纵向线性回归算法参数介绍文档

|  |  |  |  |
| --- | --- | --- | --- |
| 参数名 | 参数含义 | 默认值 | 可选值 |
| penalty | 正则化，防止模型过拟合。通过在损失函数后加L1或者L2范式的倍数实现。这个增加的范式，被称为“正则项”，也被称为"惩罚项"。 | L2 | L1,L2 |
| tol | 残差收敛条件,即容忍出现错误的概率，；误差项达到指定值时则停止训练。 | 1e-4 | 大于0小于1的实数 |
| alpha | 正则化强度系数，当alpha越大的时候，分类器的准确性越高，但同样容错率会越低，泛化能力会变差。 | 1 | 大于0任意实数 |
| optimizer | 优化器方法 | sgd | 'sgd', 'rmsprop', 'adam', 'sqn', or 'adagrad'字符串类型 |
| batch\_size | 更新模型时的批次大小 | -1 | 小于数据量大小大于0的任意整数，为-1即一个批次使用所有数据。 |
| learning\_rate | 学习率，其决定着目标函数能否收敛到局部最小值以及何时收敛到最小值。 | 0.01 | 10e-6到1之间 |
| max\_iter | 最大迭代次数 | 20 | 大于0任意整数 |
| early\_stop | 用于判断收敛与否的方法  a) diff：用两次迭代的损失之差来判断是否收敛。  b) abs。用损失的绝对值来判断是否收敛，即如果损失<tol，则为收敛。  c) weight\_diff。使用两个连续迭代的权重之差。 | diff | weight\_diff，abs，diff纵向回归只支持weight\_diff。 |
| decay | 学习率的衰减率 | 1 | 小于1大于0的实数 |
| decay\_sqrt |  | True | True/False |
| validation\_freqs | 训练期间的验证频率 | 默认None | 任意值 |
| early\_stopping\_rounds | 在指定的训练轮次之后一个指标没有改善，将停止训练。 | 默认None | 任意正整数值 |
| metrics | 指示在训练过程中执行评估时，哪些指标将被使用。如果无设置，则为默认度量。对于二元分类默认度量为AUC和KS。 | 默认为None | 指标名或者指标列表 |
| use\_first\_metric\_only | 是否只使用第一个指标来进行early\_stopping判断。 | False | 为bool型变量，可以设置为True |
| init\_param | 初始化相关参数 |  |  |
| init\_param.init\_method | init\_method:初始化模式 | 'random\_uniform' | 'random\_normal', 'one', 'zeros' or 'const' |
| init\_param.init\_const | init\_const:当init\_method为'const'时需要指定常数。 | 1 | 任意正整数 |
| init\_param.fit\_intercept | fit\_intercept:是否初始化截距 | True | False |
| encrypt\_param | 加密相关参数 |  |  |
| encrypt\_param.method | 加密使用的方法 | Paillier | None,IterativeAffine，Random\_IterativeAffine |
| encrypt\_param.key\_length | 指定加密方法种的密钥长度 | 1024 | 任意正整数 |
| predict\_param | 预测相关参数 |  |  |
| predict\_param.threshold | threshold:用于区分正负样本的阈值 | 0.5 | 0到1之间的任意值 |
| cv\_param | 交叉验证相关参数 |  |  |
| cv\_param.n\_splits | n\_splits:指定在KFold中使用多少个分割点 | 5 | 大于0任意正整数 |
| cv\_param.mode | mode:指明当前的任务模式 | Hetero | Homo |
| cv\_param.role | role:在当前任务中扮演什么角色 | Guest | Host,Arbiter |
| cv\_param.shuffle | shuffle:在KFold之前是否打乱顺序 | True | False |
| cv\_param.random\_seed | random\_seed:随机种子 | 1 | 任意整数 |
| cv\_param.need\_cv | need\_cv:是否要运行该模块 | False | True |
| cv\_param.output\_fold\_history | output\_fold\_history:是否输出每个fold使用的每个id表 | True | False |
| cv\_param.history\_value\_type | history\_value\_type:指明在output\_fold\_history中包括原始实例或预测分数 | score | instance |
| encrypted\_mode\_calculator\_param | 加密模式计算参数 |  |  |
| encrypted\_mode\_calculator\_param.mode | 可选模型 | 'strict', 'fast', 'balance', 'confusion\_opt' | strict |
| encrypted\_mode\_calculator\_param.re\_encrypted\_rate | 一个数字，当使用blance时有用 | [0,1] | 1 |
| sqn\_param | 设置优化器为sqn时可设置的参数 |  |  |
| sqn\_param.update\_interval\_L | update\_interval\_L:设置更新海森矩阵的迭代次数 | 3 | 大于0任意整数 |
| sqn\_param.memory\_M | 曲率信息的堆栈大小 | 5 | 大于0任意整数 |
| sqn\_param.sample\_size | sample\_size:用于更新海森矩阵的数据的样本大小 | 5000 | 大于0任意整数 |
| stepwise\_param | 逐步回归相关参数 |  |  |
| stepwise\_param.score\_name | score\_name:指定要使用的模型选择标准 | 'AIC' | 'BIC' |
| stepwise\_param.mode | mode:指明当前的任务模式 | Hetero | Homo |
| stepwise\_param.role | role:在当前任务中扮演什么角色 | Guest | Host,Arbiter |
| stepwise\_param.direction | direction:分步走的方向。 | 'both' | 'forward' , 'backward' |
| stepwise\_param.max\_step | max\_step:停止前运行的总步数 | 10 | 大于0任意正整数 |
| stepwise\_param.nvmin | nvmin:最终模型的最小子集大小 | 2 | 大于2任意正整数 |
| stepwise\_param.nvmax | nvmax:最终模型的最大子集大小 | 无 | 大于nvmin的整数 |
| stepwise\_param.need\_stepwise | need\_stepwise:该模块是否需要运行 | False | True |