

# 基于神经网络模型的信贷策略研究

## 摘 要

商业银行本质上属于风险经营，由于中小微企业规模相对较小，风险抵御能力较弱，商业银行必须准确评估各企业的风险值，从而给出相适应的信贷政策。本文根据企业相关历史数据进行预处理后获取有效的基础数据，从风险度量、风险分类、风险评估三个方面展开研究，从而制定并完善针对各中小微企业的信贷策略。

首先本文对附件所给出的数据进行预处理，包括冗余性和相关性分析，这是后续建模的基础。

针对问题 1，首先，对附件 1 中部分数据进行量化处理；其次，选取 2 种评价指标，使用 K-means 聚类法<sup>[1]</sup>将附件 1 中的 123 家中小微企业分成三大类：A 类，B 类，C 类，实现各类别之间的定量化识别。接着，对附件 1 中进/销项发票信息进行统计与分析，得出 4 种新的指标。然后，基于 K-means 聚类结果和 4 种新指标构建 BP 神经网络模型。最后，对附件 3 的内容进行可视化处理，综合可视化、聚类以及神经网络模型的结果得出：对于 A 类和 B 类企业，银行可以予以贷款，贷款额度可达最大贷款额的 70%，但 A 类企业利率不易过高。对于 C 类企业，银行可以不予贷款。

针对问题 2，首先对附件 2 中进/销项发票信息进行统计与分析，得出 4 组新数据；其次运用问题 1 中构建的 BP 神经网络模型<sup>[2]</sup>，对附件 2 中 303 家企业的类别进行预测；最后，综合预测结果以及问题 1 中的策略得出：对于 A 类企业，予以最高 75 万最高利率为 0.0545 的贷款，对于 B 类企业，予以最高 50 万最高利率为 0.0585 的贷款，对于 C 类企业，可以不予以贷款。

针对问题 3，对附件 2 中 302 家无信贷记录企业月收支进行统计后可视化处理，挖掘出相关行业的收支对时间的敏感程度。综合信贷风险和企业应对突发事件的能力得出：对于运输行业相关的企业，银行可以考虑提高贷款金额，降低贷款年利率；对于个体、科技、建筑行业相关的企业，银行可以考虑降低贷款金额。

**关键词：**信贷风险评估；K-means 聚类分析；BP 神经网络模型；信贷策略

## 一. 问题重述

### 1.1 研究背景

企业的发展离不开资金的支持，中小微企业具有投资少、收益快、灵活等优点，但规模相对较小，产品或服务单一，分散风险的能力弱，自有资金少，融资主要依靠商业银行贷款，具有借贷频率高和单笔借贷小的特点。同时，中小企业的组织机构不像大企业一样层次分明，信息透明度低，受管理者影响大，对其信用评价的主观性大于客观性，商业银行往往对中小微企业缺乏信任。因此商业银行高度重视各企业信贷风险的科学评估，以做出合理的信贷策略调整。

### 1.2 数据来源

采用数据均来自题目附录。

### 1.3 问题 1

银行目前是根据信贷政策、企业的交易票据信息和上下游企业的影响力评估企业的，判断出怎样的企业是强、供求关系稳定的企业，并对其中好的企业给予利率优惠。银行对确定要放贷企业的贷款额度为 10~100 万元；年利率 4%~15%；贷款期限为 1 年。

题目说明了银行具体的评估方式。第一步：对实力和信誉做出评估。第二步：根据评估结果，根据一些因素来确定具体包括放贷及贷款额度、利率和期限等信贷策略。根据附件所提供的 123 家企业的信贷风险进行合理量化，进行有效的风险评估，当年度信贷总额一定时，提供各中小微企业相应的信贷策略。

### 1.4 问题 2

结合问题 1 的运算结果，根据附件 2 所提供的 302 家企业的信贷风险进行量化分析，根据分析结果，给出银行信贷总额 1 亿时，针对相应中小微企业的信贷策略。

### 1.5 问题 3

企业经营效益在某种程度上受到一些突发性因素的影响，预测附件 2 中各因素对相应企业的影响，当年度信贷总额为 1 亿时，给出相应的信贷调整策略。

## 二. 问题分析

### 2.1 问题 1

问题 1 要求我们针对“附件 1：123 家有信贷记录企业的相关数据”中的 123 家企业，量化其信贷风险，并给出在银行在固定年度信贷总额的情况下的信贷策略。在对 123 家有信贷记录企业的相关数据进行量化分析前，我们首先对数据进行预处理，删除不必要的数数据，降低数据冗余度。其次，我们对各中小微企业的信誉评级和是否违约两类指标进行量化，运用 k-means 方法进行聚类，将企业的信贷风险划分为三类。各中小微企业的进/销项发票信息的有关数据作为输入层，将聚类结果作为输出层，构建起 BP 神经网络模型。最后使用数据可视化方法对附件 3 的数据进行可视化，综合可视化的结果以及 BP 神经网络模型的结果，给出银行最优的信贷策略<sup>[3]</sup>。

### 2.2 问题 2

问题 2 要求我们量化“附件 2：302 家无信贷记录企业的相关数据”中 302 家企业的信贷风险，并给出该银行在年度信贷总额为 1 亿元时对这些企业的信贷策略。通过对“附件 2：302 家无信贷记录企业的相关数据”与“附件 1：123 家有信贷记录企业的相关数据”观察分析发现：附件 2 相较于附件 1 仅仅少了各个企业的“信誉评级”和“是否违约”的相关数据。本文对附件 2 的进/销发票信息进行处理与统计，运用问题 1 中构建起的 BP 神经网络对附件 2 中 302 家中小微企业类别进行预测，根据预测结果，为每类企业制定特定的信贷策略<sup>[4]</sup>。

### 2.3 问题 3

问题 3 中提及突发风险因素不同行业的影响，要求我们综合突发因素和风险因素<sup>[5]</sup>，给出银行在年度信贷总额为 1 亿元时对附件 2 中的 302 家企业的信贷调整策略。本文将“附件 2：302 家无信贷记录企业的相关数据”中的月收入与月支出进行数据可视化处理，得出各个企业进三年来月收入与月支出随时间变化的曲线图，并对曲线图进行分析，挖掘出 302 家企业中各行业的收支对于某个时间段的敏感程度。最后综合结果和问题 2 的信贷策略，给出该银行在年度信贷总额为 1 亿元时的信贷调整策略<sup>[6]</sup>。

### 三. 模型假设

- (1) 假设附件提供的数据真实准确。
- (2) 假设各企业进/销项发票数据绝对真实。

### 四. 符号说明

符号	含义
$x_1$	企业总收入
$x_2$	企业总支出
$x_3$	企业盈利
$x_4$	企业收入标准差
$x_5$	企业支出标准差
$r$	相关常数
$BP\_loss$	BP 神经网络损失值
$d$	样本间距离
$X$	样本参数

### 五. 模型建立

#### 5.1 问题 1

##### 5.1.1 数据预处理

由于附件给定的数据较多，我们有必要将冗余数据进行处理，同时对无效数据进行剔除。在附件 1 和附件 2 中，进/销项发票信息页中的发票号码所在列和发票状态为作废发票所在行暂时不予以考虑，故可以将这些数据剔除。

##### 5.1.2 量化评价指标

观察企业信息发现，各中小微企业的信誉评级分为 A、B、C、D 四类，本文按照等级评定量化表法对信誉评级进行量化处理。已知银行对信誉评级为 D 的企业原则上不予放贷，故本文将 D 级量化为 20。

表 1 评级量化值

评级	A	B	C	D
量化值	100	80	60	20

各中小微企业的违约情况分为两类，违约情况如下：

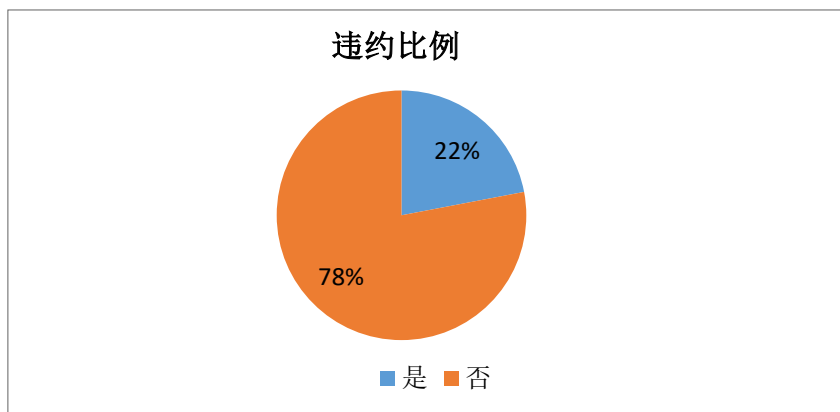


图 1 企业违约情况

因此，本文按照违约比例作为量化标准，结果如表 2 所示：

表 2 违约情况量化值

是否违约	是	否
量化值	22	78

### 5.1.3 K-means 聚类分析模型建立

#### (1) 算法介绍

K-Means 算法是一种简单的迭代型聚类算法，采用距离作为相似性指标，从而发现给定数据集中的 K 个类，且每个类的中心是根据类中所有数值的均值得到的，每个类的中心用聚类中心来描述。

#### ① 绝对距离：

$$d_{ij}^{(1)} = \sum_{k=1}^p |X_{ik} - X_{jk}| \quad (\text{式 5-1})$$

#### ② 欧氏距离：

$$d_{ij}^{(2)} = \left( \sum_{k=1}^p |X_{ik} - X_{jk}|^2 \right)^{\frac{1}{2}} \quad (\text{式 5-2})$$

欧式距离是最常用的距离，本文在使用 Python 进行聚类分析时使用的就是欧式距离。

(2) 聚类图生成步骤:

- ① 读取数据, 本文选取量化后的信誉评级和是否违约作为聚类数据;
- ② 运用 Python 软件使用 K-Means 函数, 并设置分类个数;
- ③ 使用 Matplotlib.pyplot 库画出聚类图。

#### 5.1.4 BP 神经网络模型建立

(1) 模型介绍

BP 网络能学习和存贮大量的输入-输出模式映射关系, 而无需事前揭示描述这种映射关系的数学方程。BP 神经网络一般包含三层, 即输入层, 隐含层, 输出层如图 2 所示。

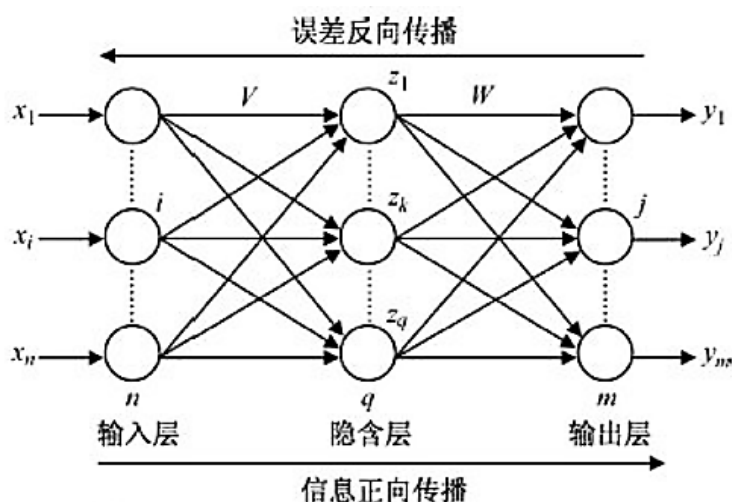


图 2 神经网络结构图

- ① 输入层: 一组连接(对应于生物神经元的突触), 连接强度由各连接上的权值表示, 权值为正表示激活, 为负表示抑制, 仅该层为外界输入。
- ② 隐含层: 该层将输入的一组连接根据连接强度(即权重)的不同映射到更多或更少的维度上去。
- ③ 输出层: 通过隐含层得到的结果进行变换, 得到输出结果。

(2) 输入层的确定

通过对附件 1 内容的观察发现, 税额和价税合计相关程度很高, 这与实际相符合, 本文从降低数据冗余度出发, 暂时不考虑税额和价税合计等相关数据。剔除上述数据后, 本文对剩余数据的操作如下:

- ① 计算各个企业在附件 1 所给的时间内的收入, 支出, 以及盈利情况, 记为  $x_1$ ,  $x_2$ ,  $x_3$

$$x_1 = \sum_{i=1}^n A_i \quad (\text{式 5-3})$$

$$x_2 = \sum_{i=1}^n B_i \quad (\text{式 5-4})$$

$$x_3 = x_1 - x \quad (\text{式 5-5})$$

② 计算各个企业在附件 1 所给的时间内(以月为单位)的收入与支出的标准差, 记为  $x_4$ ,  $x_5$ , 设  $c_1, c_2, \dots, c_n$  表示每月收入, 则标准差  $s$  计算公式为:

$$s = \sqrt{\frac{(c_1 - \bar{c})^2 + (c_2 - \bar{c})^2 + \dots + (c_n - \bar{c})^2}{n}} \quad (\text{式 5-6})$$

③ 计算  $x_1, x_2, x_3, x_4, x_5$  与 K-means 聚类的结果进行相关性比较。

1) 皮尔森相关系数计算方法:

$$R = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\left( \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \right) \left( \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2} \right)} \quad (\text{式 5-7})$$

2) 简单的相关系数的分类

(a) 0.8-1.0 极强相关

(b) 0.6-0.8 强相关

(c) 0.4-0.6 中等程度相关

(d) 0.2-0.4 弱相关

(e) 0.0-0.2 极弱相关或无相关

(3) 通过  $x_1, x_2, x_3, x_4, x_5$  与聚类结果的相关程度, 确定输入层包含的指标, 建立 BP 神经网络模型。

### 5.1.5 数据可视化处理

利用 Excel 表格, 对“附件 3: 银行贷款年利率与客户流失率关系的统计数据”表进行可视化操作, 得出不同信誉级别的公司对于银行贷款年利率的敏感程度。

将所得结果进行综合考虑, 为每类企业制定特定的信贷策略。

## 5.2 问题 2

### 5.2.1 不同类别企业信贷策略

根据问题 1 中的部分结论，给出对于不同类别企业的信贷策略如表 3：

表 3 不同类别企业信贷策略表

类别	对策		
	是否贷款	贷款金额(最大限额百分比)	贷款利率
类别 A	可以贷款	70%-100%	0.0425-0.0545
类别 B	可以贷款	30%-70%	0.0545-0.0585
类别 C	考虑不予贷款	0 或 10%-20%	0.0785-0.0865

### 5.2.2 数据处理

(1) 计算不同企业年收入  $x_1$  和收入标准差  $x_4$ ：

设公司收入为  $a_1, a_2, \dots, a_i$ ，

$$x_1 = \sum_{i=1}^n a_i \quad (\text{式 5-8})$$

$$x_4 = \sqrt{\frac{(a_1 - \bar{a})^2 + (a_2 - \bar{a})^2 + \dots + (a_n - \bar{a})^2}{n}} \quad (\text{式 5-9})$$

(2) 计算不同企业年支出  $x_2$  和标准差  $x_5$ ：

设公司支出为  $b_1, b_2, \dots, b_i$ ，

$$x_2 = \sum_{i=1}^n b_i \quad (\text{式 5-10})$$

$$x_5 = \sqrt{\frac{(b_1 - \bar{b})^2 + (b_2 - \bar{b})^2 + \dots + (b_n - \bar{b})^2}{n}} \quad (\text{式 5-11})$$

### 5.2.3 运用 BP 神经网络对 302 家企业进行分类

问题 1 求解的 BP 神经网络模型如图 3 所示：



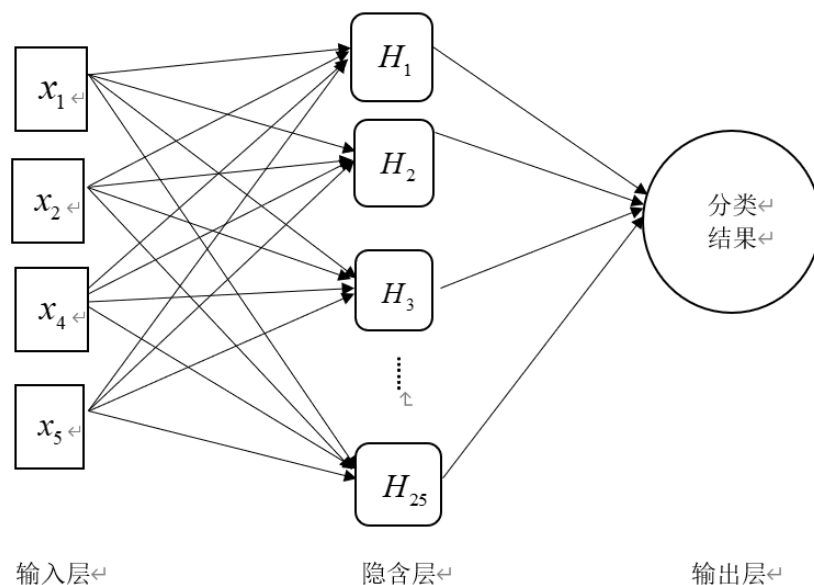


图 3 BP 神经网络模型

运用问题 1 构建的 BP 神经网络对附件 2 中的 302 家企业的类别进行预测，将预测结果与表 3 对照，得出银行在在年度信贷总额为 1 亿元时对这些企业的信贷策略。

### 5.3 问题 3

#### 5.3.1 数据统计与数据可视化

(1) 数据统计：根据附件 2 中的内容，统计出各个企业每月的收入与支出。

(2) 常见数据可视化：

① 柱状图：又称长条图，是一种以长方形的长度为变量的统计图表。柱状图用来比较两个或以上的价值（不同时间或者不同条件），只有一个变量，通常利用于较小的数据集分析。柱状图亦可横向排列，或用多维方式表达。

② 折线图：是排列在工作表的列或行中的数据可以绘制到折线图中，分类标签是文本并且代表均匀分布的数值（如月、季度或财政年度），则应该使用折线图。

③ 饼状图：饼状图常用于统计学模型饼状图显示一个数据系列中各项的大小与各项总和的比例，旨在反应出各个企业随时间收支随时间关系的变化。

本文选择折线图来进行可视化操作。

(3) 根据可视化结果，挖掘出各行业在某个时间段（如新冠疫情发生期间），收支受影响的程度，判断该行业对于突发情况的应对能力，同时结合信贷风险，给出银行信贷策略调整。

## 六. 模型求解

### 6.1 问题 1

#### 6.1.1 K-means 聚类分析模型求解

根据量化后的信誉评级和是否违约两个指标对附件 1 中 123 家有信贷记录企业进行聚类分析，样本间距离采用欧氏距离。

通过 Python 编程，得到的聚类散点图如图 4 所示：

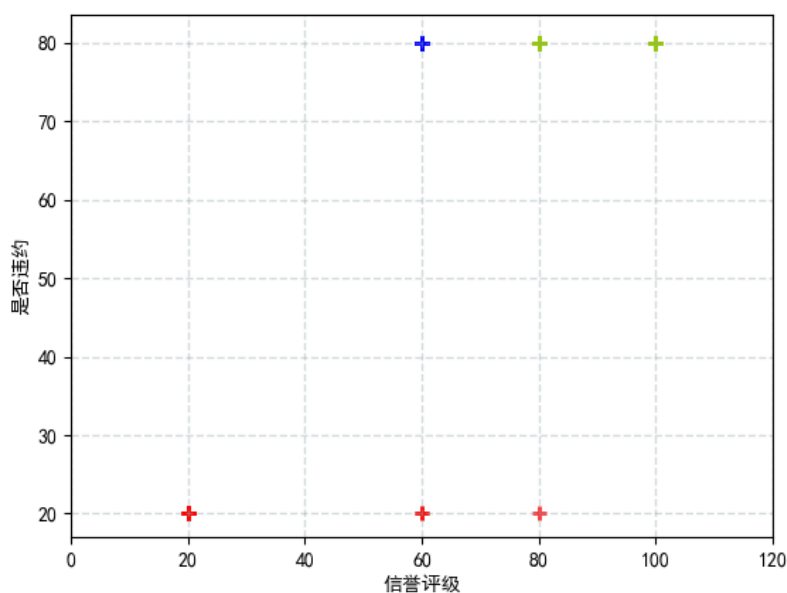


图 4 聚类散点图

根据图 4 的聚类结果，对不同企业进行分类统计，具体统计结果见表 4：

表 4 各类企业分类结果

类别	聚类中心(信誉评级, 是否违约)	包含企业(企业代号)
A 类	(88.44 , 80)	E1, E2, E5, E6, E7, E8, E9, E10, E12, E13, E15, E16, E17, E18, E19, E20, E21, E22, E23, E24, E26, E27, E28, E30, E31, E32, E33, E34, E35, E37, E38, E42, E43, E48, E51, E54, E57, E58, E59, E60, E61, E62, E63, E64, E65, E66, E67, E70, E71, E74, E76, E79, E81, E83, E84, E85, E88, E89, E91, E93, E95, E97, E98, E106
B 类	(60.00 , 80)	E3, E4, E11, E14, E25, E39, E40, E41, E44, E46, E47, E49, E50, E53, E55, E56, E68, E69, E72, E73, E75, E77, E78, E80, E86, E90, E92, E94, E96, E104, E105, E110
C 类	(25.12 , 20)	E29, E36, E45, E52, E82, E87, E99, E100, E101, E102, E103, E107, E108, E109, E111, E112, E113, E114, E115, E116, E117, E118, E119, E120, E121, E122, E123

### 6.1.2 BP 神经网络模型求解

#### (1) 数据处理

通过对 123 家企业的数据处理后产生五组新的数据如表 5(详细数据见附录)。

表 5

企业代号	$x_1$ (亿元)	$x_2$ (亿元)	$x_3$ (亿元)	$x_4$	$x_5$	类别
E5	2.02	1.97	0.04	0.05	0.05	3
E6	3.67	3.02	0.64	0.06	0.05	3
E3	5.70	0.52	5.18	0.11	0.01	2
E4	18.40	2.20	16.20	0.72	0.11	2
E103	0.017	0.006	0.011	0.0007	0.0006	1

#### (2) 计算相关系数

采用皮尔森相关系数算法，计算出  $x_1, x_2, x_3, x_4, x_5$  与类别的相关系数如表 6 所示。

表 6

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
$r$	0.834	0.705	0.005	0.762	0.623

#### (3) 求解 BP 神经网络模型

对照相关系数分类表可知： $x_1, x_2, x_3, x_4, x_5$  与类别成强相关，故本文选取  $x_1, x_2, x_3, x_4, x_5$  作为输入层。由于输入层有 4 维，本选将中间层设为 25 维，输出层设为 1 维，损失值如表 7 所示。

表 7 损失值

	输入层维度=4	隐含层维度=25	输出层维度=1
$BP\_loss$	0.53		

将输入层降维后绘出散点图如图 5 所示(蓝色是实际点，橙色是预测点)，BP 神经网络的结果见附录部分。

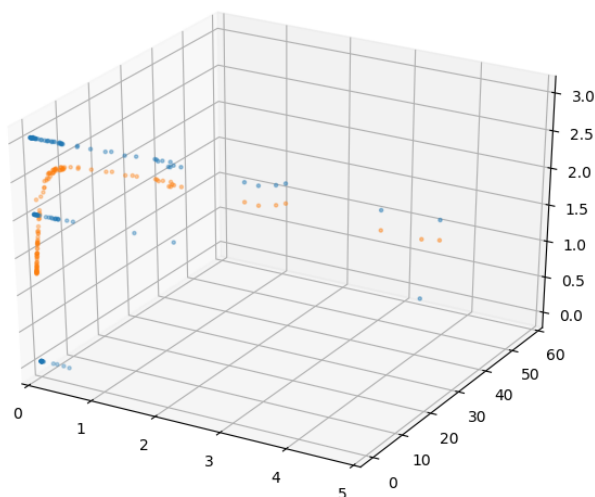


图 5 将输入层降维后得到的散点图

### 6.1.3 企业对银行利率的敏感度分析

对附录 3 的数据进行可视化操作处理后如图 6 所示：

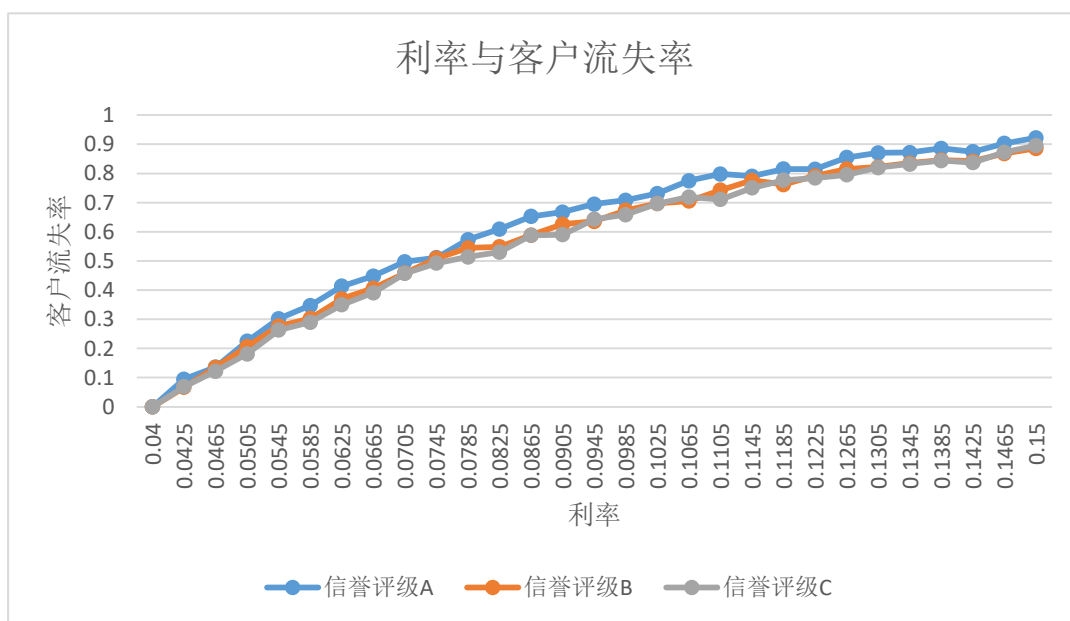


图 6 利率与客户流失率折线图

根据折线图可知，信誉级别越高的企业对银行利率的敏感度越高。

### 6.1.4 问题 1 总结与策略

对于 A 类企业（信誉评级高，没有违约记录）：这一类企业月收支流水非常大，收支稳定，企业信誉度高，并且无违约记录，但对银行利率敏感度高，该类企业贷款时银行可以考虑予以贷款，贷款金额为最大限额的 70%—100%，利率保持在 0.0425-0.0545 之间。

对于 B 类企业（信誉评级较低，没有违约记录）：这一类企业月收支流水

较大，收支较稳定，企业信誉度较高，无违约记录，对银行利率敏感度较高，该类企业贷款时银行可以考虑予以贷款，贷款金额为最大限额的 30%—70%，利率保持在 0.0545-0.0585 之间。

对于 C 类企业（有违约记录）：这一类企业月收支流水较低，企业信誉度层次不齐，但存在违约记录，对银行利率敏感度较低，该类企业贷款时银行可以考虑不予以贷款，或者贷款金额不超过最大限额的 20%，利率保持在 0.0785-0.0865 之间。

## 6.2 问题 2

### 6.2.1 对 302 家企业进/销项发票信息的处理

将处理后的数据展示如下，由于篇幅有限，此处只展示部分数据，详细见附录：

表 8

企业代号	$x_1$ (单位：亿元)	$x_2$ (单位：亿元)	$x_3$	$x_4$
E124	7.41	7.85	1.94	1.89
E144	1.04	0.08	0.33	0.06
E165	1.14	0.86	0.15	0.15
E180	0.74	0.63	0.11	0.10

### 6.2.2 运用 BP 神经网络对 302 家企业类别的预测

通过对 302 家企业相关数据的处理结果，本文使用问题 1 所建立的 BP 神经网络对该 302 家企业的类别进行预测，预测部分结果如下，详细见附录：

表 9 302 家企业类别预测结果

企业代号	类别	企业代号	类别
E124	A	E191	A
E125	A	E327	B
E126	B	E340	B
E127	B	E413	C
E128	A	E417	C

对预测结果进行统计，统计结果如下：

表 10 302 家企业类别统计

类别	包含企业(企业代号)
A 类	E124, E125, E128, E130, E131, E132, E133, E135, E136, E137, E140, E145, E146, E147, E148, E150, E151, E153, E156, E158, E159, E160, E161, E162, E163, E164, E165, E167, E169, E177, E191, E197, E202, E206, E207
B 类	E126, E127, E129, E134, E138, E139, E141, E142, E143, E144, E149, E152, E154, E155, E157, E166, E168, E170, E171, E172, E173, E174, E175, E176, E178, E179, E180, E181, E182, E183, E184, E185, E186, E187, E188, E189, E190, E192, E193, E194, E195, E196, E198, E199, E200, E201, E203, E204, E205, E208, E209, E210, E211, E212, E213, E214, E215, E216, E217, E218, E219, E220, E221, E222, E223, E224, E225, E226, E227, E228, E229, E230, E231, E232, E233, E234, E235, E236, E237, E238, E239, E240, E241, E242, E243, E244, E245, E246, E247, E248, E249, E250, E251, E252, E253, E254, E255, E256, E257, E258, E259, E260, E261, E262, E263, E264, E265, E266, E267, E268, E269, E270, E271, E272, E273, E274, E275, E276, E277, E278, E279, E280, E281, E282, E283, E284, E285, E286, E287, E288, E289, E290, E291, E292, E293, E294, E295, E296, E297, E298, E299, E300, E301, E302, E303, E304, E305, E306, E307, E308, E309, E310, E311, E312, E313, E314, E315, E316, E317, E318, E319, E320, E321, E322, E323, E324, E325, E326, E327, E328, E329, E330, E331, E332, E333, E334, E335, E336, E337, E338, E339, E340, E341, E342, E343, E344, E345, E346, E347, E348, E349, E350, E351, E352, E353, E354, E355, E356, E357, E358, E359, E360, E361, E362, E363, E364, E365, E366, E367, E368, E369, E370, E371, E372, E373, E374, E375, E376, E377, E378, E379, E380, E381, E382, E383, E384, E386, E387, E388, E390, E391, E392, E393, E394, E395, E396, E397, E398, E400, E402, E403, E404, E405, E406, E407, E408, E419, E420
C 类	E385, E389, E399, E401, E409, E410, E411, E412, E413, E414, E415, E416, E417, E418, E421, E422, E423, E424, E425

### 6.2.3 问题 2 总结与策略

对 302 家企业类别进行定量化识别后，在银行年度信贷总额为 1 亿元时对这些企业的调整如下：

表 11 302 家企业信贷策略

类别	对策		
	是否贷款	贷款金额（万元）	贷款利率
类别 A	可以贷款	70-100	0.0425-0.0545
类别 B	可以贷款	30-70	0.0545-0.0585
类别 C	考虑不予贷款	0 或 10-20	0.0785-0.0865

### 6.3 问题 3

#### 6.3.1 数据统计

根据附件 2 中的进/销项发票信息,统计出各个企业每月的收支额度部分数据如表 12 所示:

表 12 各企业月收支情况列举

企业代号	时间	收入额(单位: 百万)	支出额(单位: 百万)
E124	2017.09	19.84	19.66
	2017.10	1.54	1.24
	2017.11	18.02	20.28
	2017.12	49.40	25.00
E140	2017.02	0	0.0047
	2017.03	5.90	0.0453
	2017.04	1.28	0.0447
	2017.05	5.75	0.0080
E176	2017.01	2.49	0.14
	2017.02	2.31	0.06
	2017.03	1.52	0.24
	2017.04	2.20	0.13
E152	2016.12	3.02	0.08
	2017.01	1.31	0.02
	2017.02	1.55	0.06
	2017.03	2.09	0.03

#### 6.3.2 数据可视化操作

使用 Excel 作图工具,画出各个企业每月收支额度的折线统计图,部分统计图如下:

##### 1.个体经营类企业

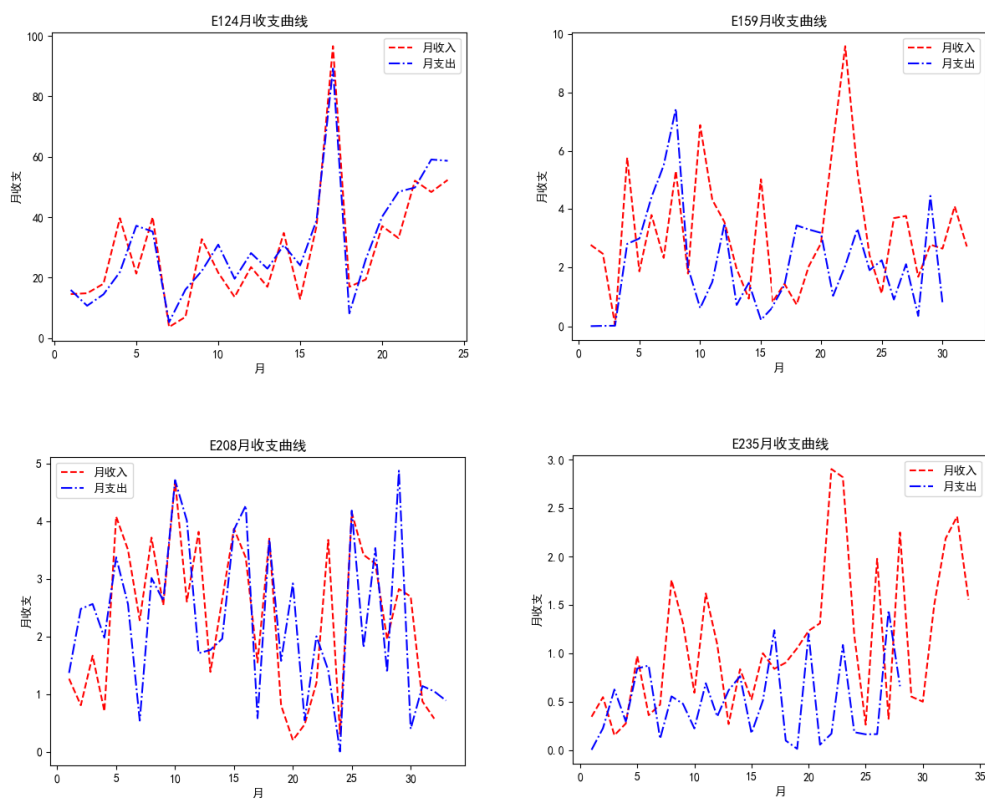


图 7 部分个体经营类企业月收支曲线

## 2.建筑劳务类企业

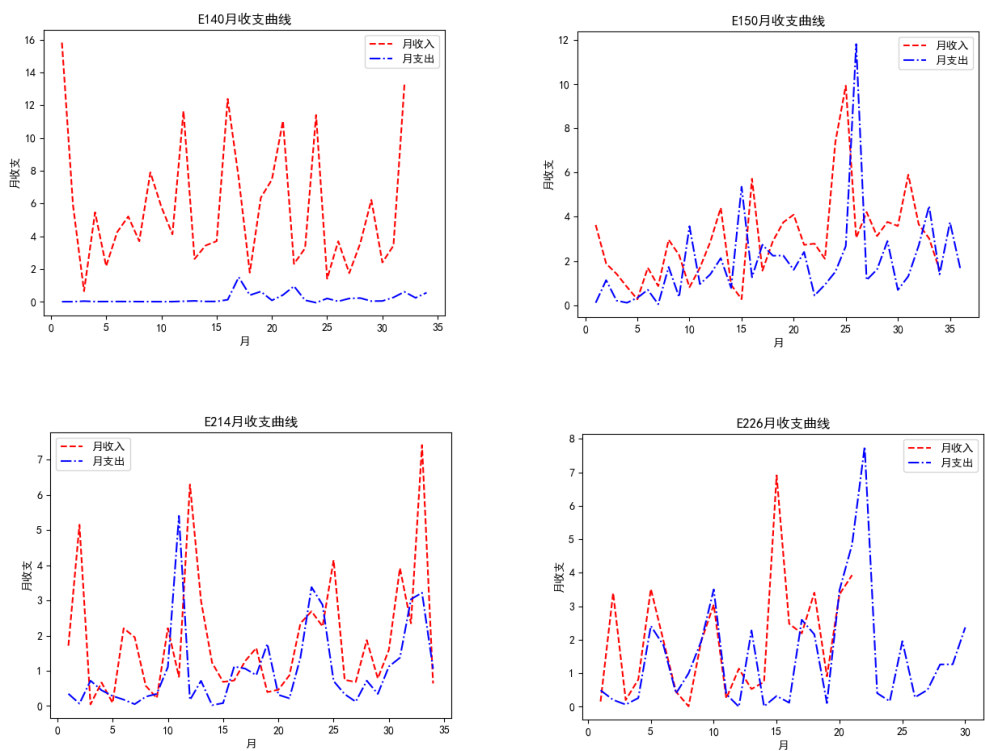


图 8 部分建筑劳务类企业月收支曲线



### 3.科技类企业

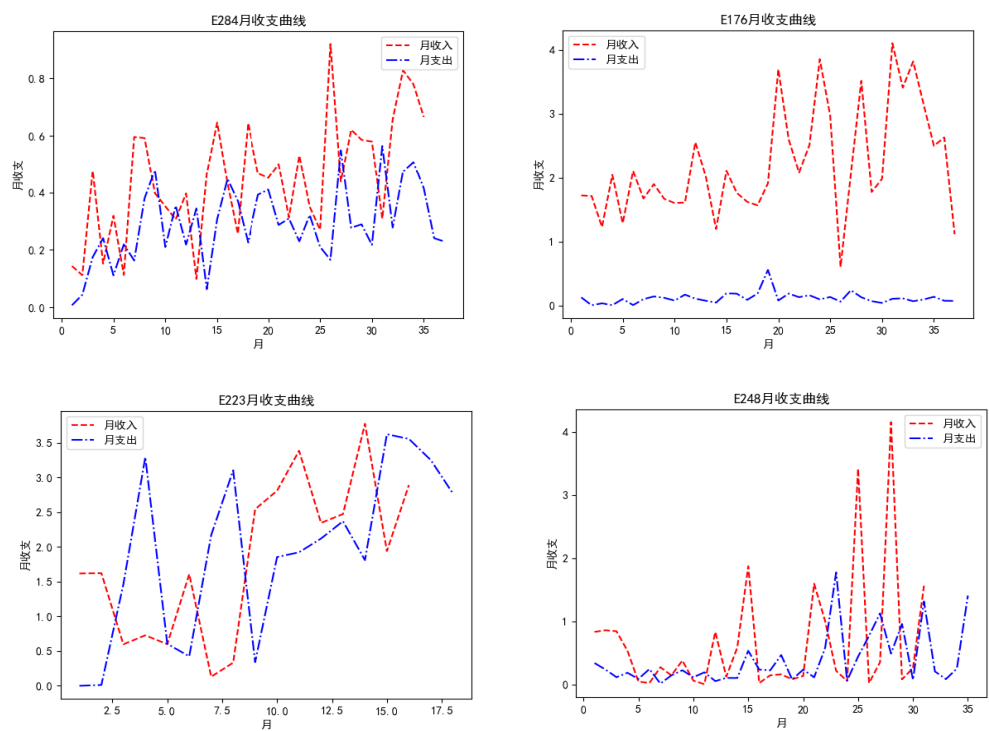


图 9 部分科技类企业月收支曲线

### 4.运输类企业

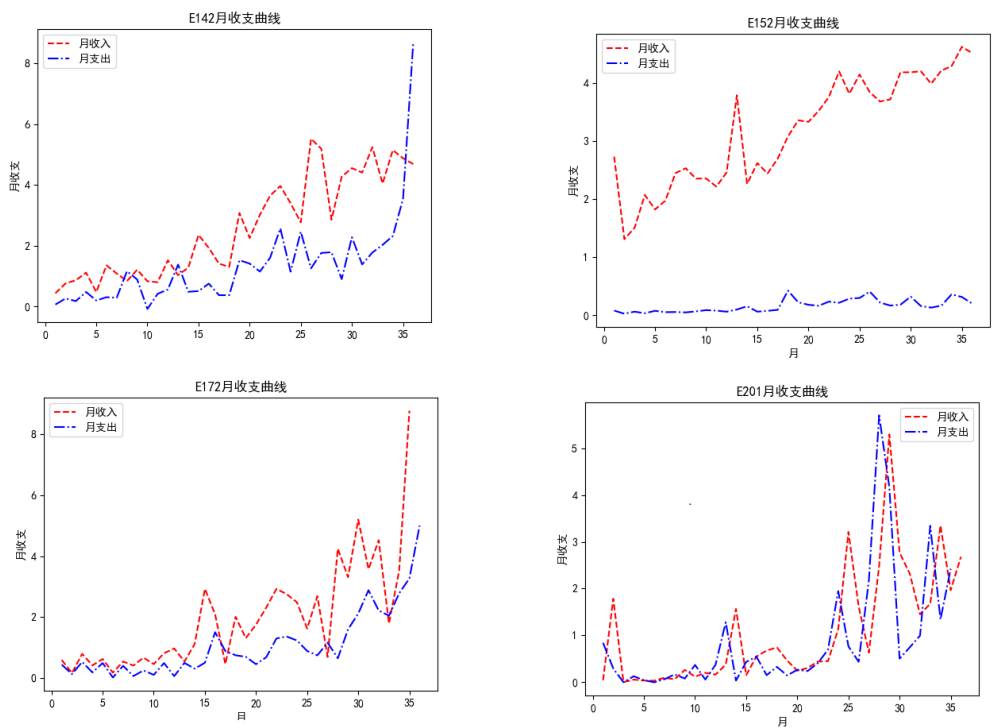


图 10 部分运输类企业月收支曲线

### 6.3.3 问题三总结与策略

对不同行业的月收入统计图分析后，发现：个体经营类行业、科技类行业以及建筑劳务行业应对突发风险的能力较差，运输类行业应对突发风险的能力较强。

基于以上得出的结论，本文给出以下调整策略：

表 13 企业信贷策略调整

类别	信贷策略			
	是否贷款	企业行业	贷款金额（万元）	
类别 A	可以贷款	运输相关	80-95	0.0425-0.0545
		个体、科技、建设	65-75	
类别 B	可以贷款	运输相关	50-75	0.0545-0.0585
		个体、科技、建设	30-45	
类别 C	考虑不予贷款	0 或 10-20	0 或 10-20	0.0785-0.0865

## 七. 模型结果检验与评价

### 7.1 模型检验

随机产生 500 组数据对求解的 BP 神经网络进行检验<sup>[7]</sup>，将得到的公式解与神经网络解进行对比，列出前四项数据如下：

表 14

$x_1$	1.13	1.82	0.02	7.41
$x_2$	0.39	0.07	0.01	7.85
$x_3$	0.70	0.38	0.01	1.94
$x_4$	1.64	0.32	0.02	1.88
公式解	2.56	2.50	1.32	3.17
神经网络解	2.42	2.37	1.38	3.10

分析上述对比结果，证明本文训练出的 BP 神经网络在预测企业类别时具有较高的准确性。

### 7.2 模型评价

本文假设各中小微企业所获取的数据真实有效，通过不断度量理想输出与实际输出的差异，不断修正权值，通过优化的神经网络来预测某一企业的风险高低，从而归属各中小微企业的企业风险类别。在风险评估中运用神经网络模型具有一

定的价值，它运算速度快，结构清晰，在进行大量数据处理时，与普通解析式计算相比，具有显著性优势。得到的 BP 预测值能够对各中小微企业的信贷风险进行有效评估。然而，也可能面临这样的情况，当初始权选择不当，出现权远离任何输入变量的死单元，权将无法得到校准。综上，我们对 BP 神经网络的优缺点进行一些总结。

**BP 神经网络的优点：**（1）BP 神经网络实现了输入到输出的映射，简化复杂性映射问题。（2）自学习和自适应性能高。它沿着逐步改善方向进行调整，通过自行训练和校准，能够修正数据间的规则。（3）具有应用泛化能力。不仅仅局限于已学习的模式进行分类，经过训练，也能保证对所需对象进行正确分类和计量。

**不足：**（1）局部极小化易致使训练失败。以不同权重进行训练，若收敛于局部极小点，易致使评估值精准度不足。（2）执行 BP 算法的收敛速度较慢。

“锯齿现象”、出现平坦区区域等都会致使算法失效。

## 八. 模型的优化

本文所采用的神经网络模型，利用 BP 算法确定权重，也就是说权重的确定过程是一个“学习过程”，这样的学习对于任意一个输入样品，其类别事先是已知的（本文就是通过问题一聚类分析先为所有样本确定了一个标准输出，构建神经网络模型，将所得的模型运用在附件 2 的数据中）。但对于未知样本类别时，BP 算法难以适用。由此，可以将 BP 算法进行改进，运用无监督有竞争的算法 LVQ 算法，可以减小错误分类的机会。

## 九. 参考文献

- [1]张馨予,门玉杰,孙晓红.基于 K-Means 和 Logistic 的寿险客户流失预测模型研究[J].中国商论,2020(17):59-61.
- [2]姜健,陈婧伊,王宇歌.基于 BP 神经网络的客户信用评级模型[J].中国商论,2018(18): 172-173.
- [3]卫伟.基于客户信用评级系统的银行信贷管理分析[J].价值工程, 2018, 37(06): 93-94.
- [4]汪晓英.工商银行中小企业信用评级指标体系的优化研究[D].华东师范大学, 2016.
- [5]刘振华.基于客户信用评级的商业银行信贷管理研究[D].湖南大学,2015.
- [6]张蕾.我国商业银行中小企业信贷风险控制策略研究[J].现代经济信息, 2019 (03): 316.
- [7]张磊. 基于 BP 神经网络的齿圈插齿工艺参数预测与优化[D].内蒙古科技大学,2020.

## 十. 附录

### 10.1 BP 神经网络结果

```
{'w1': array([[ -1.13771304, -0.65323587,  0.95667678,  0.33310259],  
              [ -1.72185864, -0.54600611,  0.01182664,  0.30806613],  
              [ -1.45525483,  0.21736632,  0.59796776,  0.95366505],  
              [ -1.00509382, -0.47522569,  0.48927447,  0.67369765],  
              [ -1.35413099, -0.8415966 ,  0.70303869,  0.61972433],  
              [ -0.71305593, -0.86080329,  0.2414866 ,  0.11545906],  
              [ -1.69867197, -0.57700364,  0.79700808,  0.58544828],  
              [ -0.12452198,  0.08332159,  0.03315084,  0.90001335],  
              [ -2.08547341, -0.78532988,  0.93235491,  0.07572871],  
              [ -0.38133481,  0.13026713,  0.93932421,  0.73625977],  
              [ -1.16364681,  0.11158717,  0.4262322 ,  0.47745652],  
              [ -1.45141814, -0.6470249 ,  0.83717369,  0.94059007],  
              [ -1.09331065, -0.00190005,  0.89376246,  0.704448  ]],
```

```

[-2.04765002, -0.8473034 , 0.33475818, 0.36733715],
[-4.1083983 , -1.86012942, 0.32573553, 0.07731364],
[-1.17247326, -0.42520421, 0.66510419, 0.40097978],
[-3.74081589, -1.35880742, 0.46223683, -0.00447904],
[-2.6002432 , -0.41340434, 0.06339235, 0.52515056],
[-0.14571579, 0.10805947, 0.52783415, -0.02806104],
[-1.9158474 , -0.0352533 , 0.29108223, 0.18376617],
[-0.53414793, -0.21788273, 0.69025932, 0.56959719],
[-0.14233586, 0.08159859, 0.59666538, 0.91397268],
[-1.79132867, -0.68502739, 0.94465959, 0.29871327],
[-1.82834506, 0.08774665, 0.49053215, 0.93018398],
[-0.89293608, -0.54556683, -0.00926777, 0.21831353]]),
'b1':array([[ -2.08270971],[ -1.94602879],[ -2.25584396],[ -2.55307307],[ -1.72619542],
[ -1.61835459],[ -1.98644531],[ -2.36040633],[ -2.2045982],[ -2.58784953],[ -2.6040799
6],[ -2.24793005],[ -2.6631406],[ -2.34779804],[ -1.63473779],[ -2.14242597],[ -1.69113
262],[ -2.1870155],[ -1.73163773],[ -2.27765731],[ -2.59932876],[ -2.41663288],[ -2.265
16786],[ -2.2320188],[ -1.63730288]]),
'w2': array([[ 0.13228312, -0.42436206, -0.1250422 , -0.14769531, 0.10000957,
0.61436796, -0.40839459, 0.87967105, -0.72708665, 0.33947413,
-0.25065612, -0.16902727, -0.23378609, -0.88813267, -2.19274254,
0.10842382, -1.95247144, -1.11554599, 1.32412529, -0.59109536,
0.17523962, 0.63731543, -0.46976646, -0.48651646, 0.48005953]]),
'b2': array([[2.5165286]])}

```

## 10.2 支撑材料目录

- (1) 302 家企业月收入折线图
- (2) Python 代码
- (3) 附件 1 聚类结果
- (4) 附件 1 数据处理结果
- (5) 附件 2 数据处理结果