二阶聚类

备注

已创建输出		05-MAY-2025 16:25:52
注解		
输入	数据	/Users/luoyiti/CodeProject/MNIST_Clustering/data/tsne_df.csv
	活动数据集	数据集1
	过滤器	<无>
	权重	<无>
	拆分文件	<无>
缺失值处理	对缺失的定义	将用户定义的缺失值视为缺失。
	使用的个案数	统计基于所有那些对于分析中的所有变量都具有有效数据的个案。
语法		TWOSTEP CLUSTER /CONTINUOUS VARIABLES=x y /DISTANCE EUCLIDEAN /NUMCLUSTERS FIXED=10 /HANDLENOISE 5 /MEMALLOCATE 64 /CRITERIA INITHRESHOLD(0) MXBRANCH(8) MXLEVEL(3) /VIEWMODEL DISPLAY=YES /PRINT IC COUNT SUMMARY /SAVE VARIABLE=TSC_3227.
资源	处理程序时间	00:00:01.31
	耗用时间	00:00:02.00
保存的文件	模型	/var/folders/0m/2sbyxxc10czg8dp3clzyzxym0000gn/T/spssJUxrf1/tsctempm.21
创建或修改的变量	TSC_3227	二阶聚类编号

聚类分布

		个案数	占组合的百分比	占总计的百分比
聚类	1	6231	10.4%	10.4%
	2	6348	10.6%	10.6%
	3	5402	9.0%	9.0%
	4	5193	8.7%	8.7%
	5	3910	6.5%	6.5%
	6	2847	4.7%	4.7%
	7	5878	9.8%	9.8%
	8	5798	9.7%	9.7%
	9	5164	8.6%	8.6%
	10	5175	8.6%	8.6%
	离群值 (-1)	8054	13.4%	13.4%
	组合	60000	100.0%	100.0%
总计		60000		100.0%

聚类概要

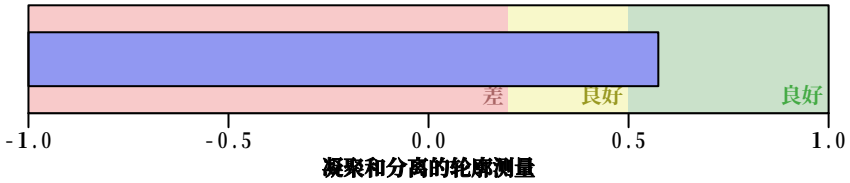
质心

		x		y	
		平均值	标准 偏差	平均值	标准 偏差
聚类	1	-46.2872364	15.0797747	-37.5093493	13.7828007
	2	2.18825569	15.6572897	-63.9323871	13.8236588
	3	-13.6028022	14.1316452	12.3582659	10.9261109
	4	-29.4238576	10.7398358	64.7344471	15.2773760
	5	-72.4813567	10.9028277	33.2867111	13.1848778
	6	-85.0215246	10.8196204	-1.70891630	9.34187642
	7	75.0720701	12.9219779	32.5647303	13.9434755
	8	20.5908844	12.5301398	59.7604733	13.7960765
	9	31.7199802	10.1219303	-1.11885340	14.8709115
	10	71.2880562	13.7259723	-32.8802362	12.2611426
	离群值 (-1)	-7.52044783	45.5607383	-30.8139914	50.3544348
	组合	.109066986	50.5131490	.040756394	46.8323590

模型概要

算法	两步
输入	2
聚类	10

聚类质量



Data written to
/Users/luoyiti/CodeProject/MNIST_Clustering/spss_examine/two_factor_cluster.csv
4 variables and 60000 cases written.
Variable x Type: Number Width: 23 Dec: 19
Variable y Type: Number Width: 23 Dec: 19
Variable label Type: String Width: 3
Variable TSC_3227 Type: Number Width: 10 Dec: 0

快速聚类

备注

已创建输出		05-MAY-2025 16:35:10
注解		
输入	数据	/Users/luoyiti/CodeProject/MNIST_Clustering/data/tsne_df.csv
	活动数据集	数据集1
	过滤器	<无>
	权重	<无>
	拆分文件	<无>
	工作数据文件中的行数	60000
缺失值处理	对缺失的定义	Text\: ^1
	使用的个案数	LISTWISE...
语法		QUICK CLUSTER x y /MISSING=LISTWISE /CRITERIA=CLUSTER(10) MXITER(300) CONVERGE(0) /METHOD=KMEANS(NOUPDATE) /SAVE CLUSTER /PRINT ID(label) INITIAL ANOVA.
资源	处理程序时间	00:00:02.83
	耗用时间	00:00:03.00
	所需工作空间	1552 字节
创建或修改的变量	QCL_1	个案聚类编号

初始聚类中心

	聚类				
	1	2	3	4	5
x	39.6161194	105.855431	31.3651848	-12.3012972	-78.6359253
y	4.57635164	24.9659405	-93.4967957	-40.5271721	-57.0617218

初始聚类中心

	聚类				
	6	7	8	9	10
x	-30.9528179	-100.408890	42.7819595	-27.6222019	90.3173904
y	97.0397263	12.5913935	79.7480621	25.9353428	-44.5295486

迭代历史记录[a]

聚类中心中的变动

迭代	1	2	3	4	5	6	7	8
1	3.618	21.810	21.692	5.929	19.819	20.214	18.693	20.621
2	7.264	7.131	5.940	1.536	5.501	5.460	4.804	8.029
3	2.045	2.641	2.286	1.413	.866	2.837	.450	2.580
4	.861	.164	1.313	1.143	.179	1.518	.074	.353
5	.513	.088	.898	1.069	.250	.933	.277	.165
6	.405	.023	.536	.788	.207	.419	.134	.125
7	.179	.000	.399	.764	.181	.242	.152	.042
8	.137	.005	.311	.699	.193	.142	.116	.010
9	.108	.000	.256	.601	.167	.067	.072	.005
10	.113	.005	.168	.531	.194	.030	.026	.013
11	.051	.005	.150	.386	.146	.027	.015	.011
12	.025	.010	.128	.258	.077	.007	.014	.005
13	.017	.005	.074	.170	.058	.000	.000	.005
14	.019	.005	.078	.165	.059	.000	.000	.010
15	.010	.005	.088	.221	.106	.000	.000	.005
16	.019	.000	.082	.216	.087	.000	.005	.000
17	.000	.000	.071	.168	.068	.000	.000	.000
18	.000	.000	.066	.146	.050	.005	.000	.000
19	.000	.000	.076	.138	.031	.000	.000	.000
20	.000	.000	.050	.098	.031	.000	.011	.000
21	.005	.000	.024	.072	.036	.000	.010	.000
22	.000	.000	.058	.114	.033	.000	.000	.000
23	.005	.000	.060	.120	.031	.000	.000	.005
24	.013	.000	.059	.137	.039	.000	.010	.000
25	.009	.000	.101	.170	.034	.005	.005	.005
26	.000	.000	.034	.121	.058	.000	.005	.010
27	.011	.000	.036	.102	.042	.005	.000	.010
28	.011	.005	.019	.052	.020	.005	.000	.010
29	.000	.000	.017	.045	.015	.005	.000	.000
30	.000	.000	.014	.035	.004	.009	.000	.000
31	.000	.000	.008	.021	.004	.005	.000	.010
32	.005	.000	.010	.015	.000	.014	.005	.000
33	.000	.000	.009	.010	.000	.000	.005	.000
34	.000	.000	.012	.015	.000	.000	.000	.000
35	.000	.000	.010	.023	.005	.000	.000	.000
36	.000	.000	.008	.021	.009	.000	.000	.000
37	.000	.000	.012	.014	.000	.000	.000	.000
38	.000	.000	.009	.015	.004	.000	.000	.000
39	.000	.000	.012	.025	.009	.000	.000	.000
40	.000	.000	.016	.030	.005	.000	.000	.000
41	.000	.000	.028	.034	.005	.000	.000	.000
42	.000	.000	.030	.041	.008	.000	.000	.000
43	.000	.000	.035	.057	.012	.000	.000	.000

迭代历史记录[a]

迭代	聚类中心中的变动	
	9	10
1	.582	20.302
2	8.083	2.415
3	4.559	.573
4	2.525	.308
5	1.282	.113
6	.638	.029
7	.236	.014
8	.108	.005
9	.055	.005
10	.032	.005
11	.032	.000
12	.025	.000
13	.015	.005
14	.024	.000
15	.012	.000
16	.005	.000
17	.005	.000
18	.015	.000
19	.015	.000
20	.013	.000
21	.006	.000
22	.016	.000
23	.024	.000
24	.020	.000
25	.019	.000
26	.035	.000
27	.026	.000
28	.024	.000
29	.015	.000
30	.023	.000
31	.020	.000
32	.022	.000
33	.006	.000
34	.000	.000
35	.005	.000
36	.000	.000
37	.000	.000
38	.000	.000
39	.000	.000
40	.005	.000
41	.000	.000
42	.005	.000
43	.000	.000

迭代历史记录[a]

	聚类中心中的变动							
迭代	1	2	3	4	5	6	7	8
44	.005	.000	.010	.055	.025	.000	.000	.000
45	.000	.000	.018	.067	.035	.000	.000	.000
46	.000	.000	.027	.077	.036	.000	.000	.000
47	.000	.000	.034	.069	.019	.000	.000	.000
48	.000	.000	.012	.043	.019	.000	.000	.000
49	.000	.000	.018	.025	.005	.000	.000	.000
50	.000	.000	.021	.025	.000	.000	.000	.000
51	.000	.000	.024	.033	.004	.000	.000	.000
52	.000	.000	.010	.016	.004	.000	.000	.000
53	.000	.000	.009	.013	.005	.000	.000	.000
54	.000	.000	.012	.023	.008	.000	.000	.000
55	.000	.000	.000	.020	.016	.000	.000	.000
56	.000	.000	.000	.006	.000	.000	.000	.000
57	.000	.000	.000	.017	.005	.000	.000	.000
58	.000	.000	.004	.005	.000	.000	.000	.000
59	.000	.000	.000	.006	.005	.000	.000	.000
60	.000	.000	.000	.000	.000	.000	.000	.000

迭代历史记录[a]

	聚类中心中的变动	
迭代	9	10
44	.010	.000
45	.005	.000
46	.000	.000
47	.006	.000
48	.005	.000
49	.000	.000
50	.000	.000
51	.000	.000
52	.000	.000
53	.000	.000
54	.000	.000
55	.000	.000
56	.006	.000
57	.010	.000
58	.000	.000
59	.000	.000
60	.000	.000

a. 由于聚类中心中不存在变动或者仅有小幅变动，因此实现了收敛。任何中心的最大绝对坐标变动为 .000。当前迭代为 60。初始中心之间的最小距离为 68.206。

最终聚类中心

	聚类				
	1	2	3	4	5
x	30.6165153	75.2698553	3.75627832	-6.60462094	-54.1313519
y	.034170158	32.6792041	-75.1477340	-38.9788261	-46.2951026

最终聚类中心

	聚类				
	6	7	8	9	10
x	-33.3296675	-77.0957114	20.4686510	-18.2573120	71.7138339
y	65.5268442	15.5228199	59.1529332	10.7398449	-32.1795367

ANOVA

	聚类		误差			
	均方	自由度	均方	自由度	F	显著性
x	15629773.5	9	207.104	59990	75468.165	<.001
y	13070696.4	9	232.668	59990	56177.551	<.001

由于已选择聚类以使不同聚类中个案之间的差异最大化，因此 F 检验只应该用于描述目的。
实测显著性水平并未因此进行修正，所以无法解释为针对“聚类平均值相等”这一假设的检验。

每个聚类中的个案数目

聚类	1	5940.000
	2	5998.000
	3	5979.000
	4	5023.000
	5	6260.000
	6	6134.000
	7	6795.000
	8	6169.000
	9	5717.000
	10	5985.000
有效		60000.000
缺失		.000

Data written to
/Users/luoyiti/CodeProject/MNIST_Clustering/spss_examine/k_means_cluster.csv
4 variables and 60000 cases written.
Variable x Type: Number Width: 23 Dec: 19
Variable y Type: Number Width: 23 Dec: 19
Variable label Type: String Width: 3
Variable QCL_1 Type: Number Width: 8 Dec: 0

图形

备注

已创建输出		05-MAY-2025 16:40:29
注解		
输入	数据	/Users/luoyiti/CodeProject/MNIST_Clustering/data/tsne_df.csv
	活动数据集	数据集1
	过滤器	< 无 >
	权重	< 无 >
	拆分文件	< 无 >
	工作数据文件中的行数	60000
语法		GRAPH /SCATTERPLOT(BIVAR) =x WITH y BY label BY QCL_1 (IDENTIFY) /MISSING=LISTWISE /TITLE='K-Means'.
资源	处理程序时间	00:00:02.94
	耗用时间	00:00:02.00