

Guidance for weka-CDT

Saisai Ma, 06/09/2015

CDT: a Program for generating Causal Decision Trees. CDT algorithm is located at weka-classifiers-causality.

Parameter settings

- **height** - The maximum height of the causal decision tree. (default: 5)
- **pruned** - Whether pruning is performed. (default: True)
- **subclass** - Whether subclassification (propensity score) is performed, exact matching otherwise. (default: False)
In this version, subclassification is performed by invoking the function *matchit(formula, data, method="subclass", ...)* in the R package *MatchIt*. Thus extra configurations need to be done:
 - Run the following commands in R (Note that keep Rserve() running when using "subclass" option)

```
> install.packages("MatchIt")
> install.packages("Rserve")
> library(Rserve)
> Rserve()
```
- **location** - The location of propensityTest_weka.R file. When subclassification selected, please indicate the location of propensityTest_weka.R. This R file is employed to do subclassification and return the subclasses to Java. This option only works when subclass=True.
- **numControl** - The maximum number of control list during the causal testing. If the number of control list is larger than the threshold L, then select the first L variables as the control. This option only works when subclass=False.