xgboost分类

> *xgboost是一种高效率boosting算法，适用回归和二分类问题，详见 https://github.com/dmlc/xgboost 。 PAI平台目前只开放树形模型（gbtree），暂不开放线性模型（gblinear），其他参数见下表；原始参数说明 https://github.com/dmlc/xgboost/blob/master/doc/parameter.md 。 PAI平台支持稠密和稀疏两种数据格式作为输入。*

PAI命令行 参数说明 具体示例 常见问题

PAI命令行

```bash pai -name xgboost -DinputTableName=wpbc -DfeatureColNames=f1,f2,f3 -DlabelColName=label -Dobjective=binary:logistic -DmodelName=algo_adult_binary_model;
```

参数说明

参数key名称	参数描述	参数value可选项	默认值
inputTableName	输入表的表名	-	-
featureColNames	输入表中用于训练的特征的列名	-	-
labelColName	输入表中标签列的列名	-	-
modelName	输出的模型名	-	-
inputPartitions	输入表中指定哪些分区参与训练，格式为: partition_name=value。如果是多级格式为name1=value1/name2=value2；如果是指定多个分区，中间用','分开	-	输入表的所有partition
enableSparse	是否稀疏数据	true，false	false
itemDelimiter	item(key value对)之间的分隔符	冒号、空格、逗号	空格
kvDelimiter	表中每个item的key和value之间的分隔符	冒号、空格、逗号	冒号
eta	为了防止过拟合，更新过程中用到的收缩步长。在每次提升计算之后，算法会直接获得新特征的权重。 eta通	[0-1]	0.3

	过缩减特征的权重使提升计算过程更加保守		
gamma	最小损失衰减	[0,无穷]	0
max_depth	树的最大深度	[1,无穷]	6
min_child_weight	孩子节点中最小的样本权重和。如果一个叶子节点的样本权重和小于min_child_weight则拆分过程结束。在线性回归模型中，这个参数是指建立每个模型所需的最小样本数	[0,无穷]	1
max_delta_step	每个树所允许的最大delta步进	[0,无穷]	0
subsample	用于训练模型的字样本占整个样本集合的比例	(0,1]	1
colsample_bytree	在建立树时对特征采样的比例	(0,1]	1
objective	定义学习任务及相应的学习目标，可选的目标函数	reg:linear reg:logistic binary:logistic	reg:linear
base_score	初始的predict score	-	0.5
seed	随机数的种子	[0,无穷]	0
num_round	树的棵树	[1,无穷]	10

具体示例

常见问题

> *输入特征和目标特征都必须是int和double类型*