

## CS 374 Assignment #8 Evaluation of Learning Algorithms

Due the week of April 7, 2008

This week you will step back from your study of individual classifier-learning algorithms in order to evaluate them both empirically and analytically. The readings and exercises will focus on evaluation methodology, but the broader goal of this assignment is to give you the opportunity to reflect on all that you've already learned this semester.

### 1 Assessing and Comparing Classifier-Learning Algorithms

#### 1.1 Reading

Please read the following:

- Mitchell, Chapter 5 and
- Alpaydin, Chapter 14 (through 14.7.2).

The Alpaydin reading covers more than the chapter in Mitchell, but the explanations in Mitchell are more clear. So while they can be read in any order, I recommend reading Mitchell first.

#### 1.2 Exercises

To check your understanding of the readings, please do the following exercises:

- 5.1-5.4 on page 152 of Mitchell;
- 2 and 3 on page 349 of Alpaydin.

#### 1.3 Report on the Evaluation of Learning Algorithms

Your central task this week is to write a report that discusses and compares the classifier-learning algorithms you have studied this semester, including:

- Naive Bayes
- Backpropagation for multilayer perceptrons
- SMO
- C4.5
- k Nearest Neighbor (specifically, 1-NN and 3-NN).

Your report should include an empirical evaluation as outlined below. It should also include a general analysis and discussion of the pros and cons of each of the methods evaluated. You might even find it useful to consider some of the major theorems that relate to each of the algorithms.

### 1.3.1 Empirical Evaluation: Comparison of Average Error Rates

Your report should include comparisons of the average error rates of each of the algorithms on several data sets. The specific data sets to be considered are:

- `vote`
- `vehicle`
- `ionosphere`
- `autos`

These can all be found in

```
~andrea/shared/cs374/UCI
```

This directory includes many data sets from the UC Irvine Machine Learning Repository, which have been converted into arff format for Weka. Your evaluation should certainly include the specific data sets named here, but feel free to extend your evaluation to other data sets as well. I selected the four data sets above for several reasons:

- They include both large and small numbers of examples.
- They include both continuous and nominal attributes.
- Some have missing attribute values.
- They include both two-class and multi-class problems.
- All of the algorithms will run to completion on them in a reasonable amount of time. (Note that “reasonable” is not the same as “fast”. You will likely have to wait at least 15 minutes for the experiments to run to completion.)

Use the Weka Experimenter (recall the exercise you did in order to find the optimal value for  $k$  for the  $k$ -Nearest Neighbor assignment). You should perform 10-fold cross validation for each algorithm for each data set. You do not need to repeat this more than one time. (So be sure to adjust the parameter in the right half of the window that asks for the number of times to repeat. The default is 10 – change it to 1.) Include tables with the results of your experiments in the report you write. How do the algorithms compare to each other? Which differences in error rates are statistically significant? Be sure to compare each algorithm to all of the others.

For more information on the Experimenter and, in particular, the “Analyze” panel, see Witten and Frank (starting on page 440).

### 1.3.2 General Analysis of Learning Algorithms

In addition to performing the tests specified above, you should take this assignment as an opportunity to reflect on the general advantages and disadvantages of each of the algorithms you have studied this semester. Alpaydin notes (on page 330) that criteria for evaluating and comparing classification algorithms include

- training time and space complexity,
- testing time and space complexity,

- interpretability (or comprehensibility of the learned model), and
- easy programmability.

Be sure to include these in your report. You might also consider the size of the learned model. As noted above, you might even want to refer to some of the major theorems you have proved.

### 1.3.3 Optimizing the Algorithms

As an extra challenge (this is not a requirement for the assignment), try out various parameter settings for each of the learning algorithms to find those that work best on the data sets above.

## 2 Challenges to Evaluation Methodology

Many researchers in the field of machine learning are concerned with making evaluation methodology better. Sometimes work toward that goal considers ways in which some of the statistics we report might be misleading. Others propose new frameworks for evaluating algorithms.

### 2.1 Reading

Please read “The Case Against Accuracy Estimation for Comparing Induction Algorithms” by Provost, Fawcett, and Kohavi. This paper appeared in ICML-98.

### 2.2 Reaction

Write a brief (one- to two-page) summary and reaction to the paper. Are you convinced by the authors’ arguments? Be prepared to discuss the paper in your tutorial session.