**Understanding regression trees and model trees**

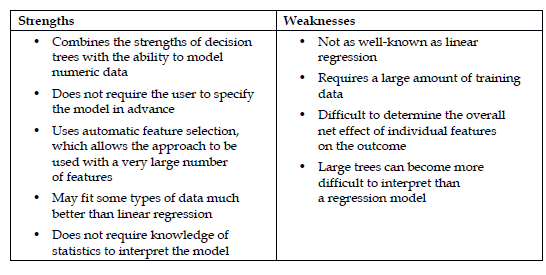
Trees for numeric prediction fall into two categories.

**Regression trees**: part of the seminal classification and regression tree (CART) algorithm. regression trees do not use linear regression methods. They make predictions based on the average value of examples that reach a leaf.

**Model trees**: Less widely known, but perhaps more powerful. The same way as regression

trees, but at each leaf, a multiple linear regression model is built from the examples

reaching that node.



Trees for numeric prediction are built in much the same way as they are for classification. Beginning at the root node, the data is partitioned using a divide and conquer strategy according to the feature that will result in the greatest increase in homogeneity in the outcome after a split is performed. In classification trees, you will recall that homogeneity is measured by entropy. This is undefined for numeric data. Instead, for numeric decision trees, homogeneity is measured by statistics such as variance, standard deviation, or absolute deviation from the mean.

