MARS3.6 with Bagging: `mars3.6-bag-1`

Carl Edward Rasmussen and Mike Revow

May 2, 1996

The Multivariate Adaptive Regression Splines (MARS) method of Friedman (Friedman 1991) has also been tested. This method is a fairly well known method for non-linear regression for high dimensional data from the statistics community. A detailed description of MARS will not be given here, see (Friedman 1991). Following is a simplistic account of MARS gives a flavor of the method. The input space is carved up into several (overlapping) regions in which splines are fit. The fit is built using first a constructive phase which introduces input regions and splines followed by a pruning phase. The final model has the form of a sum of products of univariate splines; it is a continuous function (with continuous derivatives) and is additive in the sets of variables allowed to interact.

Friedman has supplied his FORTRAN implementation of MARS (version 3.6). Since MARS is not very computationally demanding it is used in conjunction with the Bagging procedure of Breiman (Breiman 1994). Using this method one trains MARS on a number of bootstrap samples of the training set and averages the resulting predictions. The bootstrap samples are generated by sampling the original training set with replacement; samples of the same size as the original training set are used. Only one set of predictions is generated regardless of the loss function - the same predictions are used for the absolute error loss function and the squared error loss function.

The following parameter settings have been used for MARS: 50 bootstrap repetitions were used in the bagging procedure, the maximum number of basis functions was 15, the maximum number of variables allowed to interact was 8. The computational demand of applying this method is fairly modest. A training set with 32 inputs and 1024 training cases the 50 bootstrap replications take a total of 5 minutes on a 200MHz R4400.

References

Breiman, L. 1994. Bagging predictors. Technical Report 421, Berkeley, California 94720.

Friedman, J. H. 1991. Multivariate adaptive regression splines (with discussion). *Annals of Statistics*, 19:1–141.