ranking
Clustering
Regression
Decision-making

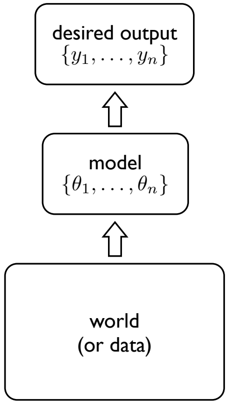
Supervised
Unsupervised
Semi-supervised
Active
Reinforcement

Techniques
Bayesian learning
Decision trees
Neural networks
Support vector machines
Boosting
Case-based reasoning
Dimensionality reduction
…

Applications
Document retrieval
Document classification
Data mining
Computer vision
Scientific discovery
Robotics
…

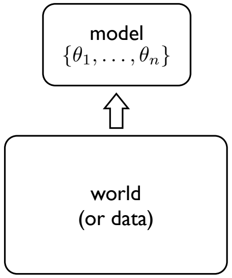
Supervised

Given pairs (x,y) with y=f(x), agent builds model to predict f(x) for new x



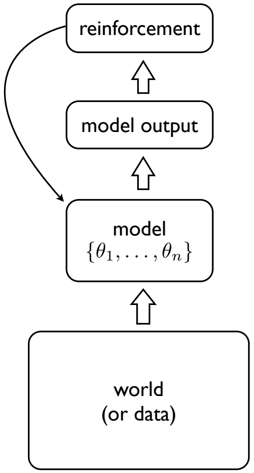
Unsupervised

Given data points x, agent learns patterns in the data (e.g. clusters)



Reinforcement

Agent performs actions a_1 … a_n, receives reward R; decides which actions



Semi-supervised

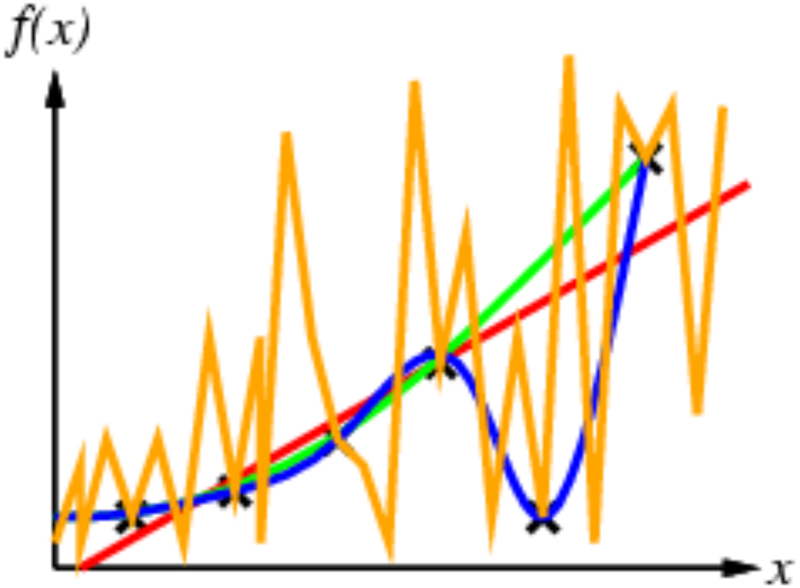
Some labels are missing in the training set, or some labels are erroneous

Active

Learner asks an oracle for correct output y=f(x) for some input points x

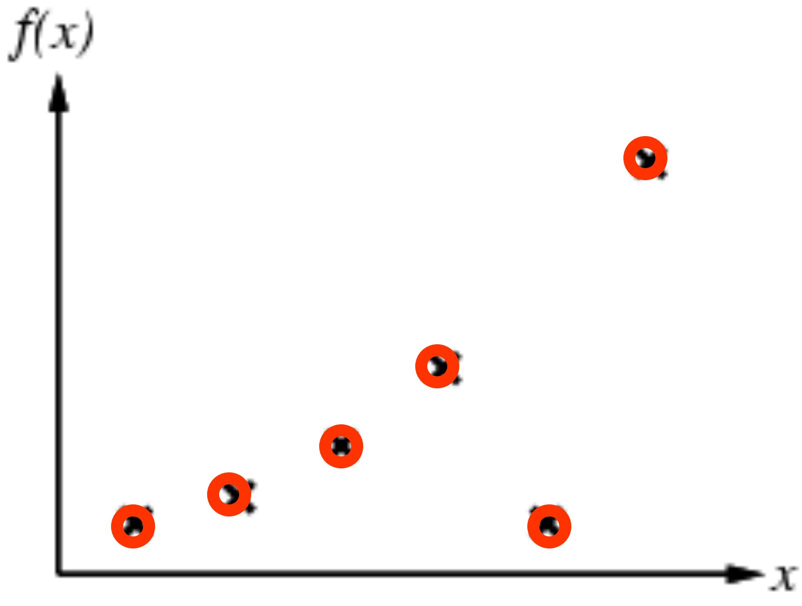
Supervised learning

- Construct/adjust *h* to agree with *f* on training set
- *h* is consistent if it agrees with *f* on all examples
- E.g., curve fitting:



Supervised learning

- Construct/adjust *h* to agree with *f* on training set
- *h* is consistent if it agrees with *f* on all examples
- E.g., curve fitting:



h=D is a trivial, but perhaps uninteresting solution (memorization)

Tasks

- Depending on the task, the goal is to produce values for f(x) of different forms:
  - **Regression:** Predict continuous values
  - **Classification:** Predict one of a discrete set of labels
  - **Binary Classification:** Predict positive or negative
  - **Structured:** Predict complex structure (e.g. tree, sequence, etc.)

Logic-Based Inductive Learning

- Here, examples (x, f(x)) take on discrete values

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Learning a Logical Predicate
(Concept Classifier)

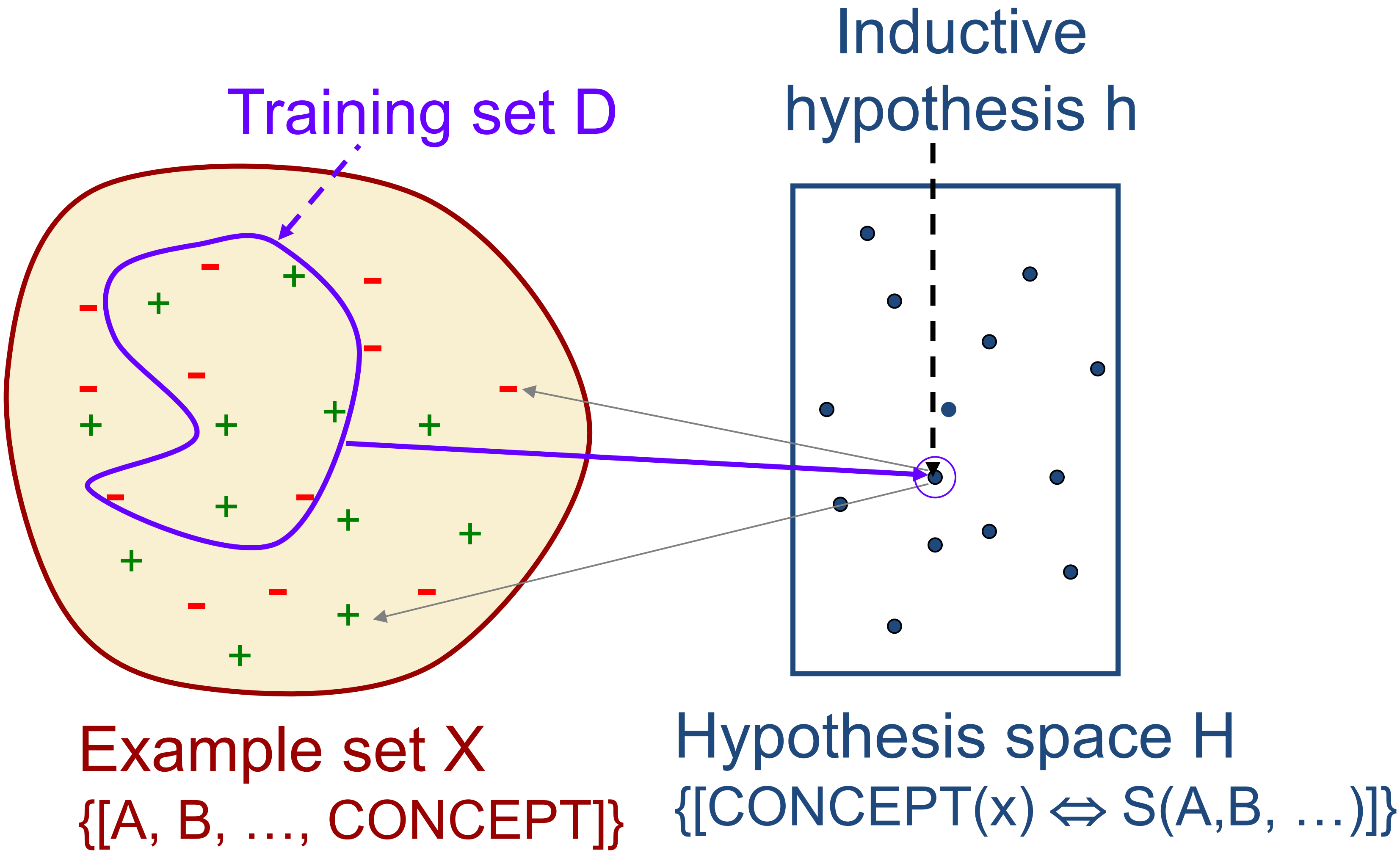
- Given set E of objects, and predicates A(x), B(x), … for x ∈ E, the **goal is to predict** CONCEPT(x) ∈ {T,F}, where CONCEPT is a sentence of predicates, e.g.:

  $$\text{CONCEPT}(x) \Leftrightarrow A(x) \wedge (\neg B(x) \vee C(x))$$

  CONCEPT(x) ⇔ A(x) and (not B(x) or C(x))
- **Training set:** values of CONCEPT for subset of predicate values
- A **hypothesis** is a possible model, CONCEPT(x) ⇔ S(A,B, …)
- The set of all hypotheses is called the **hypothesis space**
- A hypothesis **agrees** with an example if it gives the correct value of CONCEPT

- https://www.rapidtables.com/math/symbols/Logic_Symbols.html

Inductive Learning Scheme

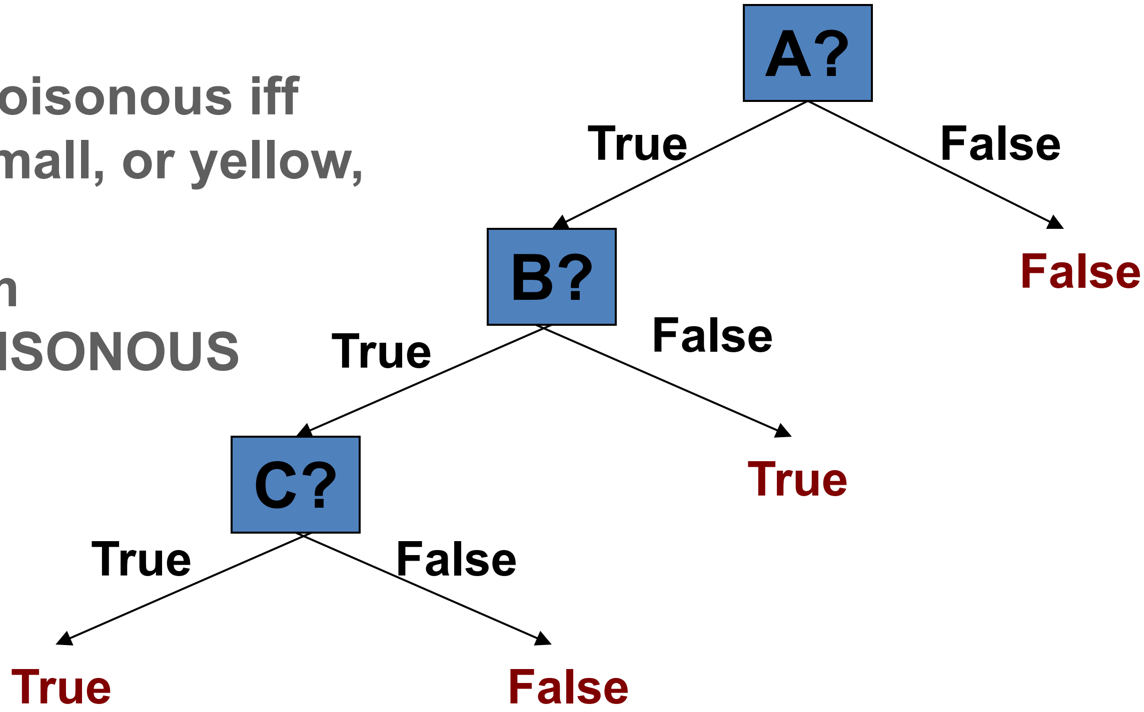


Predicate as a Decision Tree

The predicate CONCEPT(x) ⇔ A(x) ∧ (¬B(x) v C(x)) can be represented by the following decision tree:

Example:
A mushroom is poisonous iff it is yellow and small, or yellow, big and spotted

- **x is a mushroom**
- **CONCEPT = POISONOUS**
- **A = YELLOW**
- **B = BIG**
- **C = SPOTTED**

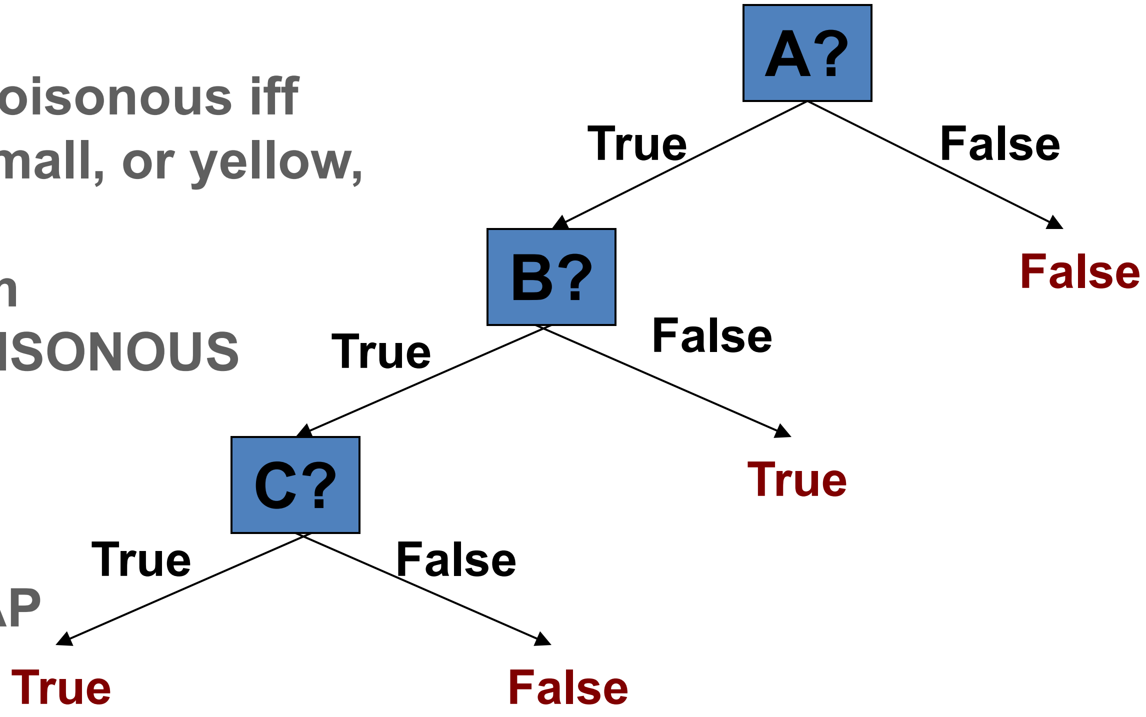


Predicate as a Decision Tree

The predicate CONCEPT(x) $\Leftrightarrow$ A(x) $\wedge$ ($\neg$B(x) v C(x)) can be represented by the following decision tree:

Example:
A mushroom is poisonous iff it is yellow and small, or yellow, big and spotted

- x is a mushroom
- CONCEPT = POISONOUS
- A = YELLOW
- B = BIG
- C = SPOTTED
- D = FUNNEL-CAP
- E = BULKY

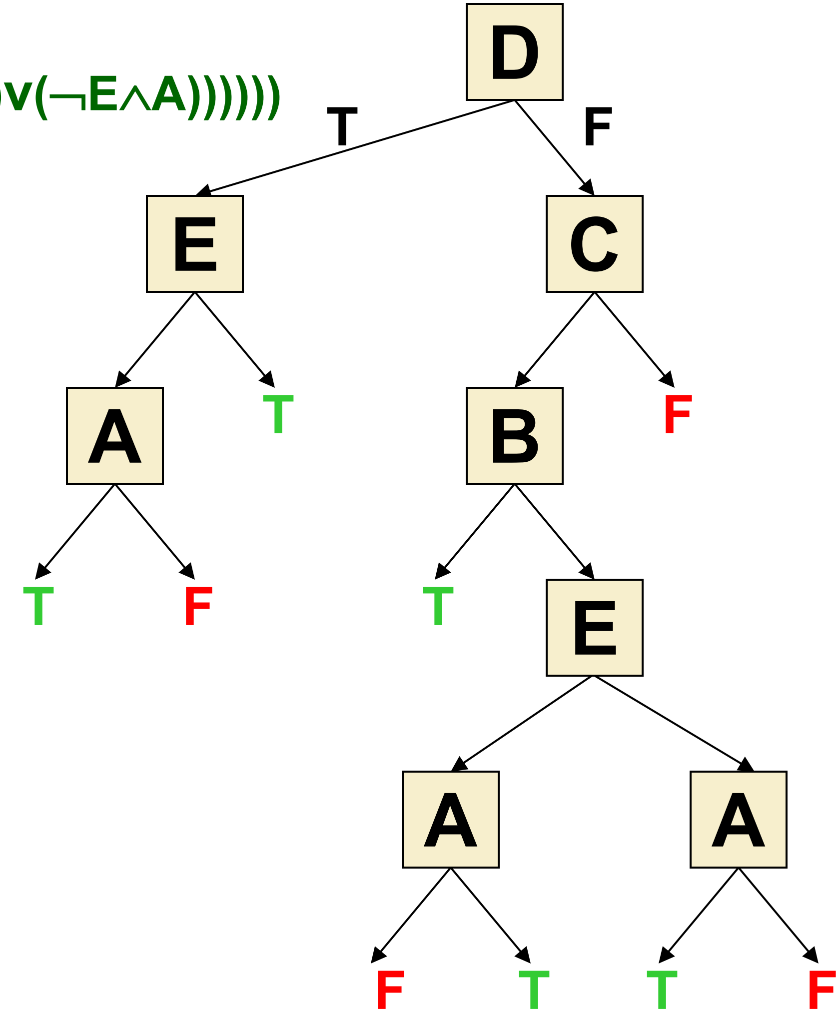


Training Set

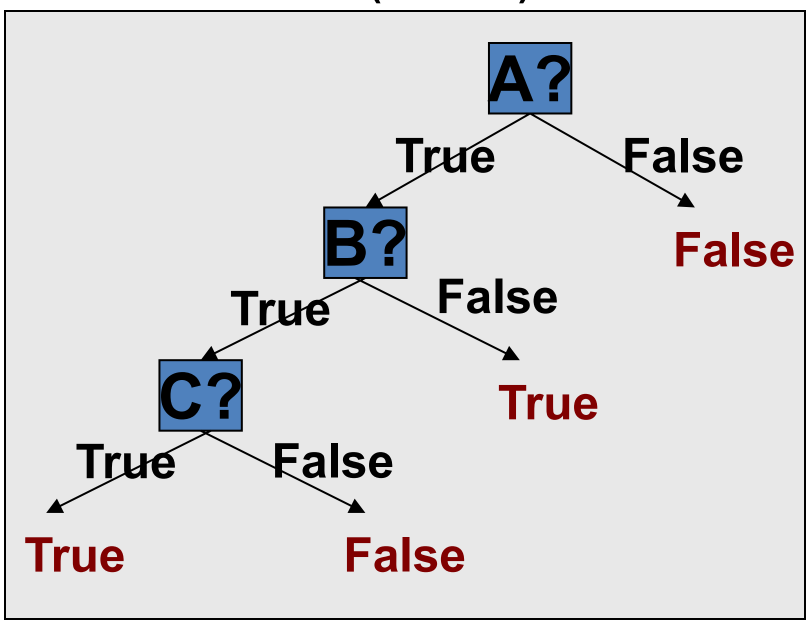
Ex. #	A	B	C	D	E	CONCEPT
1	**False**	**False**	**True**	**False**	**True**	**False**
2	**False**	**True**	**False**	**False**	**False**	**False**
3	**False**	**True**	**True**	**True**	**True**	**False**
4	**False**	**False**	**True**	**False**	**False**	**False**
5	**False**	**False**	**False**	**True**	**True**	**False**
6	**True**	**False**	**True**	**False**	**False**	**True**
7	**True**	**False**	**False**	**True**	**False**	**True**
8	**True**	**False**	**True**	**False**	**True**	**True**
9	**True**	**True**	**True**	**False**	**True**	**True**
10	**True**	**True**	**True**	**True**	**True**	**True**
11	**True**	**True**	**False**	**False**	**False**	**False**
12	**True**	**True**	**False**	**False**	**True**	**False**
13	**True**	**False**	**True**	**True**	**True**	**True**

Possible Decision Tree

CONCEPT $\Leftrightarrow$ $(D \wedge (\neg E \vee A)) \vee (\neg D \wedge (C \wedge (B \vee (\neg B \wedge ((E \wedge \neg A) \vee (\neg E \wedge A))))))$

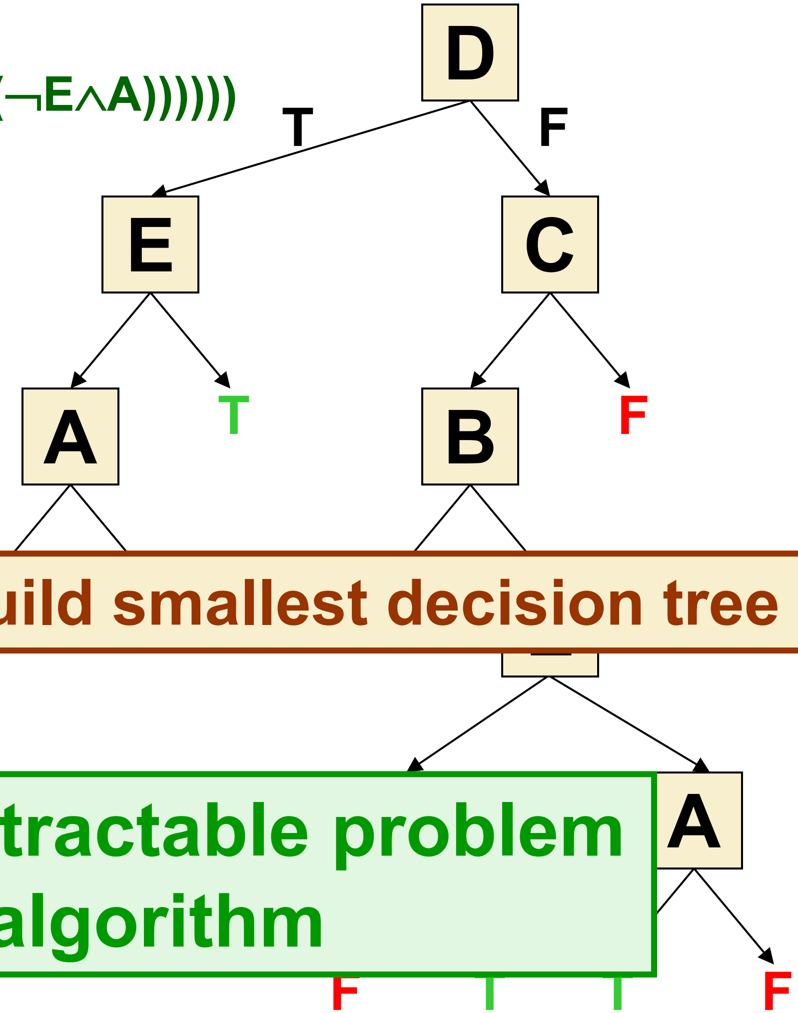


CONCEPT $\Leftrightarrow A \wedge (\neg B \vee C)$

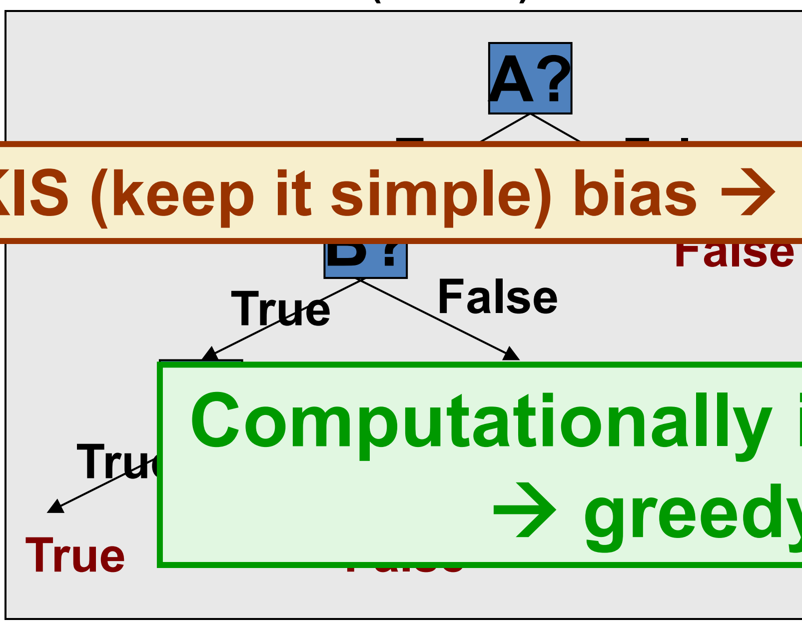


Possible Decision Tree

CONCEPT ⇔ $(D \wedge (\neg E \vee A)) \vee (\neg D \wedge (C \wedge (B \vee (\neg B \wedge ((E \wedge \neg A) \vee (\neg E \wedge A))))))$



CONCEPT ⇔ $A \wedge (\neg B \vee C)$



KIS (keep it simple) bias → Build smallest decision tree

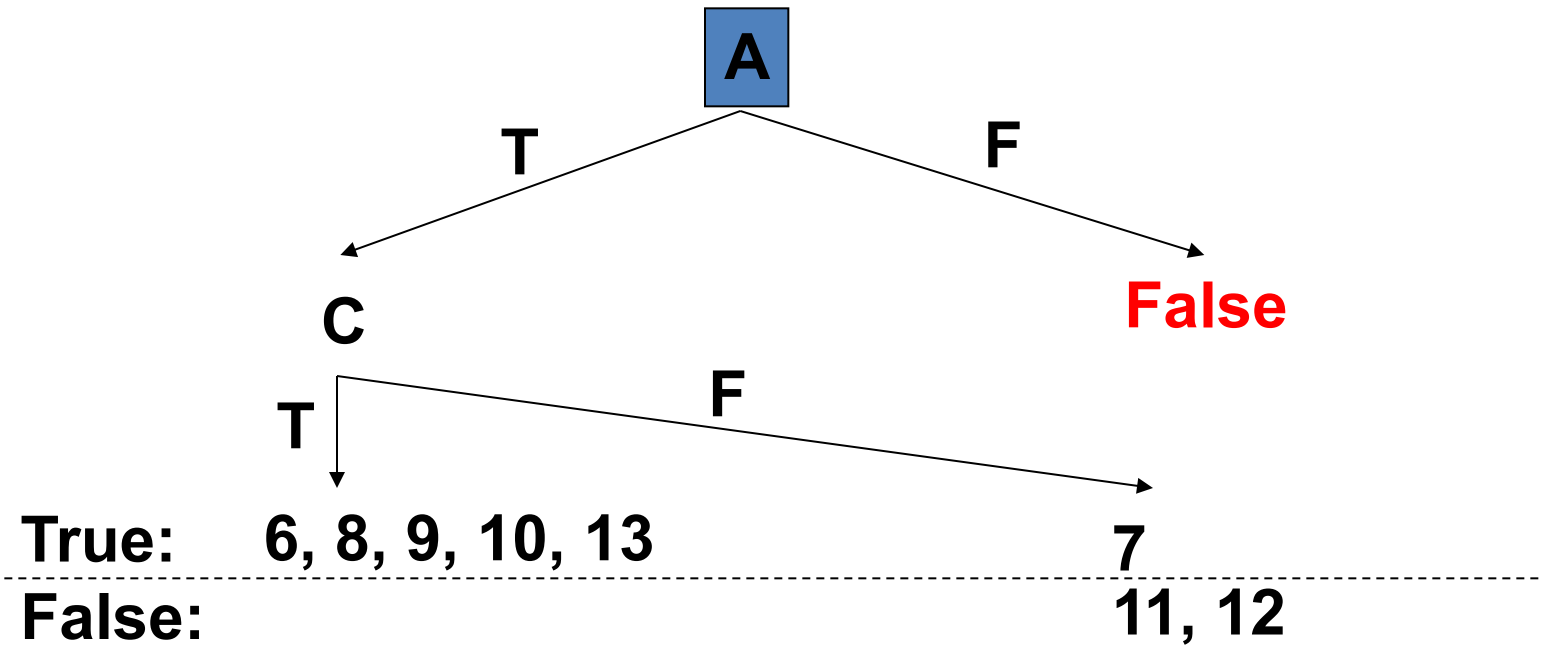
Computationally intractable problem → greedy algorithm

Top-down induction of decision tree

Ex. #	A	B	C	D	E	CONCEPT
1	False	False	True	False	True	False
2	False	True	False	False	False	False
3	False	True	True	True	True	False
4	False	False	True	False	False	False
5	False	False	False	True	True	False
6	True	False	True	False	False	True
7	True	False	False	True	False	True
8	True	False	True	False	True	True
9	True	True	True	False	True	True
10	True	True	True	True	True	True
11	True	True	False	False	False	False
12	True	True	False	False	True	False
13	True	False	True	True	True	True

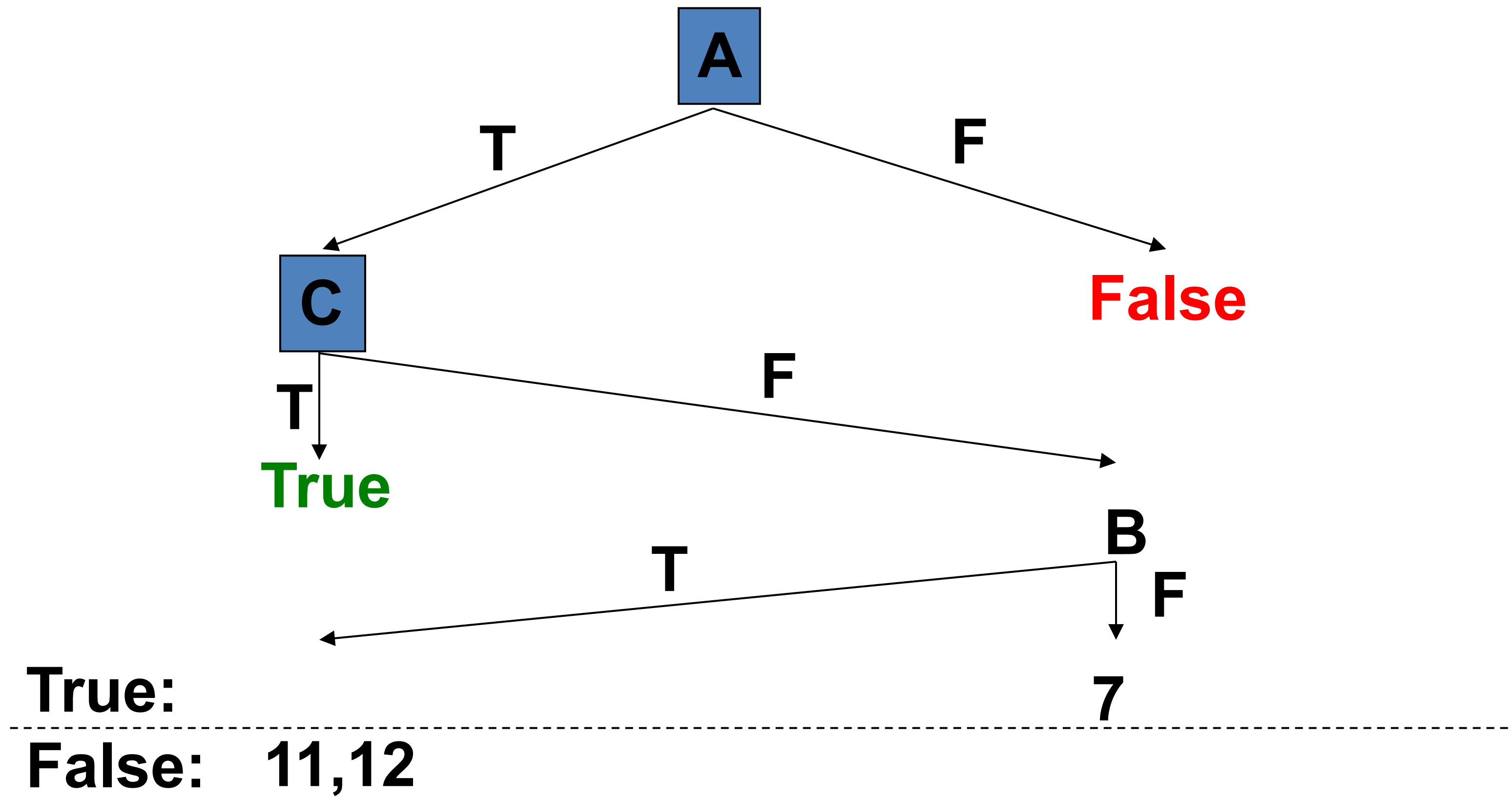
- What should the classifier do if it observes none of the predicates?
- What should it do if it can choose only one predicate?

Choice of Second Predicate

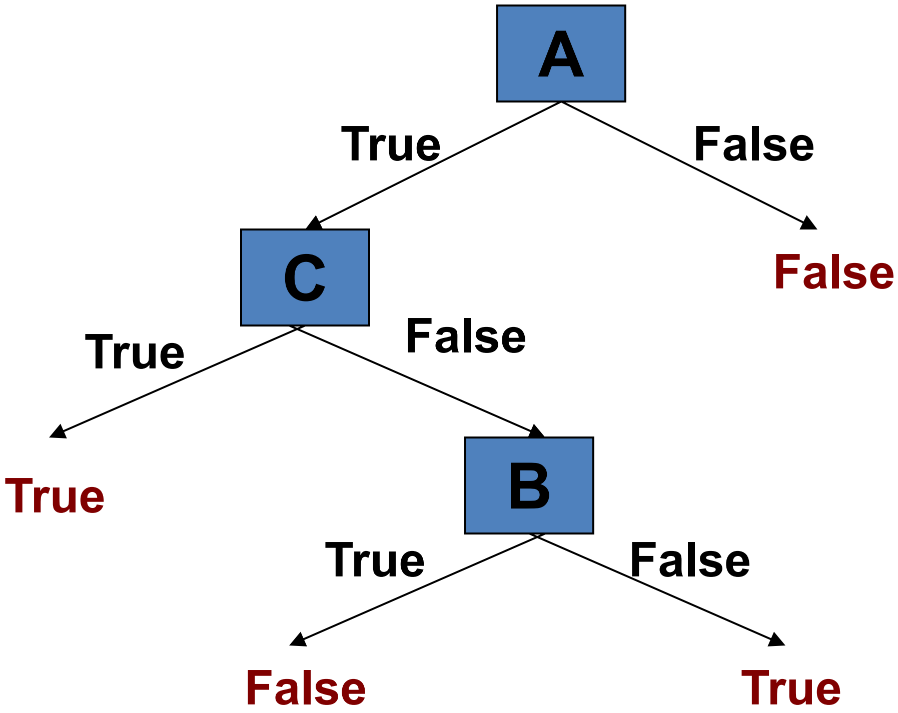


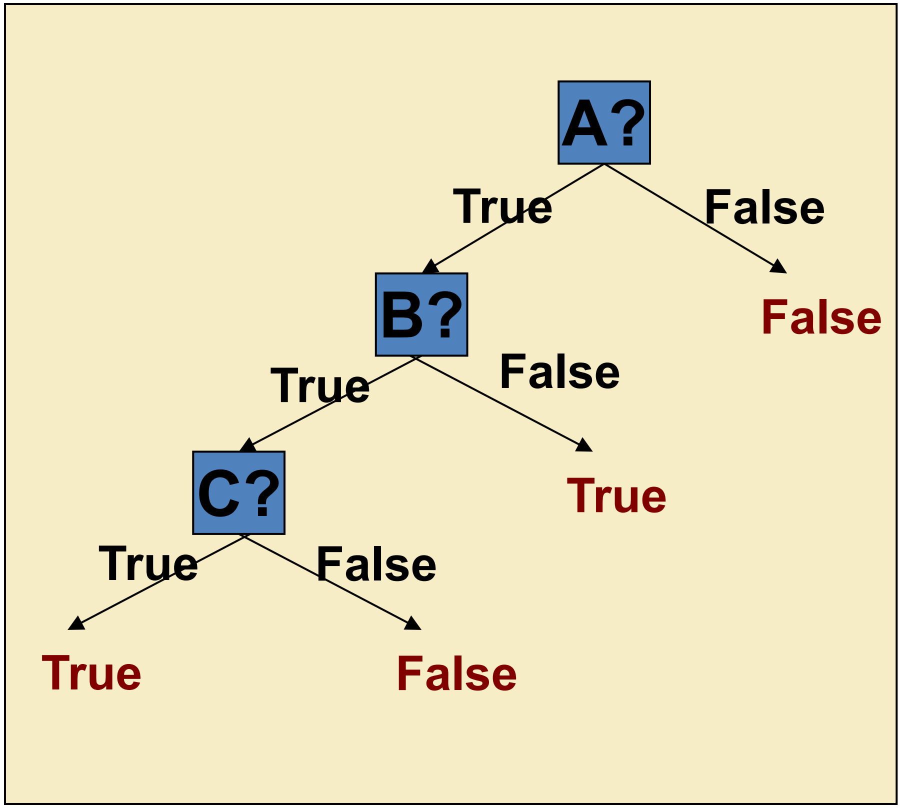
→ The number of misclassified examples from the training set is 1

Choice of Third Predicate



Final Tree

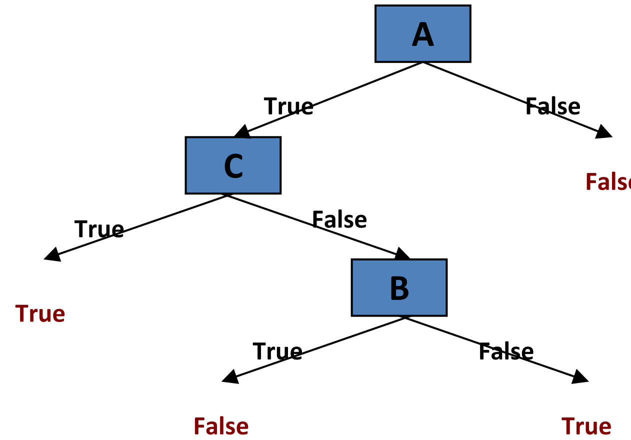




CONCEPT $\Leftrightarrow$ A $\wedge$ (C v $\neg$B)

CONCEPT $\Leftrightarrow$ A $\wedge$ ($\neg$B v C)

Top-Down Induction of a DT

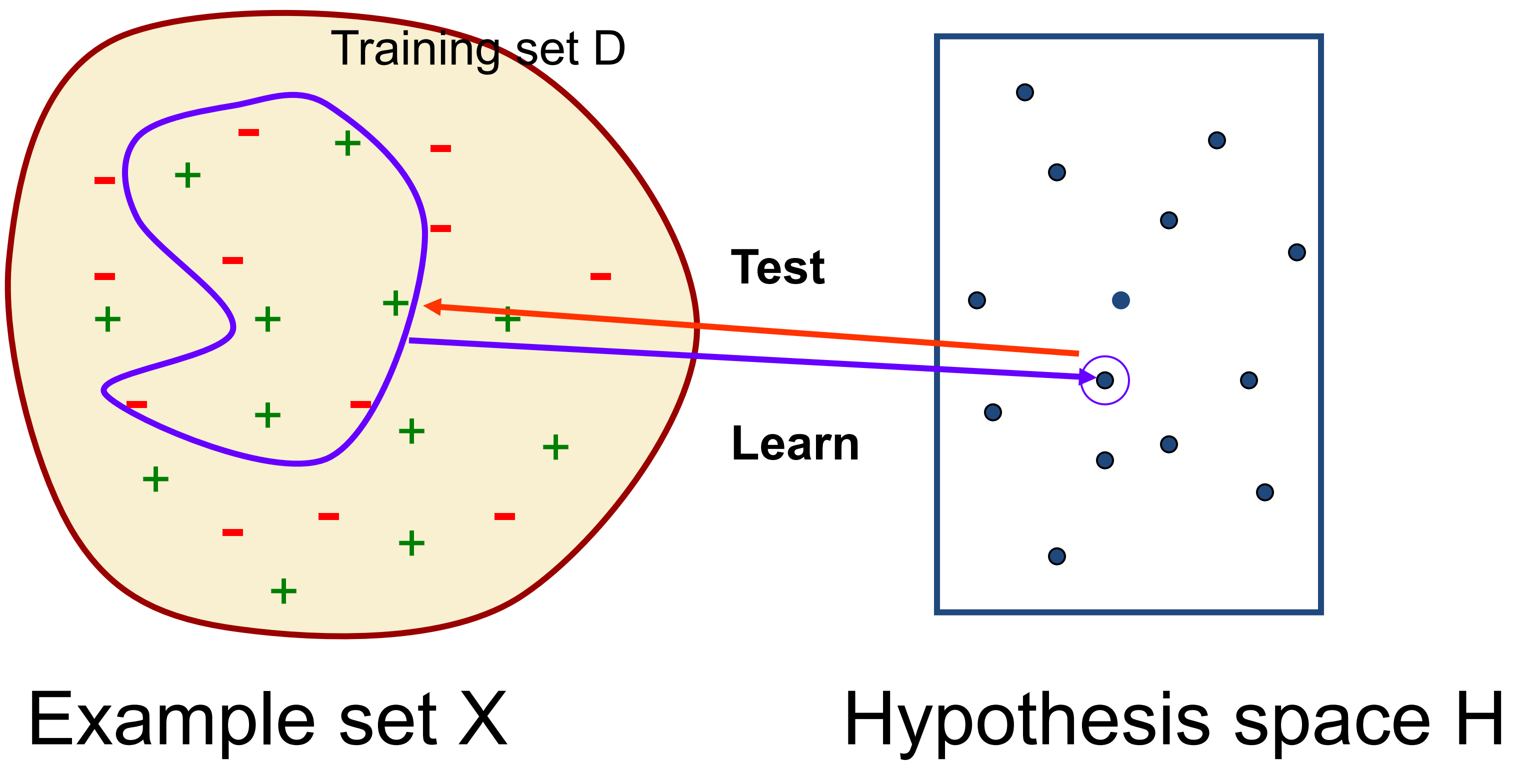


DTL(D, Predicates)

1. If all examples in D are positive then return True
2. If all examples in D are negative then return False
3. If Predicates is empty then return *failure*
4. A ← **error-minimizing** predicate in Predicates
5. Return the tree whose:
   - root is A,
   - left branch is DTL(D^{+A},Predicates-A),
   - right branch is DTL(D^{-A},Predicates-A)

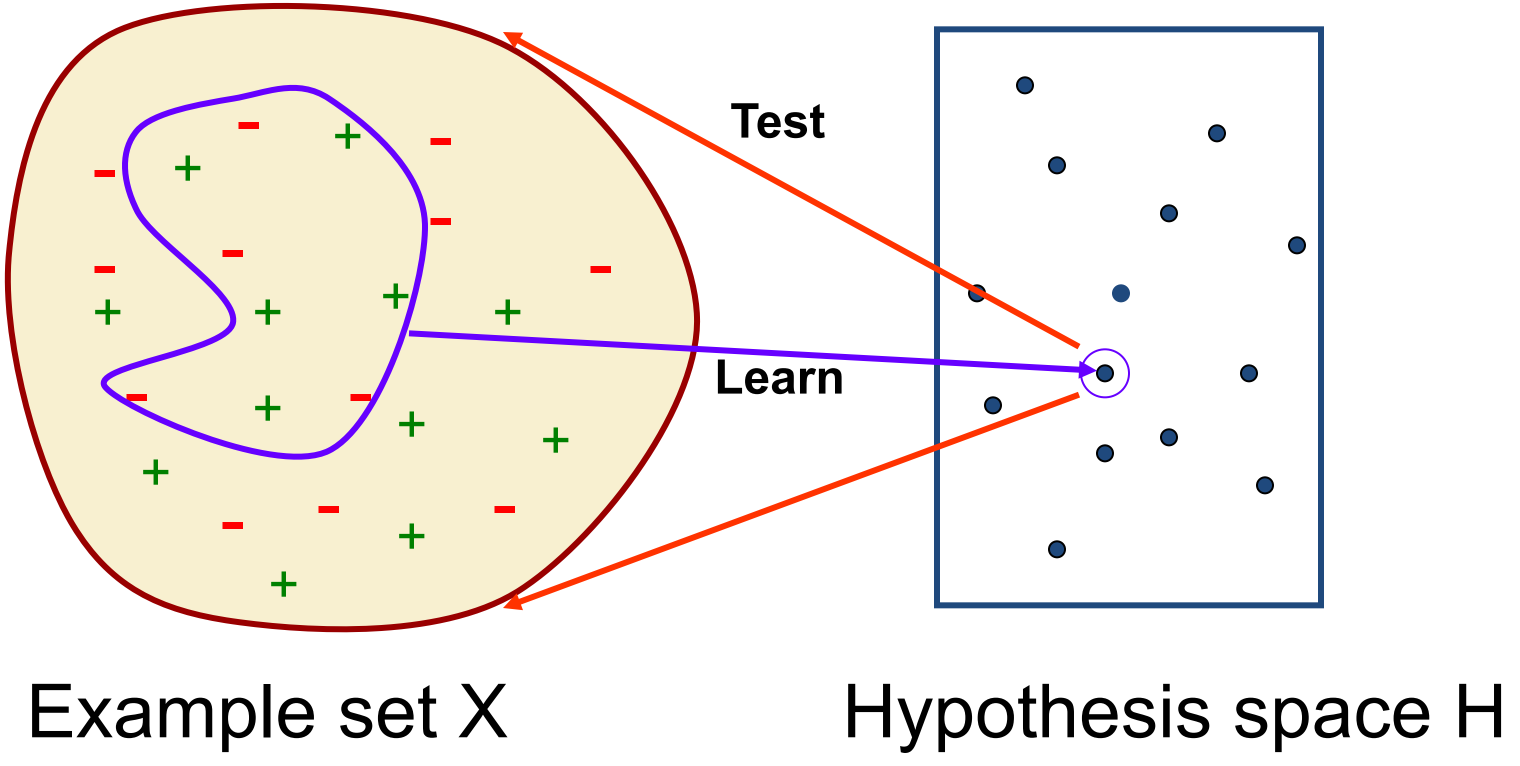
Capacity is Not the Only Criterion

- Accuracy on training set isn't the best measure of performance



Generalization Error

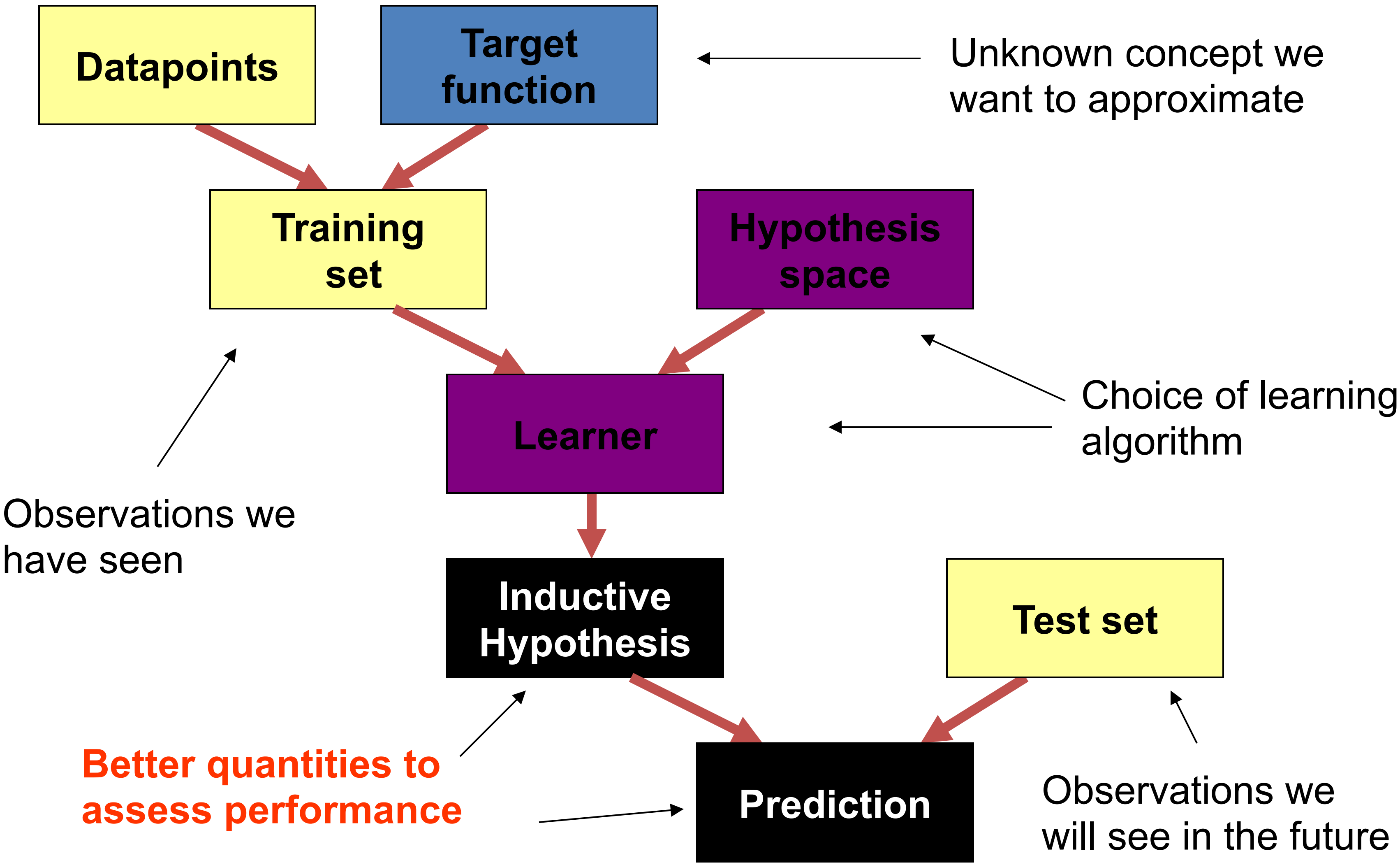
- A hypothesis h is said to *generalize* well if it achieves low error on all examples in X



Assessing Performance of a Learning Algorithm

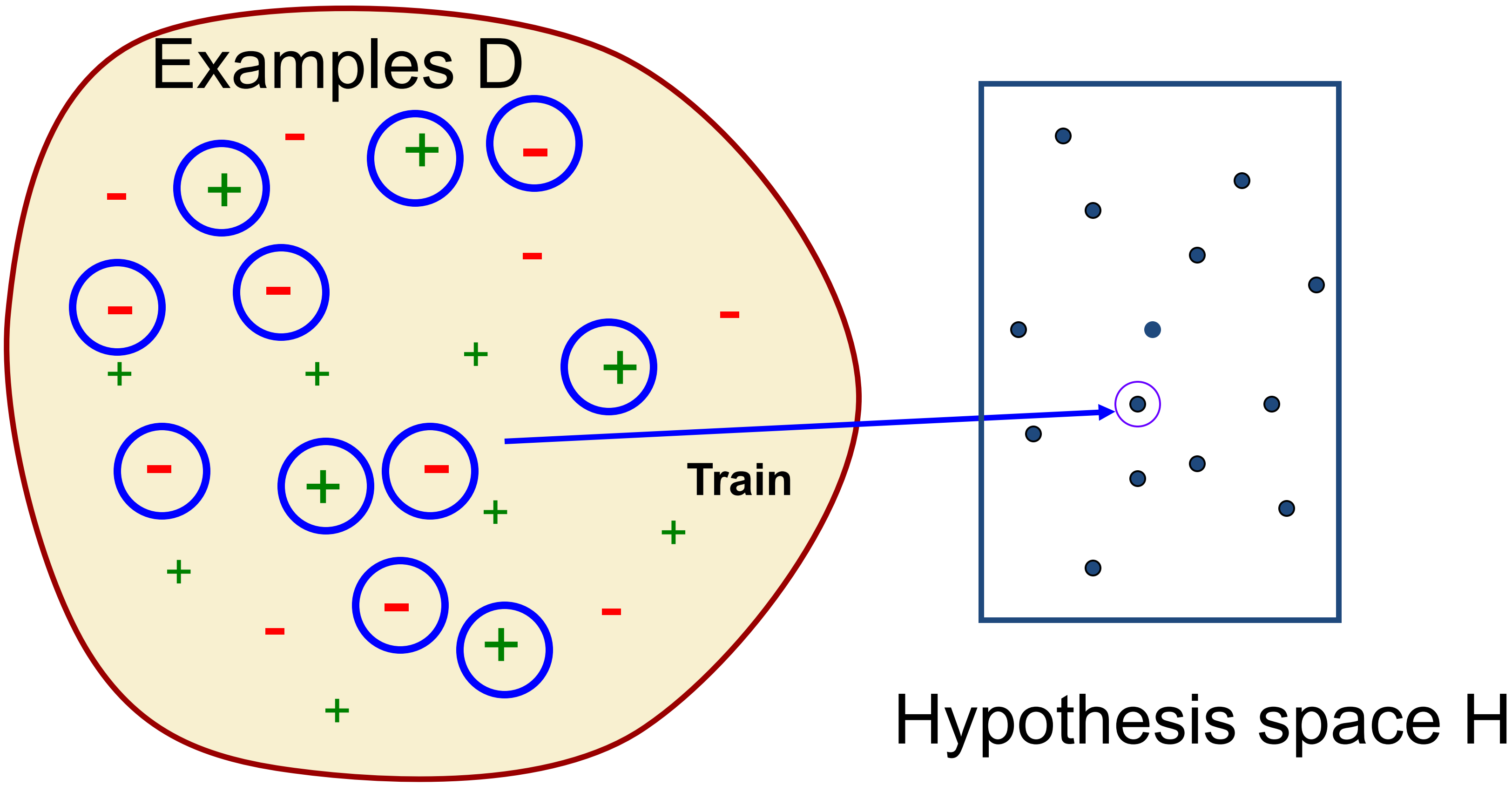
- Samples from X are typically unavailable
- Take out some of the training set
  - Train on the remaining training set
  - Test on the excluded instances
  - *Cross-validation*

Supervised Learning Flow Chart



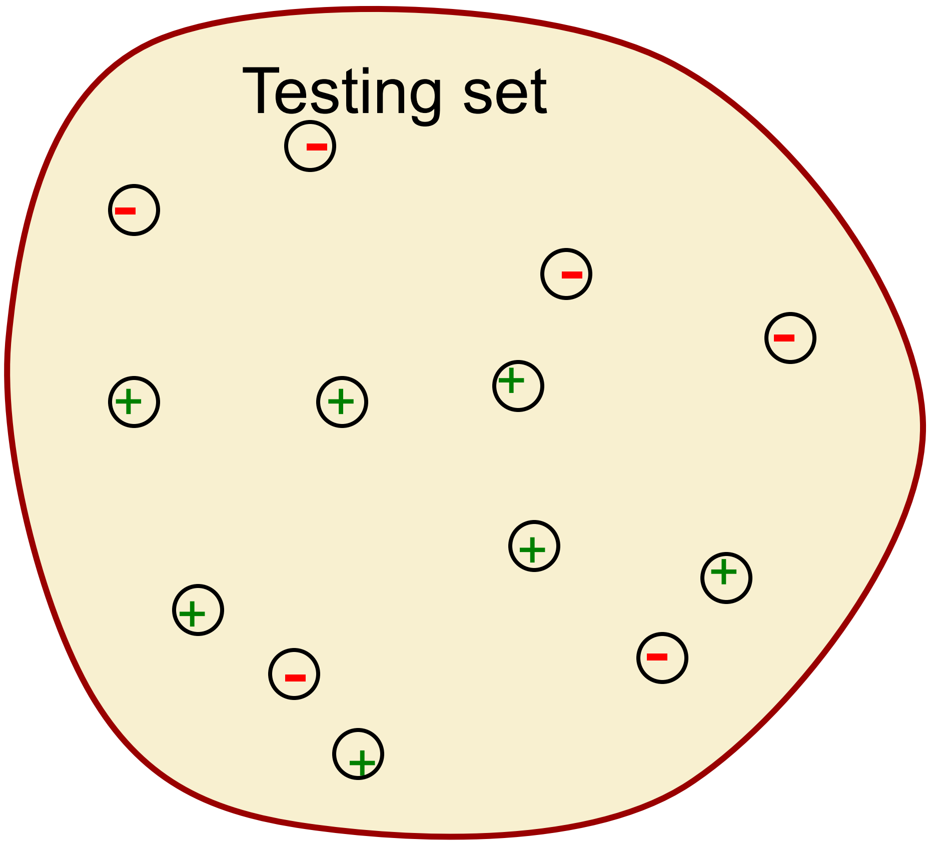
Cross-Validation

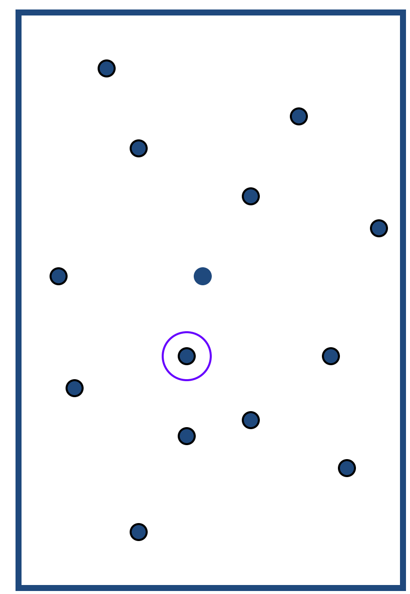
- Split original set of examples, train



Cross-Validation

- Evaluate hypothesis on testing set

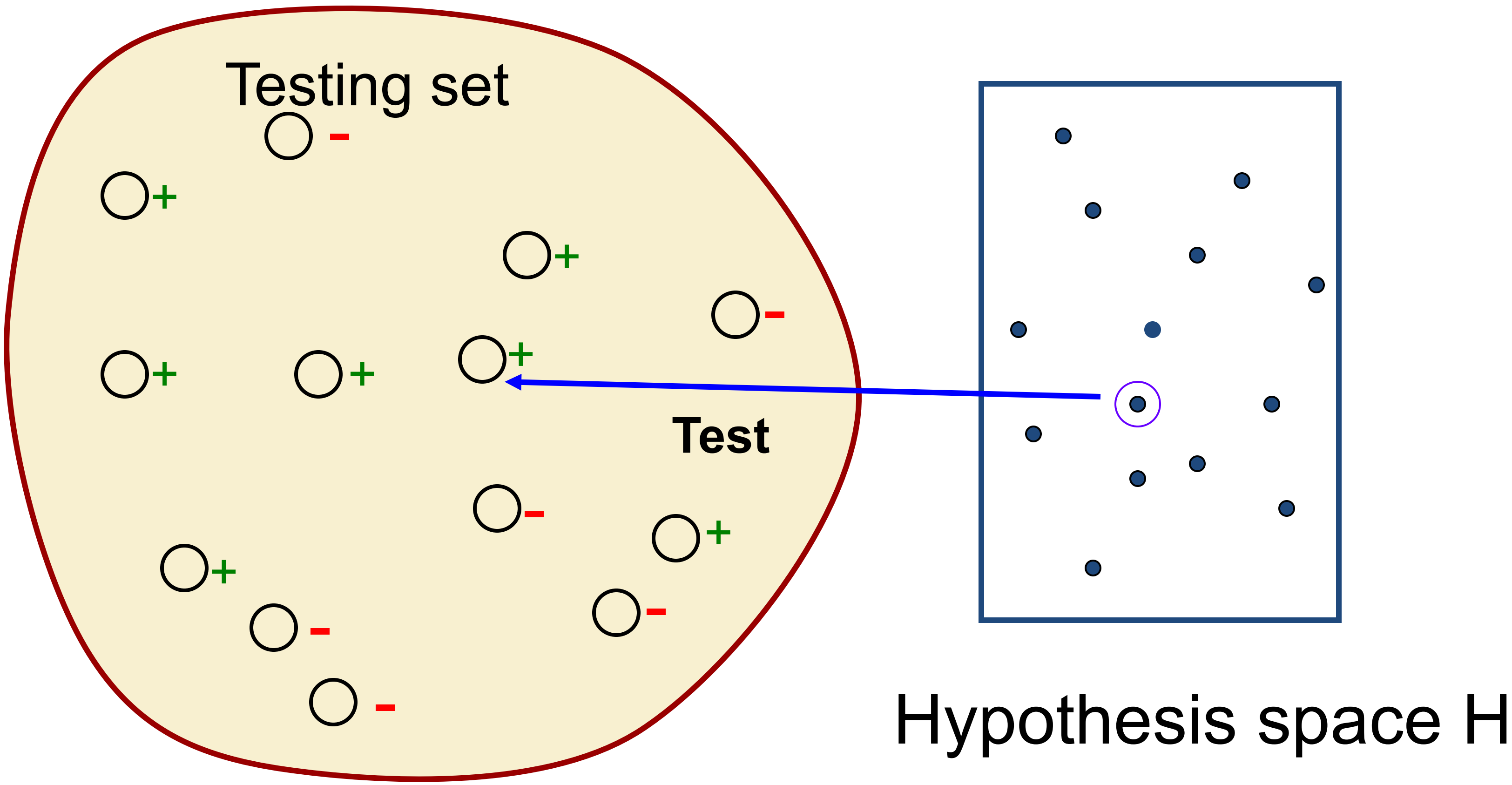




Hypothesis space H

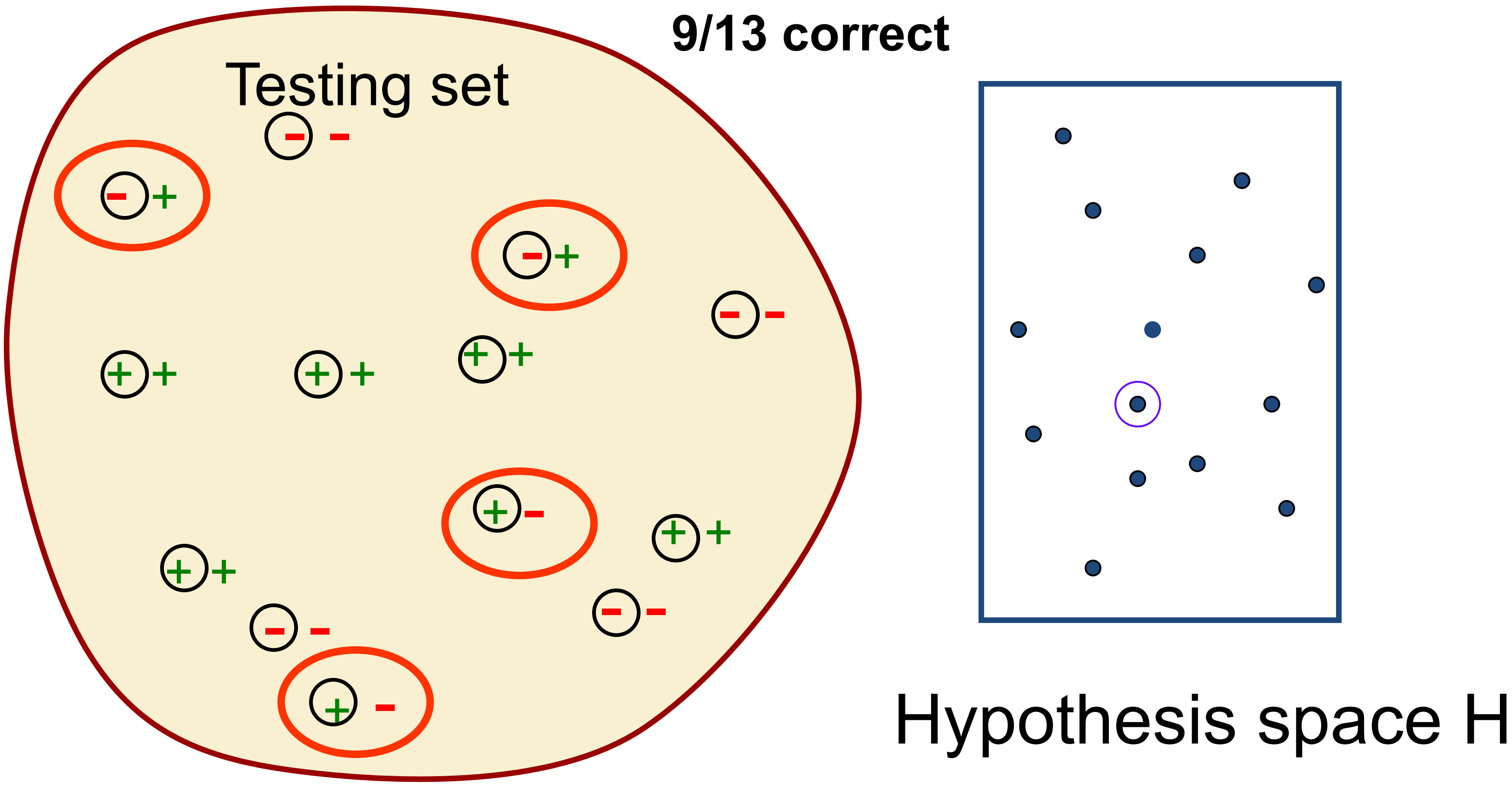
Cross-Validation

- Evaluate hypothesis on testing set



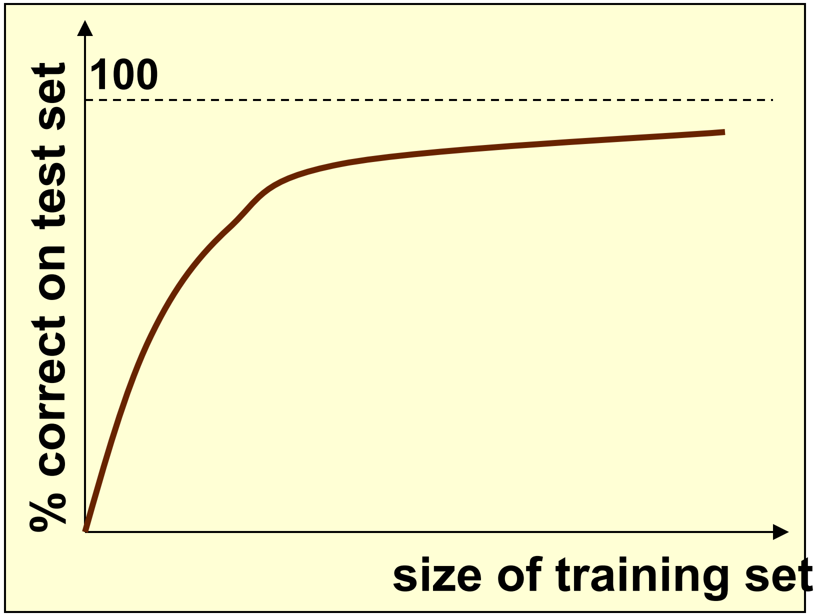
Cross-Validation

- Compare true concept against prediction



Performance Issues

- Assessing performance:
  - Training set and test set
  - Learning curve

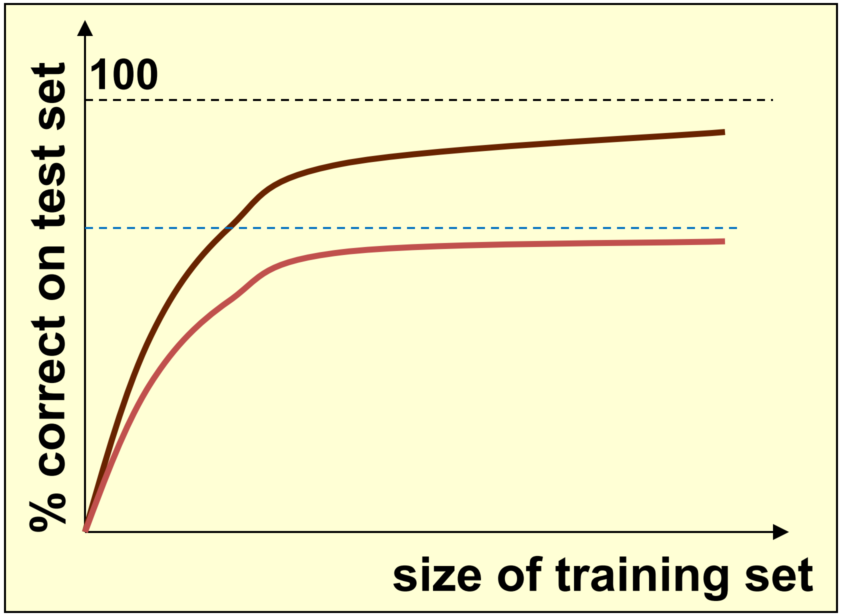


Typical learning curve

Performance Issues

- Assessing performance:
  - Training set and test set
  - Learning curve

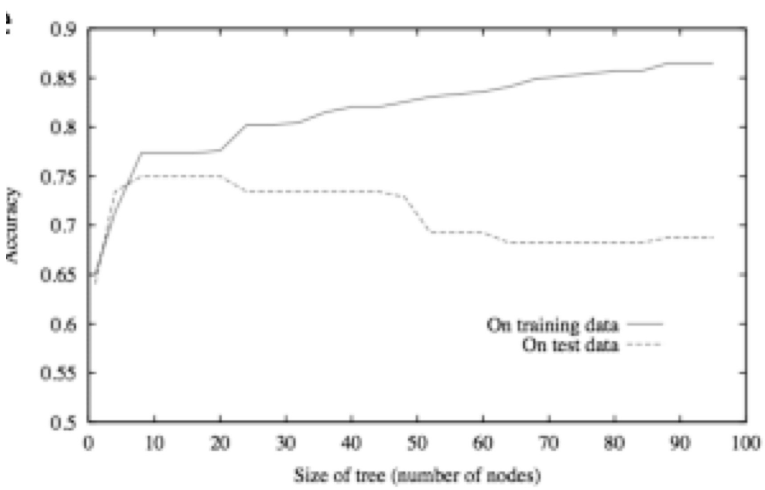
Some concepts are unrealizable within a machine's capacity



Typical learning curve

Performance Issues

- Assessing performance:
  - Training set and test set
  - Learning curve
- Overfitting
  - Classifier works well on training set, poorly on test set



- Next class: more decision trees