Probability Estimation and Aggregation for Rule Learning

Technical Report TUD-KE-2010-03

Jan-Nikolas Sulzmann, Johannes Fürnkranz
Technische Universität Darmstadt

Knowledge Engineering Group, Technische Universität Darmstadt

TECHNISCHE UNIVERSITÄT DARMSTADT

Knowledge Engineering

Abstract

Rule learning is known for its descriptive and therefore comprehensible classification models which also yield good class predictions. For different classification models, such as decision trees, a variety of techniques for obtaining good probability estimates have been proposed and evaluated. However, so far, there has been no systematic empirical study of how these techniques can be adapted to probabilistic rules and how these methods affect the probability-based rankings. In this paper we apply several basic methods for the estimation of class membership probabilities to classification rules. We also study the effect of a shrinkage technique for merging the probability estimates of rules with those of their generalizations. Finally, we compare different ways of combining probability estimates from an ensemble of rules. Our results show that for probability estimation it is beneficial to exploit the fact that rules overlap (i.e., rule averaging is preferred over rule sorting), and that individual probabilities should be combined at the level of rules and not at the level of theories.

1 Introduction

The main focus of symbolic learning algorithms such as decision tree and rule learners is to produce a comprehensible explanation for a class variable. Thus, they learn concepts in the form of crisp IF-THEN rules. On the other hand, many practical applications require a finer distinction between examples than is provided by their predicted class labels. For example, one may want to be able to provide a confidence score that estimates the certainty of a prediction, to rank the predictions according to their probability of belonging to a given class, to make a cost-sensitive prediction, or to combine multiple predictions.

All these problems can be solved straight-forwardly if we can predict a probability distribution over all classes instead of a single class value. A straight-forward approach to estimate probability distributions for classification rules is to compute the fractions of the covered examples for each class. However, this naïve approach has obvious disadvantages, such as that rules that cover only a few examples may lead to extreme probability estimates. Thus, the probability estimates need to be smoothed.

There has been quite some previous work on probability estimation from decision trees (so-called *probability-estimation trees (PETS)*). A very simple, but quite powerful technique for improving class probability estimates is the use of m-estimates, or their special case, the Laplace-estimates [3]. It has been shown that unpruned decision trees with Laplace-corrected probability estimates at the leaves produce quite reliable decision tree estimates [13]. A recursive computation of the m-estimate, which uses the probability disctribution at level l as the prior probabilities for level $l+1$, was proposed in [8]. In [15], a general shrinkage approach was used, which interpolates the estimated class distribution at the leaf nodes with the estimates in interior nodes on the path from the root to the leaf.

An interesting observation is that, contrary to classification, class probability estimation for decision trees typically works better on unpruned trees than on pruned trees. The explanation for this is simply that, as all examples in a leaf receive the same probability estimate, pruned trees provide a much coarser ranking than unpruned trees. In [11], a simple but elegant analysis of this phenomenon was provided, which shows that replacing a leaf with a subtree can only lead to an increase in the area under the ROC curve (AUC), a commonly used measure for the ranking capabilities of an algorithm. Of course, this only holds for the AUC estimate on the training data, but it still may provide a strong indication why unpruned PETs typically also outperform pruned PETs on the test set.

Despite the amount of work on probability estimation for decision trees, there has been hardly any systematic work on probability estimation for rule learning. Despite their obvious similarility, we nevertheless argue that a separate study of probability estimates for rule learning is necessary.

A key difference is that in the case of decision tree learning, probability estimates will not change the prediction for an example, because the predicted class only depends on the probabilities of a single leaf of the tree, and such local probability estimates are typically monotone in the sense that they all maintain the majority class as the class with the maximum probability. In the case of rule learning, on the other hand, each example may be classified by multiple rules, which may possibly predict different classes. As many tie breaking strategies depend on the class probabilities, a local change in the class probability of a single rule may change the global prediction of the rule-based classifier, even if the order of all local estimates is maintained.

Because of such non-local effects, it is not evident that the same methods that work well for decision tree learning will also work well for rule learning. Indeed, as we will see in this paper, our conclusions differ from those that have been drawn from similar experiments in decision tree learning. For example, the above-mentioned argument that unpruned trees will lead to a better (training-set) AUC than pruned trees, does not straight-forwardly carry over to rule learning, because the replacement of a leaf with a subtree is a local operation that only affects the examples that are covered by this leaf. In rule learning, on the other hand, each example may be covered by multiple rules, so that the effect of replacing one rule with multiple, more specific rules is less predictable. Moreover, each example will be covered by some leaf in a decision tree, whereas each rule learner needs to induce a separate default rule that covers examples that are covered by no other rule.

The rest of the paper is organized as follows: In Section 2 we briefly describe the basics of probabilistic rule learning and discuss the used estimation techniques used for rule probabilities. In Section 3 we describe the rule learning algorithm that we used in our experiments, contrast two approaches for the generation of a probabilistic rule set, and describe how they are used for classification. Experimental results, which compare the different probability estimation techniques in these two scenarios, are described in Section 4. We then discuss four techniques for obtaining rule probabilities from a bagging ensemble (Section 5), and compare them experimentally in Section 6. In the end, we summarize our conclusions in Section 7.[1]

[1] Parts of this paper have previously appeared as [14].

2 Rule Learning and Probability Estimation

This section is divided into two parts. The first one describes briefly the properties of conjunctive classification rules and of its extension to a probabilistic rule. In the second part we introduce the probability estimation techniques used in this paper. These techniques can be divided into basic methods, which can be used stand-alone for probability estimation, and the meta technique shrinkage, which can be combined with any of the techniques for probability estimation.

2.1 Probabilistic Rule Learning

In classification rule mining one searches for a set of rules that describes the data as accurately as possible. As there are many different generation approaches and types of generated classification rules, we do not go into detail and restrict ourselves to conjunctive rules. The *premise* of these rules consists of a conjunction of number of conditions, and in our case, the *conclusion* of the rule is a single class value. So a conjunctive classification rule r has basically the following form:

$$condition_1 \wedge \cdots \wedge condition_{|r|} \implies class \tag{2.1}$$

The size of a rule $|r|$ is the number of its conditions. Each of these conditions consists of an attribute, an attribute value belonging to its domain and a comparison determined by the attribute type. For our purpose, we consider only nominal and numerical attributes. For nominal attributes, this comparison is a test of equality, whereas in the case of numerical attributes, the test is either less (or equal) or greater (or equal). If all conditions are met by an instance, the instance is covered by the rule ($r \supseteq x$) and the class value of the rule is predicted for the instance. Consequently, the rule is called a *covering rule* for this instance.

This in mind, we can define some statistical values of a data set which are needed for later definitions. A data set consists of $|C|$ classes and n instances from which n^c belong to the class c respectively ($n = \sum_{c=1}^{|C|} n^c$). A rule r covers n_r instances which are distributed over the classes, so that n_r^c instances belong to class c ($n_r = \sum_{c=1}^{|C|} n_r^c$).

A probabilistic rule r is an extension of a classification rule, which does not only predict a single class value, but a set of *class probabilities*, which form a probability distribution over the classes. This probability distribution estimates all probabilities that a covered instance belongs to any of the class in the data set, so we get one class probability per class. The example is then classified with the most probable class. The probability that an instance x covered by rule r belongs to c can be viewed as a conditional probability $\Pr(c|r \supseteq x)$. Thus the set of class probabilities can be noted as vector of probabilities sorted by the class ordering:

$$\vec{\Pr}(r \supseteq x) = (\Pr(c_1|r \supseteq x), \cdots, \Pr(c_{|C|}|r \supseteq x)) \tag{2.2}$$

On the vector $\vec{\Pr}(r \supseteq x)$, abbreviated $\vec{\Pr}_r(x)$, we define the following maximum function

$$\max\left(\vec{\Pr}_r(x)\right) = \max_{c \in C} \Pr(c|r \supseteq x). \tag{2.3}$$

On sets of class probability vectors $\bigcup_{j=1}^{k} \vec{\Pr}_i(x)$ we define the average function

$$\operatorname{avg}\left(\bigcup\nolimits_{j=1}^{k} \vec{\Pr}_j(x)\right) = \frac{1}{k}\sum_{j=0}^{k} \vec{\Pr}_j(x) \tag{2.4}$$

and the multiplication

$$\operatorname{mult}\left(\bigcup\nolimits_{j=1}^{k} \vec{\Pr}_j(x)\right) = \left(\prod\nolimits_{j=0}^{k} \Pr(c_1|r_j \supseteq x), \cdots, \prod\nolimits_{j=0}^{k} \Pr(c_{|C|}|r_j \supseteq x)\right) \tag{2.5}$$

Obviously the results of these functions are also class probability vectors.

In the next section, we discuss some approaches for estimating these class probabilities.

2.2 Basic Probability Estimation

In this subsection we will review three basic methods for probability estimation. Subsequently, in Section 2.3, we will describe a technique known as shrinkage, which is known from various application areas, and show how this technique can be adapted to probabilistic rule learning.

All of the three basic methods we employed, calculate the relation between the number of instances covered by the rule n_r and the number of instances covered by the rule but also belong to a specific class n_r^c. The differences between the methods are the minor modifications of the calculation of this relation.

The simplest approach to rule probability estimation directly estimates a class probability distribution of a rule with the fraction of examples that belong to each class.

$$\Pr_{\text{naïve}}(c|r \supseteq x) = \frac{n_r^c}{n_r} \tag{2.6}$$

This naïve approach has several well-known disadvantages, most notably that rules with a low coverage may be lead to extreme probability values. For this reason, the use of the Laplace- and m-estimates was suggested in [3].

The Laplace estimate modifies the above-mentioned relation by adding one additional instance to the counts n_r^c for each class c. Hence the number of covered instances n_r is increased by the number of classes $|C|$.

$$\Pr_{\text{Laplace}}(c|r \supseteq x) = \frac{n_r^c + 1}{n_r + |C|} \tag{2.7}$$

It may be viewed as a trade-off between $\Pr_{\text{naïve}}(c|r \supseteq x)$ and an *a priori* probability of $\Pr(c) = 1/|C|$ for each class. Thus, it implicitly assumes a uniform class distribution.

The m-estimate generalizes this idea by making the dependency on the prior class distribution explicit, and introducing a parameter m, which allows to trade off the influence of the *a priori* probability and$\Pr_{\text{naïve}}$.

$$\Pr_{m}(c|r \supseteq x) = \frac{n_r^c + m \cdot \Pr(c)}{n_r + m} \tag{2.8}$$

The m-parameter may be interpreted as a number of examples that are distributed according to the prior probability, which are added to the class frequencies n_r^c. The prior probability is typically estimated from the data using $\Pr(c) = n^c/n$ (but one could, e.g., also use the above-mentioned Laplace-correction if the class distribution is very skewed). Obviously, the Laplace-estimate is a special case of the m-estimate with $m = |C|$ and $\Pr(c) = 1/|C|$.

2.3 Shrinkage

Shrinkage is a general framework for smoothing probabilities, which has been successfully applied in various research areas.[1]t'Its key idea is to "shrink" probability estimates towards the estimates of its generalized rules r_k, which cover more examples. This is quite similar to the idea of the Laplace- and m-estimates, with two main differences: First, the shrinkage happens not only with respect to the prior probability (which would correspond to a rule covering all examples) but interpolates between several different generalizations, and second the weights for the trade-off are not specified *a priori* (as with the m-parameter in the m-estimate) but estimated from the data.

In general, shrinkage estimates the probability $\Pr(c|r \supseteq x)$ as follows:

$$\Pr_{Shrink}(c|r \supseteq x) = \sum_{k=0}^{|r|} w_c^k \Pr(c|r_k) \tag{2.9}$$

where w_c^k are weights that interpolate between the probability estimates of the generalized rules r_k. In our implementation, we use only generalizations of a rule that can be obtained by deleting a final sequence of conditions. Thus, for a rule with length $|r|$, we obtain $|r|+1$ generalizations r_k, where r_0 is the rule covering all examples, and $r_{|r|} = r$.

The weights w_c^k can be estimated in various ways. We employ a shrinkage method proposed in [15] which is intended for decision tree learning but can be straight-forwardly adapted to rule learning. The authors propose to estimate the weights w_c^k with an iterative procedure which averages the probabilities obtained by removing training examples covered by this rule. In effect, we obtain two probabilities per rule generalization and class: the removal of an example of class c leads to a decreased probability $\Pr_{-}(c|r_k \supseteq x)$, whereas the removal of an example of a different class results in

[1] Shrinkage is, e.g., regularly used in statistical language processing [4, 12]

an increased probability $\Pr_+(c|r_k \supseteq x)$. Weighting these probabilities with the relative occurrence of training examples belonging to this class we obtain a smoothed probability

$$\Pr_{Smoothed}(c|r_k \supseteq x) = \frac{n_r^c}{n_r} \cdot \Pr_-(c|r_k \supseteq x) \tag{2.10}$$

$$+ \frac{n_r - n_r^c}{n_r} \cdot \Pr_+(c|r_k \supseteq x) \tag{2.11}$$

Using these smoothed probabilities, this shrinkage method computes the weights of these nodes in linear time (linear in the number of covered instances) by normalizing the smoothed probabilities separately for each class.

$$w_c^k = \frac{\Pr_{Smoothed}(c|r_k \supseteq x)}{\sum_{i=0}^{|r|} \Pr_{Smoothed}(c|r_i \supseteq x)} \tag{2.12}$$

Multiplying the weights with their corresponding probability we obtain "shrinked" class probabilities for the instance. Note that all instances which are classified by the same rule receive the same probability distribution. Therefore the probability distribution of each rule can be calculated in advance.

3 Rule Learning Algorithm

For the rule generation we employed the rule learner Ripper [5], arguably one of the most accurate rule learning algorithms today. We used Ripper both in ordered and in unordered mode:

Ordered Mode: In ordered mode, Ripper learns rules for each class, where the classes are ordered according to ascending class frequencies. For learning the rules of class c_i, examples of all classes c_j with $j > i$ are used as negative examples. No rules are learned for the last and most frequent class, but a rule that implies this class is added as the default rule. At classification time, these rules are meant to be used as a decision list, i.e., the first rule that fires is used for prediction.

Unordered Mode: In unordered mode, Ripper uses a one-against-all strategy for learning a rule set, i.e., one set of rules is learned for each class c_i, using all examples of classes $c_j, j \neq i$ as negative examples. At prediction time, all rules that cover an example are considered and the rule with the maximum probability estimate is used for classifying the example. If no rule covers the example, it classified by the default rule predicting the majority class.

We used JRip, the Weka [16] implementation of Ripper. Contrary to William Cohen's original implementation, this re-implementation does not support the unordered mode, so we had to add a re-implementation of that mode.[1] We also added a few other minor modifications which were needed for the probability estimation, e.g. the collection of statistical counts of the sub rules.

In addition, Ripper (and JRip) can turn the incremental reduced error pruning technique [10, 9] on and off. Note, however, that with turned off pruning, Ripper still performs pre-pruning using a minimum description length heuristic [5]. We use Ripper with and without pruning and in ordered and unordered mode to generate four set of rules. For each rule set, we employ several different class probability estimation techniques.

In the test phase, all covering rules are selected for a given test instance. Using this reduced rule set we determine the most probable rule. For this purpose we select the most probable class of each rule and use this class value as the prediction for the given test instance and the class probability for comparison. Ties are solved by predicting the least represented class. If no covering rules exist the class probability distribution of the default rule is used.

[1] Weka supports a general one-against-all procedure that can also be combined with JRip, but we could not use this because it did not allow us to directly access the rule probabilities.

4 Experimental Results

4.1 Experimental Setup

We performed our experiments within the WEKA framework [16]. We tried each of the four configuration of Ripper (unordered/ordered and pruning/no pruning) with 5 different probability estimation techniques, Naïve (labeled as Precision), Laplace, and m-estimate with $m \in \{2, 5, 10\}$, both used as a stand-alone probability estimate (abbreviated with B) or in combination with shrinkage (abbreviated with S). As a baseline, we also included the performance of pruned or unpruned standard JRip accordingly. Our unordered implementation of JRip using Laplace stand-alone for the probability estimation is comparable to the unordered version of Ripper (Cohen, 1995), which is not implemented in JRip.

We evaluated these methods on 33 data sets of the UCI repository [1] which differ in the number of attributes (and their categories), classes and training instances. As a performance measure, we used the weighted area under the ROC curve (AUC), as used for probabilistic decision trees in [13]. Its key idea is to extend the binary AUC to the multi-class case by computing a weighted average the AUCs of the one-against-all problems N_c, where each class c is paired with all other classes:

$$\mathrm{AUC}(N) = \sum_{c \in C} \frac{n_c}{|N|} \mathrm{AUC}(N_c) \tag{4.1}$$

For the evaluation of the results we used the Friedman test with a post-hoc Nemenyi test as proposed in [6]. The significance level was set to 5% for both tests. We only discuss summarized results here, detailed result tables can be found in the Appendix.

4.2 Ordered Rule Sets

In the first two test series, we investigated the ordered approach using the standard JRip approach for the rule generation, both with and without pruning. The basic probability methods were used standalone (B) or in combination with shrinkage (S).

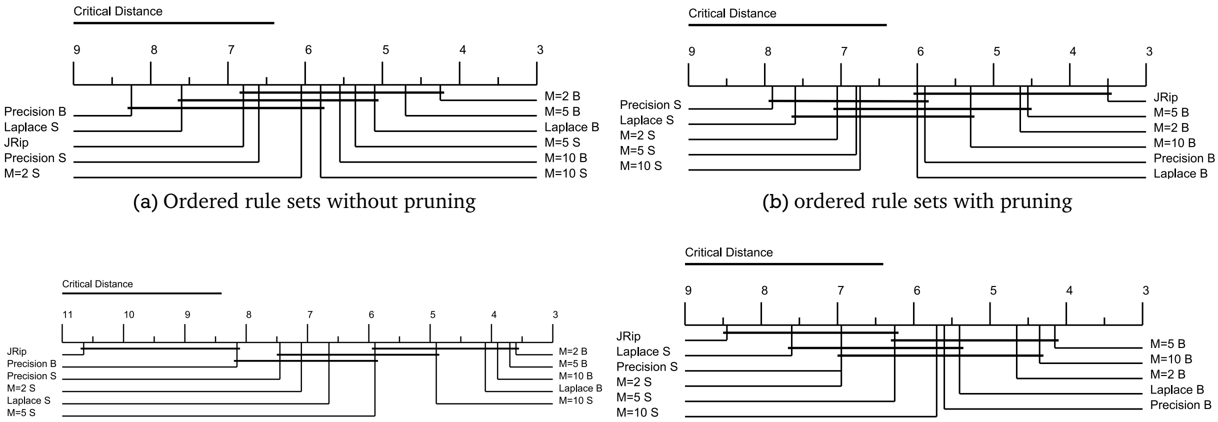
The Friedman test showed that in both test series, the employed combinations of probability estimation techniques showed significant differences. Considering the CD chart of the first test series (Figure 4.1(a)), one can identify three groups of equivalent techniques. Notable is that the two best techniques, the m-Estimate used stand-alone with $m = 2$ and $m = 5$ respectively, belong only to the best group. These two are the only methods that are significantly better than the two worst methods, Precision used stand-alone and Laplace combined with shrinkage. On the other hand, the naïve approach seems to be a bad choice as both techniques employing it rank in the lower half. However our benchmark JRip is positioned in the lower third, which means that the probability estimation techniques clearly improve over the default decision list approach implemented in JRip.

Comparing the stand-alone techniques with those employing shrinkage one can see that shrinkage is outperformed by their stand-alone counterparts. Only Precision is an exception as shrinkage yields increased performance in this case. In the end shrinkage is not a good choice for this scenario.

The CD-chart for ordered rule sets with pruning (Figure 4.1(b)) features four groups of equivalent techniques. Notable are the best and the worst group which overlap only in two techniques, Laplace and Precision used stand-alone. The first group consists of all stand-alone methods and JRip which dominates the group strongly covering no shrinkage method. The last group consists of all shrinkage methods and the overlapping methods Laplace and Precision used stand-alone. As all stand-alone methods rank before the shrinkage methods, one can conclude that they outperform the shrinkage methods in this scenario as well. Ripper performs best in this scenario, but the difference to the stand-alone methods is not significant.

4.3 Unordered Rule Sets

Test series three and four used the unordered approach employing the modified JRip which generates rules for each class. Analogous to the previous test series the basic methods are used as stand-alone methods or in combination with shrinkage (left and right column respectively). Test series three used no pruning while test series four did so. The results of the Friedman test showed that the techniques of test series three and test series four differ significantly.



(a) Ordered rule sets without pruning

(b) ordered rule sets with pruning

(c) Unordered rule sets without pruning

(d) Unordered rule sets with pruning

Figure 4.1: Critical distance charts of ordered rule sets ((a) and (b)) and unordered rule sets ((c) and (d))

Regarding the CD chart of test series three (Figure 4.1(c)), we can identify four groups of equivalent methods. The first group consists of all stand-alone techniques, except for Precision, and the m-estimates techniques combined with shrinkage and $m = 5$ and $m = 10$, respectively. Whereas the stand-alone methods dominate this group, $m = 2$ being the best representative. Apparently these methods are the best choices for this scenario. The second and third consist mostly of techniques employing shrinkage and overlap with the worst group in only one technique. However our benchmark JRip belongs to the worst group being the worst choice of this scenario. Additionally the shrinkage methods are outperformed by their stand-alone counterparts.

The CD chart of test series four (Figure 4.1(d)) shows similar results. Again four groups of equivalent techniques groups can be identified. The first group consists of all stand-alone methods and the m-estimates using shrinkage and $m = 5$ and $m = 10$ respectively. This group is dominated by the m-estimates used stand-alone with $m = 2$, $m = 5$ or $m = 10$. The shrinkage methods are distributed over the other groups, again occupying the lower half of the ranking. Our benchmark JRip is the worst method of this scenario.

4.4 Unpruned vs. Pruned Rule Sets

Rule pruning had mixed results, which are briefly summarized in Table 4.1. On the one hand, it improved the results of the ordered approach, on the other hand it worsened the results of the unordered approach. In any case, in our experiments, contrary to previous results on PETs, rule pruning was not always a bad choice. The explanation for this result is that in rule learning, contrary to decision tree learning, new examples are not necessarily covered by one of the learned rules. The more specific rules become, the higher is the chance that new examples are not covered by any of the rules and have to be classified with a default rule. As these examples will all get the same default probability, this is a bad strategy for probability estimation. Note, however, that JRip without pruning, as used in our experiments, still performs an MDL-based form of pre-pruning. We have not yet tested a rule learner that performs no pruning at all, but, because of the above deliberations, we do not expect that this would change the results with respect to pruning.

Table 4.1: Unpruned vs. pruned rule sets: Win/Loss for ordered (top) and unordered (bottom) rule sets

	JRip	Precision		Laplace		M 2		M 5		M 10	
Win	26	23	19	20	19	18	20	19	20	19	20
Loss	7	10	14	13	14	15	13	14	13	14	13
Win	26	21	9	8	8	8	8	8	8	8	6
Loss	7	12	24	25	25	25	25	25	25	25	27

5 Probability Estimates from Rule Ensembles

Instead of trying to improve the probability estimates for each individual leaf, one can also resort to averaging multiple estimates, thereby reducing the variance of the resulting probability estimates. For example, a technique based on bagging multiple unpruned decision trees was used in [7] to obtain improved probability estimates, which were subsequently used for cost-sensitive classification. An adaptation of this technique to rule learning again has the effect that an example may be covered by a varying number of rules, whereas in decision tree learning, each example will be covered by exactly s leafs (where s is the number of trees learned).

For investigating the performance ensemble-based probability estimates, we combined our rule learner with bagging [2]. We generated s bootstrap samples $S_1, \cdots, S_s$ by repeately drawing n instances with replacement. The rule algorithm described in Section 3 was applied to each sampled data set s_i obtaining s classifiers $C_1, \cdots, C_s$ and their corresponding rule sets $R_1, \cdots, R_s$. The probability estimation for each is computed using the previously introduced basic estimation methods.

For prediction, we determine the covering rules of each sampled rule set R_i for a given example x

$$R_i(x) = \{r \in R_i | r \supseteq x\} \tag{5.1}$$

Let $Cov_i(x)$ denote the set of all class probability distributions that originate from a rule in the set of covering rules

$$Cov_i(x) = \left\{\vec{\Pr}_r(x) | r \in R_i(x)\right\}. \tag{5.2}$$

From this set of class probability distributions of the covering rules, we try to estimate a class probability distribution for the given example x. For this purpose we have to decode the probability estimations of these covering rules $\vec{\Pr}_r(x)$ into a single normalized global class probability distribution $\vec{\Pr}_{global}(x)$.

For our approach we considered four decoding methods. The first three methods have in common that they average the class probability distribution of (some of) the covering rules.

Best Rule: Only the most confident covering rule $\vec{\Pr}_i(x)$ of each covering rule set $R_i(x)$ is determined.

$$\vec{\Pr}_i(x) = \underset{\vec{\Pr}(x) \in Cov_i}{\arg\max} \left(\vec{\Pr}(x)\right) \tag{5.3}$$

Afterwards the class probability distributions of these rules are averaged and the result is normalized:

$$\vec{\Pr}_{global}(x) = \frac{\text{avg}\left(\vec{\Pr}_1, \cdots, \vec{\Pr}_s\right)}{\left\|\text{avg}\left(\vec{\Pr}_1, \cdots, \vec{\Pr}_s\right)\right\|_1}. \tag{5.4}$$

Macro Averaging: All covering rules are determined and their class probability distributions are macro-averaged in two steps. First the class probability distributions $Cov_i(x)$ of each covering rule set $R_i(x)$ are averaged and normalized:

$$\vec{\Pr}_i = \frac{\text{avg}\left(Cov_i(x)\right)}{\left\|\text{avg}\left(Cov_i(x)\right)\right\|_1}. \tag{5.5}$$

These local class probabilities are then averaged as above (5.4).

Micro Averaging: All covering rules are determined and their class probability distributions are micro-averaged. Essentially, this means that all learned rules are pooled, and the average is formed over the resulting set set of rules:

$$\vec{\Pr}_{global}(x) = \frac{\text{avg}\left(Cov_1(x) \cup \cdots \cup Cov_s(x)\right)}{\left\|\text{avg}\left(Cov_1(x) \cup \cdots \cup Cov_s(x)\right)\right\|_1} \tag{5.6}$$

Bayesian Decoding: All covering rules are pooled as above, but their class probability distributions are multiplied with each other and with the vector of the a priori class probabilities

$$\vec{\Pr}_{prior} = \left(\Pr(c_1), \cdots, \Pr(c_{|C|}\right).$$

Thus $\vec{\Pr}_{global}(x)$ is calculated as follows

$$\vec{\Pr}_{global}(x) = \frac{\text{mult}\left(\bigcup_{i=1}^{s} Cov_i(x) \cup \left\{\vec{\Pr}_{prior}\right\}\right)}{\left\|\text{mult}\left(\bigcup_{i=1}^{s} Cov_i(x) \cup \left\{\vec{\Pr}_{prior}\right\}\right)\right\|_1} \tag{5.7}$$

In all cases, the resulting class probability distribution $\vec{\Pr}_{global}(x)$ is used for the prediction. For this purpose the most probable class according to $\vec{\Pr}_{global}(x)$ is selected and this class value is used as the prediction for the given test instance x. Ties are solved by predicting the least represented class. If no covering rules ($Cov_i(x) = \emptyset$) exist for a sampled rule set R_i the class probability distribution of the default rule is used accordingly.

6 Results on Ensemble-Based Probability Estimates

6.1 Experimental Setup

The above methods were integrated into the framework described in Section 4.1. For exampling, we employed the unsupervised random sampling of WEKA for the generation of the bootstrap samples and applied the Bagging implementation of WEKA to JRip. In accordance with the results obtained above, we only used unordered rule sets for these experiments because these produce better probability estimates than ordered rule sets. Furthermore we know that the unpruned rule sets perform better for the Laplace- and m-estimates than pruned rule sets if the rule sets are generated by the unordered JRip. For Precision the opposite is true. So we employed these basic probability estimation techniques on either unpruned or pruned rule sets according to these observations. As the employed shrinkage method worsened the probability estimation we abstained from using shrinkage in these experiments. All previously mentioned decoding methods - Best rule, Macro and Micro Averaging, and Bayesian Decoding - were employed. So we computed all combinations of the basic probability estimation and decoding methods on a different number of bootstrap samples - 10, 20, 50 and 100 samples accordingly. These methods were compared to the default configuration of JRip (pruned, ordered rule sets) and to its bagged version (with or without pruning) using the same bootstrap samples.

All methods were evaluated on the 33 previously used data sets of the UCI repository. As a performance measure, we used the weighted area under the ROC curve (AUC) also. For our comparison we calculated the average weighted AUC over all data sets for all combinations (combining a probability estimation technique and a decoding method to a given number of boostrap samples), the result is summarized in Table 6.1 and depicted in Figure 6.1.

6.2 Comparison to JRip

In our experiments we compared JRip and its bagged versions to the basic probability estimation techniques employing the four decoding methods. Figure 6.1 shows the results of unpruned bagged JRip and the basic probability estimation techniques. We omitted to depict the results of JRip and the pruned bagged JRip because they both had a worse performance than the unpruned bagged version which has the worst performance of the depicted methods. The weighted AUC of their best representative, unpruned bagged JRip, was always at least two absolute percentage points lower than the weighted AUC of the basic probability estimation techniques for all decoding methods and all numbers of samples. So we can conclude that the performance of the basic probability estimation techniques improve over the default probability estimation integrated in JRip. As this observation was also made in the basic rule learning experiments, we see our approach reconfirmed.

6.3 Comparison of the base probability estimation techniques

Each of the results of Figure 6.1 is the average performance over several different base probability estimation techniques: Precision applied to pruned rule sets and the Laplace and the m-estimate ($m \in \{2, 5, 10\}$) applied to unpruned rule sets. The applied decoding methods seem to have only a small impact on the ranking (according the weighted AUC) of the basic probability estimation techniques. Especially for a higher number of bootstrap samples, 50 and 100, these rankings are always the same - Precision having the best performance followed by the m-estimate using $m = 2$, $m = 5$, or $m = 10$ and the Laplace estimate in this order. For the smaller numbers of samples, 10 and 20, the ranking is a little bit more dynamic. Although the best two methods, m-estimate with $m = 2$ and $m = 5$ (in this order), and the worst method, Laplace estimate, are always the same, the methods in the center, Precision and the m-estimate using $m = 10$, switch places dependent on the decoding methods. So we can conclude that the m-estimate using $m = 2$ is the best choice for a low number of samples just as Precision is the best choice for a higher number of samples. For all probability estimation techniques holds that an increase in the number of samples leads to an improvement in the weighted AUC but the gain is a bit lower than the gain of the bagged JRip.

6.4 Comparison of the decoding methods

In this section, we want to compare the four employed decoding methods: Best Rule, Macro and Micro Averaging, and Bayesian Decoding. For this purpose we calculated the average weighted AUC over all data sets for all combinations obtained by combining a probability estimation technique, a decoding method and a number of bootstrap samples (Table 6.1). Afterwards, we determined on this data the average rank of each decoding method (Table 6.2) and used this information for a Friedman test with a significance of 5%. According to this test, the decoding methods differ significantly, so we applied a post-hoc Nemenyi test which is depicted in a CD chart (Figure 6.2).

Obviously, Micro Averaging is the best decoding methods in our experiments as it always placed first according to the average weighted AUC. This observation is reconfirmed by the CD chart since the Micro Averaging is the only member of the best group of methods. Macro Averaging and Bayesian Decoding do not differ significantly being both in the second best group of decoding methods. Nevertheless Bayesian Decoding is also in the worst group together with Best Rule which is the worst choice in our experiments.

The observed ranking of the decoding methods can be attributed to the individual exploitation of the covering rules. As the decoding method Best Rule only uses the best rule of each bootstrap sample, a great deal of evidence has no influence on the probability estimation. Thus, it is not surprising that Best Rule ranks behind the methods that make better use of the ensemble.

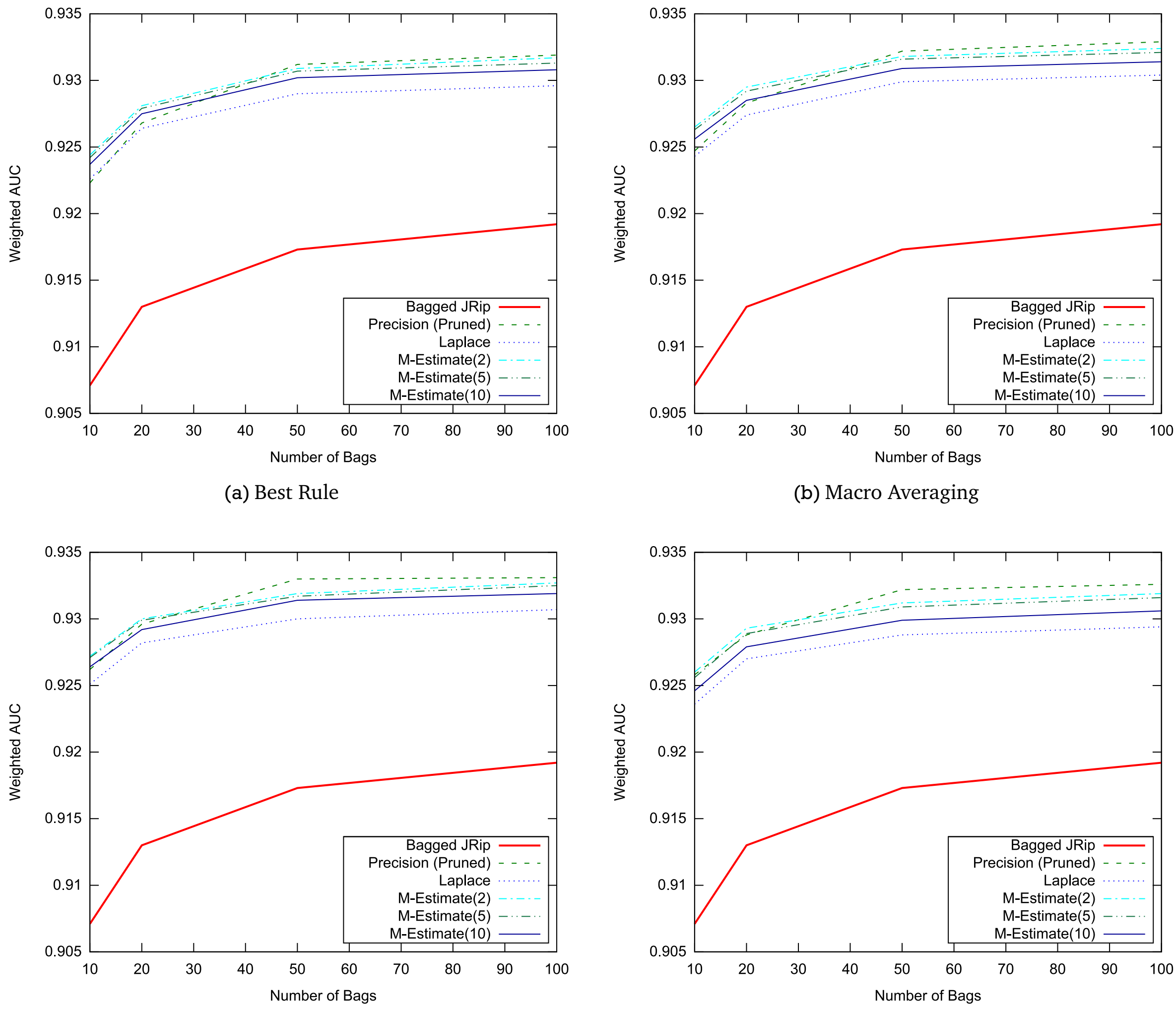
The methods Macro and Micro Averaging both average the probability distributions of a number of covering rules but their averaging approaches differ. So the influence of a high evidence for a class in a sampled rule set is also different for the two methods. For Micro Averaging the aforementioned evidence has a direct effect on resulting probability distribution as all covering rules have the same weight in its calculation. As our rule sets only have a low redundancy, this effect is desirable. For Macro Averaging a high evidence in a sampled rule set influences only the probability distribution of this bootstrap sample. So the number of covering rules has no effect on the global probability distribution. As Macro Averaging partially discards the available information Micro Averaging should perform better than Macro Averaging as observed in our experiments.

Bayesian Decoding uses all the information contained in the covering rules as Micro Averaging does. These two methods differ only the way how they combine the information of the covering rule, averaging or multiplying their probability distributions. The multiplication used in the Bayesian approach has a tendency to prefer a number of medium probabilities to a balanced number of low and high probabilities. This bias has a negative effect on the calculation of the global probability distribution. Averaging is more desirable as high probabilities have a greater impact on its calculation.

Table 6.1: Average Weighted AUC of the bagged JRip and the employed probability estimation techniques (applied decoding methods: Best Rule (BR), Macro (Mac) and Micro Averaging (Mic), and Bayesian Decoding (BD))

	10 Samples				20 Samples			
	BR	Mac	Mic	BD	BR	Mac	Mic	BD
Bagged JRip	0.9071				0.9130			
Precison (pruned)	0.9223	0.9247	0.9262	0.9258	0.9268	0.9283	0.9296	0.9288
Laplace	0.9226	0.9243	0.9251	0.9236	0.9264	0.9274	0.9282	0.9270
M 2	0.9244	0.9265	0.9272	0.9260	0.9281	0.9295	0.9300	0.9293
M 5	0.9242	0.9263	0.9271	0.9256	0.9279	0.9292	0.9299	0.9289
M 10	0.9237	0.9256	0.9264	0.9246	0.9275	0.9285	0.9292	0.9279
Average	0.9234	0.9255	0.9264	0.9251	0.9273	0.9286	0.9294	0.9284

	50 Samples				100 Samples			
	BR	Mac	Mic	BD	BR	Mac	Mic	BD
Bagged JRip	0.9173				0.9192			
Precison (pruned)	0.9312	0.9322	0.9330	0.9322	0.9319	0.9329	0.9331	0.9326
Laplace	0.9290	0.9299	0.9300	0.9288	0.9296	0.9304	0.9307	0.9294
M 2	0.9309	0.9318	0.9319	0.9312	0.9317	0.9324	0.9327	0.9319
M 5	0.9307	0.9316	0.9317	0.9309	0.9313	0.9321	0.9325	0.9316
M 10	0.9302	0.9309	0.9314	0.9299	0.9308	0.9314	0.9319	0.9306
Average	0.9304	0.9313	0.9316	0.9306	0.9311	0.9318	0.9322	0.9312



(a) Best Rule

(b) Macro Averaging

(c) Micro Averaging

(d) Bayesian Decoding

Figure 6.1: Average Weighted AUC of the employed decoding methods on unpruned rule sets (where stated the pruned rule set is used)

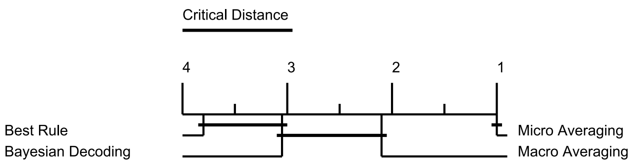


Figure 6.2: CD chart of the decoding methods

Table 6.2: Decoding Methods: Count of each rank and the average rank

Decoding Method	Ranked 1st	2nd	3rd	4th	Average Rank
Best Rule	0	0	4	16	3.80
Macro Averaging	0	17	3	0	2.15
Micro Averaging	20	0	0	0	1.00
Bayesian Decoding	0	3	13	4	3.05

7 Conclusions

The most important result of our study is that probability estimation is clearly an important part of a good rule learning algorithm. The probabilities of rules induced by JRip can be improved considerably by simple estimation techniques. In unordered mode, where one rule set is generated for each class, JRip is outperformed in every scenario. On the other hand, in the ordered setting, which essentially learns decision lists by learning subsequent rules in the context of previous rules, the results were less convincing, giving a clear indication that the unordered rule induction mode should be preferred when a probabilistic classfication is desirable.

Among the tested probability estimation techniques, the m-estimate typically outperformed the other methods in our version of the JRip algorithm. Among the tested values, $m = 5$ seemed to yield the best overall results, but the superiority of the m-estimate was not sensitive to the choice of this parameter. When combined with an ensemble-based approach, the m-estimate maintained its superiority for smaller number of bootstrap samples, but typically lower value of m performed better. For higher numbers of bootstrap samples, precision, which corresponds to a value of $m = 0$, outperformed the other methods. Thus, it seems to be the case that the use of the m-estimate primarily helps to reduce the variance of the probability estimates.

The employed shrinkage method did, in general, not improve the simple estimation techniques. It remains to be seen whether alternative ways of setting the weights could yield superior results. Rule pruning did not produce the bad results that are known from ranking with pruned decision trees, presumably because unpruned, overly specific rules will increase the number of uncovered examples, which in turn leads to bad ranking of these examples.

Ensemble-based probability estimation based on a bagging approach further improved the probability estimates. The improvement increases with the number of bootstrap samples. In every case, the probabilistic approach outperformed Bagged JRip using the same number of bootstrap samples. Amongst the employed decoding methods, Micro Averaging of the probability distributions of all covering rules was without exceptions superior to the other methods, indicating that the predictions of rule-based ensembles should be combined at the level of individual rules and not at the level of theories.

Acknowledgements

This research was supported by the *German Science Foundation (DFG)* under grant FU 580/2.

Bibliography

[1] A. Asuncion and D. J. Newman. UCI machine learning repository, 2007. 7

[2] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996. 9

[3] B. Cestnik. Estimating probabilities: A crucial task in Machine Learning. In L. Aiello, editor, *Proceedings of the 9th European Conference on Artificial Intelligence (ECAI-90)*, pages 147–150, Stockholm, Sweden, 1990. Pitman. 2, 4

[4] S. F. Chen and J. T. Goodman. An empirical study of smoothing techniques for language modeling. Technical Report TR-10-98, Computer Science Group, Harvard University, Cambridge, MA, 1998. 4

[5] W. W. Cohen. Fast effective rule induction. In A. Prieditis and S. Russell, editors, *Proceedings of the 12th International Conference on Machine Learning (ML-95)*, pages 115–123, Lake Tahoe, CA, 1995. Morgan Kaufmann. 6

[6] J. Demsar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006. 7

[7] P. Domingos. Metacost: A general method for making classifiers cost-sensitive. In *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-99)*, pages 155–164, San Diego, CA, 1999. ACM. 9

[8] C. Ferri, P. A. Flach, and J. Hernández-Orallo. Improving the AUC of probabilistic estimation trees. In N. Lavrac, D. Gamberger, L. Todorovski, and H. Blockeel, editors, *Proceedings of the 14th European Conference on Machine Learning*, pages 121–132, Cavtat-Dubrovnik, Croatia, 2003. Springer. 2

[9] J. Fürnkranz. Pruning algorithms for rule learning. *Machine Learning*, 27(2):139–171, 1997. 6

[10] J. Fürnkranz and G. Widmer. Incremental Reduced Error Pruning. In W. W. Cohen and H. Hirsh, editors, *Proceedings of the 11th International Conference on Machine Learning (ML-94)*, pages 70–77, New Brunswick, NJ, 1994. Morgan Kaufmann. 6

[11] E. Hüllermeier and S. Vanderlooy. Why fuzzy decision trees are good rankers. *IEEE Transactions on Fuzzy Systems*, 17(6):1233–1244, 2009. 2

[12] C. D. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, MA, 1999. 4

[13] F. J. Provost and P. Domingos. Tree induction for probability-based ranking. *Machine Learning*, 52(3):199–215, 2003. 2, 7

[14] J.-N. Sulzmann and J. Fürnkranz. An empirical comparison of probability estimation techniques for probabilistic rules. In J. Gama, V. Santos Costa, A. Jorge, and P. B. Brazdil, editors, *Proceedings of the 12th International Conference on Discovery Science (DS-09)*, pages 317–331. Springer-Verlag, 2009. Winner of Best Student Paper Award. 2

[15] B. Wang and H. Zhang. Improving the ranking performance of decision trees. In J. Fürnkranz, T. Scheffer, and M. Spiliopoulou, editors, *Proceedings of the 17th European Conference on Machine Learning (ECML-06)*, pages 461–472, Berlin, Germany, 2006. Springer-Verlag. 2, 4

[16] I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco, CA, 2nd edition, 2005. 6, 7

Table 7.1: Weighted AUC results with rules from ordered, unpruned JRip.

Name	Jrip	Precision		Laplace		M 2		M 5		M 10	
		B	S	B	S	B	S	B	S	B	S
Anneal	.983	.970	.971	.970	.970	.970	.970	.971	.971	.970	.970
Anneal.orig	.921	.917	.920	.920	.919	.920	.919	.920	.920	.919	.919
Audiology	.863	.845	.843	.832	.836	.840	.844	.839	.841	.831	.832
Autos	.904	.907	.901	.900	.891	.907	.902	.904	.902	.903	.898
Balance-scale	.823	.801	.812	.821	.812	.820	.811	.821	.812	.821	.815
Breast-cancer	.591	.577	.581	.578	.580	.578	.580	.578	.579	.577	.579
Breast-w	.928	.930	.929	.935	.930	.935	.931	.935	.932	.935	.933
Colic	.736	.739	.741	.747	.746	.748	.746	.748	.745	.748	.746
Credit-a	.842	.849	.857	.861	.859	.861	.859	.861	.862	.861	.864
Credit-g	.585	.587	.587	.587	.587	.587	.587	.587	.587	.587	.587
Diabetes	.642	.654	.656	.655	.656	.655	.656	.655	.656	.655	.655
Glass	.806	.803	.795	.790	.787	.794	.797	.793	.799	.792	.795
Heart-c	.762	.765	.775	.796	.777	.796	.777	.796	.780	.796	.789
Heart-h	.728	.737	.755	.758	.757	.758	.755	.758	.757	.758	.757
Heart-statlog	.763	.759	.781	.806	.782	.806	.783	.806	.790	.806	.791
Hepatitis	.679	.661	.661	.660	.663	.660	.665	.660	.663	.660	.663
Hypothyroid	.971	.973	.974	.974	.974	.974	.974	.974	.974	.973	.974
Ionosphere	.884	.885	.897	.903	.900	.903	.899	.903	.900	.903	.902
Iris	.957	.889	.876	.889	.878	.889	.878	.889	.878	.889	.878
Kr-vs-kp	.993	.994	.994	.995	.994	.995	.994	.995	.994	.995	.994
Labor	.812	.800	.810	.794	.810	.793	.806	.793	.795	.793	.783
Lymph	.750	.739	.748	.748	.745	.746	.748	.744	.746	.749	.746
Primary-tumor	.649	.636	.652	.615	.638	.645	.656	.641	.653	.642	.662
Segment	.983	.964	.944	.967	.943	.966	.944	.967	.943	.966	.943
Sick	.922	.928	.929	.929	.929	.929	.929	.929	.929	.929	.929
Sonar	.774	.771	.779	.784	.778	.783	.778	.783	.779	.783	.781
Soybean	.962	.971	.972	.966	.971	.973	.972	.967	.973	.967	.971
Splice	.938	.934	.938	.943	.938	.943	.938	.943	.938	.943	.939
Vehicle	.772	.799	.811	.811	.816	.812	.813	.811	.816	.812	.819
Vote	.952	.954	.950	.955	.949	.955	.949	.955	.952	.953	.956
Vowel	.884	.906	.909	.909	.906	.909	.910	.911	.910	.910	.907
Waveform	.847	.850	.853	.872	.854	.872	.854	.873	.855	.873	.858
Zoo	.916	.899	.916	.902	.897	.908	.900	.907	.895	.899	.890
Average	.834	.830	.834	.836	.832	.837	.834	.837	.834	.836	.834
Average Rank	6.79	8.24	6.62	5.11	7.62	4.26	6.03	4.68	5.33	5.53	5.79

Table 7.2: Weighted AUC results with rules from ordered, pruned JRip.

Name	Jrip	Precision		Laplace		M 2		M 5		M 10	
		B	S	B	S	B	S	B	S	B	S
Anneal	.984	.981	.980	.981	.981	.981	.980	.981	.980	.980	.980
Anneal.orig	.942	.938	.937	.936	.936	.937	.936	.936	.937	.935	.936
Audiology	.907	.865	.854	.810	.776	.852	.840	.839	.826	.834	.801
Autos	.850	.833	.836	.821	.829	.829	.830	.823	.830	.821	.819
Balance-scale	.852	.812	.810	.815	.810	.815	.810	.816	.811	.816	.811
Breast-cancer	.598	.596	.597	.596	.597	.596	.597	.598	.599	.598	.602
Breast-w	.973	.965	.956	.965	.956	.964	.956	.964	.957	.961	.957
Colic	.823	.801	.808	.804	.815	.809	.815	.813	.815	.816	.816
Credit-a	.874	.872	.874	.873	.874	.874	.874	.874	.873	.875	.874
Credit-g	.593	.613	.612	.613	.612	.613	.612	.613	.612	.613	.612
Diabetes	.739	.734	.736	.734	.736	.734	.736	.734	.736	.734	.736
Glass	.803	.814	.810	.822	.825	.820	.818	.820	.817	.820	.812
Heart-c	.831	.837	.818	.843	.818	.842	.818	.845	.823	.847	.825
Heart-h	.758	.739	.742	.740	.740	.740	.742	.741	.742	.742	.741
Heart-statlog	.781	.792	.776	.790	.776	.790	.776	.791	.775	.790	.773
Hepatitis	.664	.600	.596	.600	.596	.599	.596	.599	.595	.597	.586
Hypothyroid	.988	.990	.990	.990	.990	.990	.990	.990	.990	.990	.990
Ionosphere	.900	.904	.909	.907	.909	.908	.909	.910	.910	.910	.909
Iris	.974	.888	.889	.890	.891	.890	.891	.890	.891	.890	.891
Kr-vs-kp	.995	.994	.993	.994	.993	.994	.993	.994	.994	.994	.994
Labor	.779	.782	.755	.782	.761	.781	.764	.768	.759	.746	.745
Lymph	.795	.795	.767	.788	.772	.790	.773	.779	.773	.777	.774
Primary-tumor	.642	.626	.624	.622	.627	.630	.622	.627	.622	.629	.628
Segment	.988	.953	.932	.953	.933	.954	.932	.953	.932	.953	.933
Sick	.948	.949	.949	.950	.949	.950	.949	.950	.950	.950	.950
Sonar	.759	.740	.734	.742	.737	.743	.737	.746	.740	.744	.744
Soybean	.981	.980	.970	.968	.965	.978	.970	.971	.967	.969	.966
Splice	.967	.956	.953	.957	.953	.957	.953	.957	.954	.957	.954
Vehicle	.855	.843	.839	.844	.843	.844	.842	.843	.843	.842	.844
Vote	.942	.949	.947	.949	.947	.949	.947	.949	.947	.949	.947
Vowel	.910	.900	.891	.898	.891	.904	.892	.905	.893	.898	.892
Waveform	.887	.880	.862	.880	.863	.880	.862	.881	.863	.881	.863
Zoo	.925	.889	.909	.887	.895	.895	.902	.895	.901	.889	.893
Average	.855	.843	.838	.841	.836	.843	.838	.842	.838	.841	.836
Average Rank	3.52	5.88	7.92	5.98	7.62	4.65	7.06	4.55	6.79	5.29	6.74

Table 7.3: Weighted AUC results with rules from unordered, unpruned JRip.

Name	Jrip	Precision		Laplace		M 2		M 5		M 10	
		B	S	B	S	B	S	B	S	B	S
Anneal	.983	.992	.989	.992	.991	.994	.989	.994	.989	.994	.989
Anneal.orig	.921	.987	.984	.990	.983	.993	.984	.993	.984	.993	.984
Audiology	.863	.910	.887	.877	.874	.909	.895	.903	.894	.892	.889
Autos	.904	.916	.915	.926	.914	.927	.914	.929	.918	.930	.926
Balance-scale	.823	.874	.865	.908	.873	.908	.866	.909	.871	.908	.882
Breast-cancer	.591	.608	.587	.633	.605	.633	.589	.632	.606	.632	.617
Breast-w	.928	.959	.966	.953	.966	.953	.967	.953	.969	.953	.969
Colic	.736	.835	.840	.855	.851	.855	.849	.855	.849	.859	.849
Credit-a	.842	.890	.909	.913	.911	.913	.911	.913	.914	.913	.917
Credit-g	.585	.695	.717	.716	.716	.716	.716	.716	.716	.716	.718
Diabetes	.642	.760	.778	.783	.780	.783	.779	.783	.781	.783	.783
Glass	.806	.810	.826	.808	.833	.808	.825	.808	.827	.809	.830
Heart-c	.762	.790	.813	.861	.827	.861	.823	.861	.831	.861	.844
Heart-h	.728	.789	.803	.851	.839	.853	.819	.849	.835	.852	.837
Heart-statlog	.763	.788	.811	.845	.805	.841	.805	.841	.820	.841	.829
Hepatitis	.679	.774	.817	.799	.819	.802	.821	.802	.817	.802	.816
Hypothyroid	.971	.991	.994	.994	.993	.994	.994	.994	.993	.994	.993
Ionosphere	.884	.918	.932	.938	.931	.938	.931	.938	.931	.939	.935
Iris	.957	.968	.973	.978	.980	.978	.976	.978	.980	.978	.980
Kr-vs-kp	.993	.998	.997	.999	.997	.999	.997	.999	.997	.999	.997
Labor	.812	.818	.806	.777	.803	.778	.803	.778	.790	.778	.775
Lymph	.750	.843	.852	.891	.857	.887	.848	.881	.852	.884	.878
Primary-tumor	.649	.682	.707	.671	.690	.693	.712	.694	.711	.691	.711
Segment	.983	.991	.989	.997	.990	.997	.989	.997	.990	.997	.990
Sick	.922	.958	.979	.981	.984	.982	.979	.982	.980	.982	.980
Sonar	.774	.823	.826	.841	.826	.841	.826	.841	.828	.841	.836
Soybean	.962	.979	.981	.982	.979	.985	.981	.984	.981	.985	.981
Splice	.938	.964	.968	.974	.968	.974	.968	.974	.969	.974	.970
Vehicle	.772	.851	.879	.888	.881	.888	.879	.888	.881	.888	.884
Vote	.952	.973	.967	.982	.968	.983	.968	.983	.975	.983	.978
Vowel	.884	.917	.919	.922	.920	.922	.921	.922	.920	.922	.920
Waveform	.847	.872	.890	.902	.890	.902	.890	.902	.890	.902	.893
Zoo	.916	.964	.965	.965	.970	.984	.982	.984	.982	.987	.988
Average	.834	.875	.883	.891	.885	.893	.885	.893	.887	.893	.890
Average Rank	10.67	8.15	7.45	4.08	6.65	3.58	7.08	3.68	5.88	3.88	4.91

Table 7.4: Weighted AUC results with rules from unordered, pruned JRip.

Name	Jrip	Precision		Laplace		M 2		M 5		M 10	
		B	S	B	S	B	S	B	S	B	S
Anneal	.984	.987	.988	.984	.986	.987	.985	.986	.986	.986	.986
Anneal.orig	.942	.990	.983	.985	.980	.989	.983	.988	.982	.984	.982
Audiology	.907	.912	.889	.891	.878	.895	.893	.889	.885	.883	.881
Autos	.850	.889	.882	.891	.889	.894	.888	.892	.889	.891	.889
Balance-scale	.852	.888	.861	.899	.864	.895	.860	.900	.861	.901	.864
Breast-cancer	.598	.562	.555	.557	.555	.557	.555	.557	.555	.560	.558
Breast-w	.973	.962	.972	.963	.973	.963	.973	.963	.973	.961	.974
Colic	.823	.782	.831	.799	.830	.793	.836	.801	.837	.812	.837
Credit-a	.874	.876	.878	.877	.877	.877	.878	.879	.879	.881	.879
Credit-g	.593	.702	.711	.703	.711	.703	.711	.703	.711	.705	.711
Diabetes	.739	.740	.729	.742	.729	.742	.729	.741	.730	.739	.731
Glass	.803	.819	.821	.821	.826	.819	.821	.824	.824	.828	.825
Heart-c	.831	.827	.816	.827	.804	.829	.816	.828	.810	.830	.807
Heart-h	.758	.739	.740	.735	.736	.737	.738	.736	.737	.735	.736
Heart-statlog	.781	.806	.815	.816	.813	.816	.812	.823	.819	.824	.827
Hepatitis	.664	.766	.790	.769	.793	.771	.790	.764	.795	.768	.789
Hypothyroid	.988	.984	.993	.992	.993	.987	.994	.992	.993	.992	.993
Ionosphere	.900	.918	.915	.921	.917	.922	.918	.926	.923	.926	.923
Iris	.974	.975	.969	.975	.969	.975	.969	.975	.970	.975	.973
Kr-vs-kp	.995	.999	.995	.999	.995	.999	.995	.999	.996	.998	.997
Labor	.779	.837	.820	.815	.811	.812	.818	.812	.812	.809	.803
Lymph	.795	.858	.832	.849	.833	.853	.836	.851	.842	.851	.856
Primary-tumor	.642	.703	.701	.679	.694	.709	.704	.710	.706	.708	.707
Segment	.988	.991	.989	.995	.990	.995	.990	.995	.990	.995	.990
Sick	.948	.949	.934	.948	.938	.948	.935	.948	.937	.948	.937
Sonar	.759	.827	.815	.827	.814	.827	.815	.824	.813	.824	.818
Soybean	.981	.989	.981	.988	.981	.990	.981	.989	.981	.989	.981
Splice	.967	.973	.967	.974	.967	.974	.967	.974	.968	.974	.968
Vehicle	.855	.892	.891	.893	.890	.893	.890	.893	.890	.893	.890
Vote	.942	.947	.956	.961	.957	.952	.957	.960	.956	.961	.958
Vowel	.910	.921	.915	.924	.915	.925	.915	.925	.916	.924	.915
Waveform	.887	.897	.877	.899	.878	.898	.877	.899	.878	.900	.880
Zoo	.925	.973	.989	.960	.969	.987	.989	.987	.989	.987	.989
Average	.855	.875	.873	.874	.871	.876	.873	.877	.874	.877	.874
Average Rank	8.45	5.61	6.95	5.38	7.59	4.67	6.95	4.14	6.23	4.33	5.7

Table 7.5: Average Weighted AUC of the bagged JRip and the employed probability estimation techniques: Best Rule, 10 samples, unpruned rule sets (except for Precision)

Name	Bagged Jrip	Precison	Laplace	M 2	M 5	M 10
Anneal	.9948	.9975	.9975	.9983	.9981	.9979
Anneal.orig	.9826	.9946	.9925	.9937	.9937	.9937
Audiology	.9310	.9487	.9295	.9283	.9243	.9210
Autos	.9529	.9367	.9469	.9486	.9471	.9449
Balance-Scale	.9291	.9548	.9546	.9547	.9567	.9590
Breast-cancer	.6339	.6284	.6662	.6676	.6643	.6609
Breast-w	.9872	.9876	.9859	.9852	.9850	.9850
Colic	.8613	.8920	.8849	.8872	.8881	.8876
Credit-a	.9224	.9295	.9300	.9300	.9305	.9308
Credit-g	.7307	.7643	.7726	.7695	.7708	.7733
Diabetes	.7812	.8052	.8062	.8049	.8067	.8085
Glass	.8912	.8963	.8884	.8879	.8876	.8873
Heart-c	.8913	.8909	.9088	.9068	.9072	.9073
Heart-h	.8463	.8840	.8745	.8744	.8750	.8766
Heart-statlog	.8646	.8922	.8959	.8956	.8960	.8961
Hepatitis	.7622	.8471	.8434	.8487	.8505	.8512
Hypothyroid	.9929	.9937	.9954	.9957	.9957	.9957
Ionosphere	.9635	.9534	.9702	.9702	.9702	.9706
Iris	.9811	.9930	.9931	.9929	.9933	.9937
Kr-vs-kp	.9979	.9997	.9998	.9998	.9998	.9998
Labor	.8838	.8932	.9216	.9189	.9149	.9054
Lymph	.8945	.9245	.9242	.9256	.9258	.9247
Primary-tumor	.7316	.8045	.7529	.7786	.7755	.7738
Segment	.9993	.9987	.9988	.9988	.9988	.9988
Sick	.9831	.9932	.9943	.9949	.9949	.9947
Sonar	.9055	.9008	.9098	.9084	.9083	.9075
Soybean	.9937	.9955	.9914	.9941	.9940	.9937
Vehicle	.9289	.9248	.9266	.9262	.9269	.9276
Vote	.9646	.9821	.9870	.9871	.9871	.9869
Vowel	.9887	.9905	.9886	.9903	.9895	.9887
Zoo	.9495	.9936	.9703	.9928	.9936	.9919
Average	.9071	.9223	.9226	.9244	.9242	.9237
Average Rank	4.9032	3.7419	3.5645	2.9839	2.7419	3.0645

Table 7.6: Average Weighted AUC of the bagged JRip and the employed probability estimation techniques: Macro Averaging, 10 samples, unpruned rule sets (except for Precision)

Name	Bagged Jrip	Precison	Laplace	M 2	M 5	M 10
Anneal	.9948	.9977	.9975	.9981	.9981	.9980
Anneal.orig	.9826	.9974	.9960	.9972	.9972	.9972
Audiology	.9310	.9561	.9328	.9367	.9337	.9300
Autos	.9529	.9395	.9510	.9514	.9498	.9486
Balance-Scale	.9291	.9582	.9558	.9562	.9587	.9605
Breast-cancer	.6339	.6469	.6604	.6611	.6618	.6617
Breast-w	.9872	.9864	.9870	.9869	.9871	.9870
Colic	.8613	.8928	.8911	.8916	.8920	.8923
Credit-a	.9224	.9310	.9326	.9323	.9329	.9335
Credit-g	.7307	.7690	.7729	.7727	.7742	.7757
Diabetes	.7812	.8089	.8088	.8084	.8102	.8126
Glass	.8912	.8991	.8956	.9003	.8971	.8936
Heart-c	.8913	.8983	.9060	.9057	.9064	.9066
Heart-h	.8463	.8829	.8764	.8762	.8785	.8794
Heart-statlog	.8646	.8934	.8999	.8999	.9006	.9001
Hepatitis	.7622	.8417	.8498	.8509	.8521	.8504
Hypothyroid	.9929	.9939	.9960	.9963	.9962	.9962
Ionosphere	.9635	.9552	.9721	.9722	.9723	.9720
Iris	.9811	.9930	.9929	.9927	.9931	.9933
Kr-vs-kp	.9979	.9996	.9997	.9997	.9997	.9997
Labor	.8838	.8959	.9081	.9068	.9000	.8932
Lymph	.8945	.9268	.9277	.9278	.9290	.9275
Primary-tumor	.7316	.8108	.7615	.7925	.7905	.7863
Segment	.9993	.9989	.9990	.9990	.9990	.9990
Sick	.9831	.9947	.9940	.9943	.9942	.9940
Sonar	.9055	.9057	.9172	.9168	.9154	.9140
Soybean	.9937	.9957	.9929	.9949	.9948	.9946
Vehicle	.9289	.9294	.9291	.9288	.9291	.9293
Vote	.9646	.9829	.9861	.9862	.9858	.9857
Vowel	.9887	.9911	.9907	.9922	.9915	.9908
Zoo	.9495	.9944	.9723	.9943	.9941	.9913
Average	.9071	.9247	.9243	.9265	.9263	.9256
Average Rank	5.3871	3.5484	3.8226	2.7903	2.3548	3.0968

Table 7.7: Average Weighted AUC of the bagged JRip and the employed probability estimation techniques: Micro Averaging, 10 samples, unpruned rule sets (except for Precision)

Name	Bagged Jrip	Precison	Laplace	M 2	M 5	M 10
Anneal	.9948	.9975	.9971	.9979	.9978	.9977
Anneal.orig	.9826	.9970	.9954	.9966	.9968	.9969
Audiology	.9310	.9563	.9369	.9361	.9324	.9292
Autos	.9529	.9387	.9481	.9491	.9471	.9450
Balance-Scale	.9291	.9596	.9565	.9572	.9598	.9618
Breast-cancer	.6339	.6602	.6601	.6607	.6614	.6598
Breast-w	.9872	.9867	.9876	.9876	.9875	.9874
Colic	.8613	.8959	.8933	.8938	.8940	.8937
Credit-a	.9224	.9306	.9329	.9326	.9335	.9341
Credit-g	.7307	.7691	.7765	.7759	.7780	.7795
Diabetes	.7812	.8129	.8121	.8115	.8136	.8151
Glass	.8912	.9013	.8989	.9036	.9014	.8976
Heart-c	.8913	.8987	.9053	.9050	.9054	.9054
Heart-h	.8463	.8825	.8758	.8760	.8778	.8779
Heart-statlog	.8646	.8960	.9023	.9022	.9029	.9032
Hepatitis	.7622	.8422	.8504	.8529	.8552	.8526
Hypothyroid	.9929	.9938	.9962	.9963	.9963	.9962
Ionosphere	.9635	.9552	.9718	.9719	.9719	.9717
Iris	.9811	.9927	.9922	.9921	.9925	.9925
Kr-vs-kp	.9979	.9996	.9997	.9997	.9997	.9997
Labor	.8838	.9095	.9135	.9135	.9068	.9054
Lymph	.8945	.9286	.9283	.9293	.9301	.9283
Primary-tumor	.7316	.8104	.7653	.7963	.7942	.7895
Segment	.9993	.9988	.9990	.9990	.9990	.9989
Sick	.9831	.9951	.9937	.9941	.9940	.9938
Sonar	.9055	.9095	.9174	.9172	.9159	.9145
Soybean	.9937	.9952	.9928	.9945	.9946	.9945
Vehicle	.9289	.9304	.9304	.9300	.9304	.9305
Vote	.9646	.9832	.9862	.9863	.9860	.9854
Vowel	.9887	.9905	.9899	.9914	.9907	.9900
Zoo	.9495	.9939	.9720	.9943	.9938	.9913
Average	.9071	.9262	.9251	.9272	.9271	.9264
Average Rank	5.5484	3.4194	3.7419	2.5968	2.4032	3.2903

Table 7.8: Average Weighted AUC of the bagged JRip and the employed probability estimation techniques: Bayesian Decoding, 10 samples, unpruned rule sets (except for Precision)

Name	Bagged Jrip	Precison	Laplace	M 2	M 5	M 10
Anneal	.9948	.9953	.9963	.9965	.9964	.9962
Anneal.orig	.9826	.9947	.9952	.9952	.9948	.9948
Audiology	.9310	.9454	.9221	.9260	.9246	.9202
Autos	.9529	.9368	.9470	.9473	.9460	.9448
Balance-Scale	.9291	.9633	.9584	.9591	.9616	.9634
Breast-cancer	.6339	.6501	.6589	.6605	.6613	.6599
Breast-w	.9872	.9909	.9918	.9918	.9918	.9917
Colic	.8613	.8935	.8931	.8932	.8933	.8930
Credit-a	.9224	.9338	.9333	.9332	.9335	.9336
Credit-g	.7307	.7736	.7797	.7793	.7804	.7814
Diabetes	.7812	.8180	.8142	.8137	.8151	.8166
Glass	.8912	.8984	.8942	.8990	.8952	.8905
Heart-c	.8913	.9026	.9043	.9031	.9045	.9053
Heart-h	.8463	.8846	.8734	.8730	.8749	.8769
Heart-statlog	.8646	.8983	.9023	.9020	.9030	.9031
Hepatitis	.7622	.8481	.8417	.8420	.8425	.8402
Hypothyroid	.9929	.9957	.9981	.9977	.9977	.9977
Ionosphere	.9635	.9570	.9737	.9736	.9733	.9722
Iris	.9811	.9923	.9915	.9913	.9909	.9908
Kr-vs-kp	.9979	.9996	.9997	.9997	.9999	.9998
Labor	.8838	.8959	.9122	.9081	.9027	.8973
Lymph	.8945	.9273	.9229	.9258	.9247	.9235
Primary-tumor	.7316	.8034	.7490	.7884	.7849	.7776
Segment	.9993	.9986	.9988	.9989	.9989	.9988
Sick	.9831	.9964	.9965	.9963	.9963	.9963
Sonar	.9055	.9057	.9161	.9155	.9153	.9133
Soybean	.9937	.9955	.9925	.9948	.9947	.9944
Vehicle	.9289	.9301	.9302	.9300	.9301	.9302
Vote	.9646	.9925	.9928	.9928	.9928	.9926
Vowel	.9887	.9907	.9899	.9915	.9909	.9900
Zoo	.9495	.9910	.9623	.9876	.9819	.9753
Average	.9071	.9258	.9236	.9260	.9256	.9246
Average Rank	5.4516	3.2903	3.2258	2.7419	2.8065	3.4839

Table 7.9: Average Weighted AUC of the bagged JRip and the employed probability estimation techniques: Best Rule, 20 samples, unpruned rule sets (except for Precision)

Name	Bagged Jrip	Precison	Laplace	M 2	M 5	M 10
Anneal	.9944	.9976	.9975	.9993	.9990	.9986
Anneal.orig	.9821	.9950	.9929	.9937	.9938	.9938
Audiology	.9415	.9490	.9381	.9362	.9326	.9295
Autos	.9573	.9418	.9488	.9513	.9496	.9466
Balance-Scale	.9344	.9581	.9572	.9576	.9607	.9630
Breast-cancer	.6213	.6596	.6699	.6637	.6652	.6664
Breast-w	.9888	.9888	.9889	.9890	.9890	.9890
Colic	.8664	.9045	.8888	.8886	.8892	.8874
Credit-a	.9265	.9329	.9338	.9338	.9341	.9341
Credit-g	.7438	.7766	.7760	.7730	.7749	.7768
Diabetes	.7935	.8138	.8185	.8166	.8172	.8186
Glass	.8975	.9101	.8943	.8963	.8949	.8946
Heart-c	.8952	.9027	.9091	.9095	.9101	.9096
Heart-h	.8536	.8846	.8758	.8758	.8764	.8773
Heart-statlog	.8769	.8942	.8992	.8996	.8998	.8988
Hepatitis	.7671	.8438	.8562	.8577	.8598	.8615
Hypothyroid	.9922	.9940	.9956	.9957	.9957	.9957
Ionosphere	.9621	.9543	.9713	.9714	.9711	.9704
Iris	.9910	.9937	.9931	.9931	.9932	.9933
Kr-vs-kp	.9982	.9997	.9998	.9998	.9998	.9998
Labor	.9135	.8973	.9243	.9257	.9257	.9243
Lymph	.9048	.9257	.9332	.9348	.9331	.9300
Primary-tumor	.7479	.8062	.7694	.7949	.7907	.7857
Segment	.9994	.9991	.9991	.9992	.9992	.9991
Sick	.9883	.9951	.9958	.9964	.9963	.9962
Sonar	.9223	.9172	.9136	.9127	.9112	.9102
Soybean	.9934	.9957	.9936	.9946	.9945	.9943
Vehicle	.9344	.9300	.9320	.9319	.9322	.9325
Vote	.9668	.9812	.9903	.9903	.9902	.9900
Vowel	.9955	.9932	.9917	.9933	.9927	.9919
Zoo	.9536	.9939	.9715	.9947	.9943	.9934
Average	.9130	.9268	.9264	.9281	.9279	.9275
Average Rank	4.8710	3.7097	3.7097	2.6774	2.7419	3.2903

Table 7.10: Average Weighted AUC of the bagged JRip and the employed probability estimation techniques: Macro Averaging, 20 samples, unpruned rule sets (except for Precision)

Name	Bagged Jrip	Precison	Laplace	M 2	M 5	M 10
Anneal	.9944	.9978	.9976	.9994	.9992	.9989
Anneal.orig	.9821	.9974	.9964	.9972	.9973	.9975
Audiology	.9415	.9534	.9395	.9368	.9334	.9296
Autos	.9573	.9431	.9517	.9540	.9518	.9495
Balance-Scale	.9344	.9603	.9578	.9592	.9615	.9635
Breast-cancer	.6213	.6675	.6643	.6630	.6668	.6692
Breast-w	.9888	.9901	.9882	.9882	.9881	.9880
Colic	.8664	.8983	.8948	.8944	.8943	.8939
Credit-a	.9265	.9360	.9357	.9357	.9362	.9361
Credit-g	.7438	.7808	.7797	.7791	.7802	.7812
Diabetes	.7935	.8184	.8175	.8172	.8187	.8202
Glass	.8975	.9104	.9048	.9100	.9079	.9055
Heart-c	.8952	.9051	.9076	.9070	.9084	.9089
Heart-h	.8536	.8837	.8770	.8761	.8768	.8779
Heart-statlog	.8769	.8970	.9022	.9019	.9011	.9007
Hepatitis	.7671	.8509	.8565	.8608	.8600	.8585
Hypothyroid	.9922	.9940	.9960	.9960	.9960	.9959
Ionosphere	.9621	.9563	.9716	.9716	.9713	.9705
Iris	.9910	.9935	.9927	.9927	.9928	.9931
Kr-vs-kp	.9982	.9996	.9996	.9996	.9996	.9996
Labor	.9135	.8959	.9176	.9189	.9135	.9081
Lymph	.9048	.9271	.9364	.9365	.9359	.9314
Primary-tumor	.7479	.8122	.7701	.8000	.7972	.7936
Segment	.9994	.9991	.9992	.9992	.9992	.9992
Sick	.9883	.9975	.9973	.9976	.9975	.9973
Sonar	.9223	.9136	.9170	.9172	.9159	.9140
Soybean	.9934	.9959	.9949	.9955	.9954	.9952
Vehicle	.9344	.9329	.9345	.9342	.9344	.9342
Vote	.9668	.9825	.9854	.9855	.9850	.9845
Vowel	.9955	.9935	.9931	.9944	.9938	.9933
Zoo	.9536	.9948	.9719	.9948	.9944	.9931
Average	.9130	.9283	.9274	.9295	.9292	.9285
Average Rank	4.8548	3.4194	3.4839	2.8065	3.0484	3.3871

Table 7.11: Average Weighted AUC of the bagged JRip and the employed probability estimation techniques: Micro Averaging, 20 samples, unpruned rule sets (except for Precision)

Name	Bagged Jrip	Precison	Laplace	M 2	M 5	M 10
Anneal	.9944	.9976	.9971	.9992	.9989	.9985
Anneal.orig	.9821	.9969	.9958	.9968	.9968	.9971
Audiology	.9415	.9542	.9418	.9364	.9329	.9294
Autos	.9573	.9441	.9498	.9517	.9492	.9464
Balance-Scale	.9344	.9607	.9586	.9602	.9627	.9647
Breast-cancer	.6213	.6735	.6628	.6595	.6634	.6657
Breast-w	.9888	.9907	.9886	.9886	.9884	.9881
Colic	.8664	.9013	.8963	.8965	.8963	.8945
Credit-a	.9265	.9371	.9357	.9356	.9366	.9369
Credit-g	.7438	.7827	.7839	.7833	.7845	.7851
Diabetes	.7935	.8216	.8205	.8202	.8222	.8234
Glass	.8975	.9125	.9075	.9134	.9124	.9098
Heart-c	.8952	.9052	.9068	.9061	.9074	.9076
Heart-h	.8536	.8863	.8764	.8746	.8755	.8770
Heart-statlog	.8769	.8975	.9026	.9026	.9032	.9035
Hepatitis	.7671	.8519	.8598	.8610	.8626	.8626
Hypothyroid	.9922	.9940	.9961	.9961	.9961	.9960
Ionosphere	.9621	.9571	.9718	.9718	.9713	.9705
Iris	.9910	.9927	.9919	.9919	.9920	.9922
Kr-vs-kp	.9982	.9995	.9996	.9996	.9996	.9996
Labor	.9135	.9122	.9243	.9230	.9203	.9135
Lymph	.9048	.9276	.9360	.9365	.9361	.9341
Primary-tumor	.7479	.8112	.7734	.8033	.8000	.7955
Segment	.9994	.9991	.9992	.9992	.9992	.9992
Sick	.9883	.9974	.9968	.9975	.9974	.9971
Sonar	.9223	.9127	.9211	.9208	.9191	.9176
Soybean	.9934	.9956	.9946	.9950	.9950	.9948
Vehicle	.9344	.9338	.9349	.9347	.9348	.9344
Vote	.9668	.9826	.9856	.9857	.9850	.9839
Vowel	.9955	.9929	.9925	.9937	.9932	.9926
Zoo	.9536	.9944	.9717	.9949	.9943	.9929
Average	.9130	.9296	.9282	.9300	.9299	.9292
Average Rank	4.9839	3.4516	3.5484	2.7419	2.9677	3.3065

Table 7.12: Average Weighted AUC of the bagged JRip and the employed probability estimation techniques: Bayesian Decoding, 20 samples, unpruned rule sets (except for Precision)

Name	Bagged Jrip	Precison	Laplace	M 2	M 5	M 10
Anneal	.9944	.9950	.9964	.9967	.9966	.9967
Anneal.orig	.9821	.9957	.9959	.9961	.9963	.9964
Audiology	.9415	.9428	.9276	.9262	.9244	.9219
Autos	.9573	.9418	.9475	.9497	.9478	.9450
Balance-Scale	.9344	.9630	.9598	.9607	.9633	.9654
Breast-cancer	.6213	.6651	.6619	.6604	.6637	.6666
Breast-w	.9888	.9919	.9920	.9919	.9919	.9919
Colic	.8664	.8942	.8945	.8945	.8940	.8928
Credit-a	.9265	.9384	.9371	.9370	.9373	.9373
Credit-g	.7438	.7832	.7842	.7840	.7843	.7847
Diabetes	.7935	.8233	.8208	.8207	.8221	.8230
Glass	.8975	.9084	.9029	.9093	.9072	.9032
Heart-c	.8952	.9070	.9079	.9072	.9081	.9095
Heart-h	.8536	.8892	.8755	.8742	.8761	.8767
Heart-statlog	.8769	.8993	.9028	.9029	.9031	.9028
Hepatitis	.7671	.8478	.8572	.8572	.8582	.8580
Hypothyroid	.9922	.9955	.9983	.9979	.9979	.9978
Ionosphere	.9621	.9582	.9726	.9726	.9717	.9708
Iris	.9910	.9931	.9922	.9922	.9923	.9923
Kr-vs-kp	.9982	.9997	.9998	.9998	.9998	.9998
Labor	.9135	.9108	.9243	.9257	.9189	.9122
Lymph	.9048	.9276	.9335	.9345	.9328	.9292
Primary-tumor	.7479	.8044	.7560	.7929	.7894	.7829
Segment	.9994	.9989	.9990	.9991	.9991	.9990
Sick	.9883	.9973	.9967	.9965	.9962	.9957
Sonar	.9223	.9111	.9200	.9199	.9180	.9161
Soybean	.9934	.9958	.9948	.9953	.9952	.9951
Vehicle	.9344	.9331	.9345	.9345	.9345	.9344
Vote	.9668	.9927	.9923	.9924	.9921	.9919
Vowel	.9955	.9931	.9928	.9940	.9935	.9929
Zoo	.9536	.9940	.9664	.9930	.9895	.9838
Average	.9130	.9288	.9270	.9293	.9289	.9279
Average Rank	5.0645	3.4839	3.3387	2.8065	2.8387	3.4677

Table 7.13: Average Weighted AUC of the bagged JRip and the employed probability estimation techniques: Best Rule, 50 samples, unpruned rule sets (except for Precision)

Name	Bagged Jrip	Precison	Laplace	M 2	M 5	M 10
Anneal	.9941	.9979	.9972	.9992	.9988	.9983
Anneal.orig	.9898	.9953	.9933	.9936	.9938	.9940
Audiology	.9454	.9570	.9464	.9453	.9440	.9422
Autos	.9592	.9500	.9500	.9524	.9492	.9466
Balance-Scale	.9383	.9581	.9593	.9602	.9634	.9661
Breast-cancer	.6206	.6646	.6778	.6753	.6784	.6789
Breast-w	.9916	.9914	.9917	.9918	.9918	.9917
Colic	.8789	.9015	.8929	.8931	.8920	.8910
Credit-a	.9285	.9348	.9337	.9336	.9337	.9336
Credit-g	.7526	.7843	.7823	.7822	.7833	.7839
Diabetes	.8066	.8192	.8205	.8186	.8188	.8197
Glass	.8994	.9132	.9049	.9101	.9067	.9021
Heart-c	.8967	.9076	.9123	.9123	.9127	.9119
Heart-h	.8676	.8876	.8790	.8775	.8782	.8791
Heart-statlog	.8831	.9012	.9007	.8997	.9002	.8996
Hepatitis	.7887	.8450	.8496	.8514	.8549	.8565
Hypothyroid	.9927	.9940	.9958	.9960	.9960	.9960
Ionosphere	.9700	.9642	.9768	.9768	.9764	.9755
Iris	.9941	.9930	.9938	.9936	.9938	.9939
Kr-vs-kp	.9984	.9997	.9997	.9997	.9997	.9997
Labor	.9135	.9203	.9270	.9270	.9270	.9243
Lymph	.8994	.9407	.9382	.9397	.9392	.9380
Primary-tumor	.7555	.8161	.7707	.7982	.7932	.7885
Segment	.9996	.9993	.9992	.9992	.9992	.9992
Sick	.9913	.9977	.9961	.9966	.9965	.9963
Sonar	.9316	.9284	.9224	.9222	.9206	.9201
Soybean	.9947	.9957	.9940	.9948	.9947	.9944
Vehicle	.9370	.9331	.9353	.9352	.9351	.9349
Vote	.9739	.9873	.9923	.9925	.9922	.9919
Vowel	.9959	.9944	.9936	.9948	.9942	.9936
Zoo	.9479	.9948	.9739	.9952	.9950	.9942
Average	.9173	.9312	.9290	.9309	.9307	.9302
Average Rank	4.7742	3.0968	3.4677	2.8065	3.0484	3.8065

Table 7.14: Average Weighted AUC of the bagged JRip and the employed probability estimation techniques: Macro Averaging, 50 samples, unpruned rule sets (except for Precision)

Name	Bagged Jrip	Precison	Laplace	M 2	M 5	M 10
Anneal	.9941	.9982	.9972	.9993	.9991	.9986
Anneal.orig	.9898	.9975	.9964	.9971	.9973	.9974
Audiology	.9454	.9565	.9468	.9456	.9443	.9426
Autos	.9592	.9508	.9533	.9551	.9532	.9497
Balance-Scale	.9383	.9621	.9593	.9607	.9633	.9656
Breast-cancer	.6206	.6749	.6773	.6753	.6805	.6817
Breast-w	.9916	.9911	.9916	.9916	.9915	.9915
Colic	.8789	.8993	.8981	.8975	.8975	.8969
Credit-a	.9285	.9366	.9351	.9350	.9353	.9355
Credit-g	.7526	.7856	.7837	.7836	.7841	.7839
Diabetes	.8066	.8210	.8196	.8193	.8204	.8216
Glass	.8994	.9169	.9134	.9179	.9169	.9135
Heart-c	.8967	.9103	.9112	.9102	.9114	.9118
Heart-h	.8676	.8873	.8787	.8769	.8778	.8792
Heart-statlog	.8831	.9006	.8988	.8987	.8991	.8993
Hepatitis	.7887	.8537	.8544	.8554	.8549	.8506
Hypothyroid	.9927	.9967	.9960	.9960	.9960	.9959
Ionosphere	.9700	.9623	.9741	.9742	.9732	.9720
Iris	.9941	.9927	.9936	.9935	.9938	.9941
Kr-vs-kp	.9984	.9995	.9995	.9995	.9995	.9995
Labor	.9135	.9162	.9230	.9216	.9176	.9162
Lymph	.8994	.9389	.9393	.9400	.9383	.9368
Primary-tumor	.7555	.8201	.7688	.7999	.7976	.7934
Segment	.9996	.9993	.9993	.9994	.9993	.9993
Sick	.9913	.9975	.9976	.9977	.9976	.9975
Sonar	.9316	.9233	.9283	.9281	.9263	.9241
Soybean	.9947	.9960	.9948	.9955	.9952	.9948
Vehicle	.9370	.9352	.9361	.9362	.9361	.9359
Vote	.9739	.9877	.9928	.9929	.9924	.9916
Vowel	.9959	.9944	.9945	.9955	.9951	.9946
Zoo	.9479	.9948	.9741	.9953	.9949	.9942
Average	.9173	.9322	.9299	.9318	.9316	.9309
Average Rank	4.8548	3.3387	3.4032	2.8548	2.9677	3.5806

Table 7.15: Average Weighted AUC of the bagged JRip and the employed probability estimation techniques: Micro Averaging, 50 samples, unpruned rule sets (except for Precision)

Name	Bagged Jrip	Precison	Laplace	M 2	M 5	M 10
Anneal	.9941	.9977	.9966	.9991	.9987	.9980
Anneal.orig	.9898	.9970	.9958	.9967	.9969	.9970
Audiology	.9454	.9569	.9475	.9448	.9430	.9414
Autos	.9592	.9503	.9503	.9531	.9510	.9483
Balance-Scale	.9383	.9623	.9591	.9609	.9638	.9662
Breast-cancer	.6206	.6784	.6719	.6703	.6751	.6771
Breast-w	.9916	.9909	.9914	.9914	.9913	.9912
Colic	.8789	.9041	.8967	.8972	.8965	.8955
Credit-a	.9285	.9383	.9350	.9349	.9354	.9357
Credit-g	.7526	.7872	.7873	.7865	.7873	.7875
Diabetes	.8066	.8237	.8206	.8204	.8215	.8227
Glass	.8994	.9157	.9148	.9193	.9170	.9136
Heart-c	.8967	.9132	.9103	.9097	.9103	.9112
Heart-h	.8676	.8884	.8817	.8793	.8805	.8820
Heart-statlog	.8831	.9014	.8992	.8991	.9001	.9010
Hepatitis	.7887	.8577	.8526	.8542	.8534	.8532
Hypothyroid	.9927	.9967	.9961	.9961	.9961	.9960
Ionosphere	.9700	.9645	.9753	.9753	.9744	.9735
Iris	.9941	.9922	.9921	.9921	.9929	.9930
Kr-vs-kp	.9984	.9995	.9995	.9995	.9995	.9995
Labor	.9135	.9203	.9270	.9270	.9243	.9243
Lymph	.8994	.9403	.9407	.9420	.9400	.9378
Primary-tumor	.7555	.8173	.7705	.8010	.7988	.7956
Segment	.9996	.9993	.9993	.9994	.9993	.9993
Sick	.9913	.9973	.9975	.9976	.9975	.9974
Sonar	.9316	.9237	.9276	.9277	.9262	.9244
Soybean	.9947	.9957	.9949	.9951	.9950	.9948
Vehicle	.9370	.9355	.9364	.9365	.9363	.9360
Vote	.9739	.9882	.9931	.9931	.9925	.9915
Vowel	.9959	.9938	.9939	.9950	.9945	.9940
Zoo	.9479	.9942	.9739	.9950	.9948	.9941
Average	.9173	.9330	.9300	.9319	.9317	.9314
Average Rank	4.7419	3.0323	3.6613	2.8548	3.1129	3.5968

Table 7.16: Average Weighted AUC of the bagged JRip and the employed probability estimation techniques: Bayesian Decoding, 50 samples, unpruned rule sets (except for Precision)

Name	Bagged Jrip	Precison	Laplace	M 2	M 5	M 10
Anneal	.9941	.9950	.9960	.9956	.9956	.9956
Anneal.orig	.9898	.9956	.9957	.9958	.9959	.9957
Audiology	.9454	.9468	.9366	.9383	.9374	.9342
Autos	.9592	.9487	.9506	.9527	.9504	.9468
Balance-Scale	.9383	.9648	.9614	.9623	.9653	.9678
Breast-cancer	.6206	.6736	.6712	.6694	.6737	.6750
Breast-w	.9916	.9917	.9920	.9920	.9919	.9919
Colic	.8789	.8972	.8950	.8954	.8949	.8938
Credit-a	.9285	.9374	.9364	.9363	.9365	.9365
Credit-g	.7526	.7873	.7875	.7874	.7876	.7873
Diabetes	.8066	.8258	.8206	.8202	.8217	.8228
Glass	.8994	.9147	.9114	.9170	.9146	.9100
Heart-c	.8967	.9119	.9105	.9091	.9101	.9107
Heart-h	.8676	.8907	.8794	.8773	.8789	.8802
Heart-statlog	.8831	.9022	.8998	.8998	.9005	.9008
Hepatitis	.7887	.8585	.8544	.8549	.8554	.8526
Hypothyroid	.9927	.9973	.9984	.9979	.9979	.9979
Ionosphere	.9700	.9625	.9715	.9715	.9709	.9695
Iris	.9941	.9919	.9927	.9926	.9927	.9927
Kr-vs-kp	.9984	.9996	.9998	.9998	.9998	.9998
Labor	.9135	.9230	.9311	.9324	.9257	.9216
Lymph	.8994	.9377	.9382	.9392	.9368	.9352
Primary-tumor	.7555	.8120	.7550	.7926	.7898	.7846
Segment	.9996	.9993	.9991	.9991	.9991	.9991
Sick	.9913	.9973	.9969	.9963	.9961	.9959
Sonar	.9316	.9233	.9274	.9272	.9261	.9239
Soybean	.9947	.9959	.9949	.9953	.9952	.9948
Vehicle	.9370	.9351	.9361	.9362	.9360	.9357
Vote	.9739	.9933	.9931	.9932	.9930	.9926
Vowel	.9959	.9940	.9942	.9952	.9947	.9942
Zoo	.9479	.9942	.9676	.9942	.9924	.9881
Average	.9173	.9322	.9288	.9312	.9309	.9299
Average Rank	4.8065	3.0645	3.3387	2.8710	3.0645	3.8548

Table 7.17: Average Weighted AUC of the bagged JRip and the employed probability estimation techniques: Best Rule, 100 samples, unpruned rule sets (except for Precision)

Name	Bagged Jrip	Precison	Laplace	M 2	M 5	M 10
Anneal	.9944	.9980	.9974	.9994	.9992	.9988
Anneal.orig	.9903	.9953	.9936	.9945	.9944	.9945
Audiology	.9444	.9561	.9434	.9458	.9437	.9405
Autos	.9595	.9507	.9501	.9528	.9500	.9478
Balance-Scale	.9386	.9595	.9616	.9626	.9659	.9685
Breast-cancer	.6349	.6588	.6736	.6702	.6725	.6740
Breast-w	.9920	.9913	.9919	.9919	.9918	.9918
Colic	.8802	.9014	.8894	.8901	.8889	.8881
Credit-a	.9298	.9358	.9344	.9344	.9348	.9349
Credit-g	.7559	.7865	.7890	.7894	.7901	.7899
Diabetes	.8098	.8216	.8216	.8206	.8208	.8211
Glass	.9022	.9213	.9059	.9133	.9103	.9048
Heart-c	.8951	.9091	.9127	.9129	.9136	.9134
Heart-h	.8630	.8912	.8806	.8787	.8802	.8804
Heart-statlog	.8821	.9008	.9028	.9021	.9023	.9016
Hepatitis	.8161	.8458	.8488	.8486	.8516	.8544
Hypothyroid	.9953	.9940	.9958	.9962	.9963	.9962
Ionosphere	.9724	.9683	.9782	.9783	.9776	.9769
Iris	.9937	.9936	.9935	.9935	.9936	.9937
Kr-vs-kp	.9984	.9997	.9996	.9996	.9996	.9996
Labor	.9115	.9203	.9311	.9311	.9257	.9270
Lymph	.9076	.9416	.9382	.9381	.9373	.9359
Primary-tumor	.7547	.8164	.7678	.7984	.7941	.7892
Segment	.9996	.9993	.9992	.9993	.9992	.9992
Sick	.9947	.9978	.9961	.9966	.9965	.9963
Sonar	.9283	.9268	.9299	.9301	.9278	.9258
Soybean	.9953	.9957	.9941	.9949	.9948	.9947
Vehicle	.9368	.9343	.9361	.9363	.9360	.9356
Vote	.9764	.9878	.9922	.9923	.9920	.9918
Vowel	.9969	.9950	.9945	.9955	.9951	.9946
Zoo	.9452	.9948	.9750	.9952	.9950	.9946
Average	.9192	.9319	.9296	.9317	.9313	.9308
Average Rank	4.6452	3.2581	3.5645	2.6129	3.1452	3.7742

Table 7.18: Average Weighted AUC of the bagged JRip and the employed probability estimation techniques: Macro Averaging, 100 samples, unpruned rule sets (except for Precision)

Name	Bagged Jrip	Precison	Laplace	M 2	M 5	M 10
Anneal	.9944	.9982	.9976	.9994	.9993	.9990
Anneal.orig	.9903	.9975	.9964	.9972	.9973	.9974
Audiology	.9444	.9555	.9447	.9464	.9445	.9421
Autos	.9595	.9518	.9538	.9561	.9538	.9513
Balance-Scale	.9386	.9626	.9603	.9617	.9646	.9669
Breast-cancer	.6349	.6705	.6738	.6730	.6747	.6760
Breast-w	.9920	.9909	.9917	.9916	.9917	.9914
Colic	.8802	.9023	.8962	.8964	.8958	.8945
Credit-a	.9298	.9368	.9353	.9351	.9355	.9354
Credit-g	.7559	.7862	.7887	.7888	.7890	.7886
Diabetes	.8098	.8220	.8191	.8190	.8201	.8215
Glass	.9022	.9232	.9171	.9216	.9198	.9166
Heart-c	.8951	.9110	.9114	.9110	.9122	.9123
Heart-h	.8630	.8898	.8813	.8793	.8812	.8825
Heart-statlog	.8821	.9021	.8993	.8993	.8996	.9002
Hepatitis	.8161	.8554	.8557	.8542	.8552	.8516
Hypothyroid	.9953	.9966	.9960	.9960	.9960	.9959
Ionosphere	.9724	.9673	.9760	.9760	.9750	.9737
Iris	.9937	.9936	.9933	.9932	.9935	.9935
Kr-vs-kp	.9984	.9995	.9995	.9995	.9995	.9995
Labor	.9115	.9176	.9243	.9230	.9189	.9162
Lymph	.9076	.9400	.9398	.9412	.9402	.9377
Primary-tumor	.7547	.8208	.7678	.8002	.7984	.7944
Segment	.9996	.9993	.9994	.9994	.9994	.9994
Sick	.9947	.9975	.9977	.9979	.9978	.9976
Sonar	.9283	.9213	.9316	.9316	.9290	.9266
Soybean	.9953	.9959	.9948	.9954	.9951	.9949
Vehicle	.9368	.9353	.9368	.9369	.9366	.9363
Vote	.9764	.9881	.9924	.9924	.9918	.9911
Vowel	.9969	.9950	.9950	.9959	.9955	.9951
Zoo	.9452	.9948	.9751	.9953	.9952	.9944
Average	.9192	.9329	.9304	.9324	.9321	.9314
Average Rank	4.8387	3.1613	3.4839	2.8065	2.9194	3.7903

Table 7.19: Average Weighted AUC of the bagged JRip and the employed probability estimation techniques: Micro Averaging, 100 samples, unpruned rule sets (except for Precision)

Name	Bagged Jrip	Precison	Laplace	M 2	M 5	M 10
Anneal	.9944	.9978	.9968	.9992	.9990	.9985
Anneal.orig	.9903	.9970	.9959	.9967	.9969	.9970
Audiology	.9444	.9565	.9458	.9450	.9429	.9404
Autos	.9595	.9502	.9511	.9537	.9513	.9487
Balance-Scale	.9386	.9627	.9601	.9618	.9649	.9674
Breast-cancer	.6349	.6717	.6757	.6737	.6760	.6773
Breast-w	.9920	.9906	.9914	.9914	.9913	.9911
Colic	.8802	.9050	.8963	.8962	.8961	.8945
Credit-a	.9298	.9385	.9355	.9353	.9358	.9359
Credit-g	.7559	.7871	.7901	.7901	.7903	.7899
Diabetes	.8098	.8232	.8203	.8202	.8215	.8226
Glass	.9022	.9214	.9171	.9215	.9194	.9163
Heart-c	.8951	.9133	.9114	.9106	.9118	.9122
Heart-h	.8630	.8893	.8837	.8808	.8833	.8850
Heart-statlog	.8821	.9031	.9000	.8997	.8998	.8997
Hepatitis	.8161	.8565	.8554	.8554	.8562	.8526
Hypothyroid	.9953	.9967	.9962	.9961	.9961	.9961
Ionosphere	.9724	.9681	.9770	.9770	.9762	.9752
Iris	.9937	.9928	.9923	.9925	.9925	.9927
Kr-vs-kp	.9984	.9995	.9995	.9995	.9995	.9995
Labor	.9115	.9230	.9284	.9297	.9257	.9243
Lymph	.9076	.9393	.9408	.9422	.9409	.9388
Primary-tumor	.7547	.8194	.7713	.8018	.8001	.7973
Segment	.9996	.9993	.9994	.9994	.9994	.9994
Sick	.9947	.9973	.9976	.9978	.9977	.9975
Sonar	.9283	.9184	.9296	.9297	.9278	.9264
Soybean	.9953	.9956	.9948	.9951	.9949	.9946
Vehicle	.9368	.9356	.9372	.9373	.9370	.9366
Vote	.9764	.9885	.9927	.9927	.9921	.9910
Vowel	.9969	.9943	.9943	.9953	.9949	.9944
Zoo	.9452	.9947	.9750	.9951	.9948	.9943
Average	.9192	.9331	.9307	.9327	.9325	.9319
Average Rank	4.8065	3.1613	3.4516	2.7419	3.0161	3.8226

Table 7.20: Average Weighted AUC of the bagged JRip and the employed probability estimation techniques: Bayesian Decoding, 100 samples, unpruned rule sets (except for Precision)

Name	Bagged Jrip	Precison	Laplace	M 2	M 5	M 10
Anneal	.9944	.9951	.9961	.9962	.9962	.9962
Anneal.orig	.9903	.9956	.9958	.9958	.9959	.9958
Audiology	.9444	.9465	.9335	.9384	.9372	.9329
Autos	.9595	.9498	.9509	.9532	.9505	.9483
Balance-Scale	.9386	.9652	.9621	.9630	.9661	.9684
Breast-cancer	.6349	.6700	.6708	.6683	.6711	.6722
Breast-w	.9920	.9913	.9920	.9920	.9919	.9919
Colic	.8802	.8985	.8938	.8934	.8933	.8922
Credit-a	.9298	.9373	.9365	.9364	.9367	.9365
Credit-g	.7559	.7883	.7912	.7913	.7915	.7912
Diabetes	.8098	.8257	.8203	.8201	.8216	.8226
Glass	.9022	.9203	.9141	.9198	.9172	.9134
Heart-c	.8951	.9115	.9108	.9105	.9113	.9120
Heart-h	.8630	.8911	.8817	.8792	.8809	.8831
Heart-statlog	.8821	.9029	.9003	.9004	.9016	.9014
Hepatitis	.8161	.8572	.8514	.8537	.8521	.8501
Hypothyroid	.9953	.9986	.9983	.9979	.9979	.9979
Ionosphere	.9724	.9639	.9732	.9732	.9722	.9711
Iris	.9937	.9929	.9929	.9929	.9932	.9933
Kr-vs-kp	.9984	.9997	.9996	.9996	.9998	.9998
Labor	.9115	.9257	.9365	.9365	.9324	.9257
Lymph	.9076	.9375	.9378	.9391	.9388	.9364
Primary-tumor	.7547	.8138	.7553	.7943	.7909	.7857
Segment	.9996	.9993	.9994	.9994	.9994	.9994
Sick	.9947	.9973	.9977	.9977	.9976	.9974
Sonar	.9283	.9200	.9303	.9304	.9288	.9261
Soybean	.9953	.9959	.9950	.9953	.9952	.9949
Vehicle	.9368	.9354	.9367	.9368	.9366	.9363
Vote	.9764	.9938	.9933	.9933	.9930	.9928
Vowel	.9969	.9945	.9946	.9955	.9951	.9947
Zoo	.9452	.9944	.9708	.9944	.9937	.9905
Average	.9192	.9326	.9294	.9319	.9316	.9306
Average Rank	4.6774	3.2581	3.4194	2.9194	2.9355	3.7903