Auto Learn Data Analysis
CSC 522
Homework4

Rishabh Sinha(rsinha2)

Q 1. (10 points) Ensemble methods

a) (2 points) What are the two requirements for an ensemble classifier to perform better than a single classifier?

There are two necessary as well as sufficient conditions for ensemble classifiers to perform better than any of its individual members are:
1. the base classifiers should be independent of each other
2. the base classifiers should do better than a classifier that performs random guessing.

b) (2 points) What are some ways to ensure total independence among classifiers? What about partial independence?

In practice total Independence is very difficult to ensure amongst classifiers, But those to ensure partial independence that is slight correlation are:

1. Ways to ensure independence among classifiers is to divide the training data into distinct partitions and train the individual classifiers with each of the partitioned data.
2. Classifiers can also be built using different sets of input features to make them independent.
3. Classifiers can also be built using different sets of class labels, (for data sets with large number of class labels) to make them independent.
4. Classifiers can also be built different learning algorithms example, Bayesian learning, Neural Network, SVM etc and then combined for ensemble.

c) (2 points) What are unstable classifiers and why are they favored for ensemble methods?

An unstable classifier is a type of classifier where in small changes in the training data, causes large changes in the classifier model. That is when the training set changes even by a small amount, the classifiers performs a lot differently. An example of an unstable classifier is decision tree.
Unstable classifiers are favored for ensemble as, when different individual classifiers of an ensemble classifier are trained using data (ex. Through bagging, or through other measures described above in (b)), there may be only small differences in input data/or input features etc. between the classifiers, yet the difference in classifier structure must be large enough that they remain uncorrelated, thus ensuring a better performance accuracy of the ensemble classifier.

d) (4 points) Take the general procedure for constructing an ensemble classifier from page 280 (Algorithm 5.5) and modify it such that each D_i is a bootstrap sample and has only a subset of the training features. Use a set of unstable classifiers as base classifiers and perform accuracy weighted prediction to assign class label. Be as specific as possible.

Algorithm 5.5 procedure for ensemble method.
l: Let D denote the original training data, k denote the number of base classifiers, and T be the test data.
2: for i:l to k do
3: Create training set, Di ftom D using bootstrap.
4: Build a unstable base classifier Ci from Di.
5: end for
6: for each test record x € T do
7: C*(x) : accuracyWeightedVote(C1(x),C2(*),..., Ck(*))
8: end for

This algorithm, would take bootstrap samples from the set of data, thus on an average 63.2% of data will get selected as a part of each sample, for different set of features of the same data, since unstable classifier is used, Small changes in training data would lead to huge differences in the architecture of the Classifiers thus these classifiers can be considered as largely independent to each other. Further since, they would be different classifiers all together, their accuracies will also be largely different. Thus we would use weighted average based on accuracy to determine the final class label for each data element in the test dataset. That is we would assign a higher weight to classifiers with better accuracy and lower to those with lower accuracy, to ensure overall better performance of the ensemble classifier.

Q 2. (10 points) Accuracy Measures Say we build a decision tree model for a spam email classification problem, Table 1 is the classification result we got: Table 1: Results of a decision tree spam classifier on emails (Note: Spam is presented as +, not Spam is presented as -)

Email ID	Actual Label	Predicted Label
1	+	-
2	-	-
3	-	-
4	-	-
5	-	-
6	-	-
7	-	-
8	-	+
9	+	+
10	-	-
11	-	-
12	-	-
13	-	-
14	-	+

a) (6 points) Using the results in Table 1, compute (i) contingency table, (ii) precision (iii) recall (iv) F-measure (v) accuracy.

i) contingency table is:

	Predicted Class		
Actual Class		Class = Yes	Class = No
	Class = Yes	1	1
	Class = No	2	10

ii) precision = a/(a+c) = 1/(1+2) = 0.3333

iii) recall = a/(a+b) = 1/(1+1) = 0.5

iv) F-Measure = 2a/ (2a + b + c) = 2/(2+1+2) = 0.4

v) accuracy = (1+10)/14 = 0.7857

b) (4 points) From the Table 1 result, we find that spam email classification problem is a class imbalance problem (there are more Not Spam emails compared to Spam emails). List the techniques you can use to handle such class imbalance problems (at least two), and describe each in a sentence.

Techniques to solve class imbalance problems are:

1. Class-based ordering (RIPPER): Rules for rare classes have higher priority in case of RIPPER.
2. Cost-sensitive classification: – Misclassifying rare class as majority class is more expensive than misclassifying majority as rare class
3. Cost Matrix: Cost Matrix contains cost values. Larger cost refers to worse classification. Cost is computed by multiplying values in cost matrix to respective ones in confusion matrix and summing them up.
4. Sampling Based approaches by under-sampling the majority class or over sampling the minority class.

Q 3. (10 points) Association Rule Mining Consider the market basket transactions shown in Table 2.

Transaction ID	Transaction
1	Milk, Beer, Diapers
2	Bread, Butter, Milk
3	Milk, Diapers, Cookies
4	Bread, Butter, Cookies
5	Beer, Cookies, Diapers
6	Milk, Diapers, Bread, Butter
7	Bread, Butter, Diapers
8	Beer, Diapers
9	Milk, Diapers, Bread, Butter

10	Beer, Cookies

a) (2 points) What is the maximum number of association rules that can be extracted from this data (including rules that have zero support)?

The maximum number of association rules that can be extracted from this data set is:
$3^d - 2^{d+1} + 1 = 602$

b) (2 points) What is the maximum size of frequent itemsets that can be extracted from Table 2 (assuming minsup > 0)?

Maximum Size of frequent itemset that can extracted from the given dataset is 4. Since the largest size given in the dataset is 4.

c) (2 points) What are the maximum number of size-3 itemsets that can be derived from this dataset? (assuming minsup >= 0)?

Maximum number of size 3 itemsets that can be derived given minsup>=0 are 4C3 = 4.

d) (2 points) What is the confidence of the rules {Bread -> Milk} and {Milk -> Bread}

{Bread -> Milk}: Confidence is 3/5 = 0.6
{Milk -> Bread}: Confidence is 3/5 = 0.6

e) Under what conditions do the rules {a -> b} , {b->a} have the same confidence?

```
a -> b} , {b->a} have same confidence level whenever the individual occurrences of
a = individual occurrences of b, that is frequency(a) = frequency(b)
```