

ADVERTISEMENT

Decision Tree Forests



DTREG creates forests of decision trees to model and predict data



decision_trees, continuous

Decision Tree with Continuous Variables

Imran

Jan '15

Hi,
I am new to business analytics domain and started exploring methods of modeling. Recently I studied decision tree and not clear on method of handling continuous variable by it. I know decision tree handles categorical variable very well such as gender has two values male and female then it separates node in two sub-nodes based on gender but in case of continuous variable, how does it handle such as income? I know it is not possible to create sub-nodes for each values then how it decides to create bucket like <20000 , ≥ 20000 & <50000 and ≥ 50000 .

Regards,
Imran

aayushmnit

Jan '15

Hi,
For more back end information on tree algorithms to continuous variables visit the following link - <http://www.stat.cmu.edu/~cshalizi/350-2006/lecture-10.pdf>

It talks about Regression trees and it's back end working. Hope this helps.

Regards,
Aayush

kunal

Jan '15

Imran,
Decision trees work with continuous variables as well. The way they work is by principle of reduction of variance.

Let us take an example, where you have age as the target variable. So, let us say you compute the variance of age and it comes out to be x . Next, decision tree looks at various splits and calculates the

total weighted variance of each of these splits. It chooses the split which provides the minimum variance.

Hope this clarifies the algorithm and the doubt.

Regards,
Kunal

VJ1

Sep '15

Information Value & WOE statistics found to be very effective in reducing number of variables or better way to put it is to identify potential variables. refer the link pls

<http://support.sas.com/resources/papers/proceedings13/095-2013.pdf>

In the case of Income variable (any continuous) apply the concept called **Coarse classing** (make 40 groups) => **Fine classing** (condense to 5 to 9 groups), this methodology is widely used in Score Card building in retail banking Industry, above referred link would give you sufficient understanding of this concept - I applied this concept on metrics before building DT models & could make insights which made sense to business/clients.

ADVERTISEMENT

Decision Tree Forests



DTREG creates forests of decision trees to model and predict data



© Copyright 2015. Analytics Vidhya