

使用sklearn估计器构建分 类模型

sklearn库常用分类算法函数

模块名称	函数名称	算法名称
linear_model	LogisticRegression	逻辑回归
svm	SVC	支持向量机
neighbors	KNeighborsClassifier	K最近邻分类
naïve_bayes	GaussianNB	高斯朴素贝叶斯
tree	DecisionTreeClassifier	分类决策树
ensemble	RandomForestClassifier	随机森林分类
ensemble	GradientBoostingClassifier	梯度提升分类树

分类模型的评价标准

- 分类模型对测试集进行预测而得到的准确率并不能很好的反映模型的性能，为了有效判断一个预测模型的性能表现，需要结合真实值，计算出准确率、召回率、F1值和Cohen's Kappa系数等指标来衡量。
- sklearn的metrics模块还提供了一个能够输出分类模型评价报告的函数 `classification_report`.

方法名称	最佳值	sklearn函数
Precision (精确值)	1.0	<code>metrics.precision_score</code>
Recall (召回值)	1.0	<code>metrics.recall_score</code>
F1值	1.0	<code>metrics.f1_score</code>
Cohen's Kappa系数	1.0	<code>metrics.cohen_kappa_score</code>
ROC曲线	最靠近y轴	<code>metrics.roc_curve</code>