booster: gbtree 树模型做为基分类器(默认),gbliner 线性模型做为基分类器silent

silent=0时,不输出中间过程(默认);silent=1时,输 出中间过程

nthread: nthread=-1时、使用全部CPU进行并行运算( 默认), nthread=1时,使用1个CPU进行运算。

scale pos weight:正样本的权重,在二分类任务中,当 正负样本比例失衡时,设置正样本的权重,模型效果更 好。例如、当正负样本比例为1:10时、scale pos weight=10。

n estimatores:总共迭代的次数,即决策树的个数。 early\_stopping\_rounds:在验证集上,当连续n次迭 代,分数没有提高后,提前终止训练,防止overfitting。 max\_depth:树的深度,默认值为6,典型值3-10。值越 大, 越容易过拟合; 值越小, 越容易欠拟合。 min\_child\_weight:默认值为1,值越大,越容易欠拟合; 值越小, 越容易过拟合(值较大时, 避免模型学习到局部 的特殊样本)。

subsample:训练每棵树时,使用的数据占全部训练集的 比例。默认值为1,典型值为0.5-1,防止overfitting。 colsample\_bytree:训练每棵树时,使用的特征占全部特 征的比例。默认值为1,典型值为0.5-1,防止 overfitting.

learning\_rate:学习率,控制每次迭代更新权重时的步 长,默认0.3,值越小,训练越慢。典型值为0.01-0.2。 objective 目标函数:回归reg:linear (默认)\logistic ;二分 类 binary:logistic概率\logitraw类别;多分类 multi:softmax num class=n返回类别\soft prob num\_class=n返回概率;rank:pairwise\ eval\_metric; 回归任务(默认rmse):rmse--均方根误差/mae--平均绝 对误差:分类任务(默认error): auc--roc曲线下面积\ error--错误率(二分类)\ merror--错误率(多分类\ logloss--负对数似然函数(二分类)\mlogloss--负对数似 然函数(多分类)。

Gamma:惩罚项系数,指定节点分裂所需的最小损失函数 下降值

> alpha: L1正则化系数,默认为1 lambd: L2正则化系数,默认为1

思想是将许多弱分类器集成在一起形成一个强分类器。 XGBoost是一种将许多树模型集成在一起、形成一个很 强的分类器。而所用到的树模型则是CART回归树模型。

XGBoost是boosting算法的其中一种。Boosting算法的

不断地添加树,每添加一个树,其实是学习一个新函数, 去拟合上次预测的残差。

当样本的第i个特征值缺失时,无法利用该特征进行划分 时、XGBoost的想法是将该样本分别划分到左结点和右 结点,然后计算其增益,哪个大就划分到哪边。

支持并行化,虽然树与树之间是串行关系,但是同层级节 点可并行

Xgboost

Xgboost.XGBRegressor/xgboost.

XGBClassifier /xgboost.XGBRanker

总概