



## Classification and Regression Trees (CART) Documentation

**Description:** Tree-based regression and classification  
**Author:** Joshua Gould, gp-help@broad.mit.edu

### Summary:

CART builds classification and regression trees for predicting continuous dependent variables (regression) and categorical predictor variables (classification) (Breiman, et al., 1984). It works by recursively splitting the feature space into a set of non-overlapping regions (rectangles in the case of continuous features; subsets of values, in the case of categorical features), and by then predicting the most likely value of the dependent variable within each region.

A classification tree represents a set of nested logical if-then conditions on the values of the features variables that allows for the prediction of the value of the dependent categorical variable based on the observed values of the feature variables. A regression tree also represents a set of nested logical if-then conditions on the features variables, but these are used to predict the value of a continuous response variable instead.

CART can handle missing values. The model can be tested on a separately specified test set. Additionally, the model can be saved and used subsequently on additional test sets.

The table below summarizes the different options available and which parameters are required depending on the option selected.

| Parameter            | Train<br>create a<br>predictive model<br>from a training<br>dataset | Test with saved model<br>run a saved model on a<br>new test dataset | Train/Test<br>create a model on<br>training data and run it<br>on test data |
|----------------------|---|---|---|
| train.data.filename  | Required  | No  | Required  |
| train.cls.filename   | Required  | No  | Required  |
| saved.model.filename | No  | Required  | No  |
| test.data.filename   | No  | Required  | Required  |
| test.cls.filename    | No  | Required  | Required  |
| pred.results.file    | No  | Yes   | Yes   |
| tree.output.file     | Required  | No  | Required  |
| model.output.file    | Required  | No  | Required  |

### References:

- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software.

### Parameters:

| Name | Description |
|------|-------------|
|------|-------------|

# GenePattern

|                      |  |
|----------------------|--|
| train.data.filename  | The training data file - .gct, .res, ignored if a saved model (saved.model.filename) is used |
| train.cls.filename   | The training class file - .cls, ignored if a saved model (saved.model.filename) is used      |
| saved.model.filename | input CART model   |
| model.output.file    | name of output CART model file   |
| test.data.filename   | The test data file - .gct, .res  |
| test.cls.filename    | The test class file - .cls   |
| pred.results.file    | The name of the output file for prediction results   |
| tree.output.file     | The name of the file containing a plot of the classification tree                            |

## Output Files:

If test data is supplied:

1. a file containing the prediction results
2. plot of the decision tree

If training data is specified

1. a file containing the saved prediction model

## Platform dependencies:

**Module type:** Prediction  
**CPU type:** any  
**Language:** R