

CART Licensing Details

To accommodate different dataset sizes, CART is available in several different memory sizes.

A user's license sets a limit on the amount of learn sample data that can be analysed. The learn sample is the data used to grow the maximal tree. Note that there is no limit to the number of test sample data points that may be analysed.

For example, the 32MB version has a learn sample limitation of 8 MB. Each data point occupies 4 bytes. Therefore, a 8MB license will allow up to $8 \times 1024 \times 1024 / 4 = 2,097,152$ learn sample data points to be analysed. A data point is represented by a 1-variable by- 1-observation (1-row by- 1-column).

The following is a table that describes the current set of "sizes" available. Please note, the minimum required RAM is not the same as the learn sample limitation.

Size (MB)	Data Limit (MB)	Data Limit (number of values)
32	8	2,097,152
64	18	4,718,592
128	45	11,796,480
256	100	26,214,400
512	200	52,428,800
1024	400	104,857,600
2048	800	209,715,200

The number of variables CART can handle can be significantly increased if node sub-sampling is used when searching for the optimal split. In node sub-sampling, all the data are used to grow the tree, but only a sub-sample of the data is actually searched in the largest nodes near the top of the tree. Judiciously chosen sub-sampling can sometimes double the number of variables CART can search while growing the tree on all the data.