# 基于粒子群改进动态规划的菜品定价和补货模型

# 摘要

在生鲜超市中由于菜品的保质期较短，如果菜品隔夜将无法进行售卖，但由于商家通常在不明确具体单品和价格的情况下进行第二天的补货决策可能会使利益受损，因此本文中提出一种**基于粒子群改进动态规划**的菜品定价和补货模型。

**针对问题一，**本问首先对数据进行了预处理，对各品类销量数据进行描述型统计分析；其次，应用**皮尔逊系数**计算了各菜品类间的相关性，不同单品间的相关性，各菜品类销量与月份之间的相关性；最后引入**K-Means聚类算法**进一步分析了不同品类间的相关性。可以得到水生根茎类和茄子类具有较强相关性。

**针对问题二**，本文首先利用**线性估算法**分析得到各蔬菜品类的销售总量与成加成定价的函数关系; 其次，本文建立**动态规划模型**分析不同菜品补货量和定价策略，其中辣椒类7月1日补货量为78.144（kg)和定价策略为78.14%。2日- 7日具体补货量和定价策略如文中表5及附录所示。

**针对问题三**，相较第二问，本问中的数据量更大，需要对第二问中的模型增加 两个约束条件，实现对计算过程的简化。然后引入了**粒子群算法**进行动态规划的改进，得到了**优化的动态规划模型。**从而给出 7 月 1 日的单品补货量和定价策略，进而得到商超收益的最大值。

**针对问题四，**为了更好的制定蔬菜商品的补货和定价，商家可以收集在不同季节蔬菜商品的销量也可以收集促销和打折对不同种商品的销量的影响以及要重视客户的反馈信息进而动态的对商品的价格进行调整，进而提升销量。

**关键词：**皮尔逊系数 K-Means聚类算法 动态规划模型 粒子群算法

# 问题重述

## 1.1背景知识

随着国家的稳步发展，人均消费支出连年增加，表面国民在食品消费方面逐渐由满足基本生活要求转向追求更高品质，更多样化的消费结构。国务院在2020年发布的国民食品消费需求目标中指出，蔬菜类食品消费比重应占到31%[1]，而生鲜商超作为主要的居民蔬菜购入来源，在居民消费蔬菜这一环占据重要作用。对于生鲜商超常常面临为了达到补足商品的种类与数量如何平衡，由于运输，天气等原因造成商品损耗问题和如何定价使收益最大化等问题。所以能否建立有效的补货与定价策略是对于提高商超收益具有实质意义。

## 1.2数据分析

附件1给出了每种商品所对应的编号和所属品类，其中同种商品不同编码是因为货源地不同；附件2给出了每种商品从2020年9月到2023年6月每天的销售额，销售单价，促销情况，卖出时间和销售类型；附件3给出了每种商品当天的批发价；附件4记录了每种商品的平均损耗率。

## 1.3具体问题

对于问题一，根据附件2的数据，建立数学模型对不同品类或不同单价进行分析，基于上问所得到的数据，对原有模型进行优化，分析蔬菜品类对单品销售额的影响。

对于问题二，基于问题一 以品类为单位制定补货计划，定价采用“成本加成定价法”，根据福建2和附件3的数据，建立数学模型分析成本加成定价对各个品类销售总量影响，并建立模型，制定商超未来一周的补货和定价计划，使得收益最大。

对于问题三，基于问题二的模型，考虑销售空间，最小陈列量，市场等影响并根据2023年6月24日~2023年6月31日，制定七月的补货与定价策略。

对于问题四，合理进行建议并说出理由

# 二．问题分析

## 2.1问题一的分析

对于问题一第一部分，根据商品品类或各个商品的相关规律建立数学模型，文献[2]是基于图书出版数量与种类的相关性研究，蔬菜类商品虽与图书类商品不同，但文献[2]所用模型有利于本题分析，因此在进行数据预处理基础上，引入皮尔逊相关系数，建立模型对问题一第一部分进行分析；对于问题一第二部分，在基于第一部分的基础下，因对现有模型进行效率，可视化等结果的优化，因此问题一第二部分本文引入kmeans聚类模型

## 2.2问题二的分析

对于问题二第一部分，在第一问基础上，本文利用统计软件，对各商品与均值定价进行线性估算分析，得出各蔬菜拟合图像，得出销售总量与成本加成定价的关系。对于问题二第二部分，目前，有关超市补货策略大多数基于季节或地区，文献[3]分析了基于季节性的补货与定价策略，与本文针对基于成本加成定价法的补货与定价策略并不相符，本部分难点在于受到成本加成的影响，如何能在低成本的情况下如何定价与补货才能做到未来一周收益最大，针对此类问题，本文采用动态规划模型进行求解。

## 2.3问题三分析

基于问题二所得数据与模型，本问又新增加了约束条件，如若还是沿用问题二的模型，计算过程将会先当复杂，因此本文引入粒子群算法， 粒子群算法具有很强大的全局搜索能力，可以在复杂的高位搜索空间中找到全局最优解，当它与动态规划结合时可有效提高模型的效率[4]

## 2.4问题四分析

对于问题四，我国古代素有“冬吃萝卜夏吃姜”的民俗，因此可以知道，季节的变化会影响个人对蔬菜类商品购买的变化，因此本文建议商家对不同季节蔬菜效率进行统计和分析；居民在收到打折或类似消息时会更具有消费热情[5]，这在生活中非常常见，因此本文建议商家进行打折等活动，搜集数据并分析。生活中意见箱随处可见，这可以帮助商家改进缺点，因此本文建议收集不同顾客的反馈以对蔬菜价格进行动态调整

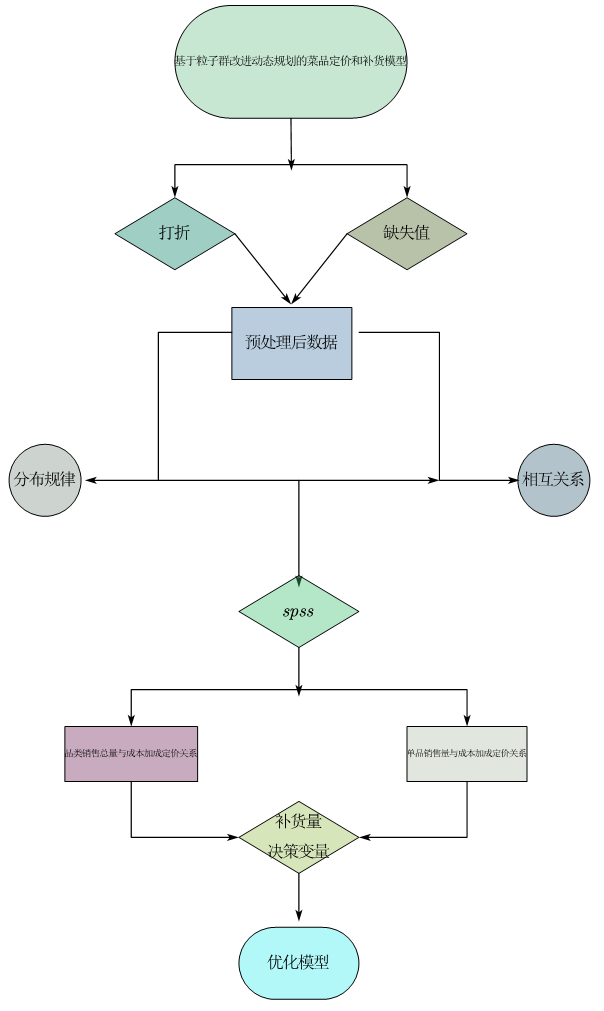


图 1 全文求解思维导图

# 模型假设

1.价格对销售量有一定的影响

2.数据在短时间内较为稳定基本不会突变

3.销售数据较为稳定，无异常值

# 符号说明

|  |  |
| --- | --- |
| 符号 | 说明 |
|  | 样本协方差 |
|  | 样本皮尔逊相关系数 |
|  | 平均值 |

# 五、模型的建立与求解

## 5.1数据预处理

### 空缺值的处理办法

在对附录二的数据进行处理的过程中，本文发现题目中给出的数据并不完整，某些品类在个别月份出现了没有销量额的问题。比如花菜类在2月和5月之间并没有给出销售量。但为了得到可靠的分析结果，本文根据相邻月份的销售额运用了中值插补法（及缺失值由该列的中值进行代替）对数据进行了填补。

### 单品打折的价格变化问题

在对附录二的数据进行处理的过程中，本文发现部分单品价格在一个月内存在因打折而导致的价格变化问题，但因打折商品出现的数量为20个远远少于未打折商品的数量因此我们采用替换的方法进行处理。

### 各种类菜品在一个月内销售额的计算方法

其中i代表不同种类，k代表天数，j代表不同的月份表示该天的这类的销量，代表不同类在每月的销量，代表单价，代表月销量

|  |  |
| --- | --- |
|  | () |

|  |  |
| --- | --- |
|  | () |

## 5.2问题一模型的建立与求解

### 5.2.1蔬菜各品类随时间的分布规律

描述性统计分析法介绍

描述性统计分析法是指通过图表或数学方法，进行对数据资料的整理和分析，并对数字的特征，随机变量和数据的分布状态的关系进行估计和描述的方法。

利用描述性统计分析法求解过程

1. 方差的计算

方差是反映曲线波动状态的一种变量，为了探寻不同种类菜品受到月份的影响程度，本文将其引入进行评估。方差的计算方法如下，其中表示方差，表示变量值，表示平均数，表示变量总数。

|  |  |
| --- | --- |
|  | （3） |

1. 基本特征值计算

最小值，最大值，均值，标准偏差是常用的描述自变量和因变量之间关系的量，本文将其引入来反应基本特征。进行计算后得到的数据如下

**表1 各类蔬菜基本特征值**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | N | 最小值 | 最大值 | 均值 | 标准偏差 | 方差 |
| 花叶类 | 36 | 2031.210 | 15678.000 | 8301.83844 | 3916.546605 | 15339337.311 |
| 花菜类 | 36 | 20.040 | 894.160 | 215.35500 | 253.370195 | 64196.456 |
| 水生根茎类 | 36 | 1895.33 | 7821.15 | 3919.1567 | 1885.99529 | 3556978.219 |
| 辣椒类 | 36 | 7417.440 | 65618.180 | 20968.68297 | 11032.193908 | 121709302.425 |
| 食用菌类 | 36 | 502.1080 | 15718.1524 | 2973.871522 | 3491.6888166 | 12191890.792 |
| 茄子类 | 36 | 1008.4100 | 10376.1375 | 4856.089206 | 2269.5126062 | 5150687.470 |

1. 绘制直方图，曲线图

在已经求解出了每类蔬菜的基本特征值的基础上，本文进一步对数据进行可视化分析，以月份为横坐标，以月销量为纵轴绘制出直方图与曲线图，结果如下所示。

图 2不同类蔬菜随时间变化的曲线图

图 3不同类蔬菜随时间变化直方图

1. 结合特征值，曲线图，直方图进行分析

由表一可以看出水生根茎类和茄子类方差较为接近，又根据图一的曲线图大致判断，两者具有相似的关系，由表一可以看出辣椒类不仅拥有最大的方差并且有最的值，因此初步判断辣椒受季节影响较大，然而花菜类方差较小说明花菜受季节的影响较小。

### 5.2.2蔬菜各品类的相关性分析

皮尔逊相关系数

皮尔逊相关系数又被称为皮尔逊积矩相关系数用于度量不同变量间的相关系数。求解皮尔逊相关系数的步骤如下。

Step1:首先对两组数据进行z分数处理，对两种数据每组取乘积求和后再除以样本数

Step2:z分数代表正态分布中数据相较于中心点距离的偏移量通常与变量与均值的差值在除样本数

Step3:标准差则等于变量减掉平均数的平方和，再除以样本数,最后再开方

可以表示为

|  |  |
| --- | --- |
|  | （4） |

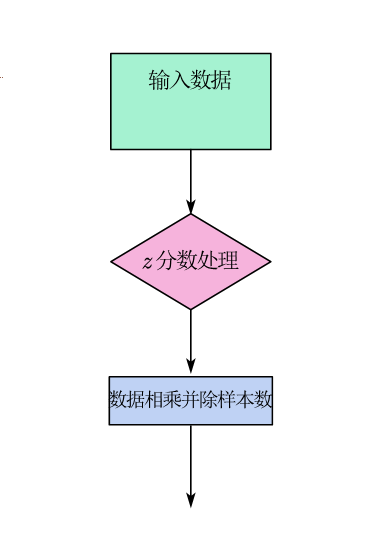


图 4皮尔逊系数求法

皮尔逊系数的数学公式如下

|  |  |
| --- | --- |
|  | （5） |

经过计算后得到数据如下

**表2各类菜品间的皮尔逊相关系数**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 花叶类 | 皮尔逊相关性 | 1 | 0.258 | 0.003 | 0.382 | 0.011 | 0.182 |
| 花菜类 | 皮尔逊相关性 | 0.258 | 1 | -0.137 | 0.196 | 0.272 | 0.263 |
| 水生根茎类 | 皮尔逊相关性 | 0.003 | -0.137 | 1 | 0.000 | -0.350 | -0.256 |
| 辣椒类 | 皮尔逊相关性 | 0.382 | 0.196 | 0.000 | 1 | 0.625 | -0.083 |
| 食用菌类 | 皮尔逊相关性 | 0.011 | 0.272 | -0.35 | 0.625 | 1 | 0.005 |
| 茄子类 | 皮尔逊相关性 | 0.182 | 0.263 | -0.27 | -0.08 | 0.005 | 1 |

### 微信图片_20230909140112

图 5皮尔逊系数热力图

皮尔逊相关系数表的解读

(1)数值解读

表格中的数值表示各蔬菜品类间的相关性，数值都在-1和1之间，该数值可以作为衡量两个变量间线性相关程度的指标，数值越接近1越表明两变量间存在正相关性关系，数值越接近-1越表明两变量间存在负相关性关系。

(2)相关性解读

花叶类和辣椒类之间存在较强的正相关关系（相关系数为0.382）。这意味着花叶类和辣椒类的特征变量可能具有一定的线性关系。

花菜类和水生根茎类之间存在较弱的负相关关系（相关系数为-0.137），这意味着花菜类和水生根茎类的特征变量之间的线性关系较弱，可能是由于随机误差引起的。

茄子类和食用菌类之间存在较弱的正相关关系（相关系数为0.263），这意味着茄子类和食用菌类的特征变量之间的线性关系较弱，可能是由于随机误差引起的。

### 5.2.3不同菜品间的相关性求解

1. 数据的处理

由题目中给出的数据可以看出各品类中蔬菜种类较多，然而再给出关系时我们更在意各品类中销量第一的商品因此在本文中选择了各品类中销量第一的蔬菜利用皮尔逊系数进行相关性分析。得到的数据如下

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | 紫茄子 | 螺丝椒 | 西陕花菇 | 水生 | 洪湖莲藕 | 西兰花 |
| 紫茄子 | 皮尔逊相关性 | 1 | 0.183 | 0.027 | 0.010 | -0.009 | -0.265 |
| 螺丝椒 | 皮尔逊相关性 | 0.183 | 1 | 0.032 | 0.001 | -0.012 | -0.312 |
| 西陕花菇 | 皮尔逊相关性 | 0.027 | 0.032 | 1 | -0.032 | -0.004 | -0.057 |
| 水生 | 皮尔逊相关性 | 0.010 | 0.001 | -0.032 | 1 | 0.022 | 0.087 |
| 洪湖莲藕 | 皮尔逊相关性 | -0.009 | -0.012 | -0.004 | 0.022 | 1 | 0.049 |
| 西兰花 | 皮尔逊相关性 | -0.265 | -0.312 | -0.057 | 0.087 | 0.049 | 1 |

**表3各品类中单品销量第一间的关系**

1. 规律分析

通过观察相关性系数的绝对值大小，我们可以判断变量之间的相关性强度。在这个数据中，紫茄子和螺丝椒之间的相关性系数为0.183，属于较弱的正相关关系。其他变量之间的相关性系数也都比较小，说明它们之间的相关性较弱。

### 5.2.4聚类分析

K-Means聚类分析法

K-means聚类算法是一种迭代求解的聚类求解算法。其中的K值为初始类簇中心点的数量，means代表簇中的点到簇中中心点的距离均值。该算法通过预先设定的k值和每类别初始质心对相似的点进行划分，并通过划分后进行迭代得到最优解。其原理公式如下，其中表示进行k次划分后到簇中心点的距离值，表示是到样本中心点的距离。

|  |  |
| --- | --- |
|  | （6） |

使用该算法步骤如下：

Step1:初始时随机选择k个点为聚类质心。

Step2:计算数据对象到质心间的距离，并根据距离远近进行点的重新分配。

Step3：针对每个聚类，分别重新进行新的质心选取，选取新的质点依据各点间的距离均值

Step4:进行n次迭代直到数据对象所属的簇不再变换，停止迭代并返回k的值

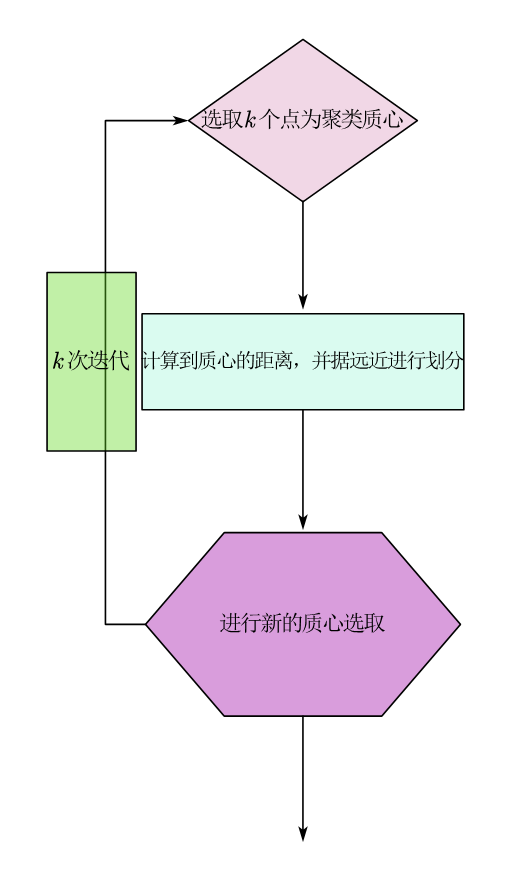


图 6 k-means算法示意图

1. k值的选择

K-Means聚类算法最后所得到的最优解与所选取的k值密切相关，k的取值直接影响了最后的结果是否为最优解，为了确定k的最优值我们引入了轮廓系数进行求解，轮廓系数结合了样本与其所分配的簇内距离和与其他簇之间的距离，可以帮助我们确定聚类结果的紧密度和分离度。计算k值的步骤如下：

Step1:对于每个样本，计算它与同簇其他样本的平均距离（a），表示该样本与自身簇的紧密度。

Step2:对于每个样本，计算它与最近其他簇中样本的平均距离（b），表示该样本与其他簇的分离度。

Step3:计算样本的轮廓系数，即 。

Step4:对所有样本的轮廓系数取平均得到聚类的整体轮廓系数。

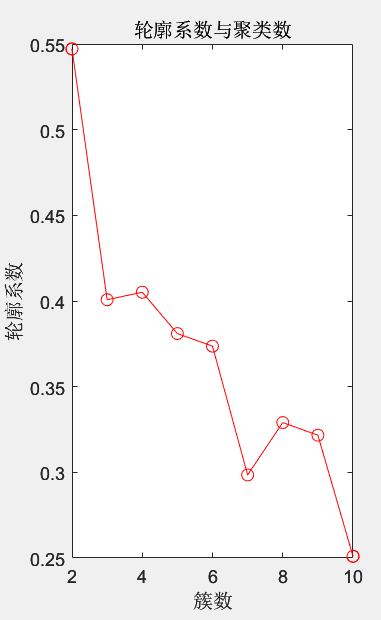


图 7轮廓系数与聚类数

廓系数的取值范围在[-1, 1]之间。如果轮廓系数接近1，则表示聚类结果较好，样本与同簇的距离远小于与其他簇的距离；如果轮廓系数接近-1，则表示聚类结果较差，样本更适合分配到其他簇；如果轮廓系数接近0，则表示样本位于两个簇的边界上据图5可知当k=2是轮廓系数最接近1，故选取k为2为划分数，为了进行验证，本文中又绘制了聚类预测图进行对照

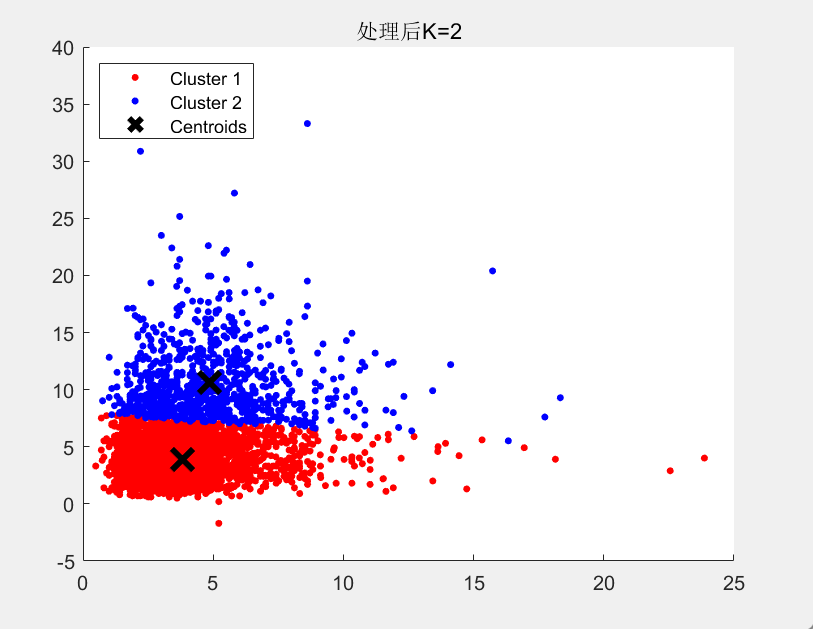


图 8聚类预测图

在选取数据时，为了更好的表现各类之间存在的相关性，因此采用选取各蔬菜类下的峰值品类作为聚类分析的要素，这样才能更直观的展现出各类之间存在的关系。如图所示，在计算轮廓系数时，我们使用的是K-Means函数的默认距离度量方法，即欧氏距离（Euclidean Distance）。欧氏距离是最常用的距离度量方法，它衡量的是样本在多维空间中的直线距离。对于两个样本点x和y，它们之间的欧氏距离可以通过以下公式计算：



其中，, , ...,和,, ..., 分别表示两个样本点在n个维度上的特征值。

由图得知，使用欧氏距离计算出距离长度后，发现图像中点的分布呈现两个部分，其中聚类1（蓝色）的重心对应的特征值为：紫茄子、螺丝椒、西陕花菇 。 聚类2的重心对应的特征值为：水生、洪湖莲藕、西兰花。由此可以从峰值的规律中体现出各个蔬菜类之间的关系。

为了更好的验证K=2，我们还构建了K=3时，各蔬菜类之间的层次聚类分析图，发现K=2与K=3拥有相同的视觉，因此最优解为K=2.

## 5.3问题二模型的建立与求解

### 5.3.1各蔬菜品类的销售总量与成本加成定价的关系

对于问题二，本文利用了spss(统计软件）进行了线性估算分析并在以定价均值为横坐标，销售总额为纵坐标的坐标系中拟合出了函数关系。各蔬菜类拟合后的函数图像如下所示

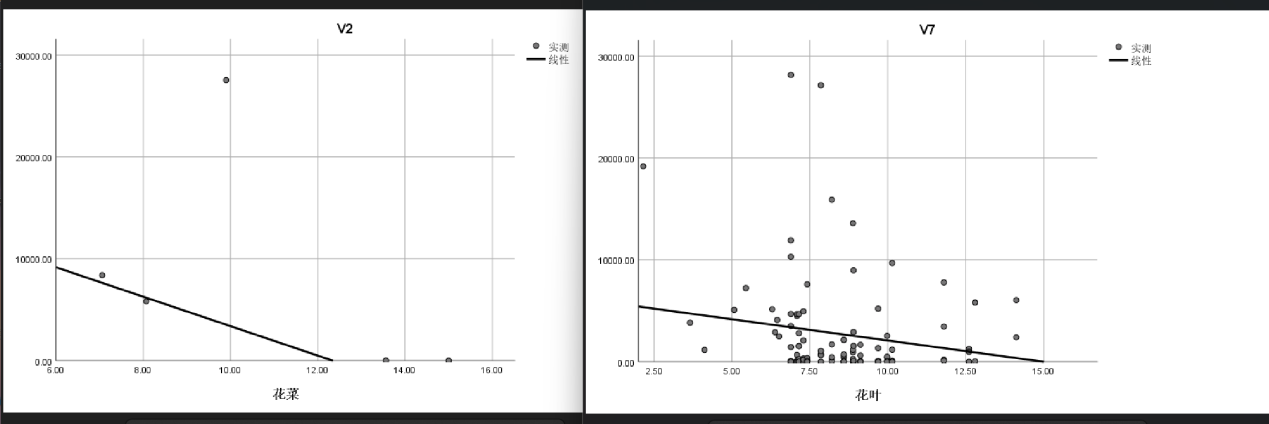


图 9花菜，花叶函数关系

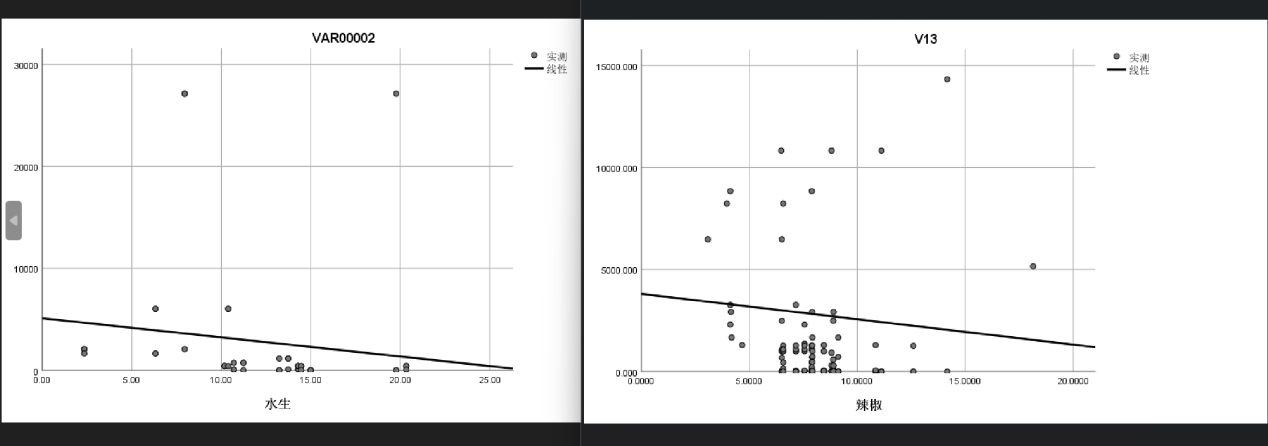


图 10水生，辣椒函数关系

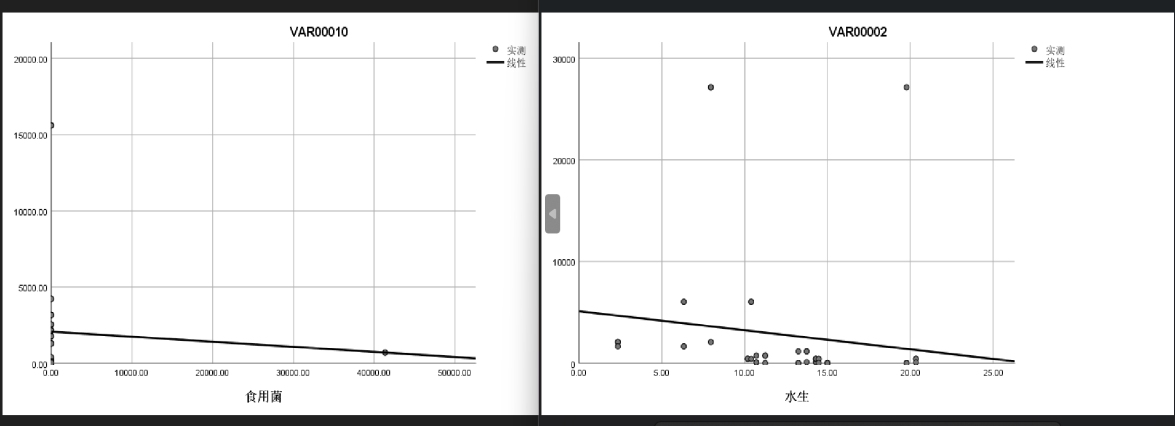


图 11食用菌，水生函数关系

**表4 各类蔬菜方程关系**

|  |  |
| --- | --- |
| 蔬菜的种类 | 方程 |
| 花叶 | y1=-417.38 x + 6262.13 |
| 辣椒 | y2=-124.2\* x + 3812 |
| 水生 | y3=-187.58x + 5130.46 |
| 茄子 | y4=-310\*x + 4732.148 |
| 食用菌 | y5=-0.0332 \* x + 2079.902 |
| 花菜 | y6=-1445.09 \* x + 17842.82 |

## 5.3.2各蔬菜品类未来一周日补货总量和定价策略

**表5数据量化处理**

|  |  |
| --- | --- |
| 量化前 | 量化后 |
| 花叶，辣椒，水生，茄子，食用菌，花菜 | 1，2，3，4，5，6 |

为了提供使商超在未来一周（7月1日到7月7日）的利润最大的补货总量和定价策略，本文建立了动态规划模型进行求解。在正文中，本文仅提供辣椒类的结果，其他类放在了附录中。

本文选取了(i从1到7，表示天数）（j从1到6,表示菜品的种类)为决策变量，及选取第i天内不同类类销售总量为决策变量。

本文中选取了成本加成定价的定价策略，设利率为，故定价可以表示为。

对于不同类第 i 日的盈利可表示。

|  |
| --- |
|  |

一周内商超的总盈利,j表示蔬菜品类。而约束条件可设为6.1中的销售总量与成本加成定价的函数关系。综上所示，动态规划模型为：

|  |  |
| --- | --- |
|  |  |
|  |  |

对辣椒类得到的结果如下

**表五辣椒类每日的日补货量与定价策略**

|  |  |  |
| --- | --- | --- |
| 日期 | 日补货总量 | 定价策略 |
| 2023-07-01 | 74.799 | 78.14% |
| 2023-07-02 | 109.071 | 107.95% |
| 2023-07-03 | 110.004 | 110.55% |
| 2023-07-04 | 42.799 | 40.09% |
| 2023-07-05 | 110.241 | 112.02% |
| 2023-07-06 | 110.912 | 110.92% |
| 2023-07-07 | 71.183 | 79.41% |

对花叶类得到的结果如下

**表五花叶类每日的日补货量与定价策略**

|  |  |  |
| --- | --- | --- |
| 日期 | 日补货总量 | 定价策略 |
| 2023-07-01 | 209.478 | 32.25% |
| 2023-07-02 | 259.814 | 28.63% |
| 2023-07-03 | 260.746 | 41.77% |
| 2023-07-04 | 165.00 | 21.44% |
| 2023-07-05 | 262.049 | 90.79% |
| 2023-07-06 | 261.375 | 39.44% |
| 2023-07-07 | 203.368 | 53.39% |

对花菜类得到的结果如下

**表五花菜类每日的日补货量与定价策略**

|  |  |  |
| --- | --- | --- |
| 日期 | 日补货总量 | 定价策略 |
| 2023-07-01 | 52.162 | 50.01% |
| 2023-07-02 | 20.275 | 25.08% |
| 2023-07-03 | 40.894 | 50.11% |
| 2023-07-04 | 39.897 | 88.12% |
| 2023-07-05 | 51.014 | 40.89% |
| 2023-07-06 | 32.021 | 39.47% |
| 2023-07-07 | 29.149 | 51.37% |

对茄子类得到的结果如下

**表五茄子类每日的日补货量与定价策略**

|  |  |  |
| --- | --- | --- |
| 日期 | 日补货总量 | 定价策略 |
| 2023-07-01 | 9.451 | 92.56% |
| 2023-07-02 | 8.83 | 40.88% |
| 2023-07-03 | 5.712 | 63.79% |
| 2023-07-04 | 11.012 | 106.71 |
| 2023-07-05 | 3.305 | 96.28% |
| 2023-07-06 | 6.514 | 110.15% |
| 2023-07-07 | 6.214 | 83.15% |

对水生根茎类得到的结果如下

**表五水生根茎类每日的日补货量与定价策略**

|  |  |  |
| --- | --- | --- |
| 日期 | 日补货总量 | 定价策略 |
| 2023-07-01 | 22.013 | 83.79% |
| 2023-07-02 | 12.046 | 43.89% |
| 2023-07-03 | 16.121 | 87.89% |
| 2023-07-04 | 10.012 | 33.71% |
| 2023-07-05 | 17.895 | 53.04% |
| 2023-07-06 | 11.114 | 34.12% |
| 2023-07-07 | 15.017 | 39.54% |

对食用菌类得到的结果如下

**表五食用菌类每日的日补货量与定价策略**

|  |  |  |
| --- | --- | --- |
| 日期 | 日补货总量 | 定价策略 |
| 2023-07-01 | 63.011 | 32.89% |
| 2023-07-02 | 65.456 | 99.79% |
| 2023-07-03 | 32.045 | 106.66% |
| 2023-07-04 | 50.406 | 85.15% |
| 2023-07-05 | 47.012 | 93.04% |
| 2023-07-06 | 59.164 | 92.30% |
| 2023-07-07 | 62.015 | 62.48% |

## 5.4问题三的求解

### 5.4.1含有约束条件的动态规划模型的建立

通过对问题二中的预报模式的求解，为每种蔬菜分别建立了预报模式。并在此基础上，建立了同一类蔬菜中不同类型蔬菜的逻辑回归模式，以此模式为依据，对问题三进行了判定。

在问题三中，利用一个整数规划的方法，得到了一个更深层次的单件补给方案。整型规划是指超市想要将可销售商品数量控制在27-33件，而且每件商品的订单数量要达到最低展示量2.5公斤，这是一个数学上的最优化问题。

在此基础上，提出了一种新的整数规划模型。对于超商获得的利润，我们沿用问题二中对成本加成定价的理解，对商品的成本加成定价进行分析。

分析前通过数据可知，部分商品在当天并未出售，因此我们需要设定一个开关变量设定 作为判断是否出售的依据，如果=0说明该商品今日并未售出，反之说明该商品今日售出。同时限定总数为27-33.



对于各个单品最小量表示，我们用表示单品订购最小约束，i = 1,2,3,...25 ,因此可以表示为:

因此用 作为开关变量。利利润率。因此，每种单品的定价可以表示为其中，  表示第 i 种单品定价，  表示第 i 种单品的成本价格。最终总盈利表示为因此有了一下的优化模型：

s.t.

### 5.4.2粒子群算法简介

微粒群算法（PSO）是除了蚁群和鱼群算法以外的另一类智能优化算法。

粒子群优化算法起源于鸟类的捕食行为，即在其附近搜索目标，寻找目标的最佳方法是寻找目标。

微粒群算法从这些物种的群体特性出发，利用微粒群的特性来解决这些问题，其中，每一个微粒都是一个可能的解决方案，每一个微粒群都有一个适合度函数。微粒的运动速度决定着微粒的运动方向和运动距离，微粒的运动速度随着微粒自己的运动和其它微粒的运动而变化，使微粒在可解空间内进行寻优。微粒群是由一组随机粒子组成的,再用迭代法求出最优解。

创建一个维度j，其中有N个粒子。表示粒子i在维度j上的速度，表示粒子在当前维度上的运动速度。代表两个随机数，用于调整粒子的速度。它们通常取值在0和1之间，用于控制粒子在当前维度上的速度调整幅度。rand() 为 0 到 1 之间随机数。表示粒子i在维度j上的个体最优解，表示粒子在当前维度上的最佳位置。粒子i在维度j上的位置，表示粒子在当前维度上的位置

在每次循环过程中，粒子都会追踪到两个“极端”。当确定了这两个最优解之后，该微粒就会根据下列公式调整自身的运动速度和运动位置。

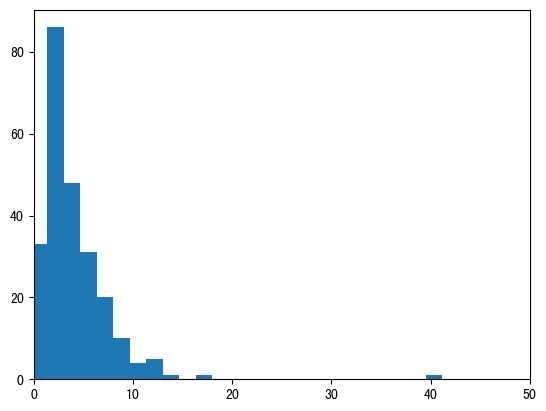


图 12优化前的单品总收益图

为惯性因子，其值非负。数值越大，则寻找最优解的能力越强，局部寻找最优解的能力越弱，反之结论正好相反。我们这里采用线性递减权值策略。

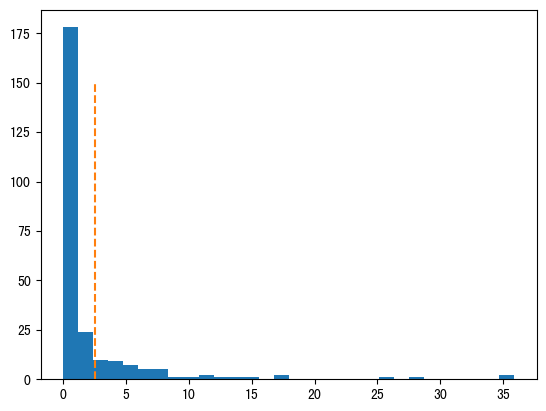


图 12优化后的单品总收益图

Gk 为最大迭代次数，ini 为初始惯性权值，end 迭代到最大进化数时的惯性权值。优化数据前利润高的产品损耗大，随着修正数量的不断增大，损耗依旧存在。对数据进行优化后，发现损耗值小于平均值，则认为优化成功，下面是粒子群模型步骤

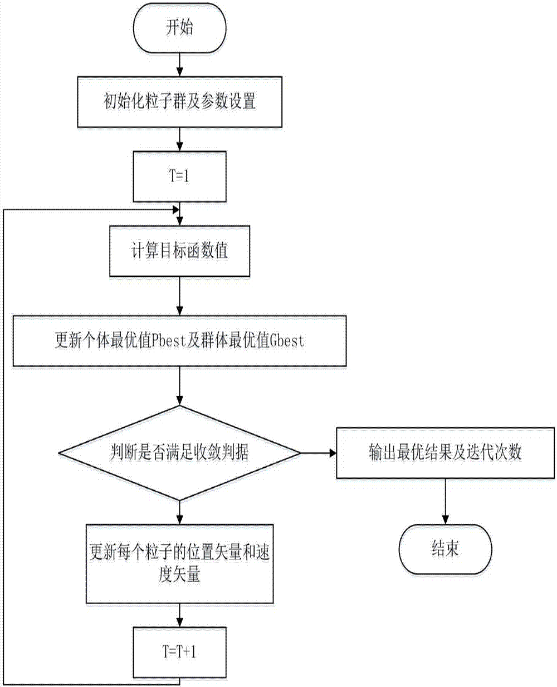


图 13粒子群算法示意图

最终完成实验三的结果：

表 5：问题三结果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 单品编码 | 日补货总量 | 定价策略 | 单品编码 | 日补货总量 | 定价策略 |
| 102900005115625 | 3.048 | 108.62% | 102900011035078 | 2.547 | 103.12% |
| 102900005115779 | 10.394 | 129.63% | 102900011035740 | 1.089 | 82.17% |
| 102900005115878 | 9.118 | 140.86% | 102900011021842 | 7.154 | 70.29% |
| 102900005115861 | 17.891 | 96.13% | 102900011023464 | 5.389 | 125.70% |
| 102900005116257 | 2.315 | 45.79% | 102900011030059 | 13.121 | 99.07% |
| 102900005116530 | 18.040 | 44.56% | 102900011030 097 | 9.198 | 94.51% |
| 102900005116714 | 17.948 | 61.05% | 102900011030110 | 15.114 | 135.71% |
| 102900005116899 | 25.24 | 127.79% | 102900011030905 | 1.123 | 62.34% |
| 102900005122654 | 2.289 | 60.45% | 102900011031100 | 19.198 | 115.42% |
| 102900011013274 | 2.121 | 119.67% | 102900011031216 | 13.7 | 116.02% |
| 102900011016701 | 27.46 | 54.79% | 106949711300259 | 20.604 | 72.11% |
| 102900011033944 | 3.342 | 132.92% | 102900051010455 | 7.994 | 95.93% |
| 102900011034231 | 13.791 | 69.94% | 102900051000463 | 4.210 | 40.37% |
| 102900011034330 | 10.187 | 54.02% | 102900011008164 | 6.501 | 36.01% |
| 102900011034439 | 5.179 | 59.11% | 1029000051 15960 | 17.523 | 88.51% |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 菠菜 | 云南生菜 | 甜白菜 | 上海青 | 菜心 | 黄白菜(2) | 云南油麦菜 | 小青菜(1) | 大白菜 |
| 补货量 | 5.446 | 16.611 | 4.892 | 7.942 | 4.695 | 8.340 | 10.759 | 3.412 | 17.523 |
| 利润（元/千克） | 3.705 | 3.141 | 3.110 | 3.055 | 2.862 | 2.822 | 2.707 | 2.430 | 0.829 |
| 利润（元） | 20.179 | 52.180 | 15.213 | 24.266 | 13.438 | 23.536 | 29.121 | 8.290 | 14.535 |
|  |  |  |  |  |  |  |  |  |  |
|  | 红薯尖 | 竹叶菜 | 奶白菜 | 苋菜 | 云南生菜(份) | 娃娃菜 | 云南油麦菜(份) | 小皱皮(份) |  |
| 补货量 | 5.964 | 7.560 | 6.073 | 5.325 | 14.956 | 16.289 | 9.238 | 5.403 |  |
| 利润（元/千克） | 2.341 | 1.963 | 1.946 | 1.798 | 1.784 | 1.766 | 1.716 | 1.656 |  |
| 利润（元） | 13.960 | 14.841 | 11.815 | 9.574 | 26.689 | 28.761 | 15.848 | 8.948 |  |
|  |  |  |  |  |  |  |  |  |  |
|  | 西兰花 | 净藕(1) | 长线茄 | 紫茄子(2) | 青茄子(1) | 红椒(1) | 螺丝椒 | 小米椒(份) |  |
| 补货量 | 19.794 | 20.423 | 2.891 | 16.848 | 3.538 | 4.720 | 10.637 | 14.788 |  |
| 利润（元/千克） | 3.273 | 2.296 | 4.105 | 2.950 | 2.459 | 6.410 | 3.617 | 2.522 |  |
| 利润（元） | 64.789 | 46.895 | 11.869 | 49.702 | 8.701 | 30.255 | 38.478 | 37.288 |  |
|  |  |  |  |  |  |  |  |  |  |
|  | 芜湖青椒(1) | 螺丝椒(份) | 西峡花菇(1) | 杏鲍菇(1) | 金针菇(1) | 白玉菇(袋) | 双孢菇(盒) | 金针菇(盒) |  |
| 补货量 | 53.653 | 11.241 | 2.686 | 3.366 | 7.346 | 3.646 | 7.997 | 29.099 |  |
| 利润（元/千克） | 2.180 | 2.145 | 7.463 | 4.112 | 3.282 | 2.703 | 1.859 | 1.340 |  |
| 利润（元） | 116.956 | 24.118 | 20.046 | 13.841 | 24.110 | 9.853 | 14.870 | 38.994 |  |

求得利润最大值为 881.955元 （其余见支撑材料：第三问最终结果.excel）

## 5.5问题四的求解

（1）蔬菜需求量可能存在季节性变化，因此本文建议商家对不同季节蔬菜商品的销量进行统计收集并进行分析，这样商家就可以根据不同季节来进行商品价格的动态调整，尽可能的提升利润。

（2）商家还可以对蔬菜商品的促销和折扣活动进行记录，可以对商品打折后的销量进行分析与记录。观察打折和促销是否会对商品的价格产生影响

（3）本文建议商家要注重对客户反馈进行分析，如调查消费者对不同种蔬菜的喜爱和对蔬菜商品的价格建议，进而进行动态的调整。

# 六、模型评价

## 6.1模型的优点

1. 本文中充分考虑到庞大的数据集对结果的影响，问题一选择k均值聚类模型，高效的完成对本题大数据集的聚类，使结论更加可靠
2. 问题二利用线性回归模型，简单直观的表述出蔬菜品类中销售总量与成本加成定价的线性关系。
3. 问题三含有多个影响因素，本文选用粒子群算法与动态规划模型结合使用，利用粒子群算法的全局搜索能力和动态规划模型的最优子结构性质，使本模型更加客观，效率更加快速。

## 6.2模型的缺点

1. 对于问题一，本文采用以皮尔逊相关系数为相关变量的聚类模型，由于在应用皮尔逊相关系数聚类之前，需进行对数据的预处理，以确保满足聚类的假设。
2. 对于问题二，本文采用动态规划模型预测未来一周的补货与定价策略，由于数据集的庞大，可能导致占用过多存储空间
3. 对于问题三，本文采用了粒子群算法与动态规划模型的结合的方法，可能会导致过度设计，导致维护成本增加。

# 参考文献

[1]王春越. 考虑保鲜方式的C公司蔬菜配送路径优化的研究[D].北京交通大学,2021.

[2]段珅,王大海,姚唐,邱琪.自我建构对消费者促销方式偏好的影响[J].管理科学,2018,31(06):128-144.

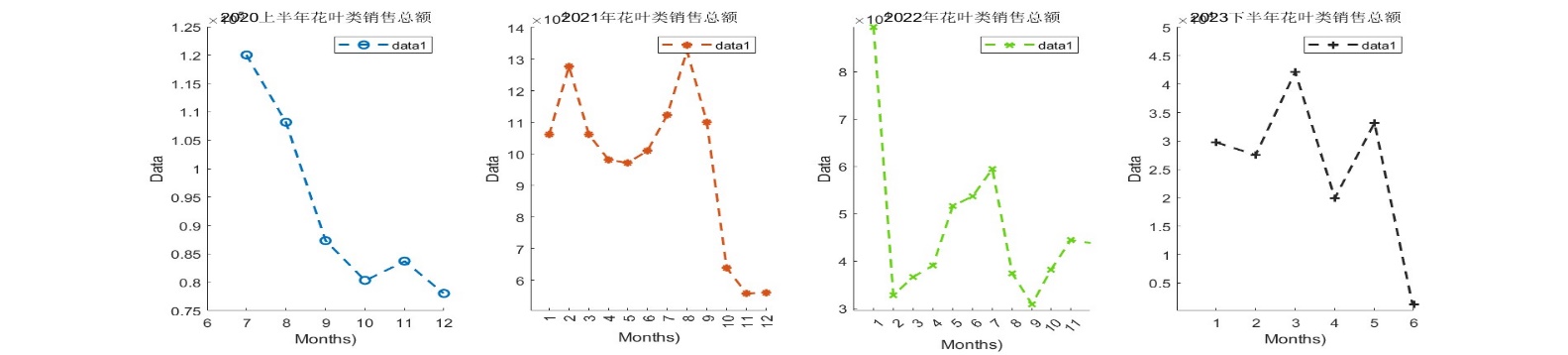
[3]汪妍,蒋多.“一带一路”图书出版现状及相关性新思考——基于“丝路书香出版工程”皮尔逊系数研究[J].科技与出版,2020(12):108-115.

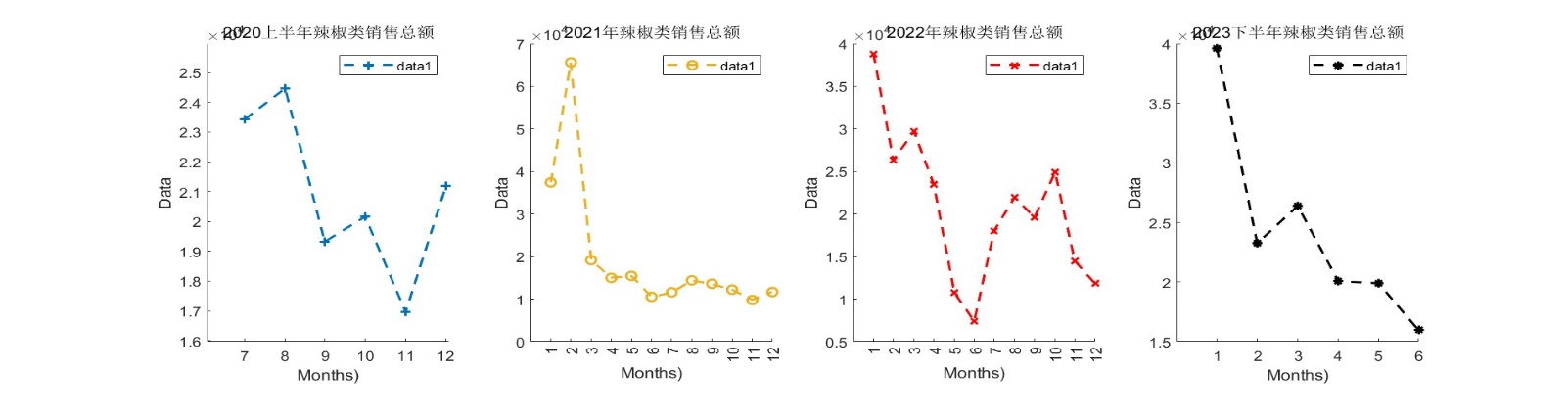
[4]曾敏敏. 基于时间情境A生鲜社区超市的动态定价策略研究[D].西南财经大学,2021.

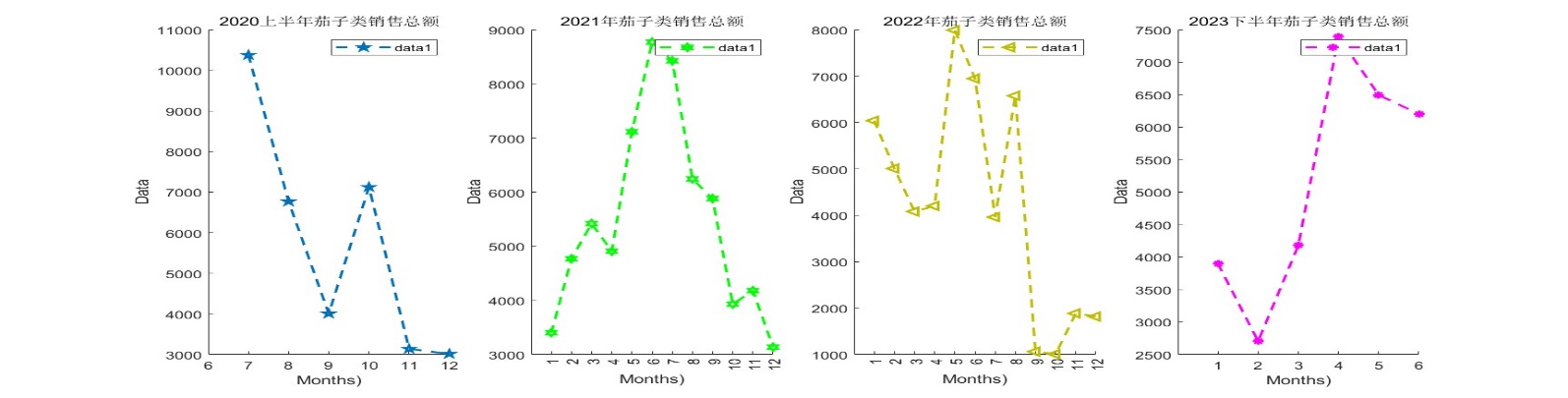
[5]冯茜. 基于改进粒子群算法的多目标优化及其应用[D].北京科技大学,2022.

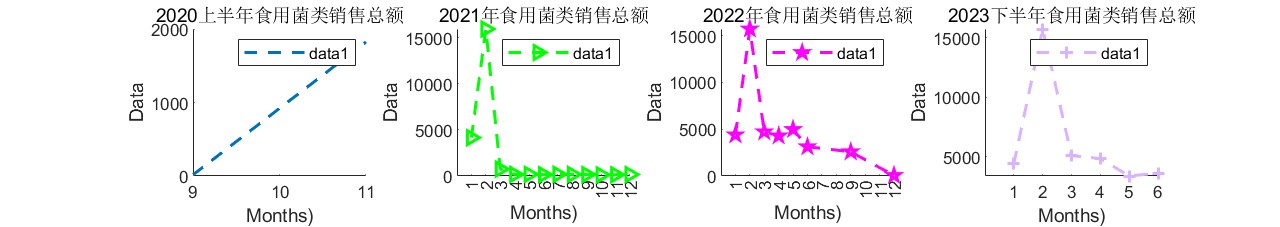
# 八.附录

附录一：8.1.1 同一品类不同年份的销售总额变化图

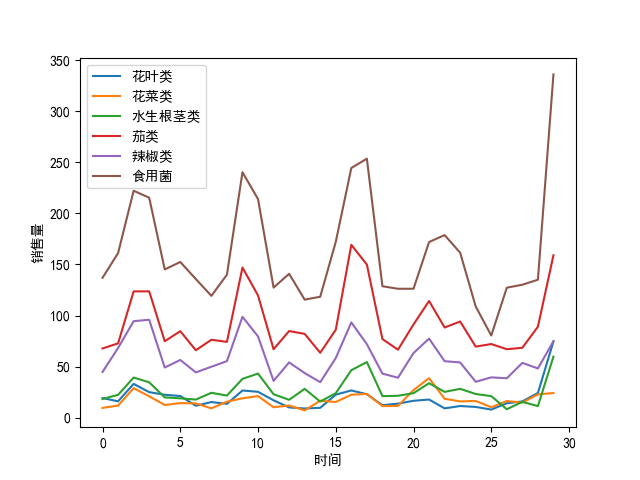
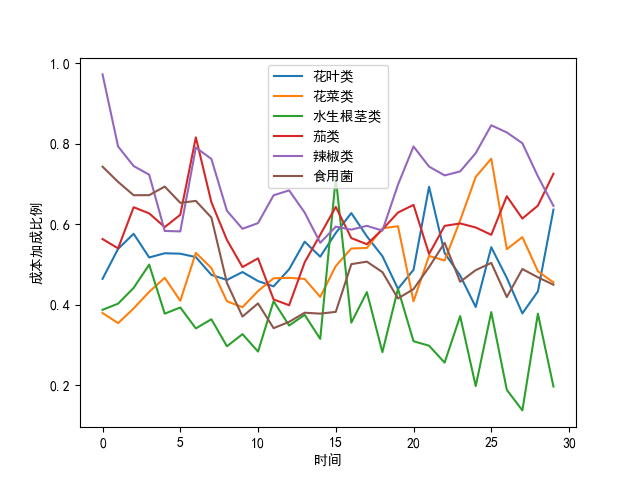
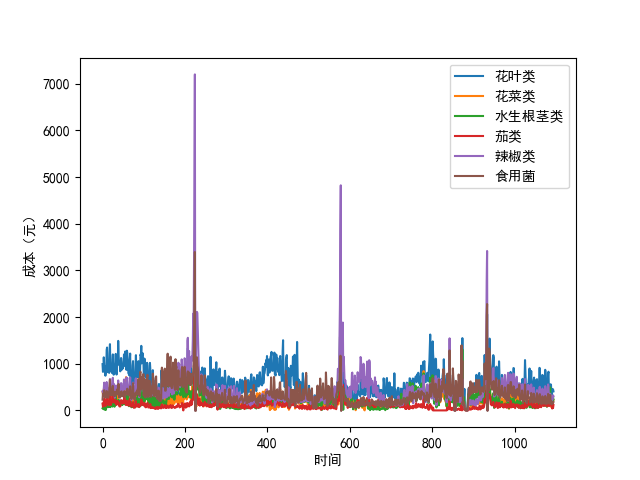
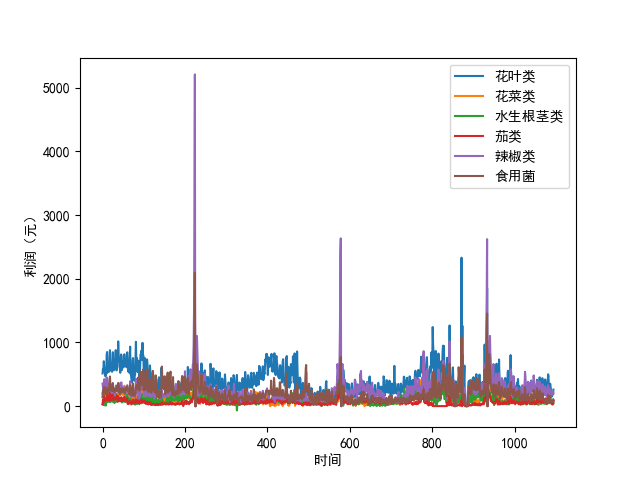


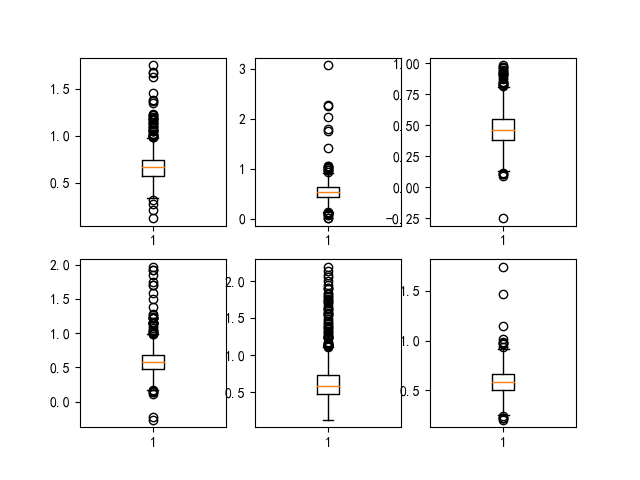
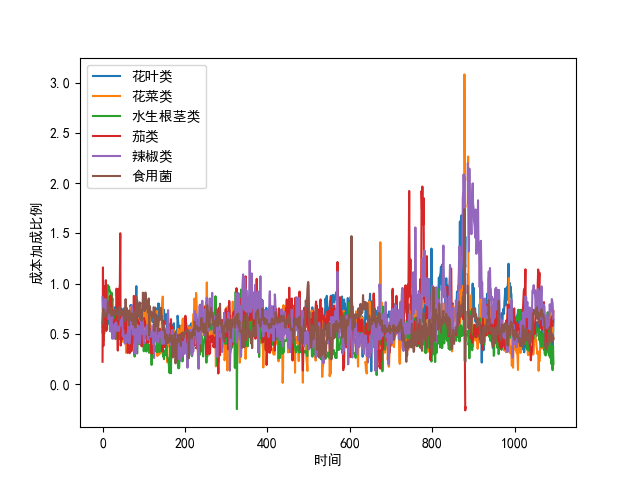






8.1.2 销售量与成本关系图





附录二：matlab代码

8.2.1 ：Kmeans曲线绘制

X = xlsread('D:\Desktop\单个元素的最大的东西.xlsx');

[idx, C] = kmeans(X, 2);

figure;

plot(X(:, 1), X(:, 2), '.', 'MarkerSize', 8);

title('对excel数据进行聚类簇团处理');

figure;

hold on;

plot(X(idx == 1, 1), X(idx == 1, 2), 'r.', 'MarkerSize', 12);

plot(X(idx == 2, 1), X(idx == 2, 2), 'b.', 'MarkerSize', 12);

plot(C(:, 1), C(:, 2), 'kx', 'MarkerSize', 15, 'LineWidth', 3);

legend('Cluster 1', 'Cluster 2', 'Centroids', 'Location', 'NW');

title('处理后K=2');

hold off;

% X = xlsread('D:\Desktop\单个元素的最大的东西.xlsx');

%

% [idx, C] = kmeans(X, 3);

%

% figure;

% plot(X(:, 1), X(:, 2), '.', 'MarkerSize', 8);

%

% title('对excel数据进行聚类簇团处理');

% figure;

% hold on;

% plot(X(idx == 1, 1), X(idx == 1, 2), 'r.', 'MarkerSize', 12);

% plot(X(idx == 2, 1), X(idx == 2, 2), 'b.', 'MarkerSize', 12);

% plot(X(idx == 3, 1), X(idx == 3, 2), 'g.', 'MarkerSize', 12);

% plot(C(:, 1), C(:, 2), 'kx', 'MarkerSize', 15, 'LineWidth', 3);

% legend('Cluster 1', 'Cluster 2', 'Cluster 3', 'Centroids', 'Location', 'NW');

% title('处理后K=3');

% hold off;

X = xlsread('D:\Desktop\单个元素的最大的东西.xlsx');

wcss = zeros(1,10);

silhouette\_vals = zeros(1,10);

for k = 2:10

[idx, C] = kmeans(X, k);

wcss(k) = sum(sum((X - C(idx,:)).^2));

silhouette\_vals(k) = mean(silhouette(X, idx));

end

figure;

subplot(1,2,1);

plot(2:10, wcss(2:10), 'bo-');

xlabel('簇数K');

ylabel('WCSS');

title('WCSS 与集群数量');

subplot(1,2,2);

plot(2:10, silhouette\_vals(2:10), 'ro-');

xlabel('簇数');

ylabel('轮廓系数');

title('轮廓系数与聚类数');

8.2.2：关系图绘制与串联

for i=1:4

%打开的fig文件酌情修改路径

hf(i)=open(['.\Mat\_pic\_qiezi\qiezi\_sum\_',num2str(i),'.fig']);%打开fig

fig(i)=get(hf(i), 'CurrentAxes');%获取绘制的图像

end

%新建一个图窗

figure

titles = {'2020上半年茄子类销售总额', '2021年茄子类销售总额', '2022年茄子类销售总额', '2023下半年茄子类销售总额'};

for i=1:4

%看要绘制多少行多少列酌情改变

subplot(1,4,i);%第i个子图

axChildren = get(fig(i),'Children');%获取绘制的图像

copyobj(axChildren, gca);%复制到当前图窗里

close(hf(i));%关掉已经复制的图像

%下面是对图片进行设置，请酌情修改

title(titles{i});

xlabel('Months)');%绘制子图x标签

ylabel('Data');%绘制子图y标签

xticks([1 2 3 4 5 6 7 8 9 10 11 12]);%绘制子图刻度

end

8.2.3 销售总额与品类关系图（花叶为例）

%导入Excel数据（1月）

data1 = xlsread('花叶类.xlsx', 'Sheet1','H623112:H626224');

%求和

total1 = sum(data1);

disp(['1月总和为：',num2str(total1)]);

%导入Excel数据（2月）

data2 = xlsread('花叶类.xlsx', 'Sheet1','H626225:H627573');

%求和

total2 = sum(data2);

disp(['2月总和为：',num2str(total2)]);

%导入Excel数据（3月）

data3 = xlsread('花叶类.xlsx', 'Sheet1','H627574:H629833');

%求和

total3 = sum(data3);

disp(['3月总和为：',num2str(total3)]);

%导入Excel数据（4月）

data4 = xlsread('花叶类.xlsx', 'Sheet1','H629834:H631849');

%求和

total4 = sum(data4);

disp(['4月总和为：',num2str(total4)]);

%导入Excel数据（5月）

data5 = xlsread('花叶类.xlsx', 'Sheet1','H631850:H634040');

%求和

total5 = sum(data5);

disp(['5月总和为：',num2str(total5)]);

%导入Excel数据（6月）

data6 = xlsread('花叶类.xlsx', 'Sheet1','H634041:H635623');

%求和

total6 = sum(data6);

disp(['6月总和为：',num2str(total6)]);

%导入Excel数据（7月）

data7 = xlsread('花叶类.xlsx', 'Sheet1','H635624:H637817');

%求和

total7 = sum(data7);

disp(['7月总和为：',num2str(total7)]);

%导入Excel数据（8月）

data8 = xlsread('花叶类.xlsx', 'Sheet1','H637818:H641047');

%求和

total8 = sum(data8);

disp(['8月总和为：',num2str(total8)]);

%导入Excel数据（9月）

data9 = xlsread('花叶类.xlsx', 'Sheet1','H641048:H642580');

%求和

total9 = sum(data9);

disp(['9月总和为：',num2str(total9)]);

%导入Excel数据（10月）

data10 = xlsread('花叶类.xlsx', 'Sheet1','H642581:H645411');

%求和

total10 = sum(data10);

disp(['10月总和为：',num2str(total10)]);

%导入Excel数据（11月）

data11 = xlsread('花叶类.xlsx', 'Sheet1','H645412:H647991');

%求和

total11 = sum(data11);

disp(['11月总和为：',num2str(total11)]);

%导入Excel数据（12月）

data12 = xlsread('花叶类.xlsx', 'Sheet1','H647992:H650020');

%求和

total12 = sum(data12);

disp(['12月总和为：',num2str(total12)]);

months = [1,2,3,4,5,6,7,8,9,10,11,12];

data = [total1,total2,total3,total4,total5,total6,total7,total8,total9,total10,total11,total12];

plot(months,data);

xlabel('Months');

ylabel('Data');

附录三：Python代码

import matplotlib.pyplot as plt  
import pandas as pd  
import datetime  
from sklearn.svm import SVR  
from sklearn.pipeline import make\_pipeline  
from sklearn.preprocessing import StandardScaler  
import numpy as np  
  
# Create a pipeline with StandardScaler and SVR  
regr = make\_pipeline(StandardScaler(), SVR(C=1.0, epsilon=0.2))  
  
# Set font and display settings for Chinese characters in plots  
plt.rcParams['font.sans-serif'] = ['SimHei']  
plt.rcParams['axes.unicode\_minus'] = False  
  
# Read data from Excel and CSV files  
data1 = pd.read\_excel("附件1.xlsx")  
data2 = pd.read\_csv("附件2.csv", encoding='ANSI')  
data3 = pd.read\_excel("附件3.xlsx")  
data4 = pd.read\_excel("附件4.xlsx")  
  
# Define a dictionary mapping class names to indices  
tinydict = {'花叶类': 0, '花菜类': 1, '水生根茎类': 2, '茄类': 3, '辣椒类': 4, '食用菌': 5}  
  
# Create arrays to store data  
class\_list = [0.0] \* 6  
all\_list = np.array([class\_list] \* 1095)  
chengben\_list = np.array([class\_list] \* 1095)  
  
# Define a function to calculate profit  
def calculate\_profit(row):  
 date = row[0]  
 product\_code = row[2]  
 quantity = row[3]  
 wholesale\_price = data3[(data3["单品编码"] == product\_code) & (data3["日期"] == date)].iloc[0, 2]  
 purchase\_price = wholesale\_price \* quantity  
 profit = (row[4] - wholesale\_price) \* quantity  
 return profit, purchase\_price  
  
# Iterate over rows in data2 dataframe  
time0 = datetime.datetime.strptime('2020-07-01', "%Y-%m-%d")  
for i, row in data2.iterrows():  
 if i % 10000 == 0:  
 print(i)  
 time\_tmp = datetime.datetime.strptime(row[0], "%Y-%m-%d")  
 index\_1 = (time\_tmp - time0).days  
 index\_2 = tinydict[data1[data1["单品编码"] == row[2]].iloc[0][3]]  
 profit, purchase\_price = calculate\_profit(row)  
 all\_list[index\_1][index\_2] += profit  
 chengben\_list[index\_1][index\_2] += purchase\_price  
  
# Read profit and cost data from CSV files  
lirun\_df = pd.read\_csv("A2\_lirun.csv")  
del lirun\_df["Unnamed: 0"]  
chengben\_df = pd.read\_csv("A2\_chengben.csv")  
del chengben\_df["Unnamed: 0"]  
lirun\_list = lirun\_df.values  
chengben\_list = chengben\_df.values  
  
# Calculate the alpha value  
alpha = (chengben\_list + lirun\_list) / chengben\_list - 1  
  
# Plot profit, cost, and alpha values  
plt.plot(lirun\_list)  
plt.ylabel("利润（元）")  
plt.xlabel("时间")  
plt.legend(['花叶类', '花菜类', '水生根茎类', '茄类', '辣椒类', '食用菌'])  
plt.show()  
  
plt.plot(chengben\_list)  
plt.ylabel("成本（元）")  
plt.xlabel("时间")  
plt.legend(['花叶类', '花菜类', '水生根茎类', '茄类', '辣椒类', '食用菌'])  
plt.show()  
  
plt.plot(alpha)  
plt.ylabel("成本加成比例")  
plt.xlabel("时间")  
plt.legend(['花叶类', '花菜类', '水生根茎类', '茄类', '辣椒类', '食用菌'])  
plt.show()