COMP4702/COMP7703 - Machine Learning

Prac 9 – Combining Learners

Aims:

- To perform some experiments to highlight issues involved in assessing the performance of predictors.
- To explore the performance of bagging and boosting ensemble techniques.
- To produce some assessable work for this subject.

Procedure:

Assessing Algorithms

In this prac we will use the Weka software. Weka has an implementation of multilayer perceptrons trained with backpropagation (in the "functions" folder of classifiers). Open this up and read the information associated with the MLP model by clicking the "More" button (including how to create/modify an MLP using the GUI). In the first part of this prac we will experiment with training MLPs on two of the datasets used previously – the diabetes and sonar datasets. Choose a model and associated parameters (e.g. number of hidden units, learning rate, momentum, number of epochs) and use this setup **unchanged** through questions 1 to 4 below.

→ Note: Weka insists that the first line of the file should be a header (containing the names of each attribute in the data file). As a quick fix to get around this issue, I labelled the attributes 1, 2, 3, … and so forth. This results in nonsensical attribute names, but we don't have this information and it's not important for our purposes.

→ **Q1:** Train your network 10 times from random weight initializations (using the default 66% split of training/test sets) and record the min, max, mean and standard deviation of your 10 results (in terms of percentage correct/incorrect) on each dataset (diabetes and sonar). Note: to do different random weight initializations, you will need to vary the random seed parameter of the MLP (left- click on the classifier "bar" once you're selected MLP). For these experiments, use random seeds 0,1,…,9.

→ Pick the experiment from Q1 that was the closest to your mean result (i.e. some random seed value). Using this setup:

- **Q2:** Perform 10-fold cross-validation and record percentage correct/incorrect result.
- **Q3:** Do 10 random partitions of the dataset (66% training/test split) and retrain 10 times and record the min, max, mean and standard deviation of your 10 results (in terms of percentage correct/incorrect). Note: to do repartition the dataset in Weka, you need to go to the Preprocess tab and use the "Random" filter (under filters-unsupervised-instance). There is another random seed value for this filter – use values 42,43,…51 here. Click Apply to apply the filter, then return to the Classifier tab to use the new split.
- **Q4:** Perform a bootstrapping process by training your classifier using resampled versions of the dataset, again recording min, max, mean and standard deviation. Note: use the "Resample" filter repeatedly in Weka to do this, random seeds 1,2,…,10.

You should also have a look at some of the output that WEKA produces and connect it back to lecture notes (confusion matrix, true/false positives/negatives, precision, recall).

Bagging and Boosting

To do this part of the prac, you will need to refer to the highly readable paper (available on the course material webpage):

D. Optiz and R. Maclin. Popular ensemble methods: an empirical study. Journal of Artificial Intelligence Research v.11 (1999) pp.169-198.

In this paper, the authors produce a large set of experimental results, including MLPs and bagged and boosted ensembles of MLPs. The datasets used include most of the datasets we have used during earlier pracs.

- → **Q5:** For the sonar, diabetes, ionosphere and glass datasets, attempt to reproduce the results reported in the paper – specifically in the first column of Table 2. You should be able to find most of the experimental details used by the authors in the paper – provide comments if you have to make any assumptions for your experiments. Record your results and comment briefly on any difference between them and the results in the paper.

Weka contains implementations of both bagging and AdaboostM1. To use them, you select one as the classifier from the “meta” folder. Once you’ve done this, left-clicking on the classifier “bar” opens a window detailing the parameters of the method, including what base classifier to use. Any of the models in Weka can be selected in the same way as for a single classifier. In our case we will build ensembles of MLPs. Read the information in Weka to make sure you understand what you’re doing!

- → **Q6:** For the datasets used in Q5, attempt to reproduce the results in columns 3 and 5 of the paper. Provide comments if you have to make any assumptions for your experiments. Record your results and comment briefly on any difference between them and the results in the paper.
- → **Q7:** Try bagging and boosting with a single-layer perceptron (set the “Hiddenlayers” parameter to 0 for an MLP in Weka). Record your results for the datasets used in Q5 and Q6 and comment briefly on how they compare to your results from Q5 and Q6.

Glass Dataset

```
1. Title: Glass Identification Database

2. Sources:
    (a) Creator: B. German
        -- Central Research Establishment
           Home Office Forensic Science Service
           Aldermaston, Reading, Berkshire RG7 4PN
    (b) Donor: Vina Spiehler, Ph.D., DABFT
               Diagnostic Products Corporation
               (213) 776-0180 (ext 3014)
    (c) Date: September, 1987

3. Past Usage:
    -- Rule Induction in Forensic Science
       -- Ian W. Evett and Ernest J. Spiehler
```

-- Central Research Establishment
   Home Office Forensic Science Service
   Aldermaston, Reading, Berkshire RG7 4PN
-- Unknown technical note number (sorry, not listed here)
-- General Results: nearest neighbor held its own with respect to the rule-based system

4. Relevant Information:n

Vina conducted a comparison test of her rule-based system, BEAGLE, the nearest-neighbor algorithm, and discriminant analysis. BEAGLE is a product available through VRS Consulting, Inc.; 4676 Admiralty Way, Suite 206; Marina Del Ray, CA 90292 (213) 827-7890 and FAX: -3189. In determining whether the glass was a type of "float" glass or not, the following results were obtained (# incorrect answers):

Type of Sample		Beagle	NN	DA
Windows that were float processed	(87)	10	12	21
Windows that were not:	(76)	19	16	22

The study of classification of types of glass was motivated by criminological investigation. At the scene of the crime, the glass left can be used as evidence...if it is correctly identified!

5. Number of Instances: 214
6. Number of Attributes: 10 (including an Id#) plus the class attribute
   -- all attributes are continuously valued
7. Attribute Information:
   1. Id number: 1 to 214
   2. RI: refractive index
   3. Na: Sodium (unit measurement: weight percent in corresponding oxide, as are attributes 4-10)
   4. Mg: Magnesium
   5. Al: Aluminum
   6. Si: Silicon
   7. K: Potassium
   8. Ca: Calcium
   9. Ba: Barium
   10. Fe: Iron
   11. Type of glass: (class attribute)
       -- 1 building_windows_float_processed
       -- 2 building_windows_non_float_processed
       -- 3 vehicle_windows_float_processed
       -- 4 vehicle_windows_non_float_processed (none in this database)
       -- 5 containers
       -- 6 tableware
       -- 7 headlamps

8. Missing Attribute Values: None

Summary Statistics:

Attribute:	Min	Max	Mean	SD	Correlation with class
2. RI:	1.5112	1.5339	1.5184	0.0030	-0.1642
3. Na:	10.73	17.38	13.4079	0.8166	0.5030
4. Mg:	0	4.49	2.6845	1.4424	-0.7447
5. Al:	0.29	3.5	1.4449	0.4993	0.5988
6. Si:	69.81	75.41	72.6509	0.7745	0.1515
7. K:	0	6.21	0.4971	0.6522	-0.0100
8. Ca:	5.43	16.19	8.9570	1.4232	0.0007
9. Ba:	0	3.15	0.1750	0.4972	0.5751
10. Fe:	0	0.51	0.0570	0.0974	-0.1879

9. Class Distribution: (out of 214 total instances)
    -- 163 Window glass (building windows and vehicle windows)
       -- 87 float processed
          -- 70 building windows
          -- 17 vehicle windows
       -- 76 non-float processed
          -- 76 building windows

```
      -- 0 vehicle windows
-- 51 Non-window glass
   -- 13 containers
   -- 9 tableware
   -- 29 headlamps
```