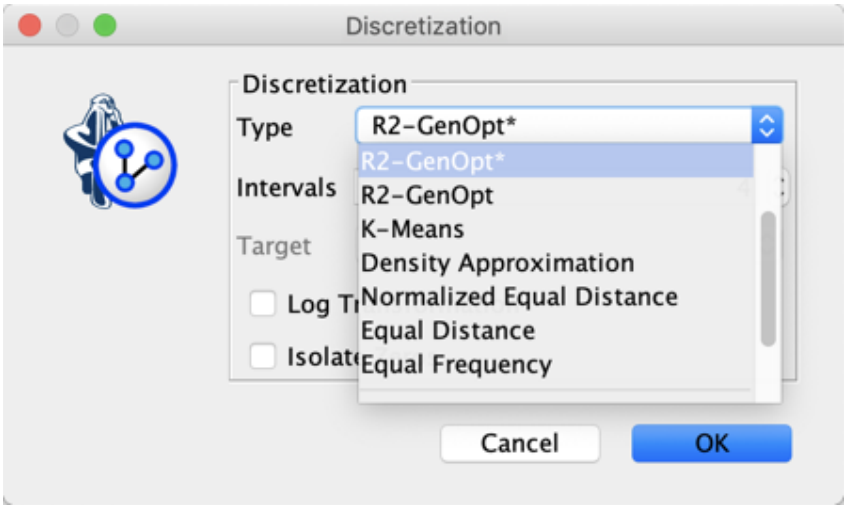
R2-GenOpt* (8.0)

Context

Data Import | Discretization and Aggregation Wizard

Learning | Discretization



R2-GenOpt is a univariate discretization algorithm that we introduced in version 6.0. This algorithm uses a genetic algorithm to find a discretization that maximizes the R^2 between the discretized variable and its corresponding continuous (hidden) variable. Therefore, it is the the optimal approach for finding an accurate representation of the continuous values of a variable.

New Feature: R2-GenOpt*

As of version 8.0, BayesiaLab offers an extended version of **R2-GenOpt** that uses a specific MDL score to choose the number of bins.

Example

100 observations
With 100 observations, even though we selected 8 bins, only 3 have been created.

Nodes 1				
	States	**Intervals**		**Discretization**
8- Wrist girth	<=15.4	13.19999981	15.39999962	Asked: R2-GenOpt* - 8 Obtained: R2-GenOpt* - 3
	<=17	15.39999962	17.0	
	>17	17.0	19.5	

1500 Observations
With 1,500 observations, even though we selected 10 bins, only 5 have been created for *AGN,* and 6 for *ALL*.

		States	**Intervals**		**Discretization**
AGN	Continuous	<=-0.04	-0.13996883	-0.03980625	Asked: R2-GenOpt* - 10 Obtained: R2-GenOpt* - 5
		<=-0.008	-0.03980625	-0.0075539132	
		<=0.01	-0.0075539132	0.0099261	
		<=0.04	0.0099261	0.0398622	
		>0.04	0.0398622	0.1258351	
		States	**Intervals**		**Discretization**
ALL	Continuous	<=-0.086	-0.23814672	-0.085690893	Asked: R2-GenOpt* - 10 Obtained: R2-GenOpt* - 6
		<=-0.032	-0.085690893	-0.032398783	
		<=-0.003	-0.032398783	-0.003224323	
		<=0.022	-0.003224323	0.0219483	
		<=0.074	0.0219483	0.0743983	
		>0.074	0.0743983	0.196468	