Data mining and Machine Learning, Mar 24, 2008

MARS: Multivariate Adaptive Regression Splines (Friedman 1991, Friedman and Silverman 1989)

An Overview

- MARS is an adaptive procedure for regression, and is well suited for high dimensional problems.
  - It can be viewed as a generalization of stepwise linear regression
  - or a modification of the CART method to improve its performance in the regression setting.
- MARS uses expansions in piecewise linear basis functions of the form in reflected pairs $(x-t)_+$ and $(t-x)_+$, with the knots at each observations. Therefore, the collection of basis functions (Fig 9.9)

$$C = \left\{(X_j - t)_+ , \ \ (t - X_j)_+\right\}_{t \in \left\{x_{1j}, x_{2j}, \ldots, x_{Nj}\right\}, \ j=1,2,\ldots p}$$

The Algorithm

- The real art is in the construction of the functions $h_m(x)$. (Fig 9.10)
  - We start with only the constant function $h_0(x) = 1$ in the model, and all functions in the set C are candidate functions.
  - At each stage, we consider as a new basis function pair, all products of the function h_m in the model set M with one of the reflected pairs in C that produces the largest decrease in the training error. Then the winning products are added to the model.
  - The procedure continues untill the model set M contains some preset maxium number of terms.
- At the end of the process we have a large model, which often overfits the data, and so backward deletion procedure is applied. The terms whose removal causes the smallest increase in residual square error is deleted from the model at each stage, producing an estimated best model $\hat{f}_\lambda$ of each size(number of terms) λ. One would use cross-validation to estimate the optimal λ, but for computional savings the MARS procedure instead uses generalized cross-validation(GCV).
- The criterion is defined as

$$GCV(\lambda) = \frac{1}{N} \frac{\sum_{i=1}^{N} \left(y_i - \hat{f}_\lambda(x_i)\right)^2}{\left(1 - M(\lambda)/N\right)^2}$$

  The value $M(\lambda)$ is the effective number of parameters in the model.

- Generalized cross-validation provides a convenient approximation to leave-one out cross-validation, for linear fitting under square-error loss.

What's the advantage of MARS?

- Why these piecewise linear basis functions?
  - A key property of the function is their ability to operate locally. Since they are zero over part of their ranges, when they are multiplied together the result is nonzero only over the small part of the feature space where both component functions are nonzero. (Fig 9.11)
  - Computational advantage.
- Why this particular model strategy?
  - The forward modeling strategy in MARS is hierarchical, in the sense that multiway products are built up from products involving terms already in the model. The philosophy here is that a high-order interaction will likely only exist if some of its lower order "footprints" exist as well. This need not to be true, but is a reasonable working assumption and avoids the search over an exponentially growing space of alternatives.
  - There is one restriction put on the formation of model terms: each input can appear at most once in a product. This prevents the formation of higher-order powers of an input, which increase or decrease too sharplynear the boundaries of the feature space. Such powers can be approximated in a more stable way with piecewise linear functions.
- A useful option in MARS procedure is to set an upper limit on the order of interaction, which can aid in the interpretation.

Relationship of MARS to CART

- Suppose we take the MARS procedure and make the following changes:
  - Replace the piecewise linear basis functions by step functions $I(x - t > 0)$ and $I(x - t \leq 0)$.

    (Multiplying a step function by a pair of reflected step functions is equivalent to splitting a node at the step.)
  - When a model term is involved in a multiplication by a candidate term, it gets replaced by the interaction, and hence is not available for further interaction. (This restriction implies that a node may not be split more than once.)

With these changes, the MARS forward procedure is the same as the CART tree-growing algorithm.

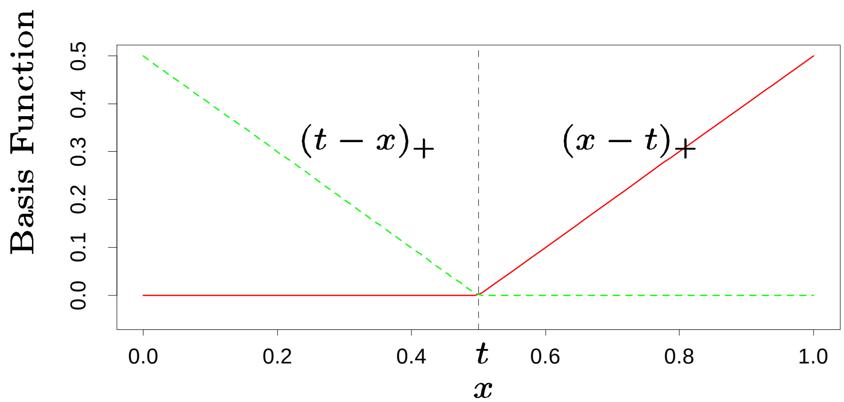

Figure 9.9: *The basis functions* $(x-t)_+$ *(solid red) and* $(t-x)_+$ *(broken green) used by MARS.*

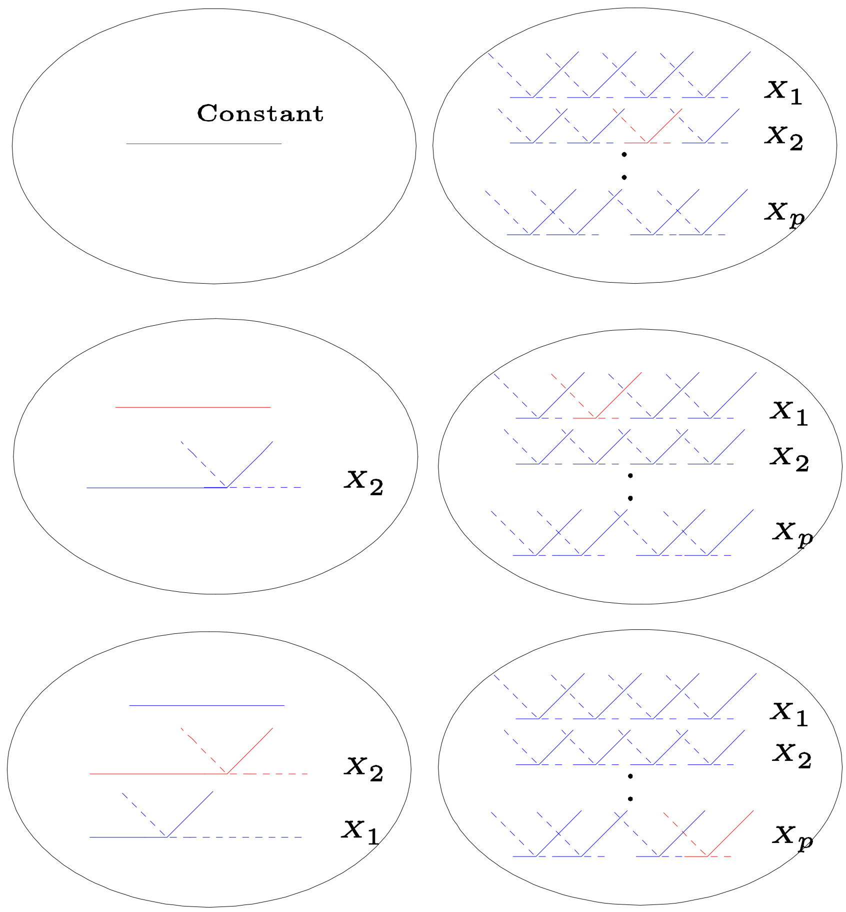

Figure 9.10: *Schematic of the MARS forward model-building procedure. On the left are the basis functions currently in the model: initially, this is the constant function $h(X) = 1$. On the right are all candidate basis functions to be considered in building the model. These are pairs of piecewise linear basis functions as in Figure 9.9, with knots t at all unique observed values x_{ij} of each predictor X_j. At each stage we consider all products of a candidate pair with a basis function in the model. The product that decreases the residual error the most is added into the current model. Above we illustrate the first three steps of the procedure, with the selected functions shown in red.*

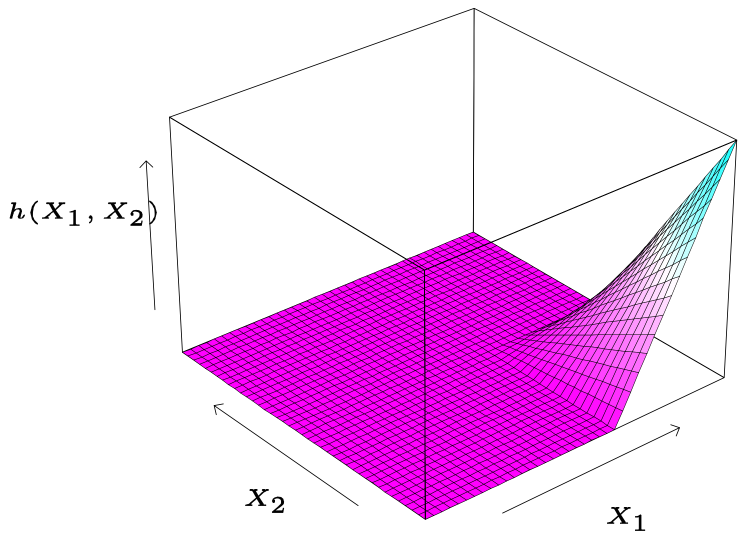

Figure 9.11: *The function $h(X_1, X_2) = (X_1 - x_{51})_+ \cdot (x_{72} - X_2)_+$, resulting from multiplication of two piecewise linear MARS basis functions.*