

数据挖掘-评估

概述

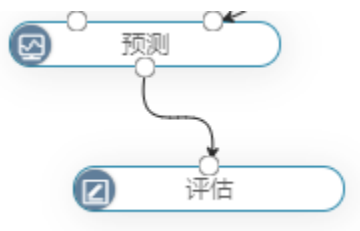
评估节点是对分类算法模型和回归模型的预测效果进行评估，检验模型在分类任务中的表现或者检验其在回归任务中的可靠性。

“评估”节点的前置节点必须是“预测”。

输入/输出

输入	只有一个输入端口，用于接收预测结果。
输出	没有输出端口。

示例



点击右键可以查看评估结果。针对分类与回归及无监督算法提供不同的评估指标。

分类预测模型评价指标如下图：

指标	值
confusion matrix(混淆矩阵)	真实\预测
	0
	1
	0
	1
accuracy(准确率)	0.9375651063543802
roc曲线	查看ROC曲线
auc	0.9713710162860738
ks	查看KS曲线
weighted precision(加权精确率)	0.9376899750459747
weighted recall(加权召回率)	0.9375651063543802
weighted F1 score(加权F1分数)	0.9375592142096183
Class 0.0 precision(精确率)	0.9450233068275294
Class 0.0 recall(召回率)	0.9288505592238243
Class 0.0 F1 score(F1分数)	0.9368671423717295
Class 0.0 falsePositiveRate(假阳率)	0.05376416169471073
Class 0.0 truePositiveRate(真阳率)	0.9288505592238243
Class 1.0 precision(精确率)	0.9303935139410718
Class 1.0 recall(召回率)	0.9462358383052892
Class 1.0 F1 score(F1分数)	0.9382478064344589
Class 1.0 falsePositiveRate(假阳率)	0.07114944077617572
Class 1.0 truePositiveRate(真阳率)	0.9462358383052892

加权F1分数：准确率、精确率、召回率综合查看结果，越大越好。

假阳率指标越小越好。

混淆矩阵：对角线数值代表模型预测样本正确的个数，越大越好。

ROC曲线：图形下方面积越大越好。

回归预测模型评价指标如下图：

指标	值
rmse(均方根误差)	0.621088252099364
mse(均方误差)	0.3857506168958432
r2(决定系数)	0.7161742587278249
mae(平均绝对误差)	0.4574047278345649
adjusted r2(校正决定系数)	0.7158958162085913

MSE（均方误差）：越小，模型预测效果越好。

RMSE(均分根误差)：越小，模型预测效果越好。

MAE(平均绝对误差)：与MSE相比，MAE指标可以减少离群点对模型指标的影响，值越小，模型预测效果越好。

R2（决定系数）：考虑预测样本方差的差异情况，取值范围为（负无穷，1]。R2越接近1代表模型预测效果越好；当R2为负值时，表明模型预测效果差；当模型预测值接近样本期望值时，R2接近0。

adjusted R2（修正R2）：修正R2取值范围是[0,1]，指标越接近1，代表模型预测效果越好。

聚类分析效果如下图：

会增加3列

特征列、
值

标准化特征列、

预测

features	featuresNormalized	prediction
[5.772814706,79.06214...	[5.772814706,79.06214338,12.79404831,...	0
[0.186630472,50.87378...	[0.186630472,50.87378641,5.001696641,...	2
[0.005277992,91.28264...	[0.005277992,91.28264402,7.277295135,...	0
[-14.97692308,56.70083...	[-14.97692308,56.70083523,13.48846154,...	2
[0.876273654,26.75022...	[0.876273654,26.75022155,13.28384279,...	1
[4.115395179,12.81268...	[4.115395179,12.81268468,25.93446905,...	1
[22.22271445,17.54116...	[22.22271445,17.54116356,53.69246445,...	1
[3.126204868,55.08163...	[3.126204868,55.08163728,12.5459244,4,...	2
[0.755752602,51.14885...	[0.755752602,51.14885298,13.56636212,...	2
[-0.329290558,79.54520...	[-0.329290558,79.54520265,28.02308241,...	0
[2.359645376,9.274966...	[2.359645376,9.274966302,13.69721024,...	1

聚类预测效果

一般通过高维数据可视化查看

