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

<table>
<thead>
<tr>
<th>Name</th>
<th>ABCDEFG</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Jeffrey B.</td>
<td>101010101</td>
</tr>
<tr>
<td>2. Paul S.</td>
<td>0110001-</td>
</tr>
<tr>
<td>3. Daniel C.</td>
<td>0010000-</td>
</tr>
<tr>
<td>4. Gregory P.</td>
<td>1010100-</td>
</tr>
<tr>
<td>5. Michael H.</td>
<td>0011000-</td>
</tr>
<tr>
<td>6. Corinne N.</td>
<td>1110101+</td>
</tr>
<tr>
<td>7. Maryam M.</td>
<td>0101001+</td>
</tr>
<tr>
<td>8. Stephanie D.</td>
<td>1111111+</td>
</tr>
<tr>
<td>9. Mary D.</td>
<td>1111111+</td>
</tr>
<tr>
<td>10. Jamie F.</td>
<td>1110011+</td>
</tr>
</tbody>
</table>

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

middle name?

Google?

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

Memorizing the Data

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.

“Learning Curve”

error

Tree size
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 \(1-\delta\), learned rule will have error smaller than \(\epsilon\).
Bounds on size of training set in terms of \(\delta, \epsilon\) “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