

3.4.4 回归

3.4.4.1 线性回归

图标: 

描述: 线性回归是利用数理统计中回归分析, 来确定两种或两种以上变量间相互依赖的定量关系的一种统计分析方法。

字段属性

特征列: 请选择数值型数据, 如果勾选了非数值类型数据, 则会自动过滤, 下个组件可能无法获取所有列。

标签列: 请选择数值型数据。

输出

表结果: 线性回归预测结果。

报告: 模型拟合效果。

示例

下列对某数据进行线性回归算法:

- 选择自变量, 因变量, 均选择数值型数据。如图 205 所示。
- 运行成功后, 选择查看数据, 如图 206 所示。
- 运行成功后, 选择查看报告, 如图 207 所示。

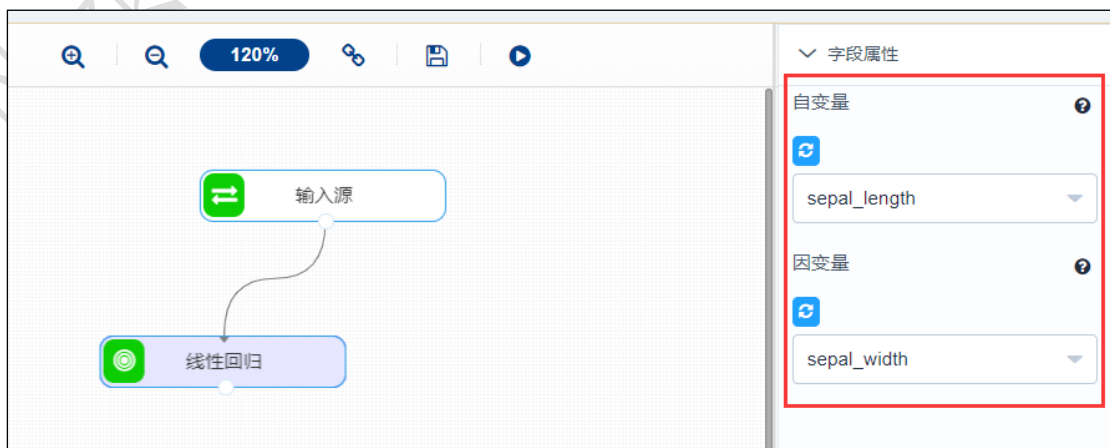


图 205

sepal_length	sepal_width	predict_label
5.1	3.5	3.103
4.9	3	3.116
4.7	3.2	3.128
4.6	3.1	3.134
5	3.6	3.11
5.4	3.9	3.085
4.6	3.4	3.134
5	3.4	3.11

共 150 条 25 条/页 < 1 2 3 4 5 6 > 前往 1 页

图 206

算法运行报告

线性回归模型评价

模型拟合效果

Coefficients(系数): [-0.0618848]
intercept(截距项): 3.418946836103816
Score(R²): 0.013822654141080748

图 207

3.4.4.2 广义最小二乘法

图标: 广义最小二乘法

描述: 广义最小二乘法是一种常见的消除异方差的方法。它的主要思想是为解释变量加上一个权重, 从而使得加上权重后的回归方程方差是相同的。因此在 GLS 方法下我们可以得到估计量的无偏和一致估计。

字段属性:

自变量: 选择自变量所在列, 请选择数值型数据。

因变量：选择响应变量所在的列，请选择数值型数据。

输出

表结果：无。

报告：GLS Regression Results。

示例

下列对某数据使用广义最小二乘法：

- 选择自变量，因变量，均选择数值型数据。如图 208 所示。
- 运行成功后，选择查看报告，如图 209、图 210 所示。



图 208

算法运行报告			
广义最小二乘法结果			
结果			
GLS Regression Results			
=====			
Dep. Variable:	sepal_width	R-squared:	0.957
Model:	GLS	Adj. R-squared:	0.956
Method:	Least Squares	F-statistic:	3277.
Date:	Wed, 07 Mar 2018	Prob (F-statistic):	2.42e-103
Time:	15:43:20	Log-Likelihood:	-146.83
No. Observations:	150	AIC:	295.7
Df Residuals:	149	BIC:	298.7
Df Model:	1		
Covariance Type:	nonrobust		

图 209

算法运行报告						
Df Model: 1						
Covariance Type: nonrobust						
=====						
	coef	std err	t	P> t	[0.025	0.975]

sepal_length	0.5118	0.009	57.246	0.000	0.494	0.529
=====						
Omnibus:	17.098	Durbin-Watson:	0.433			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	7.933			
Skew:	0.352	Prob(JB):	0.0189			
Kurtosis:	2.121	Cond. No.	1.00			
=====						
Warnings:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						

图 210

3.4.4.3 保序回归