市财政收入分析预测

1 项目背景

在我国现行的分税制财政管理体制下,地方财政收入不仅是国家财政收入的重要组成部 分,而且具有其相对独立的构成内容。如何有效的利用地方财政收入,合理的分配,来促进 地方的发展,提高市民的收入和生活质量是每个地方政府需要考虑的首要问题。因此,对地 方财政收人进行预测,不仅是必要的,而且也是可能的。科学、合理地预测地方财政收人, 对于克服年度地方预算收支规模确定的随意性和盲目性,正确处理地方财政与经济的相互关 系具有十分重要的意义。

2 实验目标

对市财政收入进行分析,挖掘其中隐藏的运行模式,并对未来两年的财政收入进行预测, 希望能够帮助政府合理的控制财政收支,优化财源建设,为制定相关决策提供依据。

3 项目步骤

3.1 工程前期准备

3.1.1 导入数据

1. 介绍影响市财政收入的因素数据

影响市财政收入的因素数据介绍如图 1 所示。

	A	В	С	D	E	F	G	H	I	J	K	L	M	N
1	x1	x 2	хЗ	x4	x5	x6	x7	x8	x 9	x10	x11	x12	x13	y
2	3831732	181.54	448.19	7571	6212.7	6370241	525.71	985.31	60.62	65.66	120	1.029	5321	64.87
3	3913824	214.63	549.97	9038.16	7601.73	6467115	618.25	1259.2	73.46	95.46	113.5	1.051	6529	99.75
4	3928907	239.56	686.44	9905.31	8092.82	6560508	638.94	1468.06	81.16	81.16	108.2	1.064	7008	88.11
5	4282130	261.58	802.59	10444.6	8767.98	6664862	656.58	1678.12	85.72	91.7	102.2	1.092	7694	106.07
6	4453911	283.14	904.57	11255.7	9422.33	6741400	758.83	1893.52	88.88	114.61	97.7	1.2	8027	137.32
7	4548852	308.58	1000.69	12018.52	9751.44	6850024	878.26	2139.18	92.85	152.78	98.5	1.198	8549	188.14
8	4962579	348.09	1121.13	13966.53	11349.47	7006896	923.67	2492.74	94.37	170.62	102.8	1.348	9566	219.91
9	5029338	387.81	1248.29	14694	11467.35	7125979	978.21	2841.65	97.28	214.53	98.9	1.467	10473	271.91
10	5070216	453.49	1370.68	13380.47	10671.78	7206229	1009.24	3203.96	103.07	202.18	97.6	1.56	11469	269.1
11	5210706	533.55	1494.27	15002.59	11570.58	7251888	1175.17	3758.62	109.91	222.51	100.1	1.456	12360	300.55
12	5407087	598.33	1677.77	16884.16	13120.83	7376720	1348.93	4450.55	117.15	249.01	101.7	1.424	14174	338.45
13	5744550	665.32	1905.84	18287.24	14468.24	7505322	1519.16	5154.23	130.22	303.41	101.5	1.456	16394	408.86
14	5994973	738.97	2199.14	19850.66	15444.93	7607220	1696.38	6081.86	128.51	356.99	102.3	1.438	17881	476.72
15	6236312	877.07	2624.24	22469.22	18951.32	7734787	1863.34	7140.32	149.87	429.36	103.4	1.474	20058	838.99
16	6529045	1005.37	3187.39	25316.72	20835.95	7841695	2105.54	8287.38	169.19	508.84	105.9	1.515	22114	843.14
17	6791495	1118.03	3615.77	27609.59	22820.89	7946154	2659.85	9138.21	172.28	557.74	97.5	1.633	24190	1107.67
18	7110695	1304.48	4476.38	30658.49	25011.61	8061370	3263.57	10748.28	188.57	664.06	103.2	1.638	29549	1399.16
19	7431755	1700.87	5243.03	34438.08	28209.74	8145797	3412.21	12423.44	204.54	710.66	105.5	1.67	34214	1535.14
20	7512997	1969.51	5977.27	38053.52	30490.44	8222969	3758.39	13551.21	213.76	760.49	103	1.825	37934	1579.68
21	7599295	2110.78	6882.85	42049.14	33156.83	8323096	4454.55	15420.14	228.46	852.56	102.6	1.906	41972	2088.14

图 1 影响市财政收入的因素数据

由于 1994 年我国对财政体制进行了重大改革,开始实行分税制财政体制,影响了财政 收入相关数据的连续性,在 1994 年前后不具有可比性。由于没有合适的方法来调整这种数 据的跃变,因此本实验仅对 1994 年及其以后的数据进行分析(本项目所用数据均来自《统 计年鉴》)。其各属性及说明如表 1 所示。

属性名称	属性说明
x1	社会从业人数
x2	在岗职工工资总额
x3	社会消费品零售总额
x4	城镇居民人均可支配收入
x5	城镇居民人均消费性支出
x6	年末总人口
x7	全社会固定资产投资额
x8	地区生产总值
x9	第一产业产值
x10	税收
x11	居民消费价格指数
x12	第三产业与第二产业产值比
x13	居民消费水平
у	市财政收入

表 1 影响市财政收入的因素数据属性及其说明

2. 上传数据到大数据分析平台

在新增数据源上,选择本地上传数据,如图 2 所示。

	Python数据挖掘建模平台	,						¢	0	
✔ 首页	我的数据源 共享数据源									
数据源	+ 新増数据源 ▲			请输入表名	选择状态 🔻	请选择创建	前间		搜索	菜
工程	▶ 数据来源于文件	创建人	数据来源	同步状态	创建时间		操作			
⑦ 个人组件	■ 数据来源于数据库									
₩										
日子				暂无数据						

图 2 本地上传数据源

在本地路径上选择文件,填写在平台新建的目标表名,如图 3 所示。

		新建数据源				
1 文件属性 上传文件	1. tinancial_data.csv 等待上传 删除	2 预览数据				3 字段设置
* 新建目标表名	299-2011 2. financial_data					
列分隔符	逗号 (,) 🔻		文件编码	UTF-8	•	
存储有效期 (天)	60 — +		预览设置	分页显示	-	
						重置 下—步

图 3 本地选择文件上传

				新建	赵据	源					
				z	2	102					3
清:【字段名】! 原字段	只能是以字母开头,由小 字段名	写英文字母、数字、下 类型	划线组成	r. K	度	NA	*	書度		备注	子段议员
x1	x1	数值	•	255	-	+	2	-	+		
x2	x2	数值	•	255	-	+	2	-	+		
x3	x3	数值	•	255	-	+	2	-	+		
x4	x4	数值	•	255	-	+	2	-	+		
x5	x5	数值	•	255	-	+	2	-	+		
x6	x6	数值	•	255	-	+	2	-	+		
XO	10	\$218	•	200		+	2		+		上一步

根据文件的数据,可以修改文件的字段名和类型,如图 4 所示。

图 4 字段设置

上传成功,可以在平台的数据源上查看 financial_data 的数据,单击数据源操作的查看 按钮如图 5 所示,平台上 financial_data 影响市财政收入的因素数据预览,如图 6 所示。

	Python数据挖掘建模型	平台						•	0	
▲ 首页	我的数据源 共享数据源	現的数据原 共享政策原								
数据源	+ 新增数据源 ▼			请输入表名	选择状态 🔻 请选	译创建时间		©	搜索	
「上程	表名	创建人	数据来源	同步状态	创建时间	操作				
	financial_data	teacher	结构化文件	同步完成	2018-12-14 10:14:24	۵				

图 5 单击预览数据按钮

	预览数据(分页加载) ×										
x1	x2	xЗ	x4	x5	x6	x7	x8				
3831732	181.54	448.19	7571	6212.7	6370241	525.71	985.31				
3913824	214.63	549.97	9038.16	7601.73	6467115	618.25	1259.2				
3928907	239.56	686.44	9905.31	8092.82	6560508	638.94	1468.06				
4282130	261.58	802.59	10444.6	8767.98	6664862	656.58	1678.12				
4453911	283.14	904.57	11255.7	9422.33	6741400	758.83	1893.52				
4548852	308.58	1000.69	12018.52	9751.44	6850024	878.26	2139.18				
4962579	348.09	1121.13	13966.53	11349.47	7006896	923.67	2492.74				
5029338	387.81	12/18/29	1/69/	11/167 35	7125979	978 21	28/1 65				
		共 20	〕条 100条/页 ▼	〈 1 〉 前往 1	页						

图 6 financial_data 影响市财政收入的因素数据预览

3.1.2 新建空白工程

右击我的工程,新建一个空白的工程,如图 7 所示。

工程信息
未选择工程
工 未送

图 7 新建工程

填写工程的信息,包括工程名称和工程描述,如图 8 所示。

	创建工程	×
* 工程名称	市财政收入分析预测	
工程描述	运用数据分析技术对市财政收入进行分析,挖掘其中隐藏的运行模式,并对未来两年的财政收入进行预测。针对某市财政收入数据,采用pearson及Lasso模型进行相关	
工程位置	▼我的工程	
	重置通定	

图 8 填写工程信息

3.2 特征工程

读取 financial_data 影响市财政收入的因素数据,步骤如下,如图 9 所示。

- 1. 选择市财政收入分析预测工程。
- 2. 选择输入源组件。
- 3. 拖入输入源组件。
- 4. 填写影响市财政收入的因素数据的数据表名。
- 5. 单击更新按钮,更新出影响市财政收入的因素数据。

	Python数据挖掘建模	莫平台	4 🔒	0
▲ 首页	<u>工程</u> ①	🛓 Q Q 120% % 🖺 O	◇ 字段属性	
数据源	1. ▲市财政收入分	34.	数据表 financial_data	•
工程		5.	字段信息	
● 个人组件	☆ 组件		字段	类型
₩ 模型	输入内容进行过滤		x1	数值
	 ▼ 赤吮垣件 ▼ 输入/输出 		x2	数值
任务	2. 🛋 输入源		x3	数值
	ご 输出源▶ 预处理		x4	数值

图 9 输入源组件

3.2.1 相关性分析

特征之间可能存在着信息的重复,分析财政收入数据特征的相关性,筛选出相关关系相 对较弱的特征。步骤如下,如图 10、图 11 相关性分析_字段属性_特征所示。

1. 选择统计分析→相关性分析组件。

2. 拖入相关性分析组件,并将数据源和相关性分析组件连接。

3. 选择字段属性,单击数据更新按钮,传入数据。

4. 选择字段属性,单击特征更新按钮,点选所有特征字段。

	Python数据挖掘建模平台	1	ļ. 7 G	0
▲ 首页	IE O 🕹 Q Q 120% % 🖹 O		◇ 字段属性	
● 数据源	▼我的工程 ▲市財政收入分 記載	3.	数据 3	Θ
工程	2. 11 11 11 11 11 11 11 11 11 11 11 11 11		字段	类型
⑦ 个人组件	≓ 输入源 ≠ ⇒ 输出源		x1	数值数值
₩	▶ 预处理 ▼ 统计分析		xЗ	数值
	▲ 因子分析 1. ▲ 上本MA		x4	数值
任务				

图 10 相关性分析_字段属性_数据

🛃 🛛 Q 🔹 🚺 120% 🗞 🛛 🖺 🔹 🖸	∨ 字段属性	
	x13	数值
输入源	у	数值
● 相关性分析	特征	0
4.		
	x8	ž
	x9 x10	
	x10	~
	x12	~
	x13	~
	у	~

图 11 相关性分析_字段属性_特征

5. 对相关性分析组件右键,选择运行该节点。运行完成后,对相关性分析组件右键,选择查看数据。查看、分析得到各属性的相关性进行特征选择,如图 12 各属性的相关性结果

所示。

		预览数据			×
ind	x1	x2	xЗ	x4	x5
x1	1	0.946	0.946	0.971	0.971
x2	0.946	1	0.997	0.992	0.99
xЗ	0.946	0.997	1	0.995	0.993
x4	0.971	0.992	0.995	1	0.999
x5	0.971	0.99	0.993	0.999	1
x6	0.994	0.92	0.919	0.95	0.948
x7	0.953	0.991	0.996	0.994	0.992
x8	0.97	0.993	0.994	0.998	0.997
	共 14	条 25条/页 💌 🤇 1	〉前往 1 页		

图 12 各属性的相关性结果

3.2.2 Lasso 回归分析

当原始特征中存在多重共线性时, Lasso 回归不失为一种很好的处理共线性的方法, 它可以有效的对存在多重共线性的特征进行筛选。步骤如下, 如图 13、图 14 Lasso 回归_字 段属性 标签和图 15 Lasso 回归 基础参数所示。

1. 选择回归→Lasso 回归组件。

2. 拖入 Lasso 回归组件,将输入源和 Lasso 回归组件连接。

3. 选择字段属性,单击特征更新按钮,勾选特征字段 x1~x13。

4. 单击标签更新按钮,选择标签字段 y。

5. 选择基础参数,设置 L1 项系数为 1000,剩余基础参数和所有高级参数设置为默认值。

	Python数据挖掘建模平台	4 8 8
▲	<u> 1程</u> 〇	◇ 字段属性
数据源	▼我的工程 ▲市財政收入分… 記載	· 特征
工程	<u> 組件 2. () 相关性分析 () LASSO回归 </u>	添加字段过滤字符串 字段
令 人组件	▼ 969 J.J.WI	 ✓ x1
₩	 ● 最近邻回归 ● 广义環小二乘 	✓ x2
任务	1. ● CART回归树 ● 多项式回归	✓ x3
	● LASSO回归 ● 支持向量回归	✓ > 基础参数

图 13 Lasso 回归_字段属性_特征

	Python数据挖掘建模平	2台		
▲ 首页	「「「「」」「「」」「「」」「」」「「」」「」」「「」」「」」「」」「」」「」	🛓 🔍 🔍 120%) % 🖹 🗅 🗘	◇ 字段属性	
	▼ 我的工程			
数据源	▲市财政收入分	(この) (注) (注) (注) (注) (注) (注) (注) (注) (注) (注	💙 x12	
工程			×13	
		C 相关性分析 O LASSO回归		
⑦ 个人组件	▶分类		У	
	▼回归		+= /7	
	⑧ 最近邻回归			
候型	◉ 广义最小三乘	4.	2	
	● CART回归树		у 👻	
任务	◉ 多项式回归			
	⊛ LASSO回归			
	⑧ 支持向量回归		/ 基础麥致	

图 14 Lasso 回归_字段属性_标签

	Python数据挖掘建模平台		4 6 0 🗆
▼ 谷 首页		초 Q Q 120% % 🖹 O	> 字段属性
● ●●● ●	▼ 我的工程 ▲ 市财政收入分…	世 输入源	 > 基础参数 L1项系数
工程	会 组件	5. 14 14 大性分析 14 14 15 14 15 14 15 15 15 15 15 15 15 15 15 15	1000
	▶ 统计分析▶ 分类		拟合截距 ● True ●
₩ 模型	▼ 回归 ⑧ 最近邻回归		最大迭代次数 ②
日子	◉ 广义最小二乘… ◉ CART回归树		

图 15 Lasso 回归_基础参数

6. 对 Lasso 回归组件右键,选择运行该节点。运行完成后,对 Lasso 回归组件右键,选择查看数据→预测值,查看标签预测结果;对 Lasso 回归组件右键,选择查看报告。如图 16
Lasso 回归预测结果预览图 17 Lasso 回归报告预览所示。

		预览数据		×
x11	x12	x13	У	predict_value
120	1.03	5321	64.87	22.19
113.5	1.05	6529	99.75	69.22
108.2	1.06	7008	88.11	90.81
102.2	1.09	7694	106.07	123.88
97.7	1.2	8027	137.32	179.91
98.5	1.2	8549	188.14	208.54
102.8	1.35	9566	219.91	270.73
98.9	1.47	10473	271.91	266.28
	共	20 条 25 条/页 💌 < 1 >	前往 1 页	

图 16 Lasso 回归预测结果预览

		算法运行报告	
(模型参数	
		需要配置的参数及其取值如下。	
参数名称	参数值		
L1项系数	1000		
拟合截距	True		
最大迭代次数	500		
归一化	False		
预计算	False		
容错率	1e-06		

图 17 Lasso 回归报告预览

3.2.3 数据筛选与标准化

根据相关性分析和 Lasso 回归分析结果,将影响市财政收入的因素数据进行特征选择和标准化处理,步骤如下,如图 18 所示。

1. 选择预处理→数据标准化组件。

2. 拖入数据标准化组件,并将 Lasso 回归的预测值端口和数据标准化组件连接。

3. 选择字段属性,单击更新数据,选择勾选特征 x1、x3、x4、x5、x6、x7、x8 和 x13 字段,标签 y 字段输出。基础参数设置为默认值。

	Python数据挖掘建模平台	↓ ▲ 🗭 0
 ▲前の)))) 一工程 ◆ 介人组 ● 操型 ● 保型 ● 保 ● 保 ● 保 ● 保 ● (1) <li< th=""><th>工程 ● ▲ Q 120% ●<</th><th> ◇ 字段属性 特征 ● ③ ⑤ ⑤ ⑦ ⑦ ⑦ ⑦ ⑦ ※ <li< th=""></li<></th></li<>	工程 ● ▲ Q 120% ●<	 ◇ 字段属性 特征 ● ③ ⑤ ⑤ ⑦ ⑦ ⑦ ⑦ ⑦ ※ <li< th=""></li<>

图 18 数据标准化组件

4. 对数据标准化组件右键,选择运行该节点。运行完成后,对数据标准化组件右键,选择查看数据,查看数据标准化的输出表结果,如图 19 所示。

		预览数据			
x1	xЗ	x4	x5	x6	x7
-1.4206941031171472	-1.0278324358747808	-1.2140853798473397	-1.2084670281198722	-1.6186555930954951	-1.0275
-1.353965401667356	-0.9734212838239051	-1.0665531636352368	-1.0346677059167826	-1.4586943313755072	-0.9474
-1.3417051448339592	-0.9004650063589514	-0.9793557473423139	-0.9732211503055178	-1.3044810015262718	-0.9294
-1.0545868859025207	-0.8383717138583697	-0.9251267274119552	-0.888743240988485	-1.1321685388889529	-0.9142
-0.9149542412630113	-0.783853642664144	-0.8435654930751163	-0.806869137129251	-1.005786691855441	-0.825
-0.837781195846349	-0.7324683023700769	-0.7668591180574521	-0.7656899725825825	-0.8264234768228009	-0.722
-0.5014820989219493	-0.6680815942446351	-0.5709743934852594	-0.565739985281946	-0.5673917134132385	-0.682
-0.44721686747325	-0.6001024029026295	-D.49782268451839934	-0.5509905096757314	-0.3707582830872344	-0.635

图 19 数据标准化的输出表结果

3.3 模型构建

3.3.1 灰色预测算法

构造、使用灰色预测算法模型组件,步骤如下,如图 20 灰色预测_字段属性、图 21 所示。

1. 选择预处理→python 脚本组件。

2. 拖入 python 脚本组件,将数据标准化和 python 脚本组件连接。

3. 选择字段属性,在脚本处填入灰度预测代码,代码如表2所示。

表 2 python 脚本-灰色预测

data_in = db_utils.query(conn, 'select * from ' + inputs['input1'])
""
载入模块
""
import numpy as np
import pandas as pd
""
def GM11(x0): #自定义灰色预测函数
""
def GM11(x0): #自定义灰色预测函数
import numpy as np
x1 = x0.cumsum() #1-AGO 序列
z1 = (x1[:len(x1)-1] + x1[1:])/2.0 #紧邻均值 (MEAN) 生成序列

```
z1 = z1.reshape((len(z1),1))
  B = np.append(-z1, np.ones\_like(z1), axis = 1)
  Yn = x0[1:].reshape((len(x0)-1, 1))
  [[a],[b]] = np.dot(np.dot(np.linalg.inv(np.dot(B.T, B)), B.T), Yn)#计算参数
  f = lambda k: (x0[0]-b/a)*np.exp(-a*(k-1))-(x0[0]-b/a)*np.exp(-a*(k-2)) #还原值
  delta = np.abs(x0 - np.array([f(i) for i in range(1, len(x0)+1)]))
  C = delta.std()/x0.std()
  P = 1.0*(np.abs(delta - delta.mean()) < 0.6745*x0.std()).sum()/len(x0)
  return f, a, b, x0[0], C, P #返回灰色预测函数、a、b、首项、方差比、小残差概率
new_reg_data = data_in.drop('y', 1) # 特征列
new_reg_data.index = range(1994, 2014)
new reg data.loc[2014] = None
new_reg_data.loc[2015] = None
# 灰色预测
1 = ['x1', 'x3', 'x4', 'x5', 'x6', 'x7', 'x8', 'x13']
for i in l:
  f = GM11(new_reg_data.loc[range(1994, 2014),i].as_matrix())[0]
  new reg data.loc[2014,i] = f(len(new reg data)-1)#2014 年预测结果
  new_reg_data.loc[2015,i] = f(len(new_reg_data)) ##2015 年预测结果
  new_reg_data[i] = new_reg_data[i].round(2) ## 保留两位小数
y=list(data in['y'].values) # 提取财政收入列,合并至新数据框中
y.extend([np.nan,np.nan])
new_reg_data['y'] = y
data_out = pd.DataFrame(new_reg_data)
return(data_out)
```

	Python数据挖掘建模	平台	A 🔒	• 0 •
▲	工程	🛓 Q Q 120% % 🖺 D	◇ 字段属性	
)) 数据源	 ▼ 我的工程 ▲ 市财政收入分… 	1 1111 11111	输入	Ø
工程	组件 繁数据筛选		input1 from 10000175_1_1 input2 from	
♀ 个人组件	國 分组聚合 國 修改列名		input3 from	
◆ 模型	 X Python脚本 X 缺失值处理 	这 数据标准化 2.	脚本	0
任务	XX 数学类函数 XX 特征构造	¥ Python脚本 3	<pre>3. 1 data_in = db_utils.query(con from ' + inputs['input1' 3 ''' a #1 b m th</pre>	nn, 'select *])
	▲ 5216高股化 & 标准化数据还质 ▶ 统计分析		6 import numpy as np 7 import pandas as pd	
	▲ 公迷 横型		9 1a 白癜的五色 55到汤弗 > 组件描述	

图 20 灰色预测_字段属性

4. 对 python 脚本组件右键,选择重命名,修改名称为灰色预测。

提示	×
请输修改后的名称	
灰色预测	
	取消 确定

图 21 python 脚本组件重命名

5. 对灰色预测组件右键,选择运行该节点。运行完成后,对灰色预测组件右键,选择查 看数据,如图 22 所示。

		预览数据		×
x6	x7	x8	x13	У
0.81	0.34	0.59	0.45	0.378992004908741
0.98	0.82	0.79	D.64	0.824457822565277
1.17	1.34	1.16	1.13	1.31532400480814
1.31	1.47	1.54	1.56	1.5443129356853
1.44	1.77	1.8	1.91	1.61931784068066
1.61	2.37	2.23	2.28	2.47555928643593
1.94	0.81	0.87	0.97	
2.1	0.85	0.91	1.03	
	Д	22 奈 25 奈/页 👻 < 1 >	前往 1 页	

冬	22	灰色预测算法的输出模型结果
---	----	---------------

3.3.2 支持向量机回归算法

选择支持向量机回归算法模型,步骤如图 23 支持向量机回归_字段属性_特征、图 24 支持向量机回归_字段属性_标签所示。

- 1. 选择回归→支持向量机回归组件。
- 2. 拖入支持向量机回归组件,将数据筛选和支持向量机回归组件连接。
- 3. 选择字段属性,单击更新特征数据,勾选 x1、x3、x4、x5、x6、x7、x8 和 x13 字段。
- 4. 单击更新标签数据,选择标签字段 y。基础参数和高级参数设置为默认值。

	Python数据挖掘建模平台	A	4 4 6 0 0
▲	「超のの	🕹 Q Q 🛛 120%) % 🖺 O	✓ 字段属性
))) 数据源	▼ 我的工程 ▲ 市财政收入分…	→ 輸入源 3.	特征 ♀
工程	组件		添加字段过速字符串
	 ▶ 分类 ▼ 回归 ● 番近郊回归 		 →FR ✓ x1
♥型	 ● 广义最小二乘… 1. ● LASSO回归 	X 数据标准化 2.	✓ x3
任务	 ● 支持向量回归 ● CART回归树 	大色技測 交色技測	✓ x4
	 多项式回归 线性回归 		▲ X5_ > 基础参数
	 ● 岐回归 		> 高级参数
	模型		> 组件描述

图 23 支持向量机回归_字段属性_特征

	Python数据挖掘建模	2	3 8 0 0 0
▲		🛓 Q Q 120% % 🖺 D	◇ 字段属性
数据源	▼ 我的工程 ▲ 市财政收入分	(二) 絶入源	×8
工程	组件	日共性分析 BLASSO回归	✓ x13
	▼ 回归 ◎ 最近邻回归		У
₩	● 广义最小二乘…● LASSO回归	文 數据标准化 4.	标签 😧
任务	● 支持向量回归● CART回归树	文色 茨角 の 支持向量回归	у —
	 多项式回归		> 基础参数
			 > 高级参数 > 组件描述

图 24 支持向量机回归_字段属性_标签

对支持向量机回归组件右键,选择运行该节点。运行完成后,对支持向量机回归组件右键,选择查看数据,查看预测结果;对支持向量机回归组件右键,选择查看报告。如图
 25 支持向量机回归预测结果预览、如图 26 支持向量机回归报告预览所示。

			预览数据		
	x7	x8	x13	У	predict_value
i	-1.02755030585191	-1.08143964382212	-1.10388923196867	-0.931606724564484	-0.77
i1	-0.947409403013852	-1.0186928597397	-0.992325565529158	-0.872869160886035	-0.79
!7	-0.929491580669791	-0.970844112711864	-0.948087985873095	-0.892470801975976	-0.79
15	-0.914215099947257	-0.922720451883234	-0.88473312231556	-0.862226345448885	-0.79
4	-0.8256652D4602636	-0.873373424641081	-0.853979230821887	-0.80960166468766	-0.77
101	-0.722237194812826	-0.817093987722291	-0.805770427939915	-0.724021303846527	-0.74
:38	-0.682911515129704	-0.736095218215626	-0.711846380945726	-0.670520948397435	-0.65
:34	-0.635679130854931	-0.656161739687029	-0.628081277087586	-0.582953479610757	-0.61
		共 20 条 25 条	\$/页 ▼ 〈 1 〉 前往	1 页	

图 25 支持向量机回归预测结果预览

		算法运行报告
模型参数		
需要配置的参数及其取	(值如下。	
参数名称	参数值	
核函数	rbf	
多项式阶数	2	
核系数	auto	
独立项	0.0	
惩罚系数	1.0	
容错率	1e-3	

图 26 支持向量机回归报告预览

3.4 模型预测

3.4.1 支持向量机回归预测

使用支持向量机回归输出的模型,对灰色预测输出的预测特征进行预测市财政收入,步骤如下,如图 27 所示。

1. 选择预处理→模型预测组件。

拖入模型预测组件,并将灰色预测组件、支持向量机回归组件分别和模型预测组件连接。

3. 单击更新按钮,勾选 x1、x3、x4、x5、x6、x7、x8 和 x13 特征字段。

	Python数据挖掘建模平台		↓ ▲ 🗭 0 🛛
· 谷 首页	工程の	📩 Q Q 120% % 🖻 O	◇ 字段属性
● ●● ● 数据源	▼我的工程	三 输入源 3.	特征 ②
工程	组件		添加字段过滤字符串
℃ 个人组件	 ▶ 紫交 ▶ 时序模型 ▶ 关联抑励 	数据标准化	 →R ×1
◆ 模型	 ▶ 模型评估 ▼ 模型预测 	x色预测 ③ 支持向量回归	✓ x3
日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日	1. <mark>₩ 模型预测</mark> ▶ 案例	2.	✓ x4
	▶ 数据探索 ▶ 协同过滤算法	便型预测	<u>x5</u>
	▶ 结巴分词 ▲ ▲ L 48//t		〉但供描述

图 27 支持向量机回归预测组件

 4. 对模型预测组件右键,选择运行该节点。运行完成后,对模型预测组件右键,选择查 看数据,查看支持向量机回归预测的输出结果,如图 28 所示。

		预览数据		×
x6	x7	x8	x13	predict_label
-1.62	-1.03	-1.08	-1.1	-0.7706347555562267
-1.46	-0.95	-1.02	-0.99	-0.7909480440643297
-1.3	-0.93	-0.97	-0.95	-0.7921624153019462
-1.13	-0.91	-0.92	-0.88	-0.7872620668551431
-1.01	-0.83	-0.87	-0.85	-0.7686182152903933
-0.83	-0.72	-0.82	-0.81	-0.7427252570328833
-0.57	-0.68	-0.74	-0.71	-0.6541674825333963
-0.37	-0.64	-0.66	-0.63	-0.6094665631797167
	д	22条 25条/页 👻 🤇 1	〉 前往 1 页	

图 28 支持向量机回归预测的输出结果

3.4.2 标准化数据还原

对预测数据进行标准化数据还原,步骤如图 29 标准化数据还原_字段属性_特征、图 30 标准化数据还原_字段属性_原数据特征所示。

1. 选择预处理→标准化数据还原组件。

2. 拖入标准化数据还原组件,并将 Lasso 回归组件、模型预测组件分别和标准化数据还原 组件连接。

3. 选择字段属性,单击特征更新按钮,勾选所有特征字段。

4. 单击原数据特征更新按钮,勾选与特征对应的 x1、x3、x4、x5、x6、x7、x8、x13 和标签 y 字段。基础参数设置为默认值。

	Python数据挖掘建树	東平台 4	1 8 0 0 0
合 首页	I程 O	🛓 🧕 🔾 🚺 🖧 🔛 🗅	◇ 字段属性
》 数据源	▼ 我的工程 ▲ 市财政收入分…	□ 输入源 3.	特征 @
工程	貧件		添加字段过滤字符串
₽ 个人组件	▲ 排序 XX 数据筛选	赵 数据标准化	 ✓ 字段 ✓ x1
₩	X 修改列名 X 修改列名	汉 灰色预测 ③ 支持向量回归	✓ x3
日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日	X 缺失值处理 X 数学类函数	使型预测	✓ x4
	X 特征构造 X 数据离散化	2.	
	 ♣ 标准化数据还质 ▶ 统计分析 	标准化数据还原	>> 基础参数
	模型		> 组件描述

图 29 标准化数据还原_字段属性_特征



图 30标准化数据还原_字段属性_原数据特征

5. 对标准化数据还原组件右键,选择运行该节点。运行完成后,对标准化数据还原组件右键,选择查看数据,查看标准化数据还原的输出结果,如图 31 所示。

		预览数据			>
x1	хЗ	x4	x5	x6	x7
3832585.9101144616	444.13541434339527	7611.627776432271	6200.448261019306	6369426.814908768	522.881
3918702.3768693823	556.3697591389257	9003.882331225232	7639.034999049492	6466324.275164693	615.258
3931004.7292629424	687.3098280670445	9898.903116449279	8118.563911726221	6563221.735420616	638.352
4287772.948676186	799.5441728625749	10396.136886018192	8757.935795295192	6666175.286942536	661.447
4460005.882186028	911.7785176581051	11291.15767124224	9397.307678864163	6738848.382134479	753.824
4546122.348940949	1005.3071383210472	11987.28494863872	9716.993620648649	6847858.024922393	880.843
4964402.330321994	1117.5414831165776	13976.22002691438	11315.423329571078	7005316.39783827	927.031
5025914.092289794	1248.4815520446964	14672.347304310859	11475.266300463321	7126438.223158175	973.220
	共2	2 条 25 条/页 👻 🧹 👔) 前往 1 页		

图 31 标准化数据还原的输出结果

3.4.3 整合预测数据

对预测的输出数据和原数据字段进行拼接,如图 32 所示。

1. 选择预处理→python 脚本组件。

2. 拖入 python 脚本组件,并将标准化数据还原组件、Lasso 回归组件组件和 python 脚本组件连接。

3. 选择字段属性, 在脚本处填入整合预测数据代码, 代码如表 3 所示。

表 3 整合预测数据代码

data_1 = db_utils.query(conn, 'select * from ' + inputs['input1'])
data_2 = db_utils.query(conn, 'select x1,x3,x4,x5,x6,x7,x8,x13,y from ' + inputs['input2'])

整合 1994~2013 与 2014、2015 年财政收入

 $y = data_2['y'].append(data_1['predict_label'].tail(2))$

import pandas as pd data_out = pd.DataFrame({'label':range(1994, 2016),'y':y})

return(data_out)

	Python数据挖掘建模	平台	4 8 9 2
▼ 谷 首页	I程 O	🛓 Q Q 120% % 🖺 D	∨ 字段属性
●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●	▼我的工程 ▲市财政收入分…	11 11 11 11 11 11 11 11 11 11 11 11 11	₩ λ 0
工程	组件 风 排序		input2 from 10000179_1_2
▶ 个人组件	X 数据筛选 X 分组聚合 X 修改列名	数据标准化	input3 from input4 from
◆ 模型	1. X Python脚本 X 缺失值处理	X 灰色颈胸 S 支持向量回归 3.	脚本
任务	■ 数学类函数■ 数字类函数■ 数 特征构造■ 数 姻 密 散化	✓ 模型预测	<pre>2 data_2 = db_utils.query(conn, 'select x1 ,x4,x5,x6,x7,x8,x13,y from ' +</pre>
	 ▲ 标准化数据还原 ▲ 统计分析 	Son the Abar Abar Abar Abar Abar Abar Abar Abar	5 data data 2['y'].append (data_1{'predict_label'].tail(2)) 6 data['label'] = consc(1984 _ 2016)
	模型		> 组件描述

图 32 整合预测数据组件

4. 对 python 脚本组件右键,选择重命名,修改名称为整合预测数据,如图 33 python 脚本组件重命名所示。

提示	\times
请输修改后的名称	
整合预测数据	
	取消 确定

图 33 python 脚本组件重命名

对整合预测数据组件右键,选择运行该节点。运行完成后,对整合预测数据组件右
 键,选择查看数据,如图 34 整合预测数据的输出结果所示。

	预览数	据
	label	У
	2008	843.14
	2009	1107.67
:	2010	1399.16
:	2011	1535.14
:	2012	1579.68
	2013	2088.14
	2014	1229.51795094269
:	2015	1247.06443840996
	共22条 25条/页 🔻	〈 1 〉 前往 1 页

图 34 整合预测数据的输出结果