实战项目 4_批发商客户分类

说明：本人可视化参考“Udacity 机器学习纳米学位项目—客户分类”，分析及折线图等内容皆作者原创，他人不得抄袭！

本章主要介绍基于 UCI 机器学习数据集[18]：Wholesale customers data.csv 先使用 KMeans 聚类算法将数据集聚类，然后利用聚类后的数据集使用 k-NN 算法进行分类，并对 2 种算法的效果从算法准确度、时间运行性、结果可解释性进行综合的对比分析比较。制定策略的有效性可以在实际中使用 A/B test[19] 来获得反馈结果，以此减少策略不当引起的不必要损失。

§1.1 商业理解

如今，许多企业都企图利用数据挖掘技术从大量的客户数据中发现数据库中隐含的有价值“知识”，以便利用这些“知识”开发更好的产品与服务来满足客户的更多需求。

近日，一家批发经销商尝试对其客户有针对性地改变其配送服务，以便提高营业利润。起初，他对某些客户将配送服务由原来的每周五次每次早上发货改为每周三次每次晚上发货，并且在短时间内没有出现负面的反馈结果。于是，他针对所有的客户群体全部采用了新的配送服务，但是，很快便收到了很多客户的投诉与取消提货的通知，使得该批发商损失巨大。

现在，该批发商雇佣你，任务就是：利用客户群体信息，将客户群分类，并针对分类结果制定有效的配送策略。

§1.2 数据理解

打开 Wholesale customers data.csv，观察数据集结构。数据集共有 440 个样本，每个样本有 6 个特征，分别是：Channel（通道）、Region（区域）、Fresh（生鲜）、Milk（牛奶）、Grocery（杂货）、Frozen（冷冻品）、Detergents_Paper（卫生纸）、Delicassen（熟食）。除了 Channel（通道）、Region（区域）2 列外，其他列的元素值都是该类产品的年度采购额（金额表示）。从项目需求分析来看，因为项目需求是：使用数据挖掘算法，观察该数据集的内在数据结构，根据不同客户群体的内在类别差异将客户进行分类，并根据不同结构的客户群体针对性指定有效的配送策略，所以，不需要 Channel（通道）、Region（区域）这 2 个特征。客户分类后，制定配送策略时，可以结合这 2 个特征来考虑。而且，该数据集中没有缺失值。

§1.3 数据准备

1、观察数据集的统计描述

	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
count	440	440	440	440	440	440
mean	12000.29773	5796.265909	7951.277273	3071.931818	2881.493182	1524.870455
std	12647.32887	5796.265909	7951.277273	3071.931818	4767.854448	2820.105937
min	3	55	3	25	3	3
25%	3127.75	1533	2153	742.25	256.75	408.25
50%	8504	3627	4755	1526	816.5	965.5
75%	16933.75	7190.25	10655.75	3554.25	3922	1820.25
max	112151	73498	92780	60869	40827	47943

图 4-1 Wholesale customers dat_统计描述

分析：从图 4-1 可以看出，（1）因为各属性值的数值范围差异很大，所以，为了减小此种差异，可以预先进行特征缩放，缩放技术结合数据分布特点与应用背景来选择。

（2）有的属性最小值、最大值差异范围太大，所以，样本集中可能含有异常点，需要先去除异常点。

（3）各特征的均值和中位数相差很大，数据集分布很有可能不是：高斯混合分布。

2、观察特征相关性

考虑 6 个特征中的一个或多个特征组合，是否对于客户分类有实际作用，即去除与客户分类不相关的特征。通过移除某一个特征，并将客户数据集按照比例 1:3 随机分成测试集与训练集，并建造一个回归树决策器对该移除的特征进行评分。评分标准是决定系数，正常值范围在[0,1]之间，0 表示不能拟合，1 表示完美拟合，负值表示不能拟合。

	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
决定系数	-0.333070534	0.173438009	0.699248197	-0.278249149	-44.49428193	-11.0236279

图 4-2 各特征的决定系数

分析：观察图 4-2 各特征的决定系数，可以发现：

（1）去除任一特征后，都不能完美拟合原始数据，所以，得出结论：原始数据集中不存在冗余特征；

（2）Detergents_Paper 决定系数最小、Delicatessen 决定系数次小，因此，可以推出：Detergents_Paper、Delicatessen 这 2 个特征对于客户分类具有重要作用。如果使用加权的算法，这 2 个特征应予以较大权重，并对最终算法得出的特征加权应予以验证，如果结果反常，应思考原因以优化建模。

（3）各特征的决定系数相差较大，所以，对于该数据集，可以使用降维方法来简化建模，如使用 PCA 来降维。

（4）各特征的决定系数相差较大，说明各特征对于客户分类的贡献度也是有较大差异的，因此，可以推测：原始数据集的分布模型很有可能不是：高斯混合分布，所以，在选择聚类模型时，倾向选择：KMeans 聚类算法，而不是：高斯混合模型聚类算法。

3、可视化特征分布

为了更好地挖掘数据集的内在结构和验证：之前对于特征的推论，可以利用可视化技术来更深入观察数据集。可以为每一个特征构建一个散布矩阵。如果该特征对于客户分类贡献度大，那么在散布矩阵中可以观察出：它与其他特征没有明显的或者没有关联关系；相反地，如果该特征对于客户分类贡献度小，那么在散布矩阵中可以观察出：它与其他某些特征的确存在较大或者明显的关联关系。

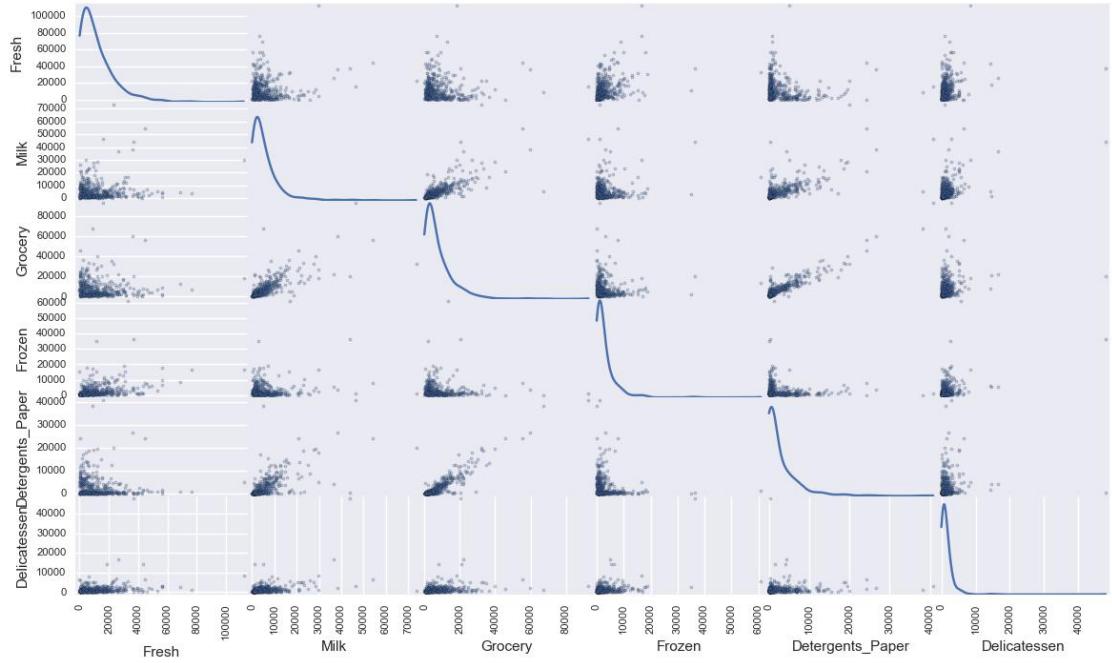

图 4-3 原始数据各特征的散布矩阵

分析：从图 4-3 各特征的散布矩阵中，可以观察出：

（1）Delicatessen 特征的确与其他特征没有明显的关联关系，所以，该特征对于客户分类的确具有重要作用；

（2）Grocery 与 Milk、Grocery 与 Detergents_Paper、Milk 与 Detergents_Paper 都具有明显的关联关系，如线性关系，又 Detergents_Paper 的决定系数最小，所以，可以采用降维技术来降维而不会对最终的客户分类结果造成明显的不良影

响；

（3）Delicatessen 不是正态分布，大部分数据集中分布在同一方向，所以，数据集整体的确很有可能不是：高斯混合分布模型，所以，采取 KMeans 聚类方法是可行的。

4、数据预处理

数据预处理是保证数据质量的关键，也是在数据挖掘中能够最终得到显著且有重要实际意义的关键。

1）特征缩放

如果数据不是正态分布的，尤其是数据非常倾斜，即数据的均值和中位数相差很大的情况时，采取非线性的缩放技术能够使得数据获得一个良好的数据结构，尤其是对于金融数据。

一种可选的缩放技术是：Box-Cox 变换，该种方法能计算出：能够最佳减小数据倾斜的指数变换方法，一般选择是：对数变换。

数据缩放后，再次为每一个特征构建散布矩阵，观察结果。

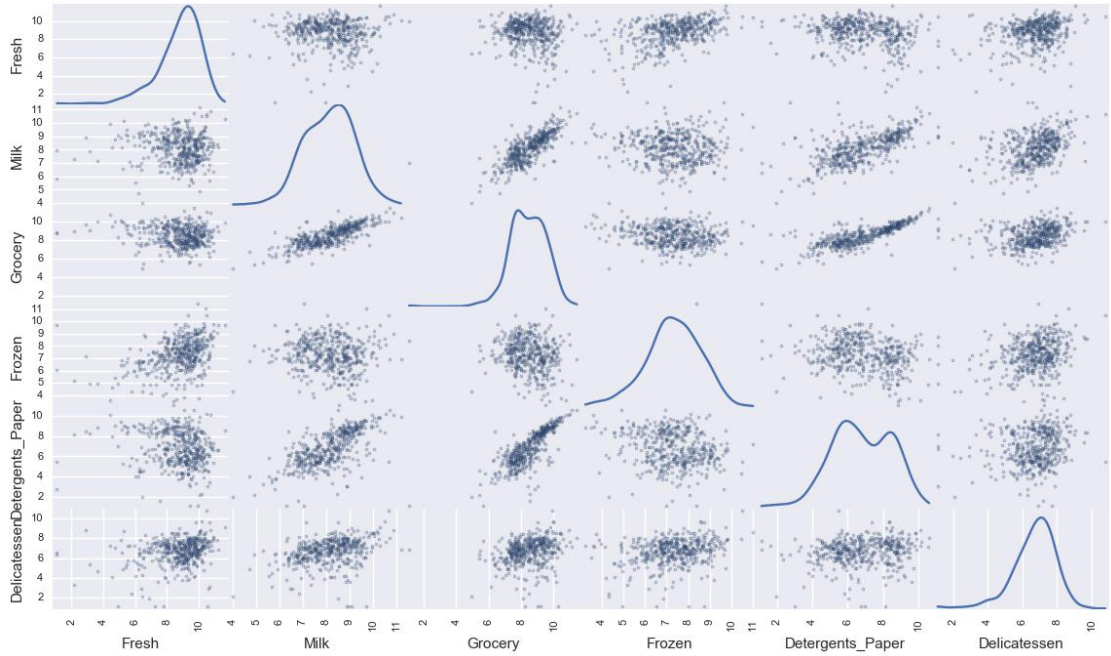

图 4-4 数据缩放后各特征的散布矩阵

分析：观察图 4-4 数据缩放后的散布矩阵，可以分析出：

（1）Grocery 与 Milk、Grocery 与 Detergents_Paper、Milk 与 Detergents_Paper 都具有明显的线性关联关系，与原始的散布矩阵比较，特征间的关联关系加强；

（2）数据集在每一个特征上分布更趋于：正态分布。所以，对于缩放后的数据，使用高斯混合模型聚类算法比较好；

（3）Delicatessen 特征的确与其他特征没有明显的关联关系，所以，该特征对于客户分类的确具有重要作用；

2）异常值检测

在数据预处理中，异常值检测是非常重要的一步。考虑进异常值的建模所得结果很有可能较大偏离实际结果，使得建模结果大打折扣。实际中有很多关于：怎样定义数据集中异常值的经验法则，这里选择：Tukey 的异常值检测方法——一个异常阶定义成 1.5 倍的四分位距（IQR）。一个数据点如果某个特征值在该特征的 IQR 范围外，那么该数据点可以被认定为异常点。

（1）根据特征 Fresh 排查出的候选异常点为：

Data points considered outliers for the feature 'Fresh':

	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
65	4.442651	9.950323	10.732651	3.583519	10.095388	7.260523
66	2.197225	7.335634	8.911530	5.164786	8.151333	3.295837
81	5.389072	9.163249	9.575192	5.645447	8.964184	5.049856
95	1.098612	7.979339	8.740657	6.086775	5.407172	6.563856
96	3.135494	7.869402	9.001839	4.976734	8.262043	5.379897
128	4.941642	9.087834	8.248791	4.955827	6.967909	1.098612
171	5.298317	10.160530	9.894245	6.478510	9.079434	8.740337
193	5.192957	8.156223	9.917982	6.865891	8.633731	6.501290
218	2.890372	8.923191	9.629380	7.158514	8.475746	8.759669
304	5.081404	8.917311	10.117510	6.424869	9.374413	7.787382
305	5.493061	9.468001	9.088399	6.683361	8.271037	5.351858
338	1.098612	5.808142	8.856661	9.655090	2.708050	6.309918
353	4.762174	8.742574	9.961898	5.429346	9.069007	7.013016
355	5.247024	6.588926	7.606885	5.501258	5.214936	4.844187
357	3.610918	7.150701	10.011086	4.919981	8.816853	4.700480
412	4.574711	8.190077	9.425452	4.584967	7.996317	4.127134

图 4-5 根据特征 Fresh 排查出的候选异常点

（2）根据特征 Milk、Grocery 排查出的候选异常点为：

Data points considered outliers for the feature 'Milk':

	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
86	10.039983	11.205013	10.377047	6.894670	9.906981	6.805723
98	6.220590	4.718499	6.656727	6.796824	4.025352	4.882802
154	6.432940	4.007333	4.919981	4.317488	1.945910	2.079442
356	10.029503	4.897840	5.384495	8.057377	2.197225	6.306275

Data points considered outliers for the feature 'Grocery':

	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
75	9.923192	7.036148	1.098612	8.390949	1.098612	6.882437
154	6.432940	4.007333	4.919981	4.317488	1.945910	2.079442

图 4-6 根据特征 Milk、Grocery 排查出的候选异常点

（3）根据特征 Frozen 排查出的候选异常点为：

Data points considered outliers for the feature 'Frozen':

	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
38	8.431853	9.663261	9.723703	3.496508	8.847360	6.070738
57	8.597297	9.203618	9.257892	3.637586	8.932213	7.156177
65	4.442651	9.950323	10.732651	3.583519	10.095388	7.260523
145	10.000569	9.034080	10.457143	3.737670	9.440738	8.396155
175	7.759187	8.967632	9.382106	3.951244	8.341887	7.436617
264	6.978214	9.177714	9.645041	4.110874	8.696176	7.142827
325	10.395650	9.728181	9.519735	11.016479	7.148346	8.632128
420	8.402007	8.569026	9.490015	3.218876	8.827321	7.239215
429	9.060331	7.467371	8.183118	3.850148	4.430817	7.824446
439	7.932721	7.437206	7.828038	4.174387	6.167516	3.951244

图 4-7 根据特征 Frozen 排查出的候选异常点

（4）根据特征 Detergents_Paper、Delicatessen 排查出的候选异常点为：

Data points considered outliers for the feature 'Detergents_Paper':

	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
75	9.923192	7.036148	1.098612	8.390949	1.098612	6.882437
161	9.428190	6.291569	5.645447	6.995766	1.098612	7.711101

Data points considered outliers for the feature 'Delicatessen':

	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
66	2.197225	7.335634	8.911530	5.164786	8.151333	3.295837
109	7.248504	9.724899	10.274568	6.511745	6.728629	1.098612
128	4.941642	9.087834	8.248791	4.955827	6.967909	1.098612
137	8.034955	8.997147	9.021840	6.493754	6.580639	3.583519
142	10.519646	8.875147	9.018332	8.004700	2.995732	1.098612
154	6.432940	4.007333	4.919981	4.317488	1.945910	2.079442
183	10.514529	10.690808	9.911952	10.505999	5.476464	10.777768
184	5.789960	6.822197	8.457443	4.304065	5.811141	2.397895
187	7.798933	8.987447	9.192075	8.743372	8.148735	1.098612
203	6.368187	6.529419	7.703459	6.150603	6.860664	2.890372
233	6.871091	8.513988	8.106515	6.842683	6.013715	1.945910
285	10.602965	6.461468	8.188689	6.948897	6.077642	2.890372
289	10.663966	5.655992	6.154858	7.235619	3.465736	3.091042
343	7.431892	8.848509	10.177932	7.283448	9.646593	3.610918

图 4-8 根据特征 Detergents_Paper、Delicatessen 排查出的候选异常点

（5）计算出重复候选异常点：

```
154    3
66     2
75     2
128    2
65     2
183    1
dtype: int64
```

图 4-9 重复候选异常点

分析：观察图 4-9 可得：一共有 5 个异常点，索引分别为：Outliers =[65,66,75,128,154]（在原始数据集、缩放后的数据集中都要去除异常点）。

3）选择最佳样本代表点

虽然对缩放后的数据集使用高斯混合模型聚类算法比较好，但是本文着重对比分析 KMeans 聚类和 k-NN 分类算法，所以，本文仍然使用 KMeans 聚类算法。决定 KMeans 聚类质量的 2 个关键因素为：初始化聚类代表点和聚类数目。所以，在去除异常点后，为了获得更好的聚类效果，要注意选取：具有“好”代表性的

数据点。因为数据缩放后，各特征的数据范围差异被缩小，相对不容易找出最佳聚类代表点，所以，在去除异常点后的原始数据集更方便寻找。又综合之前的数据统计性描述分析和特征分析，寻找最佳聚类代表点要突出重要特征的作用，如，按照特征的决定系数从小到大排序对特征依次排序，即按照特征对客户分类贡献度从大到小对特征依次排序为：Detergents_Paper、Delicatessen、Fresh、Frozen、Milk、Grocery。又由于 Detergents_Paper、Delicatessen、Fresh 的决定系数分别为：-44.494281934712、-11.0236279004667、-0.333070533604667 分别处于 3 个差异明显的数值段，所以，可以分别突出这 3 个特征值来选择 3 个最佳聚类代表点。

本文选择的 3 个聚类代表点分别为：prensents = [3,48,218],直方图如下：

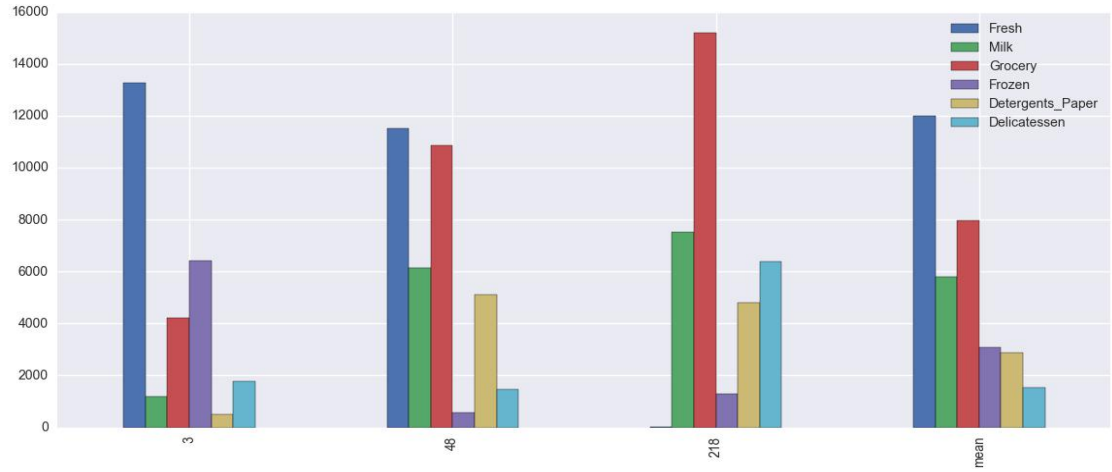

图 4-9 聚类代表点直方图

注意：在之前的数据预处理中已经对代表点统一处理过，后续无须再重复处理！

4）特征变换

由之前的特征分析可知，因为特征对客户分类的贡献度差异较大和某些特征之间确实存在明显的关联关系，所以对该数据集使用降维技术可以在不明显影响最终分类结果的前提下获得简化建模的良好效果。常用的降维技术是 PCA，但是凡是降维处理必定会带来原始数据集信息的损失。一种有效的特征约简技术要么使得原始数据信息损失最小，要么在维护原始数据内在结构与信息损失量之间取得某种平衡。经验来说，使用 PCA 降维一般使用累计贡献率来选择主成分——前 D’个主成分的方差在总方差中所占的比重大于或等于 80%。sklearn.decompisition.PCA 函数不仅可以降维，还会报告新的特征空间中每一个新维度(特征)的方差解释比——该数据集有多少方差能够用当前维度来解释。

对之前预处理好的数据集 good_data 使用 sklearn.decompisition.PCA，将其转换成与当前数据集同样的维度数并保存到 pca 中。

```
In [176]: print(pca.components_)
[[-0.16770551  0.39133933  0.45394827 -0.17713265  0.7492642   0.14101965]
 [ 0.68299312  0.17018417  0.07150195  0.50110959  0.04573645  0.49622835]
 [-0.67027699  0.04507731 -0.0302354   0.27522223 -0.21311802  0.65315852]
 [-0.23512595  0.00495026  0.06099841  0.80029318  0.19355498 -0.51287208]
 [ 0.004736   -0.71309766 -0.3631735   0.03068107  0.56171906  0.20761667]
 [-0.02858995  0.55437004 -0.80764377 -0.01885254  0.19519276 -0.03335238]]
```

图 4-10 主成分详细信息

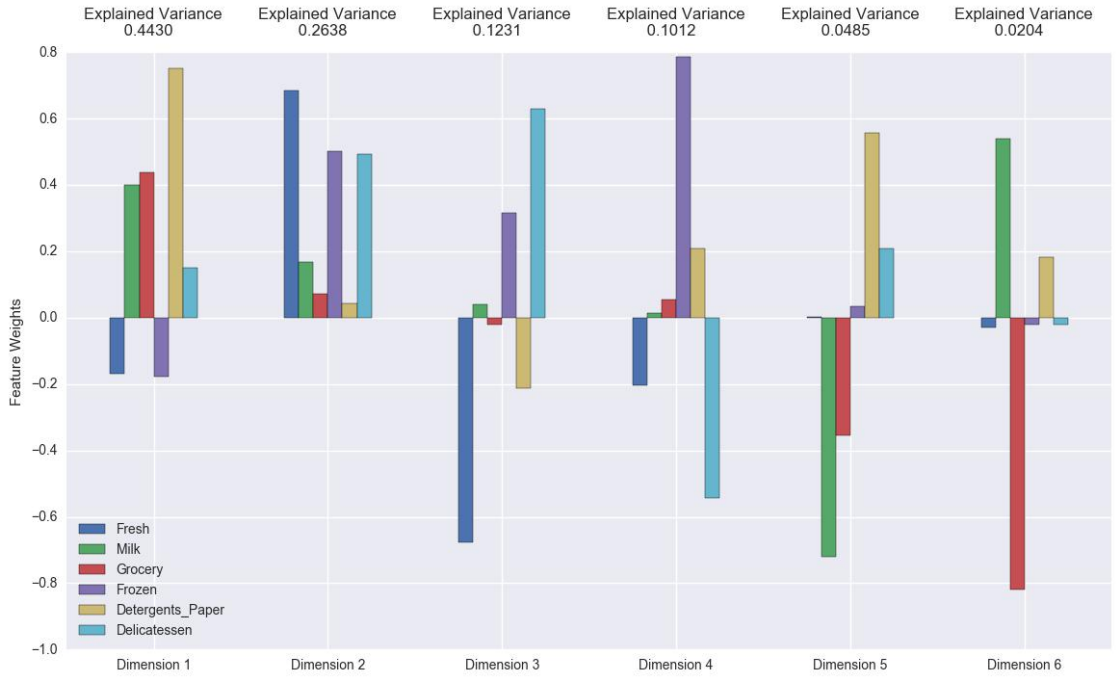

图 4-11 主成分详细信息柱状图

注意：（1）PCA 变换的主成分个数不得超过当前数据集的维度数；

（2）以此数据集为例，不管 PCA 变换的主成分个数是多少（符合（1）个数约束），前 6 个主成分的数值始终是一致的。

分析：观察图 4-10 和图 4-11，可以得出：

（1）前 6 个主成分的解释方差比从大到小依次为：0.4430、0.2638、0.1231、0.1012、0.0485、0.0204，由 PCA 选取主成分的经验知识可得：因为 0.4430+0.2638+0.1231=0.8299，0.4430+0.2638=0.7068，所以选取主成分 Dimension1、Dimension2、Dimension3 作为降维后的新特征。

（2）每一个主成分都是原始特征的权重组合，直方图纵坐标显示的是原始特征权重，并且特征权重有正负：正负代表的是原始特征在该主成分的构成中所起的正、负向作用。权重的大小与该原始特征正、负向增长的速率正相关。

（3）尝试分析维度蕴含的信息：

a.Dmension1 中 Detergents_Paper 具有最大权重，其次是 Grocery、Milk，并且 Detergents_Paper、Grocery 在所有维度中权重是最大的，所以 Dimension1 很有可能代表：零售商；

b.Dimension2 中 Fresh 权重最大，其次较大权重有 Frozen、Delicatessen，并且 Fresh 在所有新维度中权重是最大的，Frozen、Delicatessen 在所有新维度中权值也较大，所以 Dimension2 很有可能代表：餐馆；

c.Dimension3 中 Delicatessen 权值最大、次大是 Frozen，Fresh 权值最低，并且 Delicatessen 在所有新维度中权值最大，Fresh 在所有新维度中权值最小，所以 Dimension3 很有可能代表：小吃店；

d.Dimension4 中 Frozen 权值最大，Delicatessen 权值最低，并且在所有维度依然如此，Detergents_Paper 权值次较大，所以，Dimension4 很有可能代表：超市。

（4）a.Dmension1 中 Detergents_Paper 具有最大权重；

e.Dimension2 中 Fresh 权重最大，其次较大权重有 Frozen、Delicatessen，并且 Fresh 在所有新维度中权重是最大的，Frozen、Delicatessen 在所有新维度中权值也较大；

f.Dimension3 中 Delicatessen 权值最大、次大是 Frozen，并且 Delicatessen 在所有新维度中权值最大。

不难发现：对客户分类贡献度大的原始特征在 PCA 主成分的构成中都起到了主要作用甚至一个原始特征就决定了一个主成分，这个结果与 PCA 的原理是一致的：PCA 采用贪心策略，总是选取最大化数据集方差的维度，以保留更多原始数据集信息。

§1.4 数据建模

本节主要使用原始 KMeans 算法、二分快速 KMeans 算法对处理好的数据集 good_data 进行聚类建模，然后使用原始 k-NN 分类算法、改进 k-NN 算法对数据集进行分类，并从算法准确率、运行时间综合对比分析 4 种算法的效果。

针对不同情况，某些情况下聚类数目是已知的。但是，有些情况下，不仅不知道原始数据集的内在数据结构，还不知道聚类的数目，所以，仅仅依靠聚类算法是不能保证当前聚类数目对该数据集是最优的。因此，必须得采用一种度量标准来衡量聚类的质量，通过优化此评价标准来确定最佳聚类数目。轮廓系数是衡量聚类质量的常用标准。数据点的轮廓系数衡量了它与类属簇（算法分配给它的簇）的相似度，值范围在-1（不相似）与 1（相似）之间。而平均轮廓系数又是一种简易的度量聚类质量的方法。

§1.4.1 原始 KMeans 算法聚类

（1）因为我们已经选择 3 个主成分，所以，如果考虑每一个主成分代表 1 个类别，那么聚类数目可以先选择 3，然后在数目 3 附近浮动选择聚类数目。本次试验中选择聚类数目分别为 k=2、k=3、k=4，试验结果如下图：

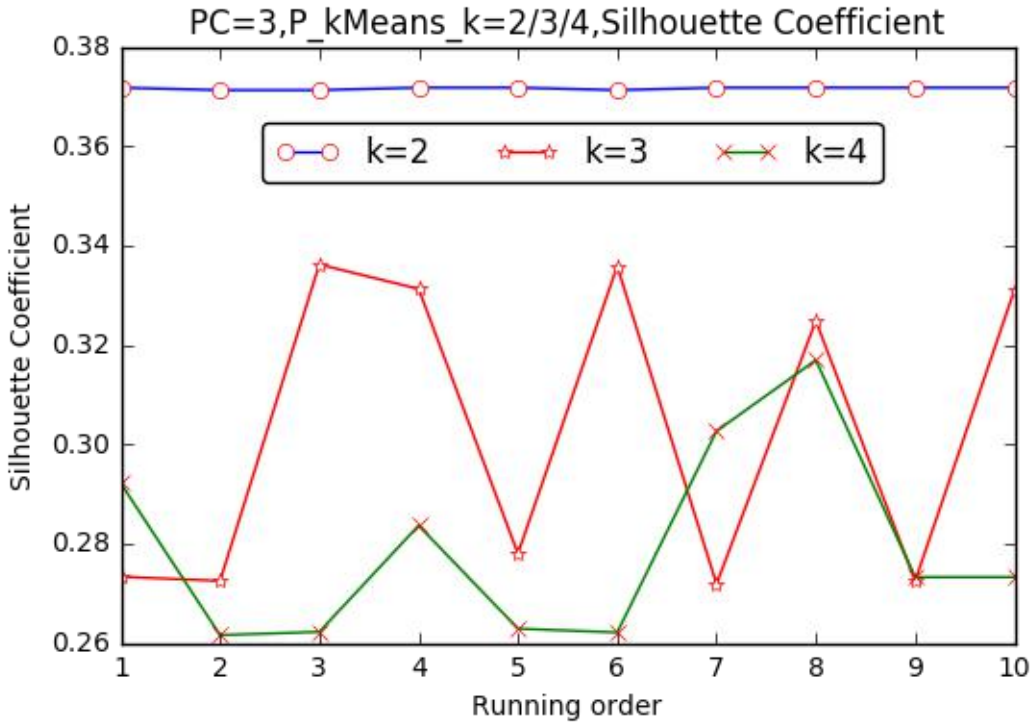

图 4-12 3 个主成分时，聚类数目分别为 2、3、4 时原始 kMeans 算法聚类轮廓系数折线图

分析：从图 4-12 可以得出：

主成分个数为 3，原始 kMeans 聚类算法聚类数目为 2 时聚类轮廓系数最高，聚类效果最好。

（2）有些情况下，由 PCA 主成分个数选取的经验法则确定的主成分个数不一定是最优的，所以，实际中，往往也要测试主成分个数更少时的算法效果，以此来对比分析（选取的主成分个数已经能保持原有数据结构的足够信息，所以不予测试主成分个数更多的情况，当然也可以进行测试。）。本次试验选取主成分个数为 2，原始 kMeans 算法聚类结果如下图：

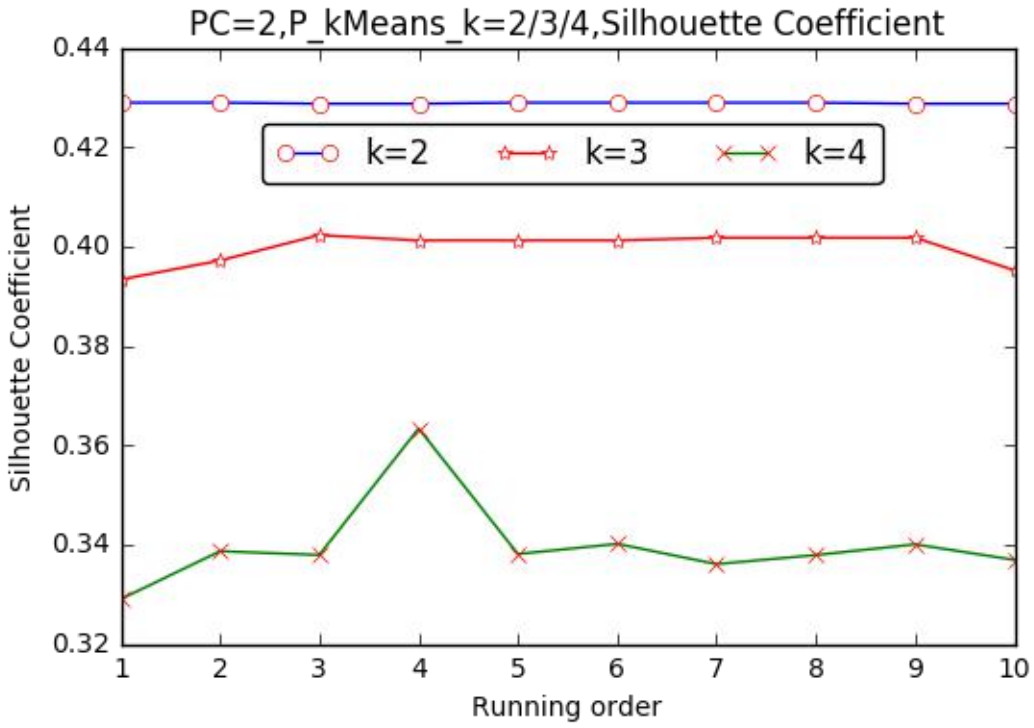

图 4-13　2 个主成分时，聚类数目分别为 2、3、4 时原始 kMeans 算法聚类轮廓系数折线图

分析：从图 4-13 可以看出：

主成分个数为 2，原始 kMeans 聚类算法聚类数目为 2 时聚类轮廓系数最高，聚类效果最好。

§1.4.2 二分快速 KMeans 聚类

（1）因为我们已经选择 3 个主成分，所以，如果考虑每一个主成分代表 1 个类别，那么聚类数目可以先选择 3，然后在数目 3 附近浮动选择聚类数目。本次试验中选择聚类数目分别为 k=2、k=3、k=4，试验结果如下图：

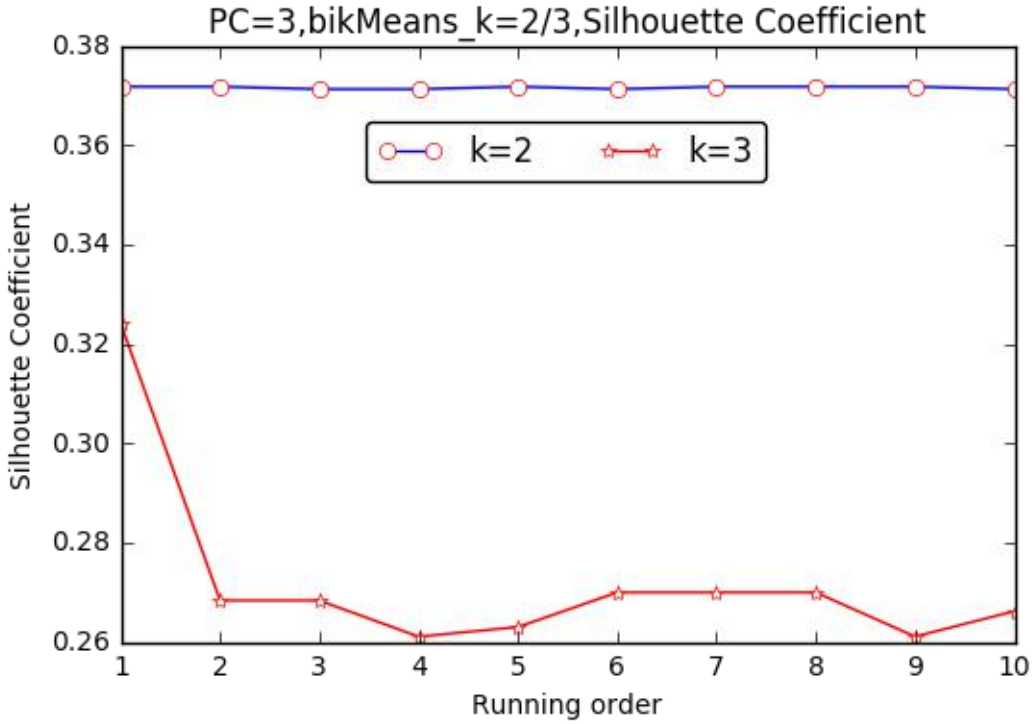

图 4-14　3 个主成分时，聚类数目分别为 2、3 时 bikMeans 算法聚类轮廓系数折线图

分析：观察图 4-14，可以得出：主成分个数为 3，聚类数目为 2 时，bikMeans 算法聚类轮廓系数最高，聚类效果最好。

（3）有些情况下，由 PCA 主成分个数选取的经验法则确定的主成分个数不一定是最优的，所以，实际中，往往也要测试主成分个数更少时的算法效果，以此来对比分析（选取的主成分个数已经能保持原有数据结构的足够信息，所以不予测试主成分个数更多的情况，当然也可以进行测试。）。本次试验选取主成分个数为 2，原始 kMeans 算法聚类结果如下图：

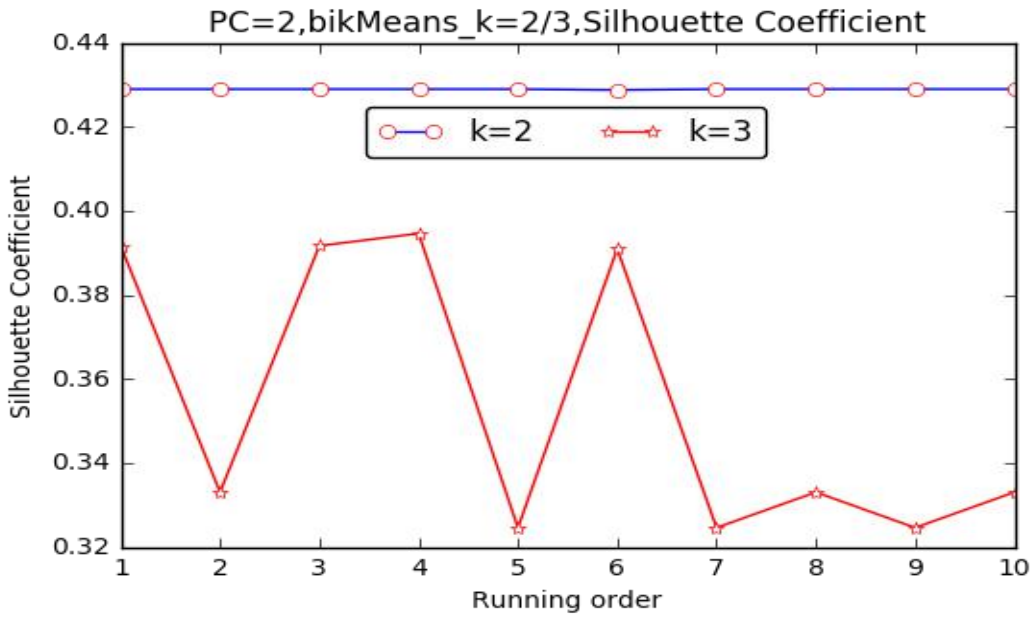

图 4-15　2 个主成分时，聚类数目分别为 2、3 时 bikMeans 算法聚类轮廓系数折线图

分析：观察图 4-15 可以得出：

主成分个数为 2，bikMeans 聚类算法聚类数目为 2 时聚类轮廓系数最高，聚类效果最好。

（3）主成分个数分别为 2、3，聚类数目为 2 时，原始 kMeans 算法、二分快速 kMeans 算法的聚类轮廓系数折线图如下：

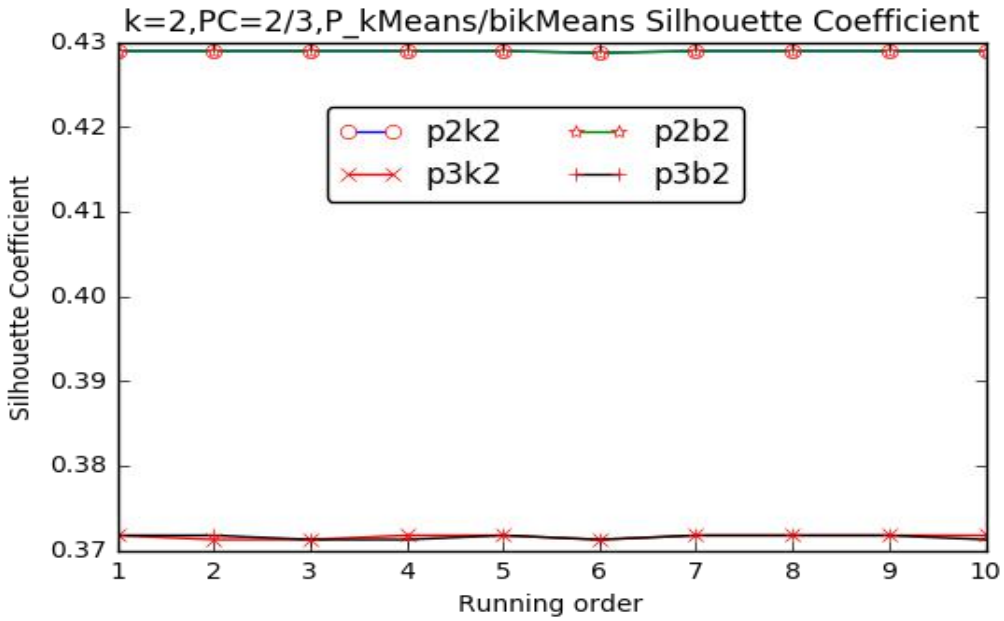

图 4-16 聚类数目分别为 2，主成分个数分别为 2、3 时，原始 kMeans、bikMeans 算法聚类轮廓系数折线图

分析：观察图 4-16，可以得出：

（1）在同等情况下，原始 kMeans 算法、bikMeans 算法的聚类效果几乎等同。下面比较 2 同等情况下，算法的运行时间：

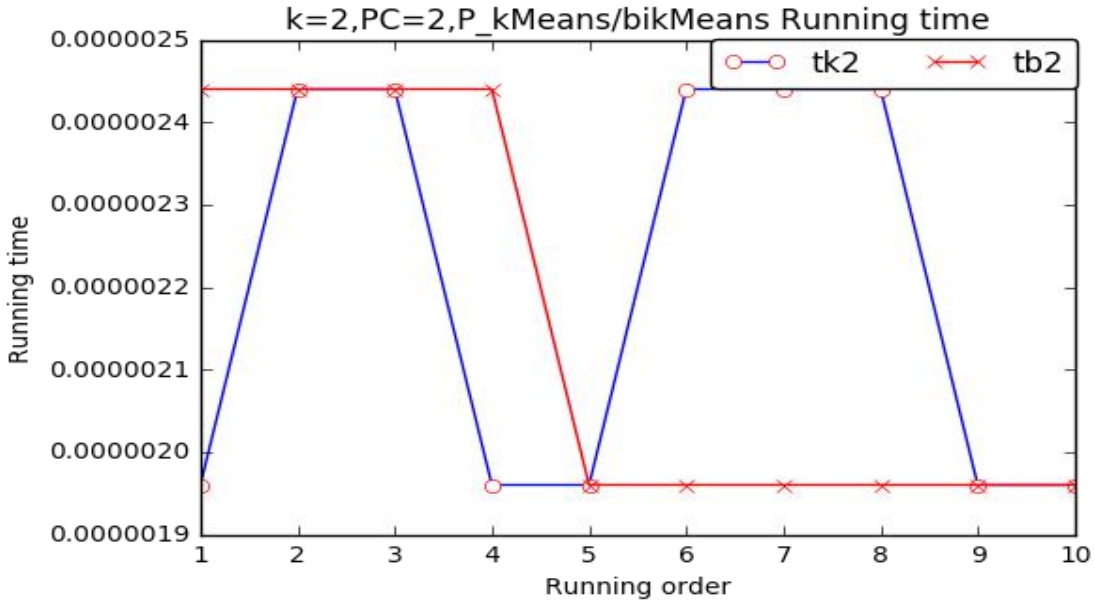

图 4-17 主成分个数为 2，聚类数目为 2 时，原始 kMeans、bikMeans 算法运行时间图

分析：观察图 4-17，可以得出：

（1）整体来说，bikMeans 算法运行时间比原始 kMeans 算法要少，又因为 2 种算法分类效果几乎等同（在同样条件下），所以，实际中选择 bikMeans 算法聚类。

§1.4.3 原始 k-NN、简化 k-NN、改进 k-NN 算法分类

本节主要从算法分类效果、运行时间来综合对比分析原始 k-NN、简化 k-NN、改进 k-NN 算法分类效果。

（1）分类错误率对比分析图：

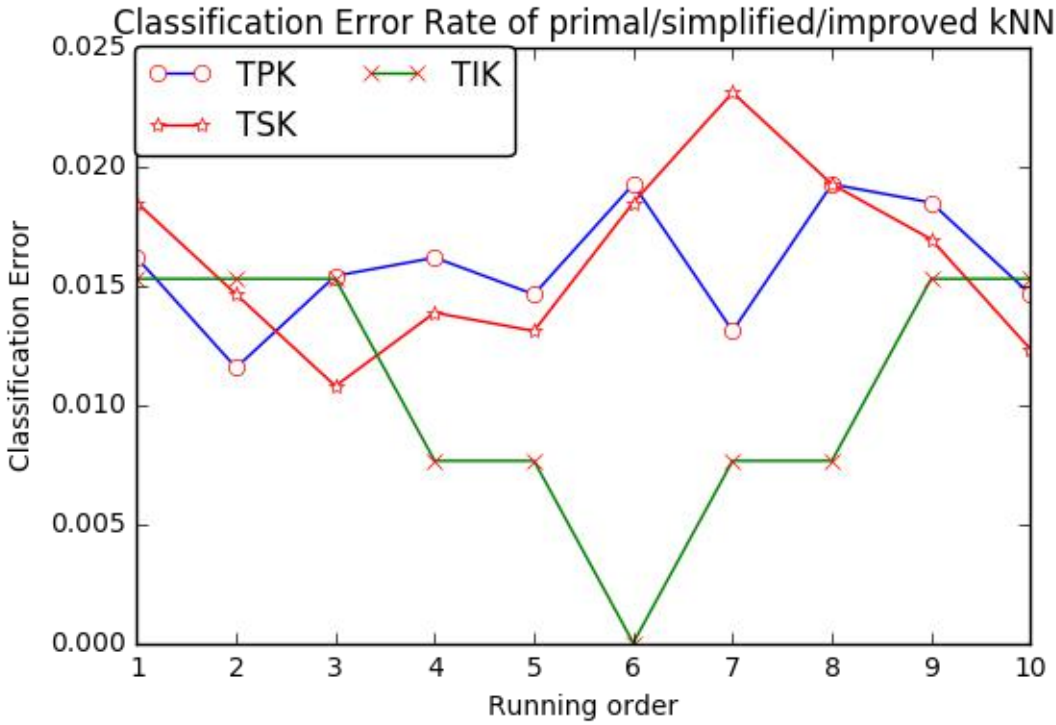

图 4-18　3 种 k-NN 分类算法分类错误率折线图

分析：观察图 4-18，可以得出：

（1）从整体上来说，改进 k-NN 算法的分类错误率最低，平均分类效果最好。

（2）运行时间对比对比分析图：

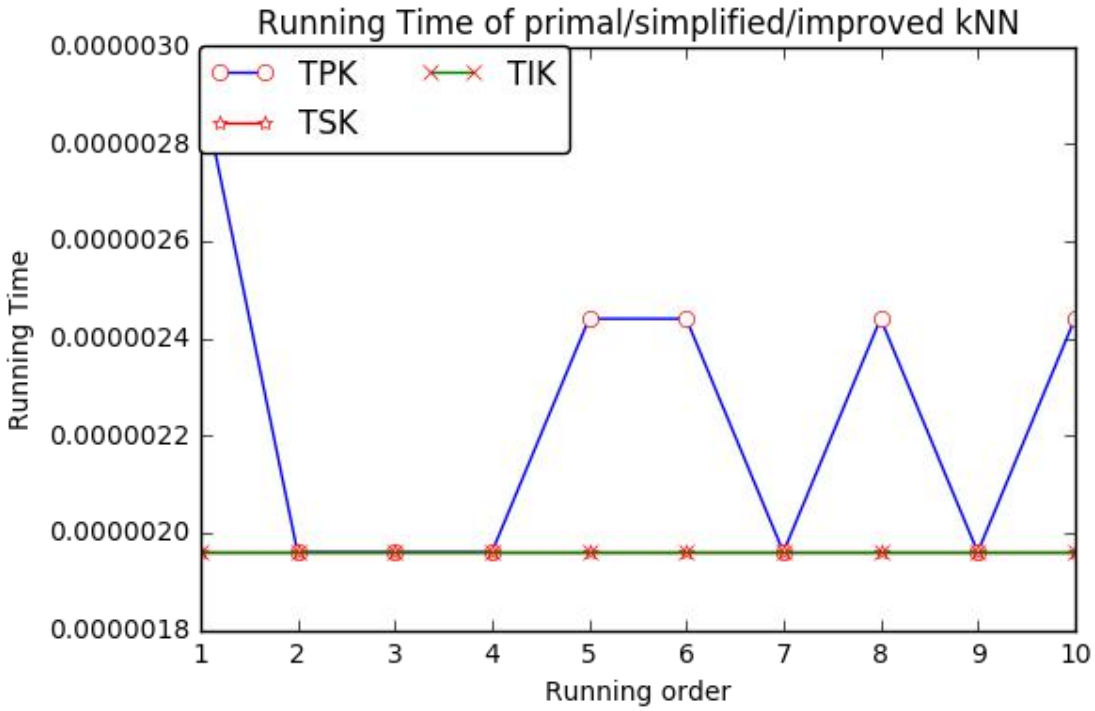

图 4-19 3 种 k-NN 分类算法运行时间折线图

分析：观察图 4-19，可以得出：

（1）改进 k-NN 算法、简化 k-NN 算法运行时间比原始 k-NN 算法运行时间低，从平均时间来看，运行时间至少减少 10%.

§1.5 结果分析

由聚类的结果，再使用 k-NN 算法进行分类，就可以将新的样本实例进行归类。在实际应用中，针对不同客户群体分别设计一种新的配送方案，并将客户群体分为原始配送方案、新配送方案 2 个对照组，注意客户群体应尽可能相似，以免测试受到人为因素的较大干扰。经过一段测试时间后，由具体的客户群反馈结果调整优化方案。这就是实际应用中，企业经常采取的 A/B test 测试方法。该项目同样可以使用 A/B test 测试方法来测试新配送方案的有效性。