

Competency Assessment Assignments

May 31, 2022

Hao ZHANG

Task 1_a

Choose of approaches

- Because the original cancer datasets do not contain annotations, the effect of joint dimensionality reduction approaches on factors with clinical annotations or biological annotations does not need to be considered when processing such datasets. Just need to consider the effect of factors on survival.
- In general, MCIA, RGCCA, and JIVE achieved the best performances, finding factors significantly associated with survival in seven out of ten cancer types.

Task 1_a

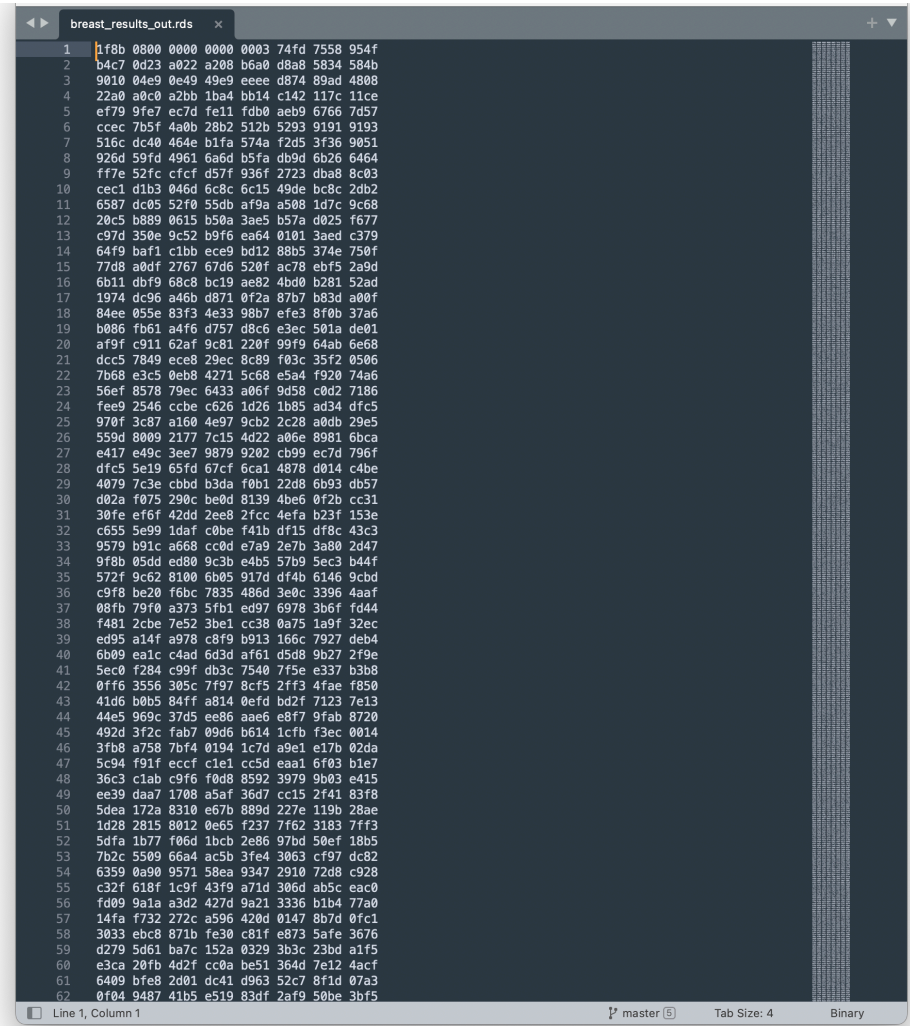
Choose of datasets

- Between the five multi-omics cancer datasets, I'd live to choose BIC (dataset of breast), LIHC (dataset of the liver) and SKCM (dataset of melanoma).
Because MCIA performed the best for melanoma; JIVE performed the best in liver cancer ; RGCCA performed the best in breast cancer.

Task 1_a

Results sample

The jDR factorization will decompose the P omics matrices into a product of a single factor matrix F and multiple weight/projection matrices A_i , $i=1 \dots P$.



```
breast_results_out.rds
1 1f8b 0800 0000 0000 0003 74fd 7558 954f
2 b4c7 0d23 a022 a208 b6a0 d8a8 5834 584b
3 9010 04e9 0e49 49e9 eeee d874 89ad 4808
4 22a0 a0c0 a2bb 1ba4 bb14 c142 117c 11ce
5 ef79 9fe7 ec7d fe11 fdb0 aeb9 6766 7d57
6 ccec 7b5f 4a0b 28b2 512b 5293 9191 9193
7 516c dc40 464e b1fa 574a f2d5 3f36 9051
8 926d 59fd 4961 6a6d b5fa db9d 6b26 6464
9 ff7e 52fc cfcf d57f 936f 2723 dba8 8c03
10 cec1 d1b3 046d 6c8c 6c15 49de bc8c 2db2
11 6587 dc05 52f0 55db af9a a508 1d7c 9c68
12 20c5 b889 0615 b50a 3ae5 b57a d025 f677
13 c97d 350e 9c52 b9f6 ea64 0101 3aed c379
14 64f9 ba1f c1bb ece9 bd12 88b5 374e 750f
15 77d8 a0df 2767 67d6 520f ac78 ebf5 2a9d
16 6b11 dbf9 68c8 bc19 ae82 4bd0 b281 52ad
17 1974 dc96 a46b d871 0f2a 87b7 b83d a00f
18 84ee 055e 83f3 4a33 98b7 efa3 8f0b 37a6
19 b086 fb61 a4f6 d757 d8c6 e3ec 501a de01
20 af9f c911 62af 9c81 220f 99f9 64ab 6e68
21 dcc5 7849 ece8 29ec 8c89 f03c 35f2 0506
22 7b68 e3c5 0eb8 4271 5c68 e5a4 f920 74a6
23 56ef 8578 79ec 6433 a06f 9d58 c0d2 7186
24 fee9 2546 ccbe c626 1d26 1b85 ad34 dfc5
25 970f 3c87 a160 4e97 9cb2 2c28 a0db 29e5
26 559d 8009 2177 7c15 4d22 a06e 8981 6bca
27 e417 e49c 3ee7 9879 9202 cb99 ec7d 796f
28 dfc5 5e19 65fd 67cf 6ca1 4878 d014 c4be
29 4079 7c3e cbbd b3da f0b1 22d8 6b93 db57
30 d02a f075 290c be0d 8139 4be6 0f2b cc31
31 30fe ef6f 42dd 2ee8 2fcc 4efa b23f 153e
32 c655 5e99 1daf c0be f41b df15 df8c 43c3
33 9579 b91c a668 cc0d e7a9 2e7b 3a80 2d47
34 9f8b 05dd ed80 9c3b e4b5 57b9 5ec3 b44f
35 572f 9c62 8100 6b05 917d df4b 6146 9cbd
36 c9f8 be20 f6bc 7835 486d 3e0c 3396 aaaf
37 08fb 79f0 a373 5fb1 ed97 6978 3b6f fd44
38 f481 2cbe 7e52 3be1 cc38 0a75 1a9f 32ec
39 ed95 a14f a978 c8f9 b913 166c 7927 deb4
40 6b09 ea1c c4ad 6d3d af61 d5d8 9b27 2f9e
41 5ec0 f284 c99f db3c 7540 7f5e e337 b3b8
42 0ff6 3556 305c 7f97 8cf5 2ff3 4fae f850
43 41d6 b0b5 84ff a814 0efd bd2f 7123 7e13
44 44e5 969c 37d5 ee86 aae6 e8f7 9fab 8720
45 492d 3f2c fab7 09d6 b614 1cfb f3ec 0014
46 3fb8 a758 7bf4 0194 1c7d a9e1 e17b 02da
47 5c94 f91f eccf c1e1 cc5d eaa1 6f03 b1e7
48 36c3 c1ab c9f6 f0d8 8592 3979 9b03 e415
49 ee39 daa7 1708 a5af 36d7 cc15 2f41 83f8
50 5dea 172a 8310 e67b 889d 227e 119b 28ae
51 1d28 2815 8012 0e65 f237 7f62 3183 7ff3
52 5dfa 1b77 f06d 1bcb 2e86 97bd 50ef 18b5
53 7b2c 5509 66a4 ac5b 3fe4 3063 cf97 dc82
54 6359 0a90 9571 58ea 9347 2910 72d8 c928
55 c32f 618f 1c9f 43f9 a71d 306d ab5c eac0
56 fd09 9a1a a3d2 427d 9a21 3336 b1b4 77a0
57 14fa f732 272c a596 420d 0147 8b7d 0fc1
58 3033 ebc8 871b fe30 c81f e873 5afe 3676
59 d279 5d61 ba7c 152a 0329 3b3c 23bd a1f5
60 e3ca 20fb 4d2f cc0e ba51 3640 7e12 4ac4
61 6400 bfe8 2d01 dc41 d963 52c7 8f1d 07a3
62 0f04 9487 41b5 e519 83df 2af9 50be 3bf5
```

Task 1_b

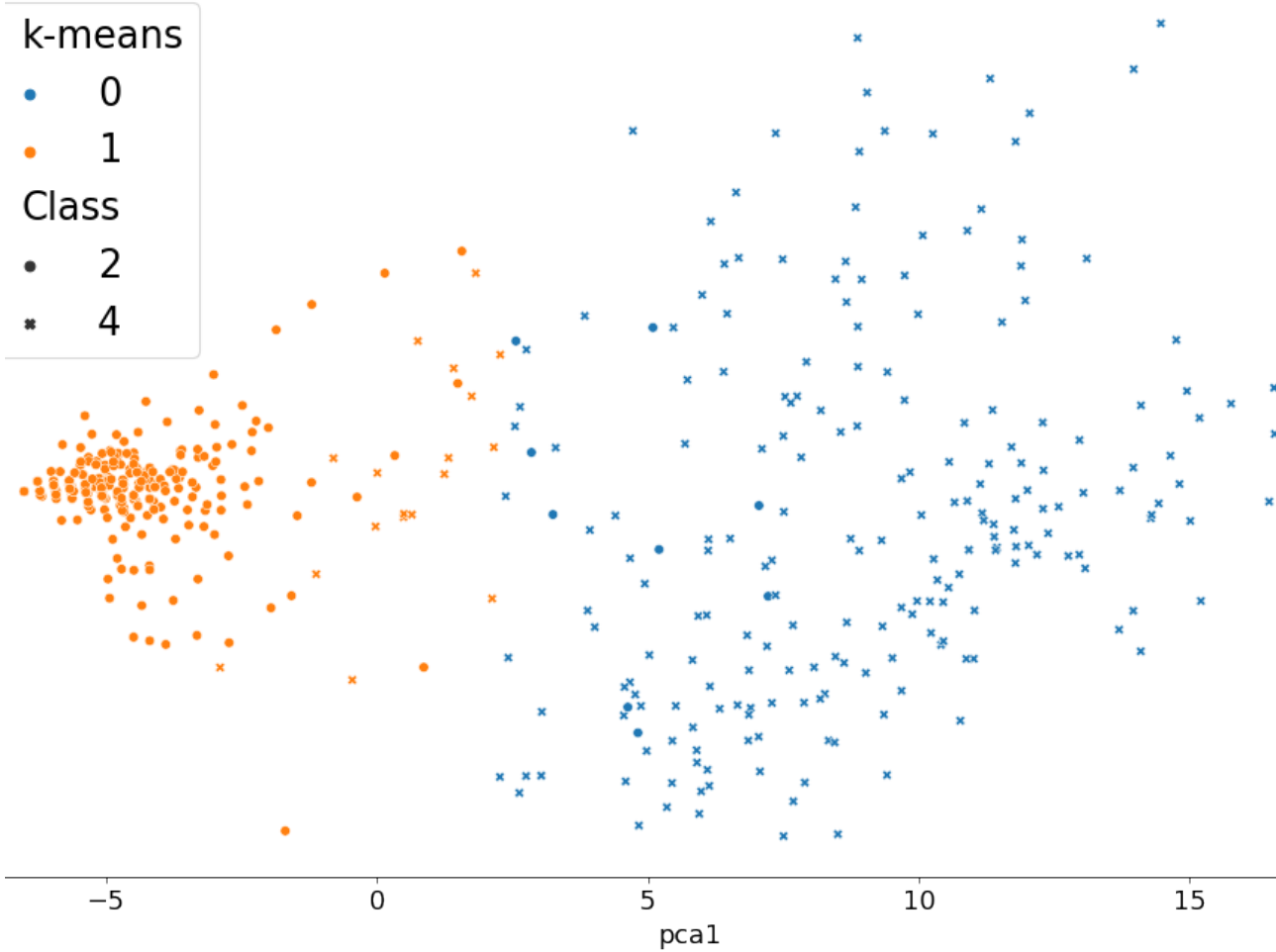
Choose of approaches

- No algorithm consistently outperformed all others in either differential survival or enriched clinical parameters. With respect to survival, MCCA had the best prognostic value, while MultiNMF was second and LRACluster third.
 - Since my system currently fails to install PMA (MCCA) successfully, I install SNFtool (SNF) instead.
- Unfortunately, due to a limited time, I couldn't test the result of task1_b in the existing time.

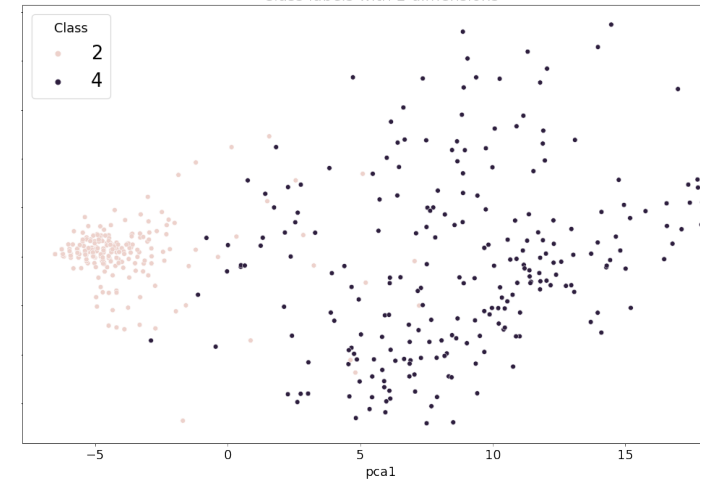
Task2

**Initial classification and clustering
effects**

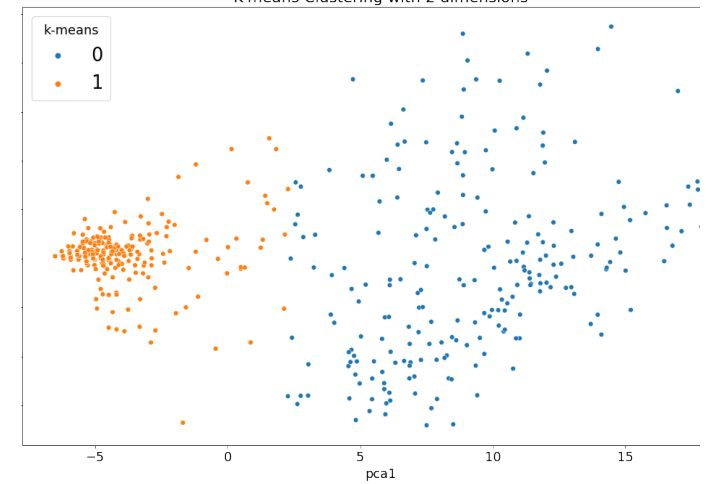
K-means Clustering and class labels with 2 dimensions



Class labels with 2 dimensions



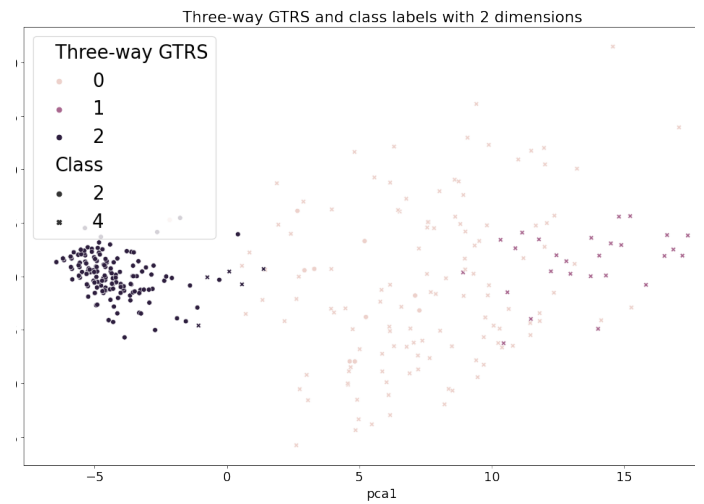
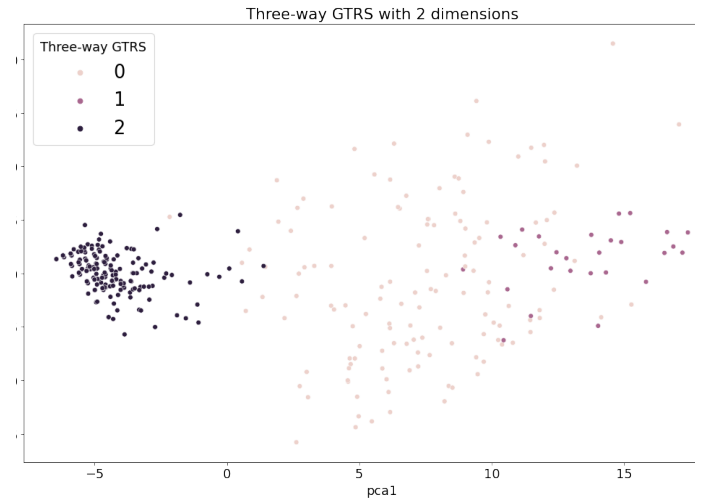
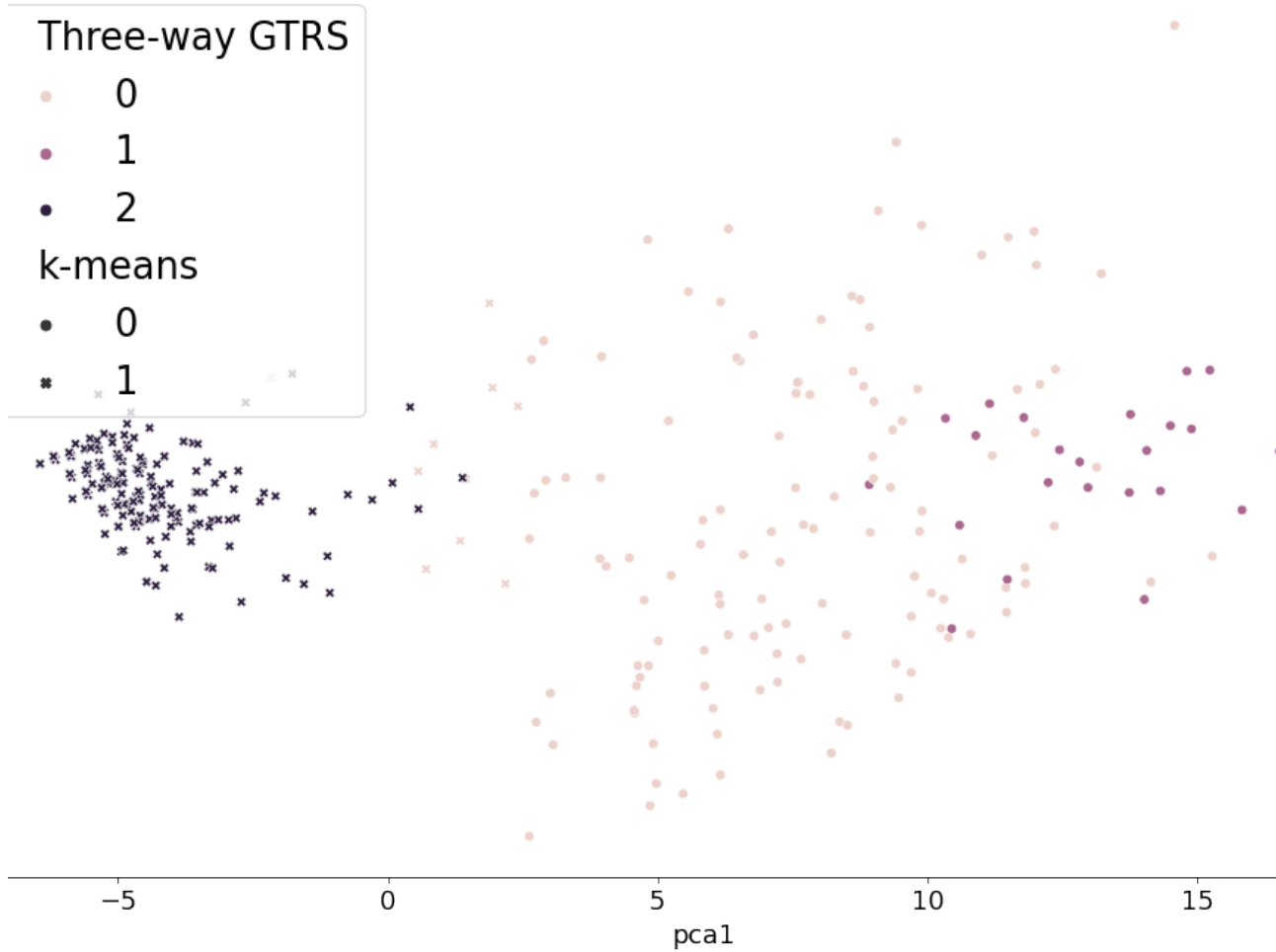
K-means Clustering with 2 dimensions



Task2

**Classification and clustering effects after
implementing the GTRS three-way clustering
algorithms**

Three-way GTRS and K-means Clustering with 2 dimensions



Task 2

Discussion of the limitations of the algorithm

- The GTRS three-way clustering algorithm can effectively distinguish the data distributed in the two extreme parts (inside and outside), but once the algorithm is overfitting or underfitting in the process of use, it may cause the operation time to be too long or face the large amount of data that deals with partial parts cannot be differentiated.