

این بخش مروری اجمالی بر الگوریتم‌های موجود در Analysis Services و پارامترهای قابل تنظیم و مقدار پیش فرض هر پارامتر می‌باشد، به منظور بررسی بیشتر هر یک به لینک‌های زیر مراجعه کنید:

[Data Mining Algorithms \(Analysis Services - Data Mining \(Algorithm Parameters \(SQL Server Data Mining Add-ins\)](#)

1 - Microsoft Association Rules

به منظور ایجاد قوانینی که توصیف کننده این موضوع باشد که چه مواردی احتمالاً با یکدیگر در تراکنش‌ها ظاهر می‌شوند، استفاده می‌شود.

Parameter	Default	Range
MAXIMUM_ITEMSET_COUNT	200000	(...,1]
MAXIMUM_ITEMSET_SIZE	3	[0,500]
MAXIMUM SUPPORT	1.0	(...,0.0)
MINIMUM IMPORTANCE	999999999 -	(...,...)
MINIMUM_ITEMSET_SIZE	1	[1,500]
MINIMUM PROBABILITY	0.4	[0.0,1.0]
MINIMUM SUPPORT	0.0	(...,0.0]

2 - Microsoft Clustering

به منظور شناسایی روابطی که در یک مجموعه داده ممکن است از طریق مشاهده منطقی به نظر نرسد، استفاده می‌شود. در واقع این الگوریتم با استفاده از تکنیک‌های تکرار شونده رکوردها را در خوشه‌هایی که حاوی ویژگی‌های مشابه هستند گروه بندی می‌کند.

Parameter	Default	Range
CLUSTER COUNT	10	(...,0]
CLUSTER SEED	0	(...,0]
CLUSTERING METHOD	1	1,2,3,4
MAXIMUM_INPUT_ATTRIBUTES	255	[0,65535]
MAXIMUM STATES	100	[2,65535],0
MINIMUM SUPPORT	1	(...,0)
MODELLING_CARDINALITY	10	[1,50]
SAMPLE SIZE	50000	(...,100], 0
STOPPING TOLERANCE	10	(...,0)

3 - Microsoft Decision Trees

مبتنی بر روابط بین ستون‌های یک مجموعه داده ای باعث پیش بینی روابط مدل‌ها می‌شود، که به صورت یک سری درختوار ویژگی‌ها در آن شکسته می‌شوند. به منظور انجام پیش بینی از هر دو ویژگی پیوسته و گسسته پشتیبانی می‌شود.

Parameter	Default	Range
COMPLEXITY_PENALTY		(0.0,1.0)
FORCE REGRESSOR		
MAXIMUM_INPUT_ATTRIBUTES	255	[0,65535]
MAXIMUM_OUTPUT_ATTRIBUTES	255	[0,65535]
MINIMUM SUPPORT	10.0	(...,0.0)
SCORE METHOD	4	1,3,4
SPLIT METHOD	3	[1,3]

Microsoft Linear Regression - 4

چنانچه یک وابستگی خطی میان متغیر هدف و متغیرهای مورد بررسی وجود داشته باشد، کارآمدترین رابطه میان متغیر هدف و ورودی ها را پیدا می کند. به منظور انجام پیش بینی از ویژگی پیوسته پشتیبانی می کند.

Parameter	Default	Range
FORCE REGRESSOR		
MAXIMUM_INPUT_ATTRIBUTES	255	[0,65535]
MAXIMUM_OUTPUT_ATTRIBUTES	255	[0,65535]

Microsoft Logistic Regression - 5

به منظور تجزیه و تحلیل عواملی که در یک تصمیم گیری مشارکت دارند که پی آمد آن به وقوع یا عدم وقوع یک رویداد می انجامد از این الگوریتم استفاده می شود. جهت انجام پیش بینی از هر دو ویژگی پیوسته و گسسته پشتیبانی می کند.

Parameter	Default	Range
HOLDOUT_PERCENTAGE	30	(0,100)
HOLDOUT_SEED	0	(...,...)
MAXIMUM_INPUT_ATTRIBUTES	255	[0,65535]
MAXIMUM_OUTPUT_ATTRIBUTES	255	[0,65535]
MAXIMUM_STATES	100	[2,65535], 0
SAMPLE_SIZE	10000	(...,0]

Microsoft Naïve Bayes - 6

احتمال ارتباط میان تمامی ستون های ورودی و ستون های قابل پیش بینی را پیدا می کند. همچنین این الگوریتم برای تولید سریع مدل کاوش به منظور کشف ارتباطات بسیار سودمند می باشد. تنها از ویژگی های گسسته یا گسسته شده پشتیبانی می کند و با تمامی ویژگی های ورودی به شکل مستقل رفتار می کند.

Parameter	Default	Range
MAXIMUM_INPUT_ATTRIBUTES	255	[0,65535]
MAXIMUM_OUTPUT_ATTRIBUTES	255	[0,65535]
MAXIMUM_STATES	100	[2,65535], 0
MINIMUM_DEPENDENCY_PROBABILITY	0.5	(0,1)

Microsoft Neural Network - 7

به منظور تجزیه و تحلیل داده‌های ورودی پیچیده یا مسائل بیزنسی که برای آنها مقدار قابل توجهی داده آموزشی در دسترس می‌باشد اما به آسانی نمی‌توان با استفاده از الگوریتم‌های دیگر این قوانین را بدست آورد، استفاده می‌شود. با استفاده از این الگوریتم می‌توان چندین ویژگی را پیش بینی نمود. همچنین این الگوریتم می‌تواند به منظور طبقه بندی برای ویژگی‌های گسسته و ویژگی‌های پیوسته رگرسیون مورد استفاده قرار گیرد.

Parameter	Default	Range
HIDDEN_NODE_RATIO	4.0	(...,0]
HOLDOUT_PERCENTAGE	30	(0,100)
HOLDOUT_SEED	0	(...,...)
MAXIMUM_INPUT_ATTRIBUTES	255	[0,65535]
MAXIMUM_OUTPUT_ATTRIBUTES	255	[0,65535]
MAXIMUM_STATES	100	[2,65535], 0
SAMPLE_SIZE	10000	(...,0]

Microsoft Sequence Clustering - 8

به منظور شناسایی ترتیب رخداد‌های مشابه در یک دنباله استفاده می‌شود. در واقع این الگوریتم ترکیبی از تجزیه تحلیل توالی و خوشه را فراهم می‌کند.

Parameter	Default	Range
CLUSTER_COUNT	10	(...,0]
MAXIMUM_SEQUENCE_STATES	64	[2,65535], 0
MAXIMUM_STATES	100	[2,65535], 0
MINIMUM_SUPPORT	10	(...,0]

Microsoft Time Series - 9

به منظور تجزیه و تحلیل داده‌های زمانی (داده‌های مرتبط با زمان) در یک درخت تصمیم گیری خطی استفاده می‌شود. الگوهای کشف شده می‌توانند به منظور پیش بینی مقادیر آینده در سری‌های زمانی استفاده شوند.

Parameter	Default	Range
AUTO_DETECT_PERIODICITY	0.6	[0.0,1.0]
COMPLEXITY_PENALTY	0.1	(1.0,...)
FORECAST_METHOD	MIXED	ARIMA,ARTXP,MIXED

Parameter	Default	Range
HISTORIC_MODEL_COUNT	1	[0,100]
HISTORIC_MODEL_GAP	10	(...,1]
INSTABILITY_SENSITIVITY	1.0	[0.0,1.0]
MAXIMUM_SERIES_VALUE	1E308 +	[...,column maximum]
MINIMUM_SERIES_VALUE	1E308 -	[column minimum,...]
MINIMUM_SUPPORT	10	(...,1]
MISSING_VALUE_SUBSTITUTION	None	None,Previous,Mean
PERIODICITY_HINT	{1}	{...list of integers...}
PREDICTION_SMOOTHING	0.5	[0.0,1.0]