

Paper 112-2009

Variable Selection and Transformation of Variables in SAS® Enterprise Miner™ 5.3

Kattamuri Sarma, Ecostat Research Corp.

ABSTRACT

In predictive modeling, one is often confronted with a large number of inputs (explanatory variables). The number of potential inputs to choose from may be as large as 2000 or higher. The Variable Selection node of SAS® Enterprise Miner™ provides alternative methods for eliminating irrelevant variables and selecting variables which have predictive power. This paper demonstrates the various variable selection methods available in SAS Enterprise Miner 5.3.

The predictive power of the inputs can sometimes be enhanced by making suitable transformations. One can use the Transform Variables node to select the best mathematical transformation for any given input. The Transform Variables node can also be used for optimally binning the interval inputs and creating dummy variables from categorical inputs.

Variable selection and transformation is also done by the Decision Tree node. The inputs that give significant splits in creating a decision tree are selected by the Decision Tree node and passed to the next node. This paper discusses the details of the variable selection methods, transformations and the options available in SAS Enterprise Miner 5.3.

No paper was submitted for publication.

CONTACT INFORMATION

Kattamuri Sarma
Ecostat Research Corp.
kssarma@worldnet.att.net

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.

Other brand and product names are trademarks of their respective companies.