BigML Logistic Regression Cheat Sheet

Logistic Regression Configuration

Logistic regression Configuration Options

Option	Description	Default	API Name
Objective field	The field you want to predict. It needs to be a categorical 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
Missing numerics	Allows the logistic regression to consider missing values for the numeric fields as valid values. If you do not activate this option or set a **Default numeric value**, your instances with missing numeric values will be ignored.	True	missing_numerics
Eps	Sets the stopping criteria for the solver. If the difference between the current results and the last iteration results is smaller then **Eps**, then the solver is finished. You can set positive float values smaller than 1.	0.0001	eps
Stats	Defines whether to compute statistics to test the predictive power of the model and the coefficient estimates: likelihood ratio, standard error, Z score, *p*-value, confidence intervals.	False	compute_stats
Balance objective	Sets instance weights so that each class has equal influence on the logistic regrssion.	False	balance_objective
Objective weights	Sets a specific weight for each class of the objective field. If a class is not listed, it is assumed to have a weight of 1. Weights of 0 are also valid.	False	objective_weights
Scale bias	Scales the intercept term. Setting it to 0 will exclude the bias term from the solution. Must be greater than or equal to 0.	1	bias
Auto-scale fields	Scales numeric fields such that their standard deviations are 1, based on the field summary statistics at training time.	False	balance_fields
Regularization	Sets L1 or L2 regularization in order to avoid overfitting. L1 norm causes more coefficients to be zero, while using the L2 norm forces the magnitude of all coefficients towards zero.	L2	regularization
Strenght (c)	The inverse of the regularization strength. Higher values indicate less regularization. Must be a postive number greater than 0.	1	c

Field Coding Configuration Options

Option	Description	Default	API Name
One-hot coding	Creates numeric variables with values 0-1 per each class, plus an additional one for missing values. This is the coding BigML uses by default to convert your categorical fields into numeric values.	True	-
Dummy coding	Sets one class as the control class to compare against the rest of the classes. The control class will be 0 for all variables.	[ ]	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

Probability threshold

Option	Description	Default	API Name
Probability threshold	A percentage between 0% and 100% that is commonly used with unbalanced classification models. When the probability of the positive class is above the established threshold, the non-positive class with the highest probability is predicted instead.	Null	operating_point

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 logistic 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 logistic regression's objective field.	Objective Field Name	prediction_name
Include probability	Allows you to include an additional column with the probability of the predicted class per instance.	False	probability
Probability column name	Allows you to set the name you want for the probability column.	Probability	probability _name
All class probability	Allows you to include all class probabilities per instance.	False	probabilities

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