BigML Linear Regression Cheat Sheet

Linear Regression Configuration

Linear regression Configuration Options

Option	Description	Default	API Name
Objective field	The field you want to predict. It needs to be a numeric field.	Last valid field in dataset	objective_field
Default numeric value	Replaces missing numeric values in your dataset by the field's maximum, mean, median, minimum, or zero. If you do not activate this option or **Missing numerics** option, your instances with missing numeric values will be ignored.	Null	default_numeric_value
Weight field	Sets instance weights using the values of the given field. The selected field must be numeric and it must not contain missing values. The value in the weight field specifies the number of times that row should be replicated when including it in the model's training set	False	weight_field
Bias	Sets whether to include bias term. Boolean. Setting it to False will exclude the bias term from the solution.	True	bias

Field Coding Configuration Options

Option	Description	Default	API Name
Dummy coding	Sets one class as the control class or dummy class to compare against the rest of the classes. The dummy class will be 0 for all variables. This is the coding BigML uses by default to convert your categorical fields into numeric values, and the default dummy class will be the first by its class name in lexicographic order.	[True]	field_codings coding:dummy
Contrast coding	Allows you to set different values for the field classes instead of the 0-1 values of **One-hot** coding. The sum of all values must equal 0. Higher values for a class assume this class has more influence on the objective field than the others. A positive value indicates a positive relationship between the classs and the objective field while a negative value indicates a negative relationship. A coefficient of 0 will exclude the class from the model.	[ ]	field_codings coding:contrast
Other coding	Allows you to set different values for the field classes the same way as with contrast coding, but it provides more freedom since the values do not need to sum 0.	[ ]	field_codings coding:other

Sampling

Option	Description	Default	API Name
Rate	Sets the proportion of the dataset you want to consider between 0% and 100%.	100%	sample_rate
Range	Specifies a subset of instances from which to sample, e.g., from instance 5 to instance 1,000. The **Rate** you set will be computed over the **Range** configured.	(1, max. rows in dataset)	range
Sampling	Allows you to choose between a random sampling or a deterministic sampling. When using deterministic sampling the random-number generator will always use the same seed, producing repeatable results.	Random	seed
Replacement	Allows a single instance to be selected multiple times. Sampling without replacement ensures that each instance cannot be selected more than once.	False	replacement
Out of bag	Selects only the out-of-bag instances for the currently defined sample. If an instance is not selected as part of a sample, it is considered out of bag. It is only selectable when a sample is deterministic and the sample rate is less than 100%.	False	out_of_bag

Prediction Configuration

Default Numeric Values

Option	Description	Default	API Name
Default numeric value	Replaces missing numeric values in your dataset by the field's maximum, mean, median, minimum, or zero. If you do not activate this option and you didn't trained your linear regression with **Missing numerics**, your instances with missing numeric values will be ignored and you will not get a prediction for them.	Null	default_numeric_value

Output File Options

Option	Description	Default	API Name
Fields separator	Allows you to choose the best separator for your fields.	Comma	separator
New line	Sets the character to use as the line break in the generated csv file: "LF", "CRLF".	LF	newline
Show/hide fields	Allows you to show or hide the rest of the fields in your output file.	True	output_fields
Headers	Allows you to show or hide the names of your columns in the output file.	True	header
Prediction column name	Allows you to set the name you want for the objective field. By default BigML takes the name of the linear regression's objective field.	Objective Field Name	prediction_name
Confidence bounds	Allows you to include include two additional columns with the confidence interval and prediction interval. By default they are not included in your ouput file.	False	confidence_bounds
Confidence interval column name	Allows you to set the name you want for the confidence interval column.	confidence interval	confidence_interval_name
Prediction interval column name	Allows you to set the name you want for the prediction interval column.	prediction interval	prediction_interval_name

Output Dataset

Option	Description	Default	API Name
Output dataset	Defines whether a dataset with the results should be automatically created or not.	True	output_dataset