Supervised Learning

Introduction to Artificial Intelligence

COS302

Michael L. Littman

Fall 2001

Administration

Exams graded!

http://www.cs.princeton.edu/courses/archive/fall01/cs302/whats-new.html

Project groups.

Supervised Learning

Most studied in machine learning.

http://www1.ics.uci.edu/~mlearn/MLRepository.html

Set of examples (usually numeric vectors). Split into:

Training: Allowed to see it

Test: Want to minimize error here

Another Significant App

Name	A	B	C	D	E	F	G	
1. Jeffrey B.	1	0	1	0	1	0	1	-
2. Paul S.	0	1	1	0	0	0	1	-
3. Daniel C.	0	0	1	0	0	0	0	-
4. Gregory P.	1	0	1	0	1	0	0	-
5. Michael N.	0	0	1	1	0	0	0	-
6. Corinne N.	1	1	1	0	1	0	1	+
7. Mariyam M.	0	1	0	1	0	0	1	+
8. Stephany D.	1	1	1	1	1	1	1	+
9. Mary D.	1	1	1	1	1	1	1	+
10. Jamie F.	1	1	1	0	0	1	1	+

Features

A: First name ends in a vowel?

B: Neat handwriting? (Lisa test.)

C: Middle name listed?

D: Senior?

E: Got extra-extra credit?

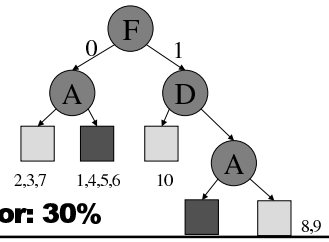
F: Google brings up home page?

G: Google brings up reference?

Decision Tree

Internal nodes: features

Leaves: classification



Error: 30%

Search

Given a set of training data, pick a decision tree: search problem!

Challenges:

- Scoring function?
- Large space of trees.

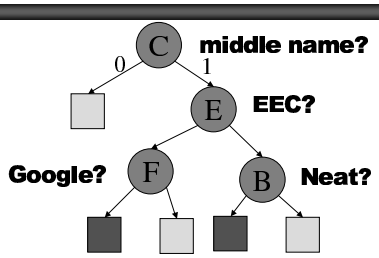
Scoring Function

What's a good tree?

- Low error on training data
- Small

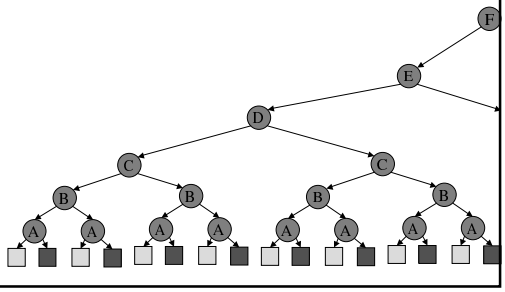
Small tree is obviously not enough, why isn't low error?

Low Error Not Enough

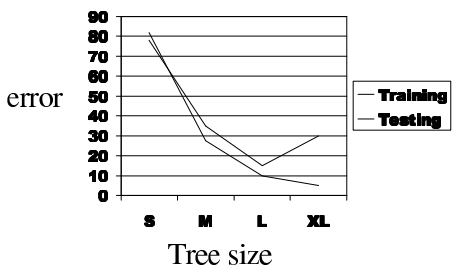


Training set Error: 0% (can always do this?)

Memorizing the Data



"Learning Curve"



What's the Problem?

Memorization w/o generalization

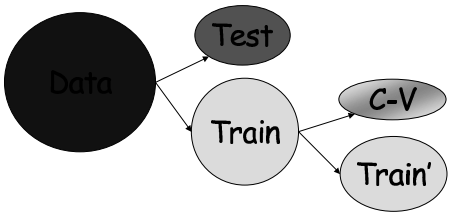
Want a tree big enough to be correct, but not so big that it gets distracted by particulars.

But, how can we know?

(Weak) theoretical bounds exist.

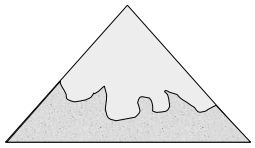
Cross-validation

Simple, effective ~~hack~~ method.



Concrete Idea: Pruning

Use Train' to find tree w/ no error.
Use C-V to score prunings of tree.
Return pruned tree w/ max score.



How Find the Tree?

Lots to choose from.
Could use local search.
Greedy search…

Why Might This Fail?

No target function, just noise
Target function too complex (2^{2^n} possibilities, parity)
Training data doesn't match target function (PAC bounds)

Theory: PAC Learning

***Probably Approximately* Correct**
Training/testing from distribution.
With probability 1-δ, learned rule will have error smaller than ε.
Bounds on size of training set in terms of δ, ε, "dimensionality" of the target concept.

Classification

Naïve Bayes classifier
Differentiation vs. modeling
More on this later.

What to Learn

- Decision tree representation
- Memorization problem: causes and cures (cross-validation, pruning)
- Greedy heuristic for finding small trees with low error

Homework 9 (due 12/5)

1. Write a program that decides if a pair of words are synonyms using wordnet. I'll send you the list, you send me the answers.
2. Draw a decision tree that represents (a) $f_1+f_2+\ldots+f_n$ (or), (b) $f_1 f_2 \ldots f_n$ (and), (c) parity (odd number of features "on").
3. More soon