**MARS Algorithm Overview**

The Multivariate Adaptive Regression Splines (MARS) technique is a non-parametric regression technique introduced by Jerome H. Friedman of Stanford University in a famous paper bearing the name of the technique (Friedman, 1991, Vol. 19, No 1.). Like all regression techniques, it seeks to estimate the relationship amongst various variables. However, the MARS technique seeks to estimate the relationship of high dimensional data with strong nonlinearities in a manner that is both computational reasonable as well as with an easily understandable output.

The MARS technique differs in its novel application of regression splines in the form of hinge functions.

Broadly speaking, a “spline” is simply a function constructed in various segments from other polynomial functions (Racine, 2012). The MARS technique constructs these splines via hinge functions, which can be represented in the form

H[+-(x-t)], where H[n] indicates to take the positive part.

The value value x corresponds to the value of the variable x and t is a constant, referred to as a “knot”. This technique is used to divide the region of data into mutually exclusive segments in such a manner as to minimize some objective function. Typically, the residual sum of squares is used.

<RSS function here>

The MARS modeling building process can be thought of as two main phases. In the first phase, called the forward pass, the mean of the independent variable is calculated. Then hinge functions are added and the residual error is calculated. This is repeated until the placement of hinge functions is found such that the objective function is minimized. This step is then repeated until the objective function is below some threshold, or a maximum number of terms is obtained (where the maximum number of terms is specified by the user).

The first stage of the MARS modeling building process will typically result in a function that matches the data extremely closely, however this model will likely not generalize well to new data (a condition frequently known as “over-fitting”). Thus, the second process, called backward pass, is used to select terms for removal. The can be done many ways, however a commonly used method is to generate a random subset of the data. Then each hinge function is removed and the objective function is recalculated. This is repeated for all possible combinations of objective functions. The combination with the highest predictive power is selected as the final model.

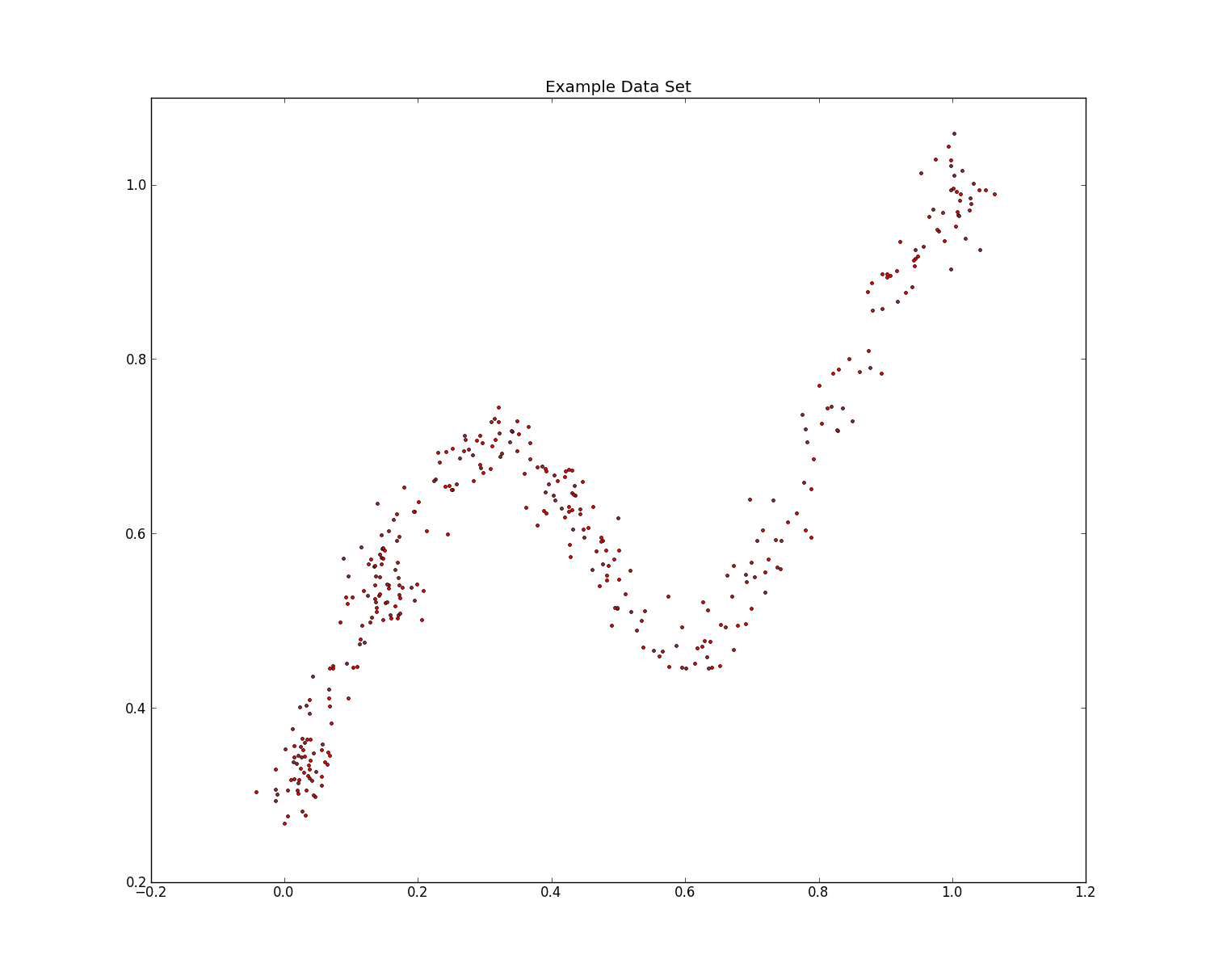
**Earth Software Overview**

The Earth software library is an implementation of the Multivariate Adaptive Regression Splines (MARS) technique. The name MARS is copy written, but the algorithm is published in the public domain. Thus, as a clever pun, open-source implementations of this algorithm are referred to as Earth.

The implementation used in this paper

Simple Example:

To illustrate the usage of the MARS software program, a simple data set was created using the paint-data tool provided in the Orange package. The data set was intentionally created to have irregular density of points and a nonlinear form. The goal is then to use the MARS software to predict the Y values when given the X value.



We use the provided method in the Orange package to fit the data using the MARS algorithm. We then obtain the final resulting equation.



Upon plotting this equation on the data, we can visually verify the accuracy of the algorithm. As one can see from the blue line (representing the above equation), that the algorithm produces a very close fit of the irregular data set.

