

图 275

free	wifi	e_pay	package_type	predict_label
0	0	0	1	0
0	0	0	3	0
19.25	0	0	3	0
28.5	0	0	3	0
0	0	0	2	0
0	0	1	1	1
22	0	0	3	0
0	0	0	1	0

图 276

3.4.5.2 SVM

图标:  SVM

描述: SVM (支持向量机) 方法是通过一个非线性映射, 把样本空间映射到一个高维乃至无穷维的特征空间中使得在原来的样本空间中非线性可分的问题转化为在特征空间中的线性可分的问题。

字段属性

特征列: 通过勾选的方式选择特征所在列。

标签列：选择分类标签所在的列，仅支持字符型数据。

参数设置

罚项系数：浮点型，默认 1.0。

核函数：支持线性核、多项式核、高斯核、sigmoid，默认 rbf。

输出

表结果：SVM 算法分类结果。

报告：Confusion Matrix、Receiver Operating Characteristic(ROC)、Precision-Recall

示例

下列对某数据进行 SVM 算法分类：

- 选择特征数列，标签列。如图 277 所示。
- 保留默认参数，罚项系数为 1.0，核函数为高斯核，如图 278 所示。
- SVM 运行成功后，可选择查看结果，如图 279 所示。
- SVM 运行成功后，可选择查看报告，如图 280 所示。
- 模型评估配置如图 281 所示。
- 模型评估运行成功后，选择查看数据，如图 282 所示。
- 模型评估报告运行成功后，选择查看报告，如图 283 所示。
- 模型预测配置如图 284 所示。
- 模型预测运行成功后，选择查看数据，如图 285 所示。



图 277

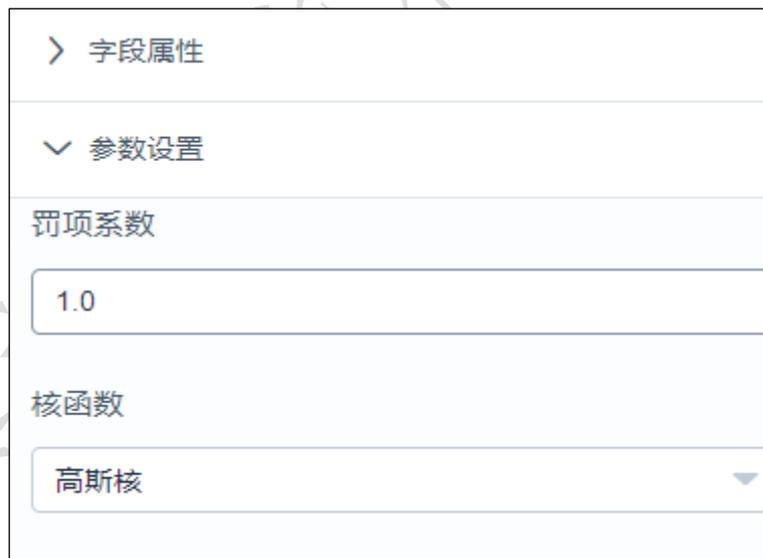


图 278

预览数据

e_pay	package_type	leave	educational_level	predict_label
1	2	1	3	3
1	4	1	3	3
0	3	1	3	3
0	3	0	4	4
0	3	0	1	1
1	1	0	3	3
1	2	1	3	3
0	2	0	3	3

共 597 条 25 条/页 < 1 2 3 4 5 6 ... 24 > 前往 1 页

图 279



图 280



图 281

预览数据

e_pay	package_type	leave	educational_level	predict_label
0	3	0	1	2
0	1	0	3	2
1	2	1	3	4
0	4	0	5	2
0	3	0	1	2
1	1	0	1	2
1	4	0	3	2
1	2	0	2	2

共 199 条 25 条/页 < 1 2 3 4 5 6 ... 8 > 前往 1 页

图 282

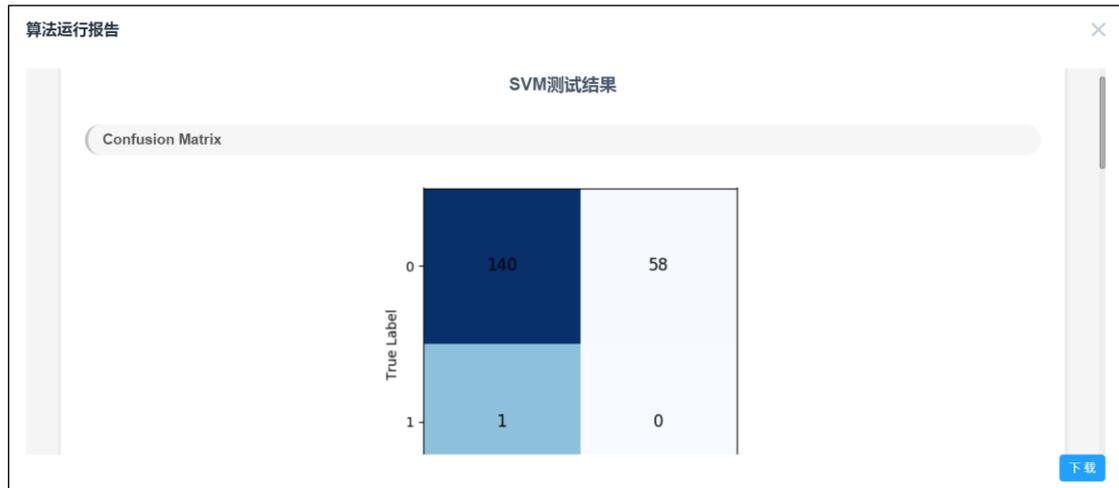


图 283



图 284

预览数据

wifi	e_pay	package_type	leave	predict_label
0	0	1	1	4
0	0	3	0	2
0	0	3	0	2
0	0	3	0	4
0	0	2	0	1
0	1	1	1	4
0	0	3	0	2
0	0	1	0	2

共 796 条 25 条/页 < 1 2 3 4 5 6 ... 32 > 前往 1 页

图 285

3.4.5.3 KNN

图标:



描述: KNN 算法的核心思想是如果一个样本在特征空间中的 k 个最相邻的样本中的大多数属于某一个类别, 则该样本也属于这个类别, 并具有这个类别上样本的特性。

字段属性

特征列: 通过勾选的方式选择特征所在列。

标签列: 选择分类标签所在的列, 一般为字符型数据。

参数设置:

最近邻个数 K : 整数型, 通常不大于 20, 默认 5。

投票权重类型: 权重相等或权重与距离成反比, 默认权重相等。

计算最近邻的算法: 包括 自动、BallTree、KDTree、暴力搜索法, 默认自动。

输出

表结果: KNN 算法分类结果。

报告: Confusion Matrix、Receiver Operating Characteristic(ROC)、Precision-Recall

示例

下列对某数据进行 KNN 算法分类:

- 选择特征数列, 标签列, 如图 286 所示。