

● 分析したデータセット

分析に使用した実世界データセットは、人間の歩行情報に関するデータセットである。このデータセットは、人間の歩行時の 9 か所の関節の角度(内積)が 1 秒間の間に 1/30 秒ごとに記録されたデータである。そのため、データの属性数は 270 である。データ数は一人当たり 30 個で、3 人分のデータが含まれている。

なお、このデータは私が高等専門学校在学中に卒業研究に使用するために作成したデータセットであり、規約上他の人に公開することはできない。また、身長や性別・年齢を始めとする各個人の情報の記録も規約上許されていない。このデータセットは、1-30 番目が一人目のデータ、31-60 番目が 2 人目のデータ、61-90 番目が 3 人目のデータになっている。

● 実装アルゴリズム

講義内で紹介されたアルゴリズムの内、私は Fuzzy c-Lines を実装しデータセットの分析を行った。ハイパーパラメータは $\theta = 2$ に設定した。また、アルゴリズムの終了条件は、損失関数の値が変化しないか、指定した繰り返し回数(ソースコード内の *itr*)に到達することとした。

● 実験結果

各人のデータに対するメンバシップ値を表 1(a)-(c)に示す。表内の太字は、各行において最大値をとったメンバシップ値を示す。表より、1 人目がクラスタ 1 に、2 人目がクラスタ 2 に、3 人目がクラスタ 3 におよそ割り当てられていることがわかる。しかし、各データについて、特定のクラスタに対してメンバシップ値が非常に大きな値となることはなかった。

上述の結果より、今回分析に使用したデータセットでは、各人のデータが類似していることが明らかとなった。特に、1 人目と 2 人目の間で類似性が高くなった。すなわち、1 人目のデータにおいてもクラスタ 2 に対するメンバシップ値が高く、同様に 2 人目のデータにおいてもクラスタ 1 に対するメンバシップ値が高かった。

一方、3 人目のデータは、1 人目、2 人目と比較すると若干類似性が低いと考えられる。これは、表 1(c)より、クラスタ 3 のメンバシップ値が高いデータが複数確認される、また各データにおいて、メンバシップ値が最大値をとるクラスタがクラスタ 3 であることが多いためである。

表 1(a) 1 人目の各データに対するメンバシップ値

data ID	Cluster 1	Cluster 2	Cluster 3
1	0.2209	0.2232	0.556
2	0.2285	0.2308	0.541
3	0.4566	0.4721	0.071
4	0.1561	0.1579	0.686
5	0.4489	0.4441	0.107
6	0.4616	0.4644	0.074
7	0.4827	0.4797	0.038
8	0.4433	0.4522	0.105
9	0.4377	0.4346	0.128
10	0.4300	0.4250	0.145
11	0.4568	0.4568	0.086
12	0.4698	0.4591	0.071
13	0.4626	0.4756	0.062
14	0.4694	0.4643	0.066
15	0.4733	0.4774	0.049
16	0.4638	0.4655	0.071
17	0.3176	0.3229	0.360
18	0.4347	0.4258	0.139
19	0.4147	0.4055	0.180
20	0.3935	0.3905	0.216
21	0.3589	0.3556	0.286
22	0.3719	0.3696	0.259
23	0.3618	0.3595	0.279
24	0.3644	0.3616	0.274
25	0.3368	0.3370	0.326
26	0.3368	0.3368	0.326
27	0.3545	0.3527	0.293
28	0.3343	0.3336	0.332
29	0.3397	0.3395	0.321
30	0.3345	0.3341	0.331

表 1(b) 2 人目の各データに対するメンバシップ値

data ID	Cluster 1	Cluster 2	Cluster 3
31	0.295	0.296	0.409
32	0.286	0.288	0.426
33	0.440	0.431	0.129
34	0.362	0.363	0.275
35	0.418	0.417	0.165
36	0.433	0.437	0.130
37	0.422	0.418	0.160
38	0.324	0.326	0.350
39	0.228	0.229	0.544
40	0.341	0.340	0.319
41	0.335	0.334	0.332
42	0.373	0.385	0.242
43	0.427	0.418	0.155
44	0.420	0.413	0.167
45	0.251	0.255	0.494
46	0.424	0.429	0.147
47	0.453	0.467	0.080
48	0.444	0.456	0.100
49	0.337	0.343	0.320
50	0.442	0.450	0.108
51	0.212	0.216	0.572
52	0.438	0.452	0.110
53	0.417	0.426	0.157
54	0.349	0.359	0.292
55	0.434	0.458	0.107
56	0.409	0.421	0.170
57	0.310	0.320	0.370
58	0.388	0.402	0.210
59	0.393	0.407	0.199
60	0.410	0.421	0.169

表 1(c) 3 人目の各データに対するメンバシップ値

data ID	Cluster 1	Cluster 2	Cluster 3
61	0.248	0.251	0.501
62	0.248	0.249	0.503
63	0.336	0.337	0.327
64	0.325	0.328	0.347
65	0.304	0.308	0.388
66	0.226	0.229	0.546
67	0.228	0.227	0.545
68	0.306	0.305	0.389
69	0.289	0.289	0.422
70	0.343	0.343	0.315
71	0.296	0.297	0.407
72	0.119	0.121	0.760
73	0.187	0.188	0.625
74	0.418	0.414	0.168
75	0.249	0.250	0.500
76	0.256	0.256	0.488
77	0.222	0.223	0.555
78	0.256	0.256	0.488
79	0.302	0.305	0.393
80	0.236	0.236	0.528
81	0.398	0.395	0.207
82	0.425	0.418	0.157
83	0.448	0.472	0.080
84	0.324	0.327	0.349
85	0.439	0.440	0.121
86	0.433	0.424	0.143
87	0.196	0.198	0.606
88	0.187	0.188	0.624
89	0.438	0.431	0.131
90	0.404	0.412	0.184