

Decision Tree Learning

CS4780/5780 – Machine Learning
Fall 2012

Thorsten Joachims
Cornell University

Reading: Mitchell Sections 2.1, 2.2, 2.5-2.5.2, 2.7, Chapter 3

Outline

- Hypothesis space
- Version space
- Inductive learning hypothesis
- List-then-eliminate algorithm
- Decision tree representation
- Classifying with a decision tree
- ID3 decision tree learning algorithm
- Entropy, Information gain
- Overfitting

Concept Learning

- **Definition:**

Acquire an operational definition of a general category of objects given positive and negative training examples.

	correct (3)	color (2)	original (2)	presentation (3)	binder (2)	A+
1	complete	yes	yes	clear	no	yes / +1
2	complete	no	yes	clear	no	yes / +1
3	partial	yes	no	unclear	no	no / -1
4	complete	yes	yes	clear	yes	yes / +1

Hypothesis Space

	correct (3)	color (2)	original (2)	presentation (3)	binder (2)	A+
1	complete	yes	yes	clear	no	yes / +1
2	complete	no	yes	clear	no	yes / +1
3	partial	yes	no	unclear	no	no / -1
4	complete	yes	yes	clear	yes	yes / +1

Instance Space X: Set of all possible objects described by attributes.

Target Function f: Maps each instance $x \in X$ to target label $y \in Y$ (hidden).

Hypothesis h: Function that approximates f.

Hypothesis Space H: Set of functions we allow for approximating f.

Training Data S: Set of instances labeled with target function f.

Consistency

Definition: A hypothesis h is consistent with a set of training examples S of target concept f if and only if $h(x) = y$ for each training example $(x, y) \in S$.

$$\text{Consistent}(h, S) \equiv [\forall (x, y) \in S : h(x) = y]$$

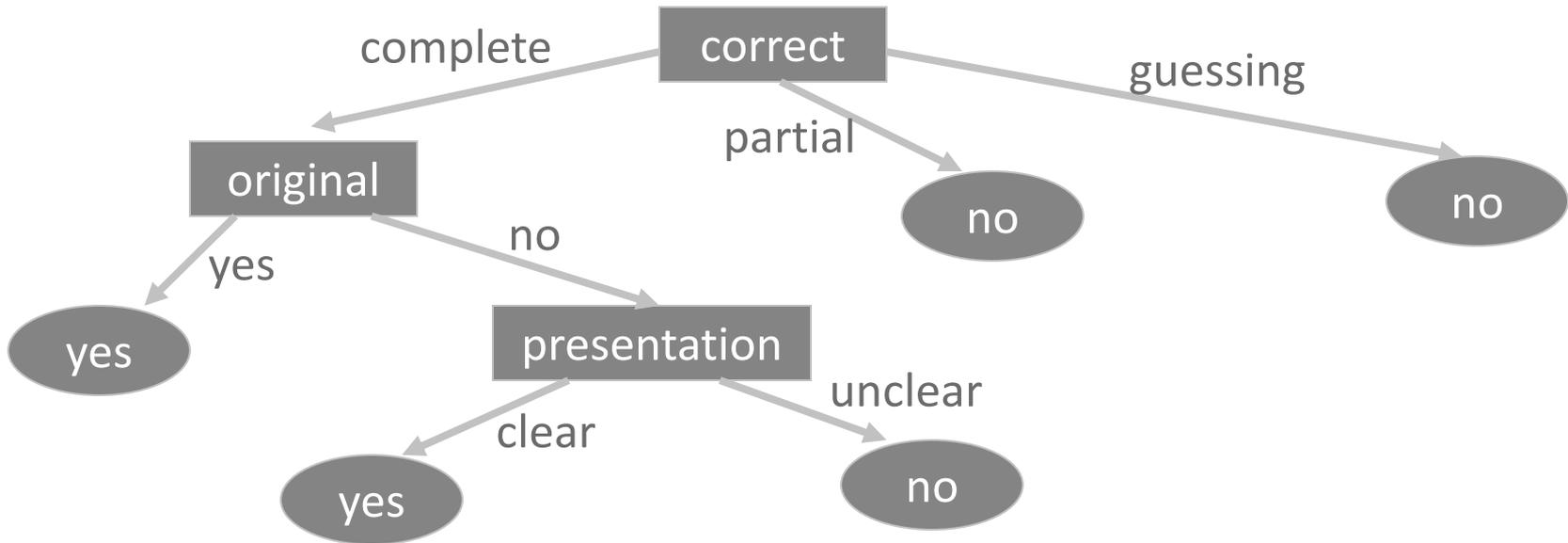
	correct (3)	color (2)	original (2)	presentation (3)	binder (2)	A+
1	complete	yes	yes	clear	no	yes / +1
2	complete	no	yes	clear	no	yes / +1
3	partial	yes	no	unclear	no	no / -1
4	complete	yes	yes	clear	yes	yes / +1

Version Space

Definition: *The version space, $VS_{H,S}$, with respect to hypothesis space H and training examples S , is the subset of hypotheses from H consistent with all training examples in S .*

$$VS_{H,S} \equiv \{h \in H \mid \text{Consistent}(h, S)\}$$

Decision Tree Example: A+ Homework



	correct (3)	color (2)	original (2)	presentation (3)	binder (2)	A+
1	complete	yes	yes	clear	no	yes / +1
2	complete	no	yes	clear	no	yes / +1
3	partial	yes	no	unclear	no	no / -1
4	complete	yes	yes	clear	yes	yes / +1

Top-Down Induction of DT (simplified)

Training Data: $S = ((\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n))$

TDIDT(S, y_{def})

- IF(all examples in S have same class y)
 - Return leaf with class y (or class y_{def} , if S is empty)
- ELSE
 - Pick A as the “best” decision attribute for next node
 - FOR each value v_i of A create a new descendent of node
 - $S_i = \{ (\vec{x}, y) \in S : \text{attribute } A \text{ of } \vec{x} \text{ has value } v_i \}$
 - Subtree t_i for v_i is TDIDT(S_i, y_{def})
 - RETURN tree with A as root and t_i as subtrees

Example: TDIDT

TDIDT(S, y_{def})

- IF (all ex in S have same y)

- Return leaf with class y
(or class y_{def} if S is empty)

- ELSE

- Pick A as the “best” decision attribute for next node

- FOR each value v_i of A create a new descendent of node

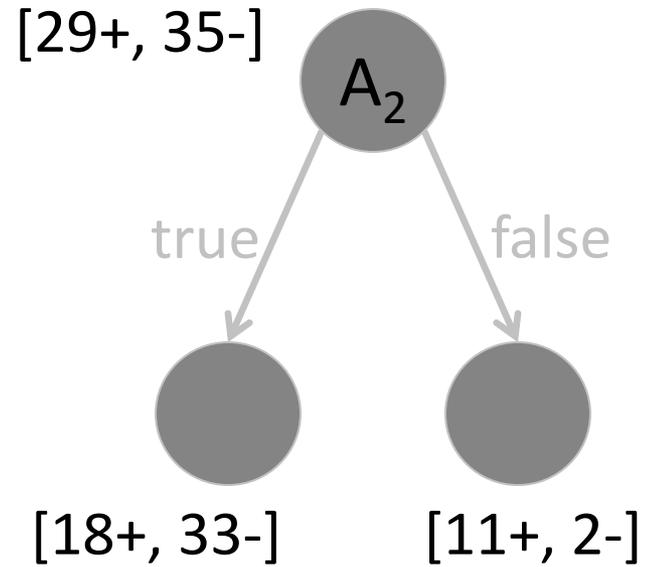
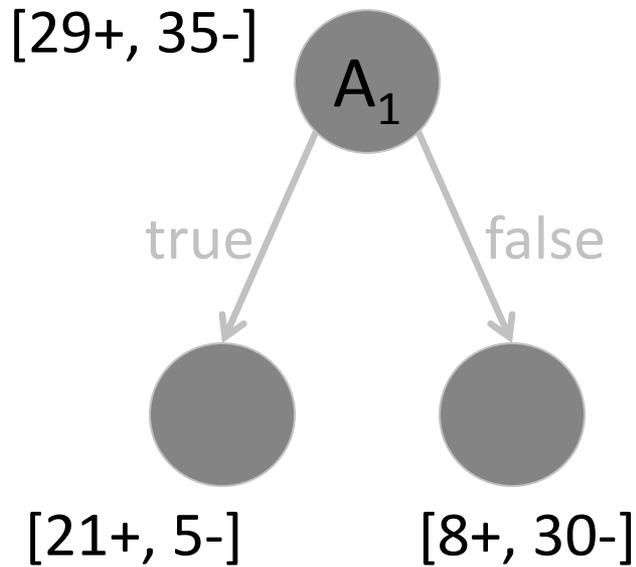
- $S_i = \{ (\vec{x}, y) \in S : \text{attr } A \text{ of } \vec{x} \text{ has val } v_i \}$

- Subtree t_i for v_i is TDIDT(S_i, y_{def})

- RETURN tree with A as root and t_i as subtrees

	CO P	A^+
$\vec{x}_1 = (c, y, c)$		$y_1 = +1$
$\vec{x}_2 = (c, n, u)$		$y_2 = -1$
$\vec{x}_3 = (c, y, u)$		$y_3 = +1$
$\vec{x}_4 = (c, n, c)$		$y_4 = +1$
$\vec{x}_5 = (p, y, c)$		$y_5 = -1$
$\vec{x}_6 = (g, y, c)$		$y_6 = -1$
$\vec{x}_7 = (c, y, c)$		$y_7 = +1$
$\vec{x}_8 = (c, y, u)$		$y_8 = +1$
$\vec{x}_9 = (p, y, c)$		$y_9 = -1$
$\vec{x}_{10} = (c, y, c)$		$y_{10} = +1$

Which Attribute is “Best”?



Example: Text Classification

- Task: Learn rule that classifies Reuters Business News
 - Class +: “Corporate Acquisitions”
 - Class -: Other articles
 - 2000 training instances
- Representation:
 - Boolean attributes, indicating presence of a keyword in article
 - 9947 such keywords (more accurately, word “stems”)

LAROCHE STARTS BID FOR NECO SHARES

Investor David F. La Roche of North Kingstown, R.I., said he is offering to purchase 170,000 common shares of NECO Enterprises Inc at 26 dlrs each. He said the successful completion of the offer, plus shares he already owns, would give him 50.5 pct of NECO's 962,016 common shares. La Roche said he may buy more, and possible all NECO shares. He said the offer and withdrawal rights will expire at 1630 EST/2130 gmt, March 30, 1987.

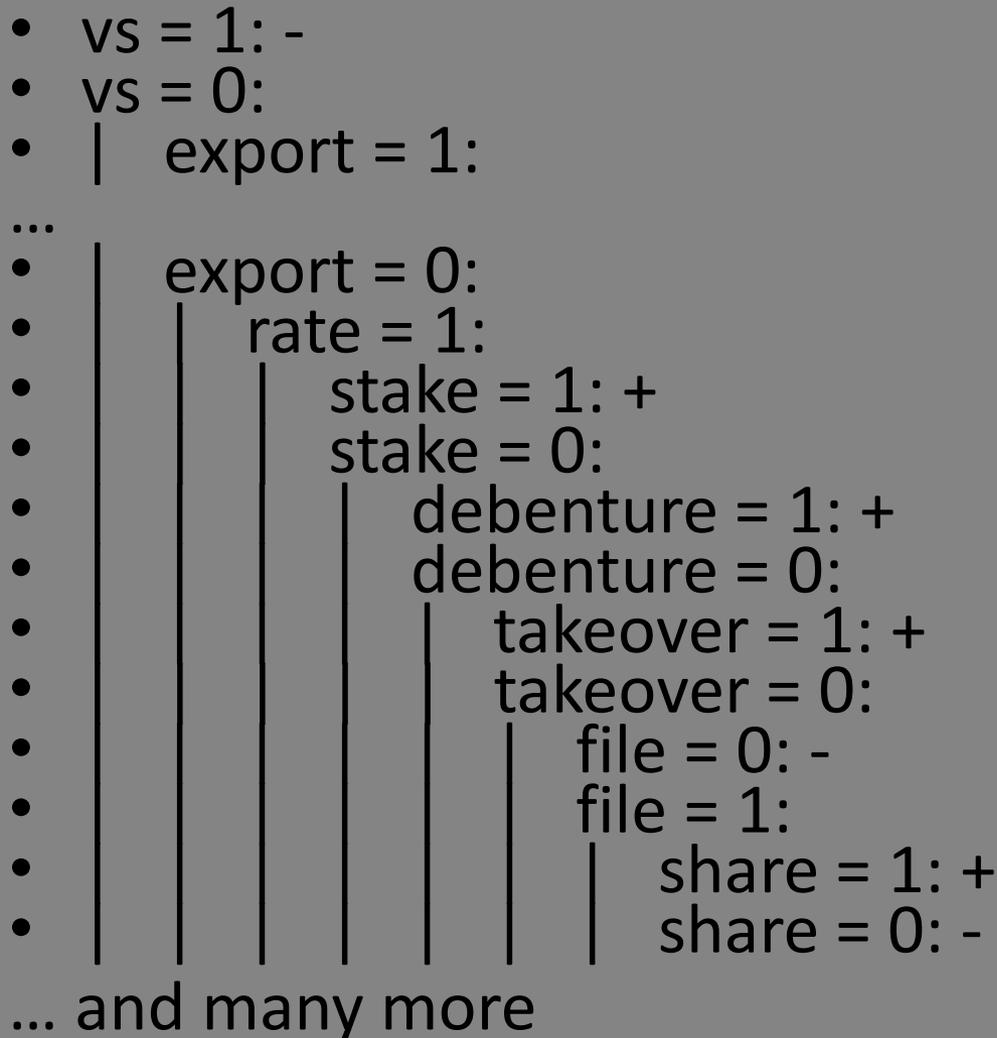


SALANT CORP 1ST QTR FEB 28 NET

Oper shr profit seven cts vs loss 12 cts.
Oper net profit 216,000 vs loss 401,000. Sales 21.4 mln vs 24.9 mln.
NOTE: Current year net excludes 142,000 dlr tax credit. Company operating in Chapter 11 bankruptcy.



Decision Tree for “Corporate Acq.”



Total size of tree:

- 299 nodes

Note: word stems expanded for improved readability.