Employing Linear Regression in Regression Tree Leaves

Aram Karalič*
Jožef Stefan Institute, Jamova 39, 61111 Ljubljana, Slovenia
aram.karalic@ijs.si

Abstract. The advantage of using linear regression in the leaves of a regression tree is analysed in the paper. It is carried out how this modification affects the construction, pruning and interpretation of a regression tree. The modification is tested on artificial and real-life domains. The results show that the modification is beneficial as it leads to smaller classification errors of induced regression trees. Keywords: machine learning, TDIDT, regression, linear regression, Bayesian approach.

1 Introduction

Regression trees, similar to classification trees, are used when learning a relation between attributes and a *continuous* class. Their inner nodes are labelled with a test on the value of an attribute, and their leaves are labelled with a function prescribing a value to the class. A regression tree actually implements a function $y(x_1, x_2, \ldots, x_n)$ of n continuous or discrete attributes. In the basic CART algorithm for the construction of regression trees [1], the class value in the leaves of a tree is estimated as a constant function. The definition of the leaves is extended in this paper, so as to allow a function in the leaf to be a linear function of continuous attributes.

2 Construction of Regression Trees

The algorithm for regression tree construction belongs to TDIDT (top-down induction of decision trees) family of algorithms. These algorithms split an example set representing the node of a tree into two subsets from which they recursively form subtrees. The vital part of the algorithm is the measure of split goodness, derived from the measure of the impurity of an example set. Since the class is continuous, an estimate of variance of the class values is used as the impurity measure. The best split of examples in a node is taken to be the one that minimises the expected impurity I_{exp}, given by

$$I_{exp} = p_l I_l + p_r I_r \tag{1}$$

where p_l, p_r denote probabilities of transitions into the left and the right son of the node, and I_l, I_r are the corresponding impurities of the nodes. The variance of the class values of example set E, computed by:

$$\sigma^2(E) = \frac{1}{W(E)} \sum_{e_i \in E} w_i (y_i - \mu(E))^2 \tag{2}$$

is used as the impurity measure in original CART algorithm. In (2), w_i is the weight of the i-th example, $W(E)$ the sum of example weights of the example set E, and $\mu(E)$ is the mean class value of example set E.

The quality of the tree is measured by the *mean squared error* R of a tree T, defined with $R(T) = \frac{1}{N}\sum_{i=1}^{N}(y_i - y(x_{i1}, x_{i2}, \ldots, x_{in}))^2$ where N is the number of examples used for testing the tree, y_i is the value of the class of the i-th example, $x_{i1}, x_{i2}, \ldots, x_{in}$ are values of example's attributes, and $y(\ldots)$ represents the value of the class estimated by the tree. To compare the quality of trees from different domains the *relative mean squared error* defined as $RE(T) = \frac{R(T)}{R(M)}$ is used. Here, M is a predictor which always predicts the mean value of the training example set.

Bayesian Approach

Bayesian approach to estimating class distribution in the tree-structured regression is based on Good's notion of Bayesian analysis [2]. The application of the approach to tree-structured regression was developed in [4], and is used in our experiments because the approach introdues parameter m, which enables us to adapt the algorithm of tree building and post-pruning to the expected level of noise in learning examples.

3 Local Linear Regression (LLR)

The definition of a leaf of a regression tree is extended in this section so as to allow function in the leaf to be

*The work has been supported by Ministry of Science and Technology, Republic of Slovenia.

a linear function of continuous attributes.

LLR During the Construction of a Regression Tree

When using local linear regression, variance (2) is not an appropriate measure for impurity of an example set since example sets with large variance and very low impurity can arise. For that reason, the expression

$$I(E) = \frac{1}{W(E)} \sum_{e_i \in E} (y_i - g(\vec{x}_i))^2 \qquad (3)$$

is used as an impurity measure of an example set E during the construction of a regression tree. Function g represents the regression plane through the examples.

LLR During Post-Pruning of a Regression Tree

For post-pruning of generated regression trees an algorithm based on the Niblett-Bratko post-pruning method [5] was used. Pruning is based on estimation of the classification error on test examples. In each node, the algorithm makes an estimate of *static error* e_s, which represents the expected error on unknown examples if this node was a leaf, and *backed-up error* e_b, which is an estimation of the expected error in the case the tree was not pruned at that node. When $e_s \leq e_b$ the subtree is pruned and the node is converted to a leaf.

When employing local linear regression, the impurity is estimated with (3) instead with (2).

LLR During Interpretation of a Regression Tree

When classifying a new example with a regression tree, the class assigned to the example is determined by the value of the linear function in the leaf to which the example came during the process of classification.

4 Experiments

The developed concepts were implemented in RETIS — a system for regression tree construction. The system was used to evaluate the behaviour of local linear regression in six domains, three of them being artificial and three of them real-life domains, describing the problems of steel grinding, yarn strength prediction and prediction of concrete properties. More about the domains and previous work can be found in [3]. Experiment for each domain consisted of 10 repetitions of the sequence:

1. Split the set of all examples into learning example set and testing example set in proportion $70 : 30$.
2. Construct a tree using learning examples and $m = 0$.
3. For each value of m (from a given set of values) do:
   - prune the tree using that value of m,
   - test (on testing examples) the pruned tree's classification error.
4. Select the best tree in the sequence of pruned trees.

Two sets of experiments were performed: with and without employing LLR. Table 1 gives the t-tests of significance, comparing the two sets of experiments.

Domain	hypothesis H_0	significance of H_0
LINE	$\sqrt{RE}_{m+} < \sqrt{RE}_{m-}$	>99.9
LEXP	$\sqrt{RE}_{m+} < \sqrt{RE}_{m-}$	>99.9
LOSC	$\sqrt{RE}_{m+} < \sqrt{RE}_{m-}$	99.9
GRIN	$\sqrt{RE}_{m+} < \sqrt{RE}_{m-}$	>99.9
YARN	$\sqrt{RE}_{m+} > \sqrt{RE}_{m-}$	90.1
CONC	$\sqrt{RE}_{m+} < \sqrt{RE}_{m-}$	99.9

Table 1: Results of t-tests for hypotheses comparing error committed using local linear regression ($\sqrt{RE}_{m+}$) and error committed without the use of LLR ($\sqrt{RE}_{m-}$).

5 Discussion and Conclusions

From the Table 1 it can be seen that in most cases local linear regression helped to achieve better results. However, if we examine each of the ten repetitions of the experiment separately, we will find cases where local linear regression decreased performance.

The overall conclusion is that local linear regression is beneficial in the sense of the quality of the induced regression tree, although it should be used cautiously, possibly in combination with expert knowledge about the domain.

References

[1] L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. *Classification and Regression Trees.* Wadsworth Int. Group, Belmont, California, USA, 1984.

[2] I.J. Good. *The Estimation of Probabilities.* M.I.T. Press, Cambridge, Massachusetts, USA, 1965.

[3] A. Karalič. Employing Linear Regression in Regression Tree Leaves. Technical Report IJS DP-6450, Jožef Stefan Institute, Ljubljana, Slovenia, 1992.

[4] A. Karalič and B. Cestnik. The Bayesian Approach to Tree-Structured Regression. In *Proc. of ITI-91, Cavtat, Croatia*, pages 155–160, 1991.

[5] T. Niblett and I. Bratko. Learning Decision Rules in Noisy Domains. In M. Bramer, editor, *Development in Expert Systems.* Cambridge University Press, 1986.