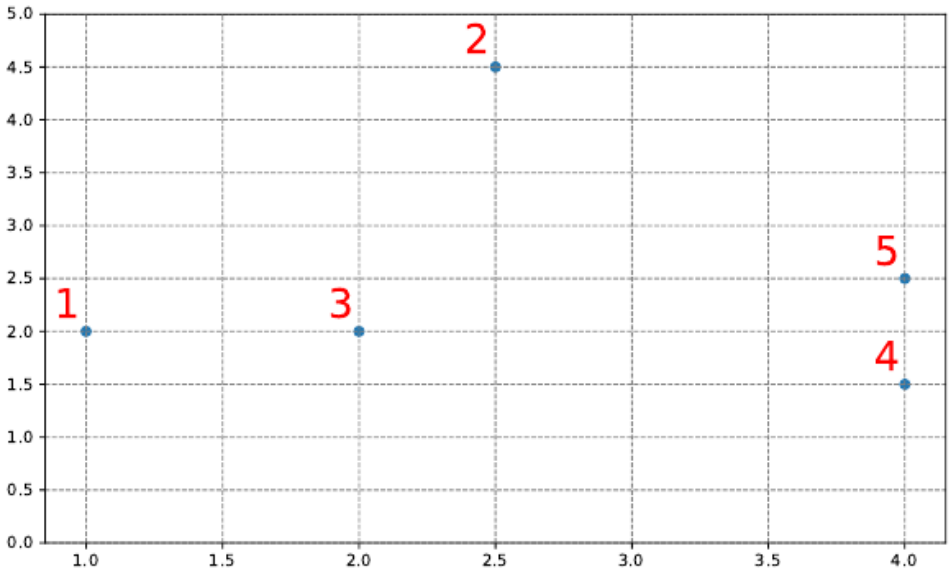
Machine Learning per la Fisica Applicata e la Fisica delle Alte Energie

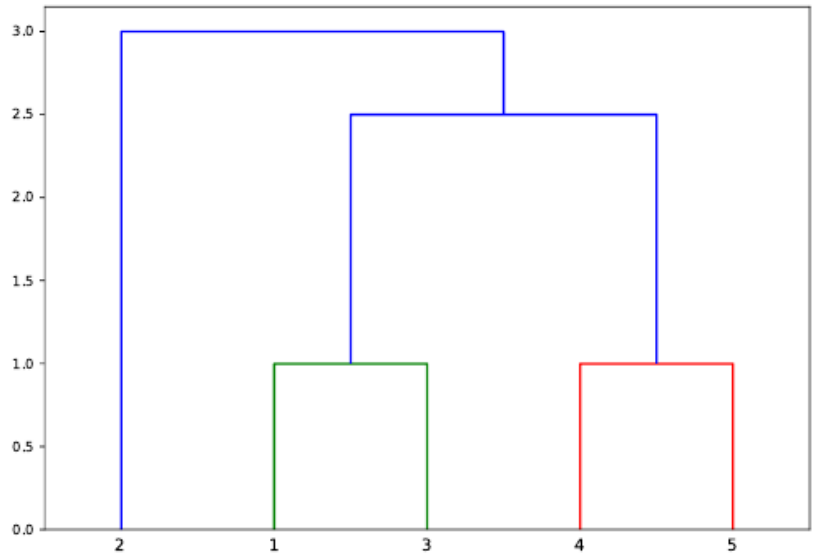
Lezione 21: Clustering

Emanuele R. Nocera

19 dicembre 2022

UNIVERSITÀ DI TORINO


1
2
3
4
5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0


3.0
2.5
2.0
1.5
1.0
0.5
0.0
2
1
3
4
5

1 Initialize clusters as singletons: **for** $i \leftarrow 1$ **to** n **do** $C_i \leftarrow \{i\}$

2

3 Initialize set of clusters available for merging: $S \leftarrow \{1, \ldots, n\}$; **repeat**

4 Pick 2 most similar clusters to merge: $(j, k) \leftarrow \arg\min_{j,k \in S} d_{j,k}$

5 Create new cluster $C_\ell \leftarrow C_j \cup C_k$

6 Mark j and k as unavailable: $S \leftarrow S \setminus \{j, k\}$

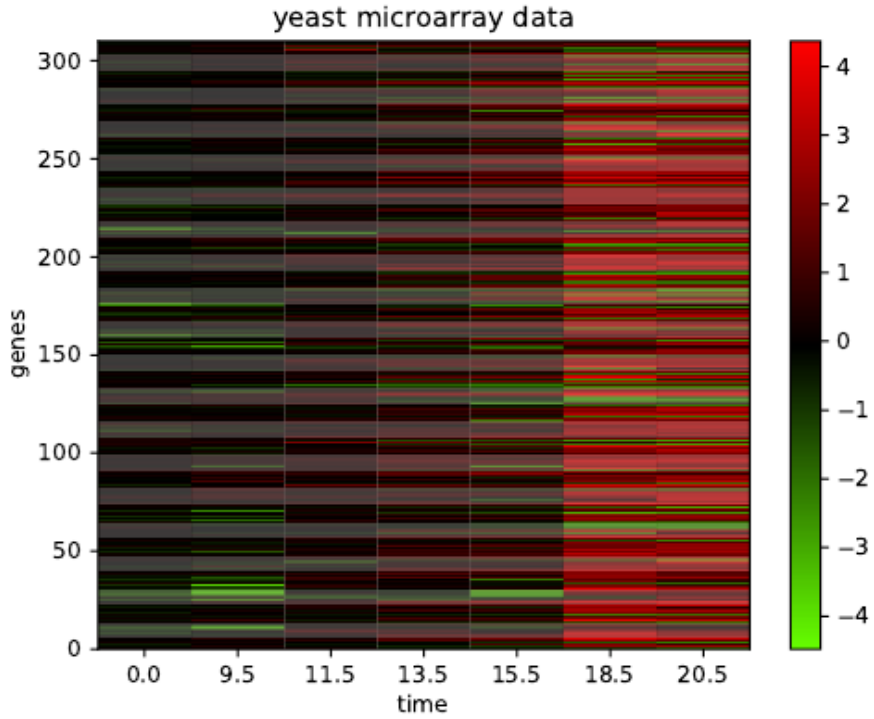
7 **if** $C_\ell \neq \{1, \ldots, n\}$ **then**

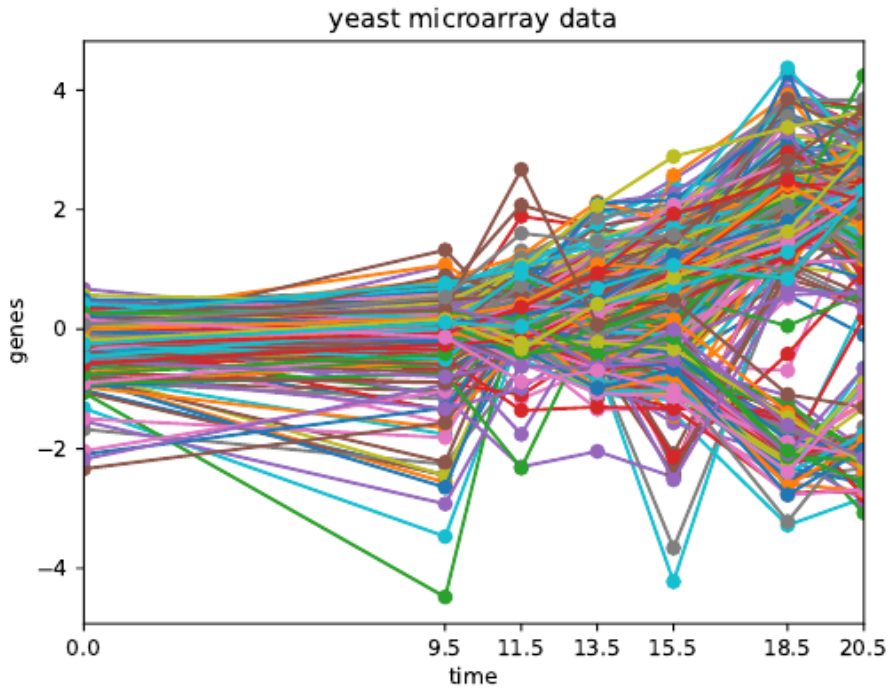
8 Mark ℓ as available, $S \leftarrow S \cup \{\ell\}$

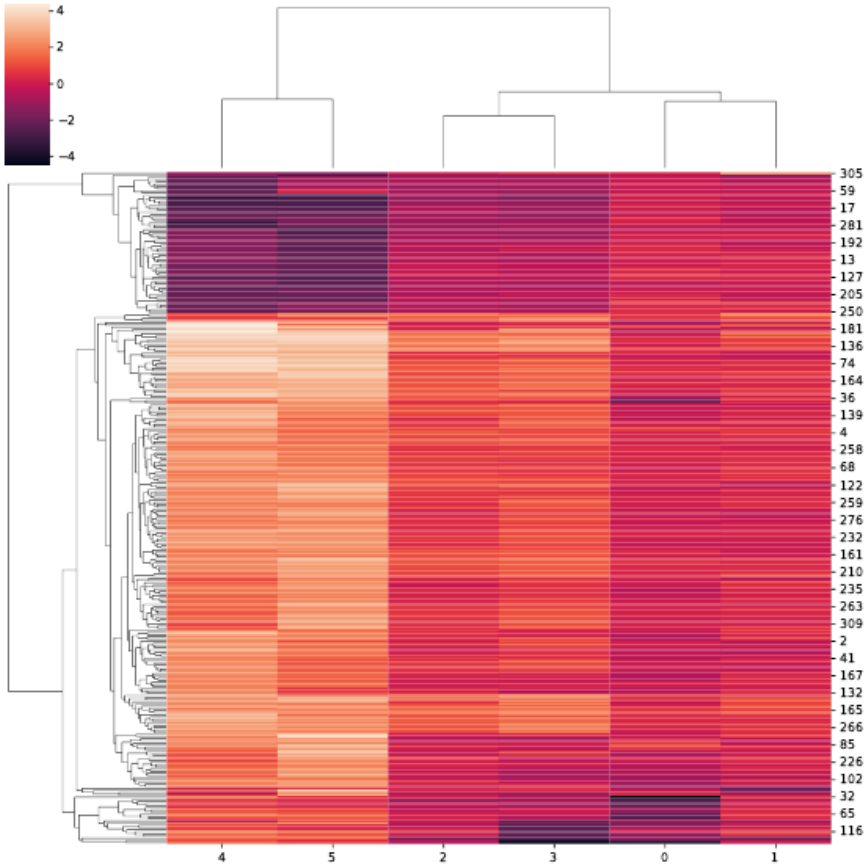
9 **foreach** $i \in S$ **do**

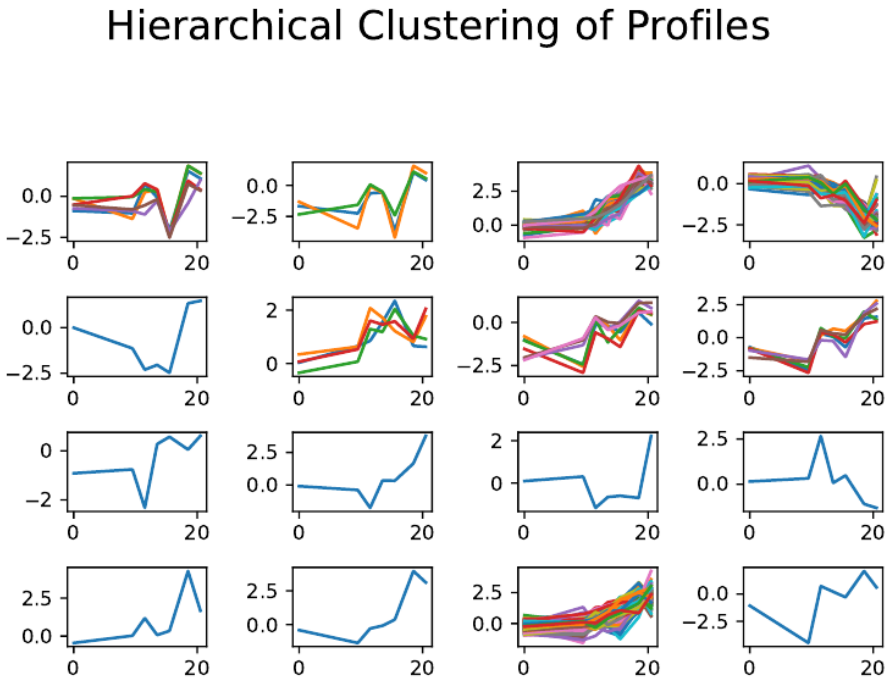
10 Update dissimilarity matrix $d(i, \ell)$

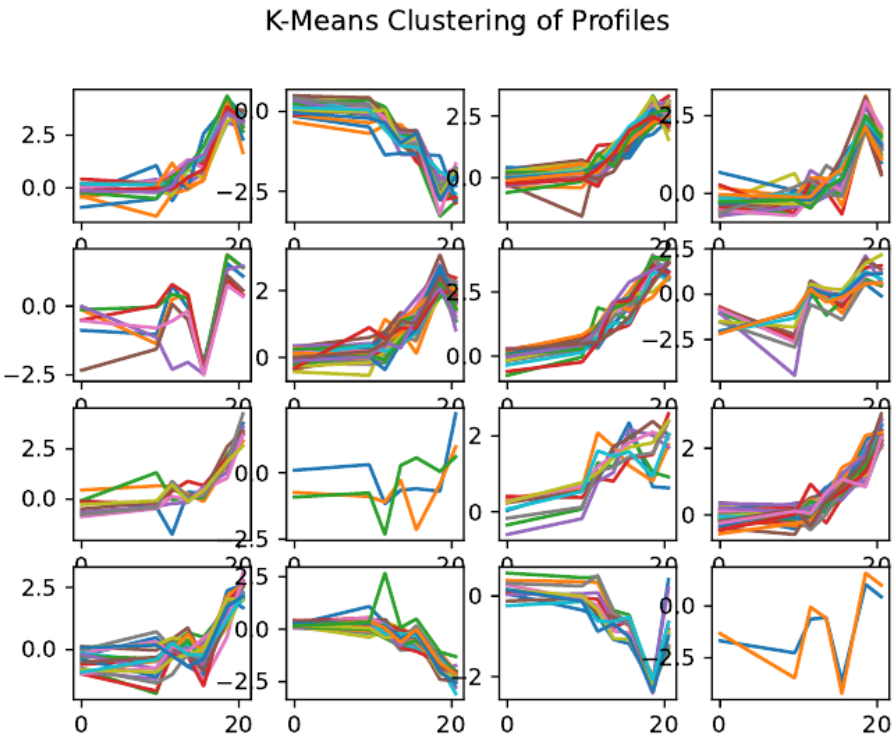
11 **until** *no more clusters are available for merging*

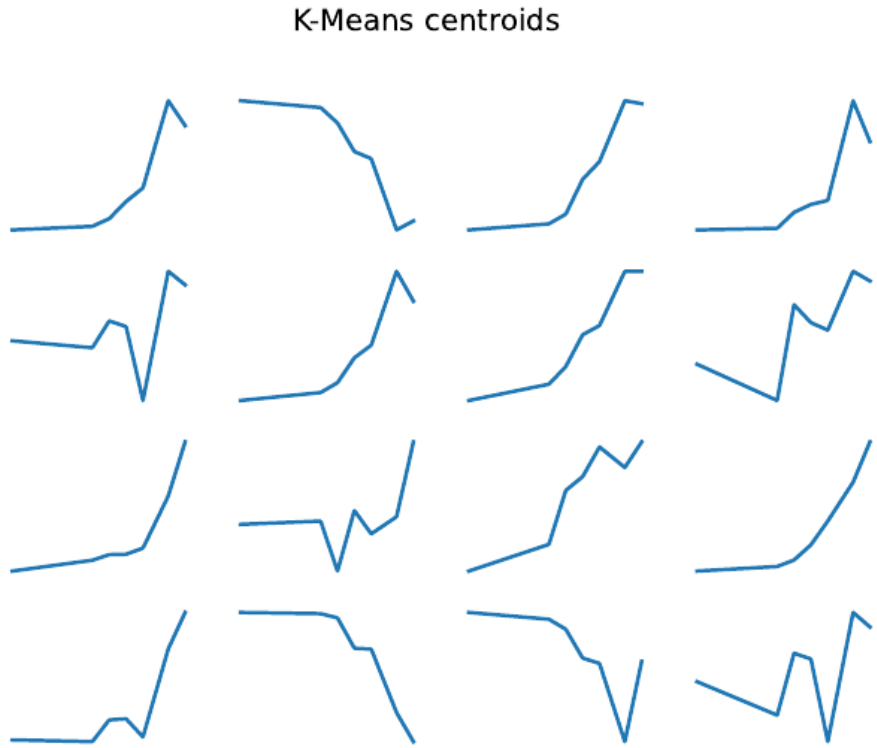

yeast microarray data
genes
time
0.0
9.5
11.5
13.5
15.5
18.5
20.5
0
50
100
150
200
250
300
4
3
2
1
0
−1
−2
−3
−4

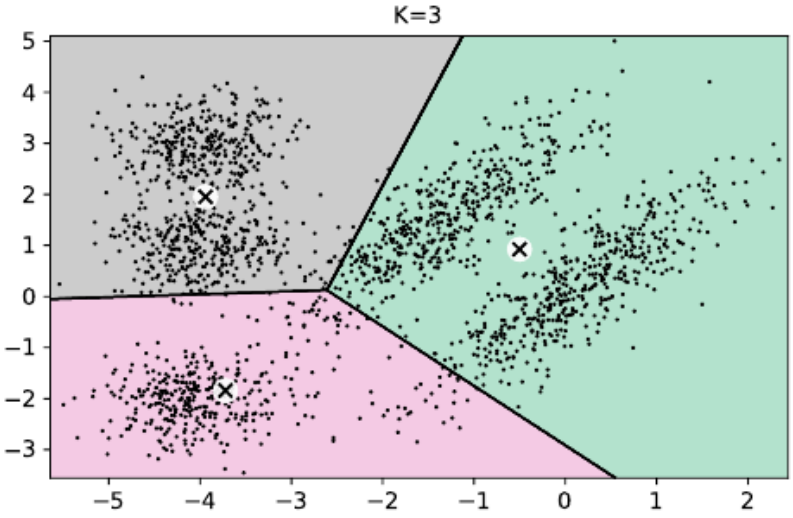

yeast microarray data
genes
time
0.0
9.5
11.5
13.5
15.5
18.5
20.5
4
2
0
−2
−4


4
2
0
−2
−4
305
59
17
281
192
13
127
205
250
181
136
74
164
36
139
4
258
68
122
259
276
232
161
210
235
263
309
2
41
167
132
165
266
85
226
102
32
65
116
4
5
2
3
0
1

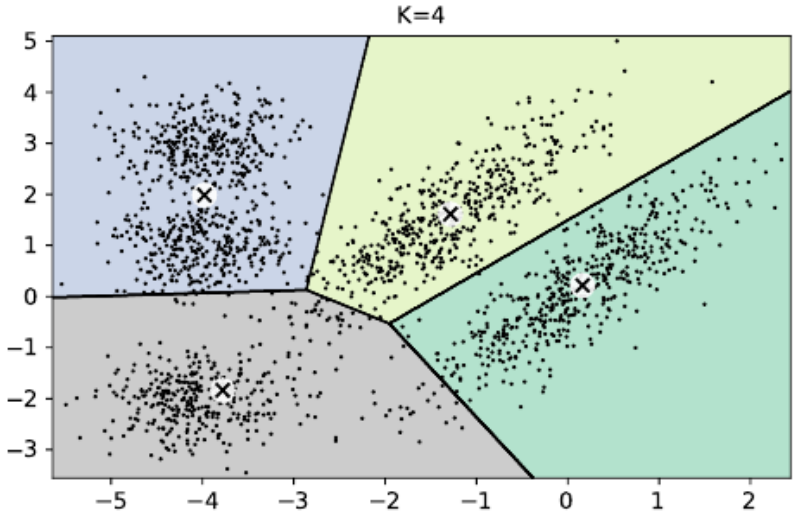

Hierarchical Clustering of Profiles


K-Means Clustering of Profiles

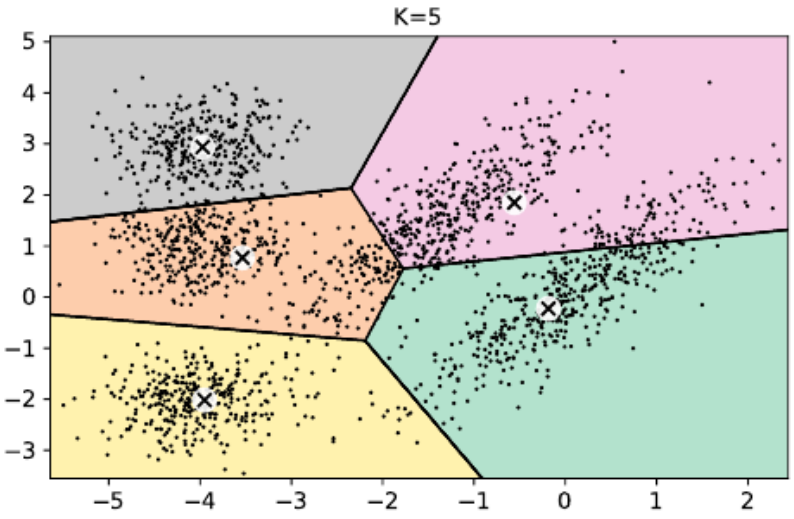

K-Means centroids



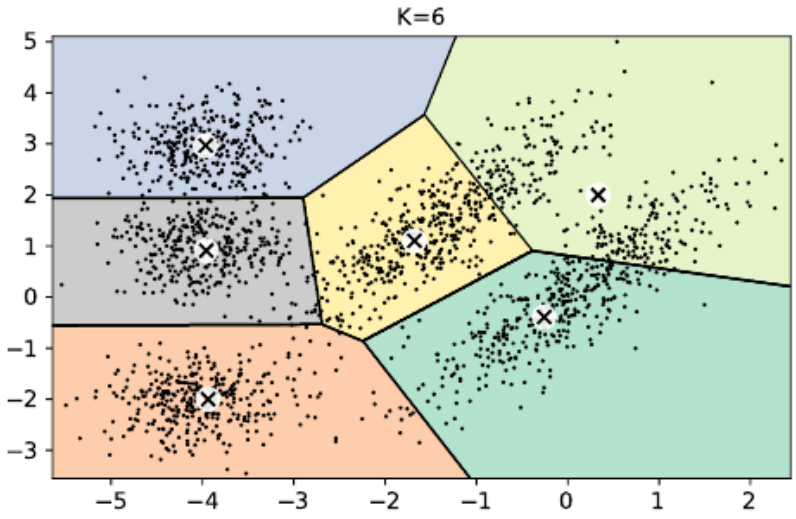
(a)



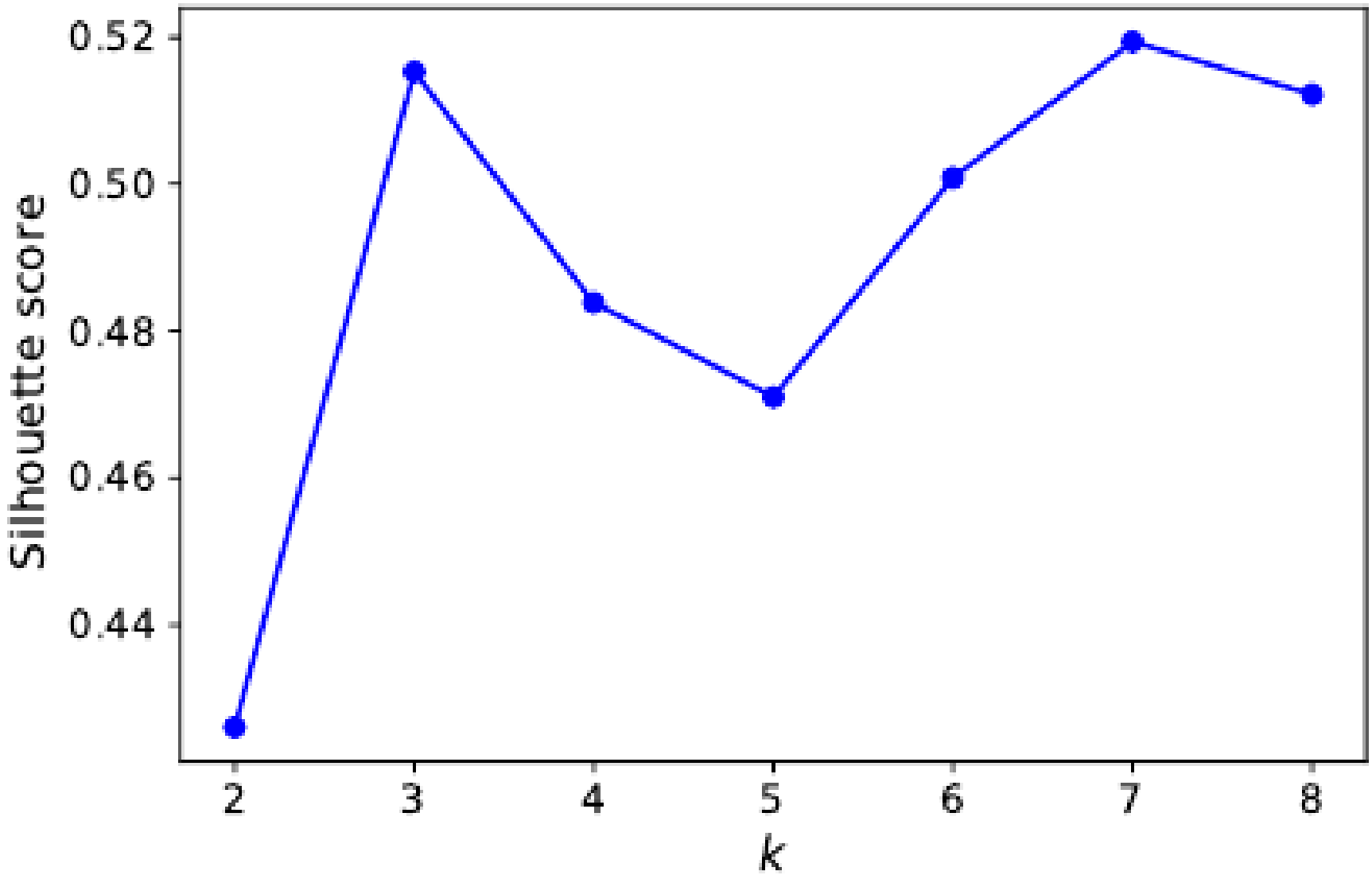
(b)



(c)



(d)


Silhouette score
0.52
0.50
0.48
0.46
0.44
2
3
4
5
6
7
8
k