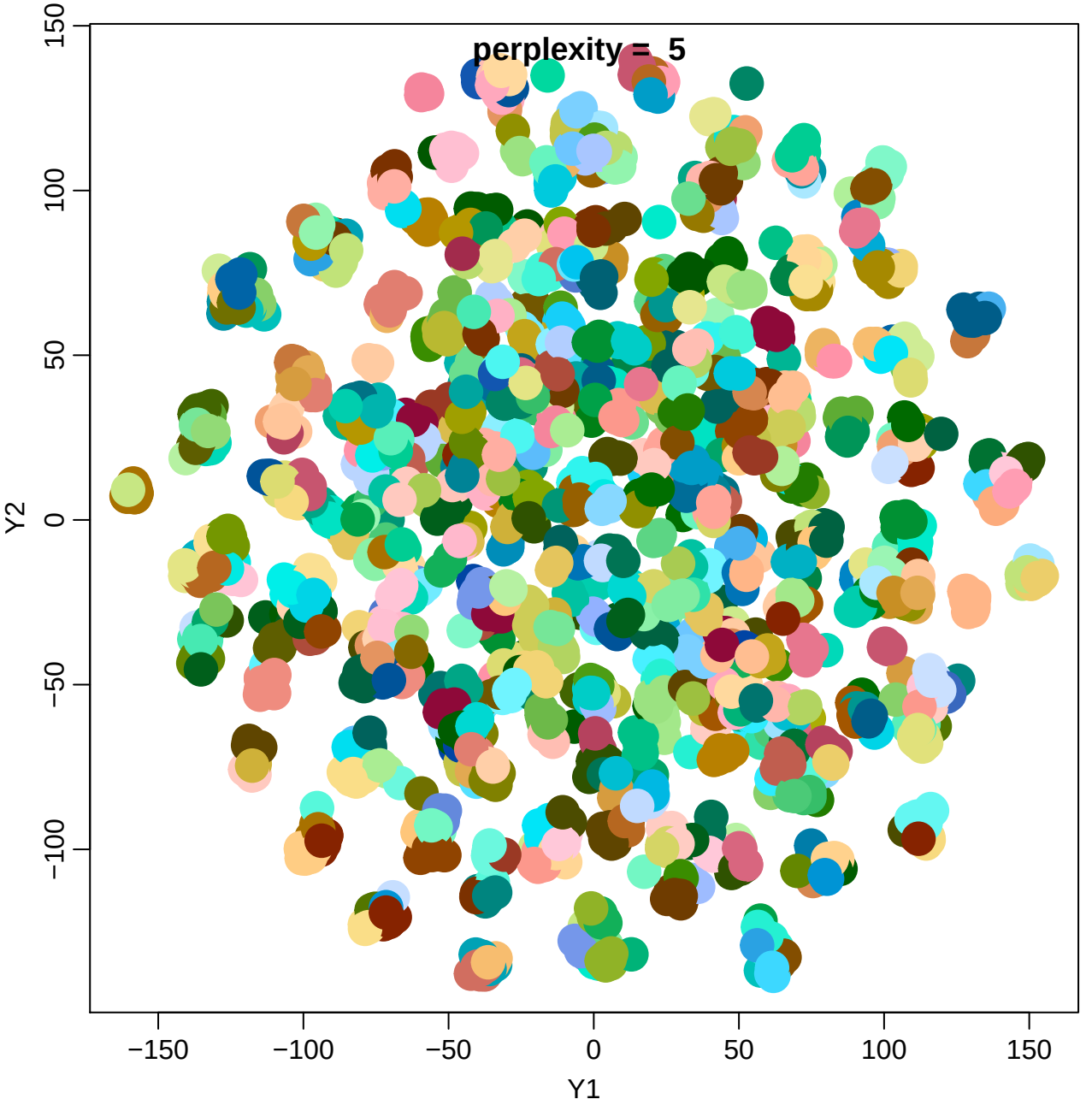
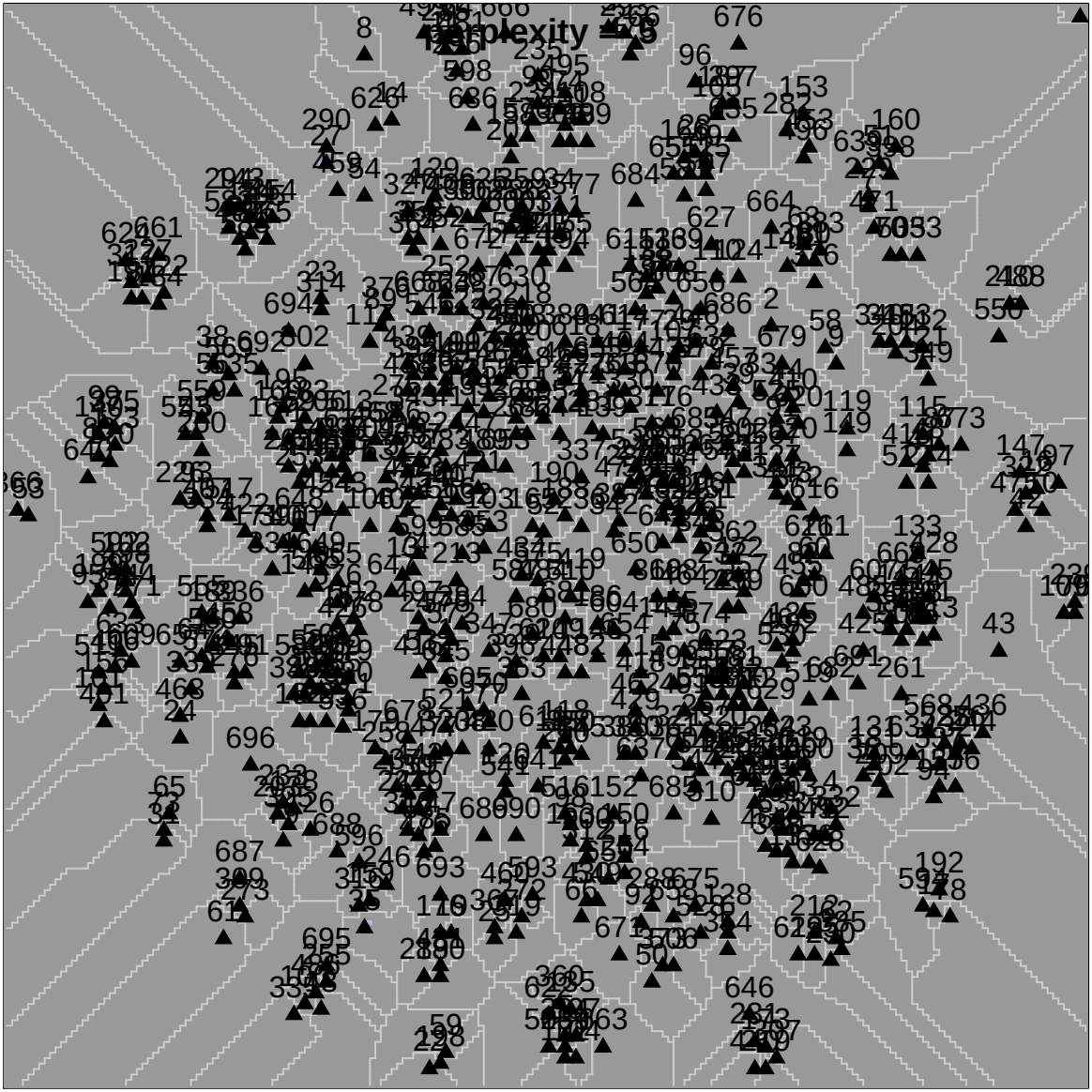
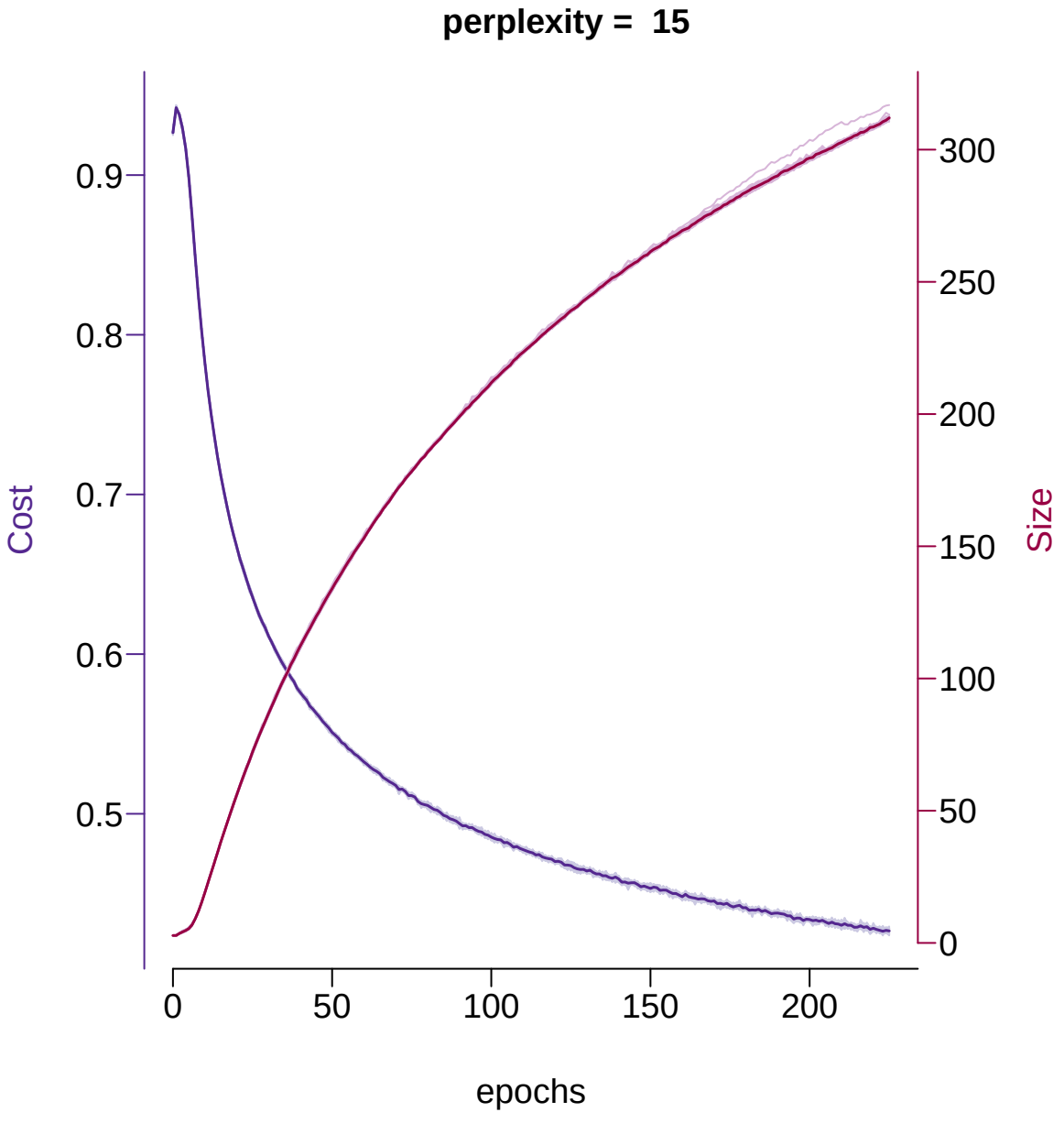
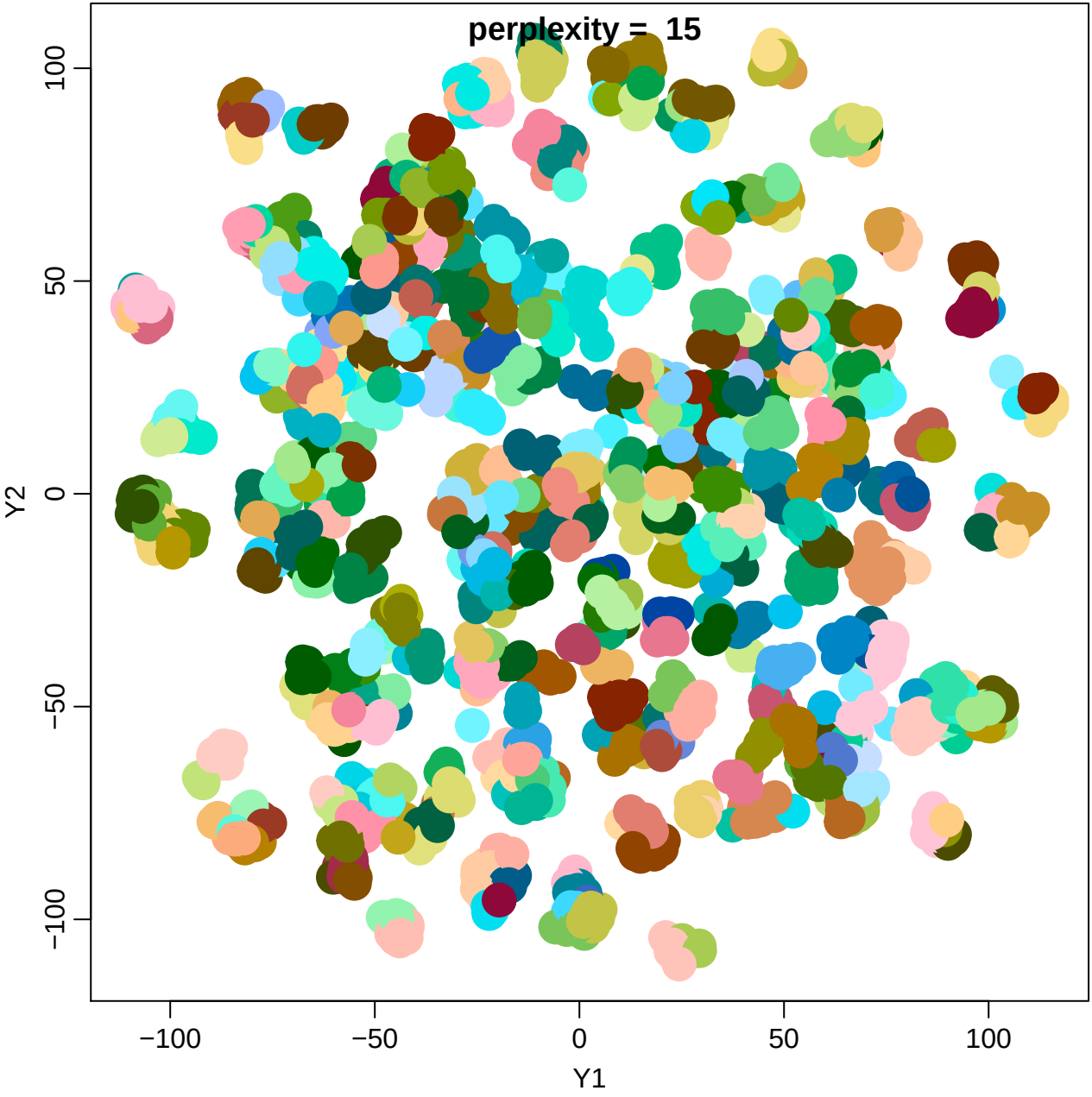
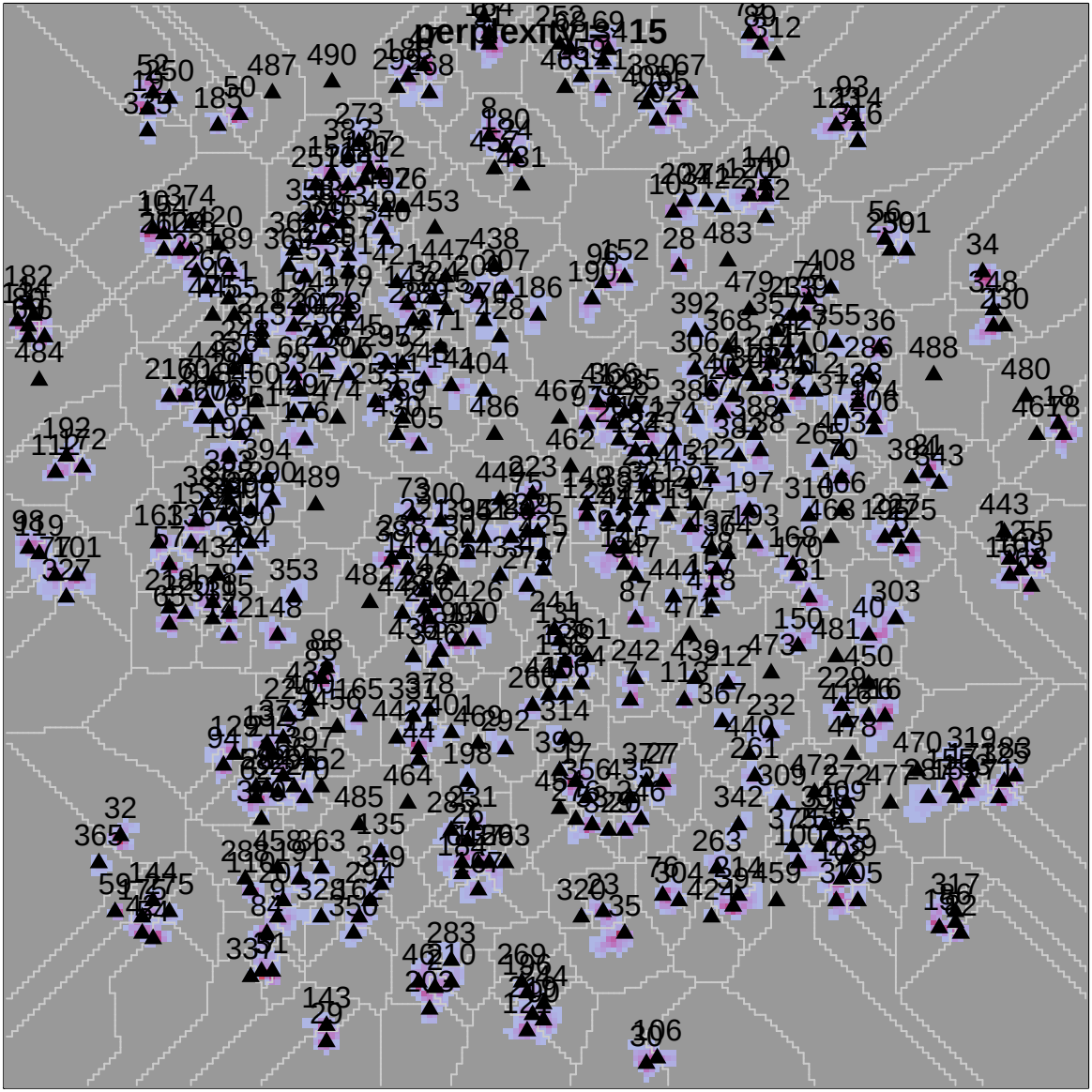
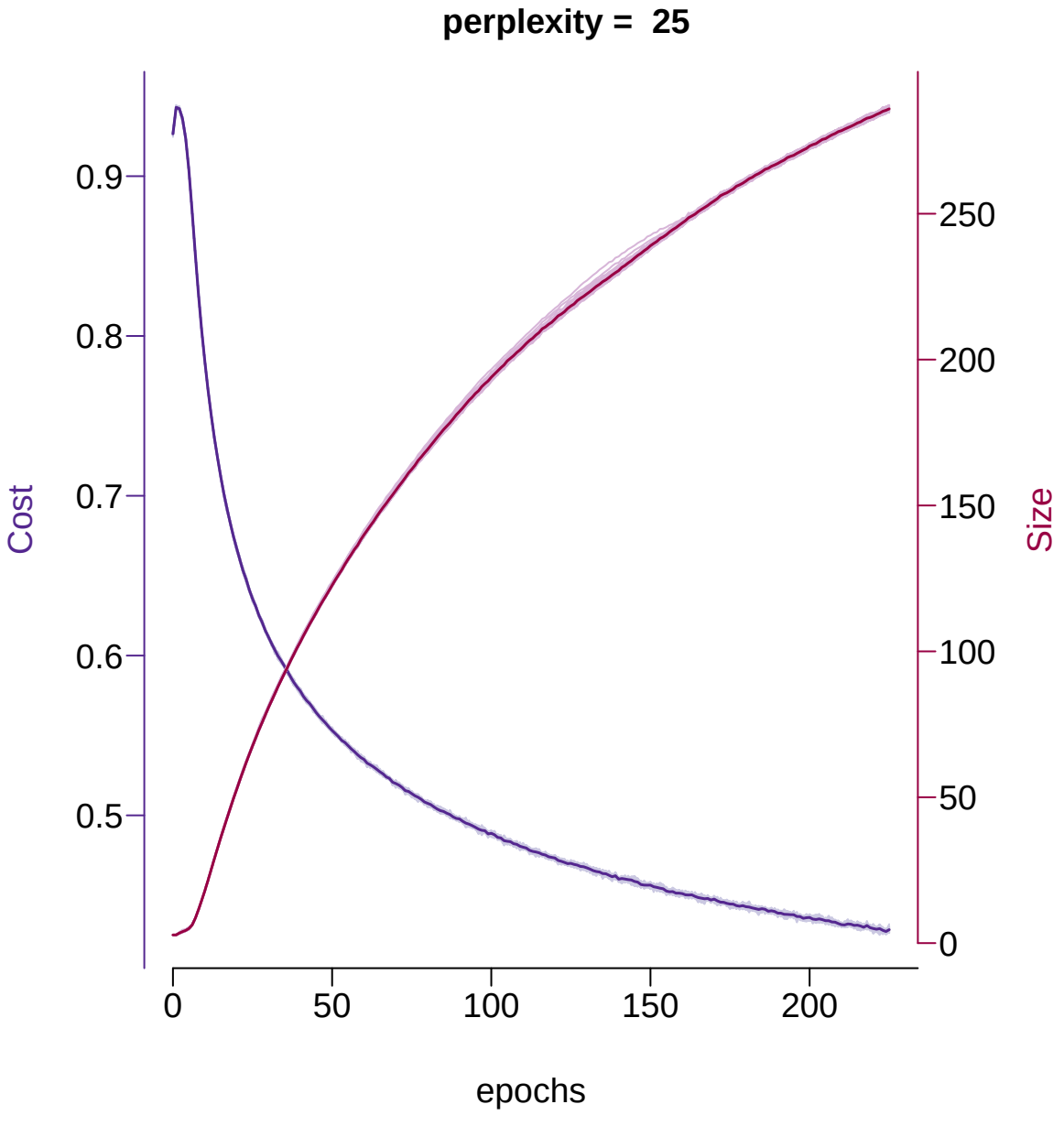

perplexity = 5
Y2
Y1
150
100
50
0
−50
−100
−150
−100
−50
0
50
100
150

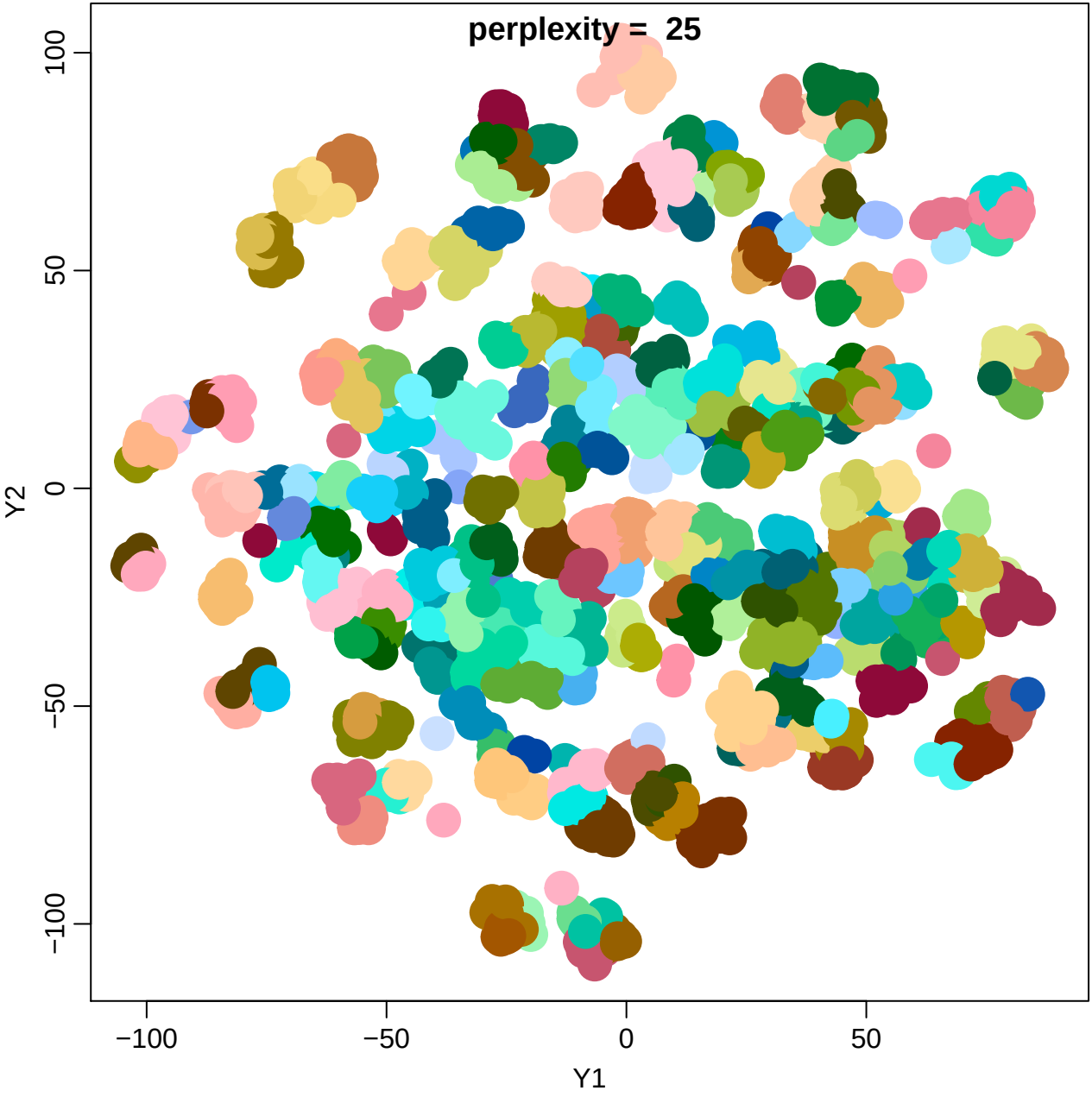

complexity = 5

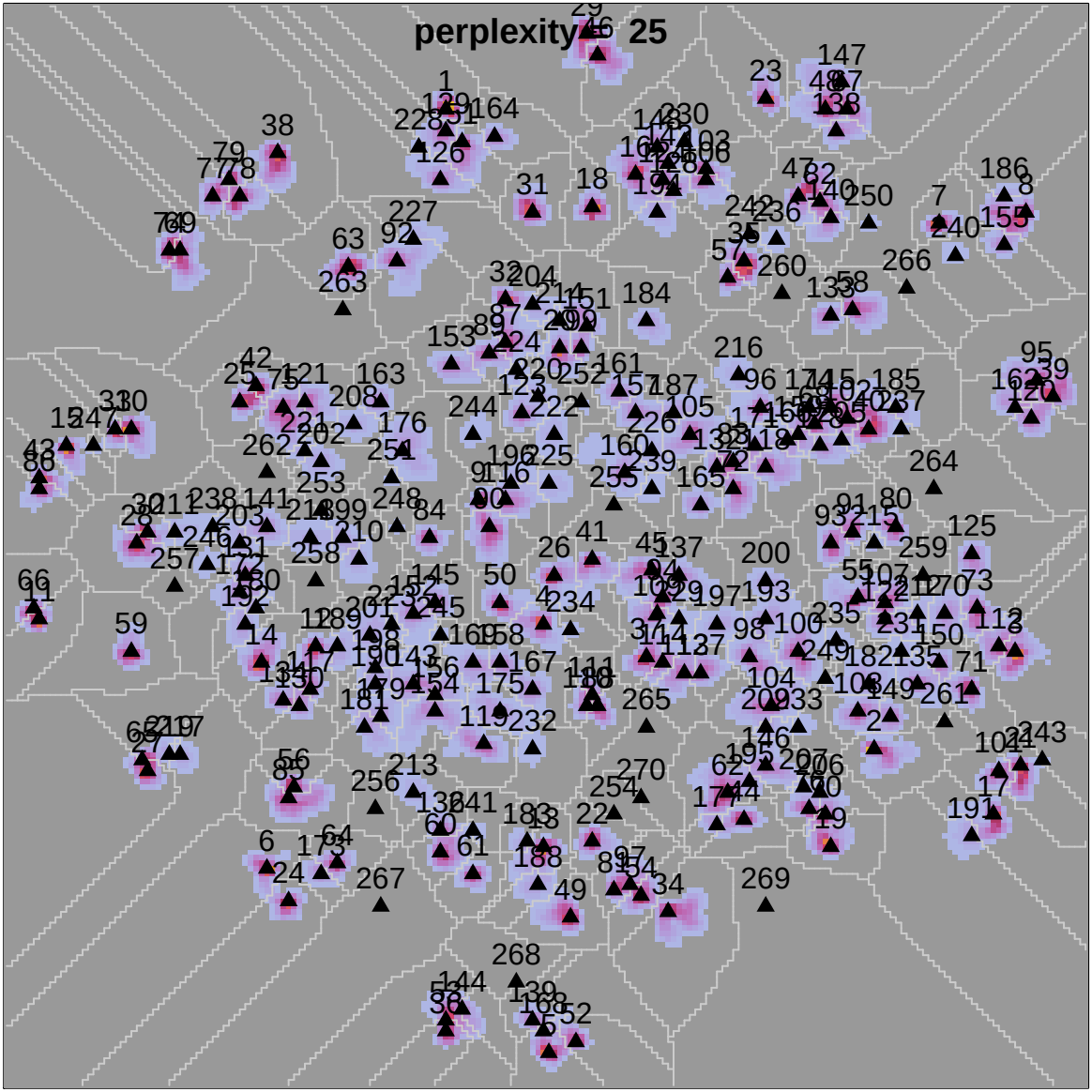

perplexity = 15
Cost
0.9
0.8
0.7
0.6
0.5
Size
300
250
200
150
100
50
0
0
50
100
150
200
epochs

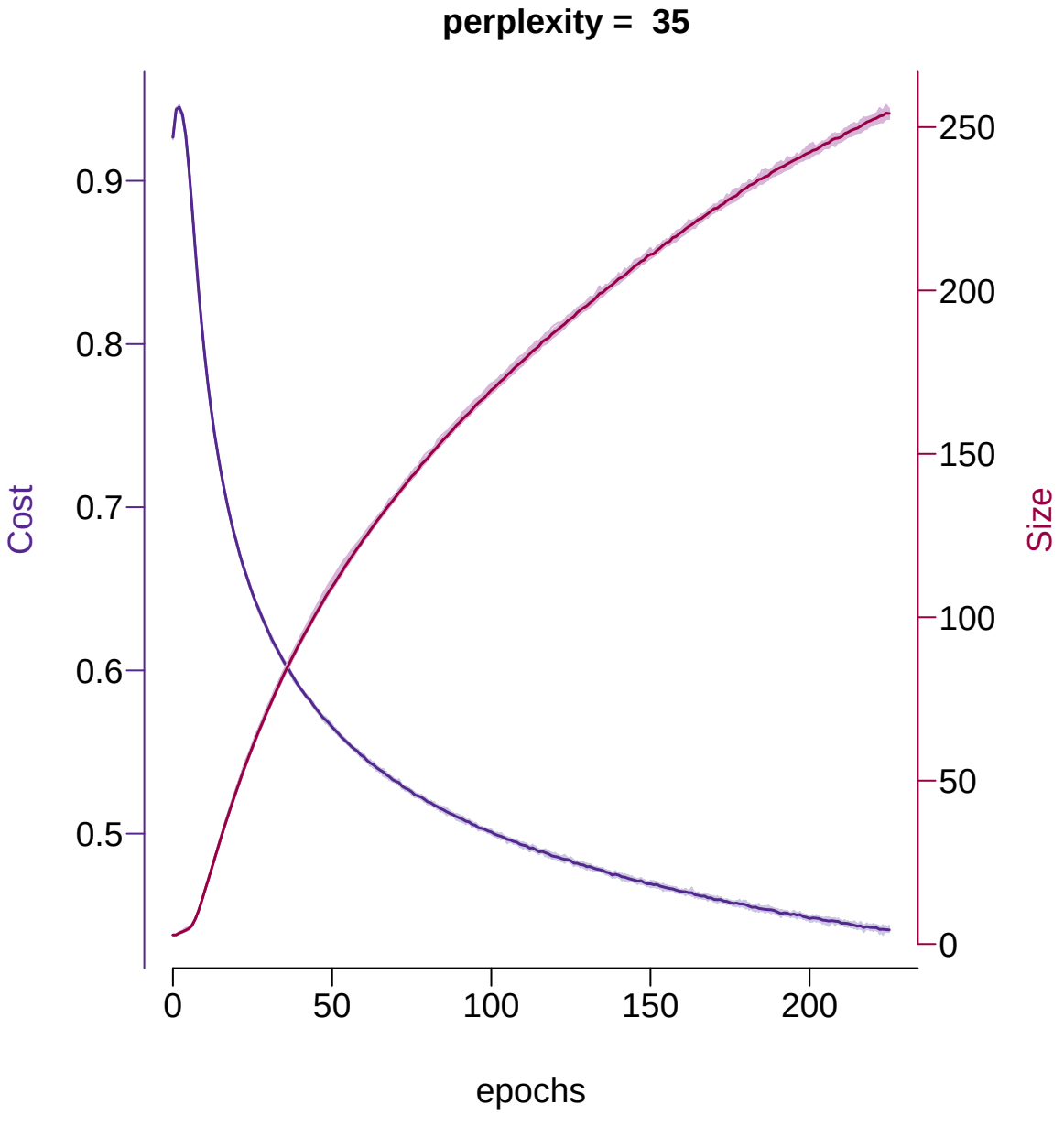

perplexity = 15
Y2
100
50
0
−50
−100
−100
−50
0
50
100
Y1

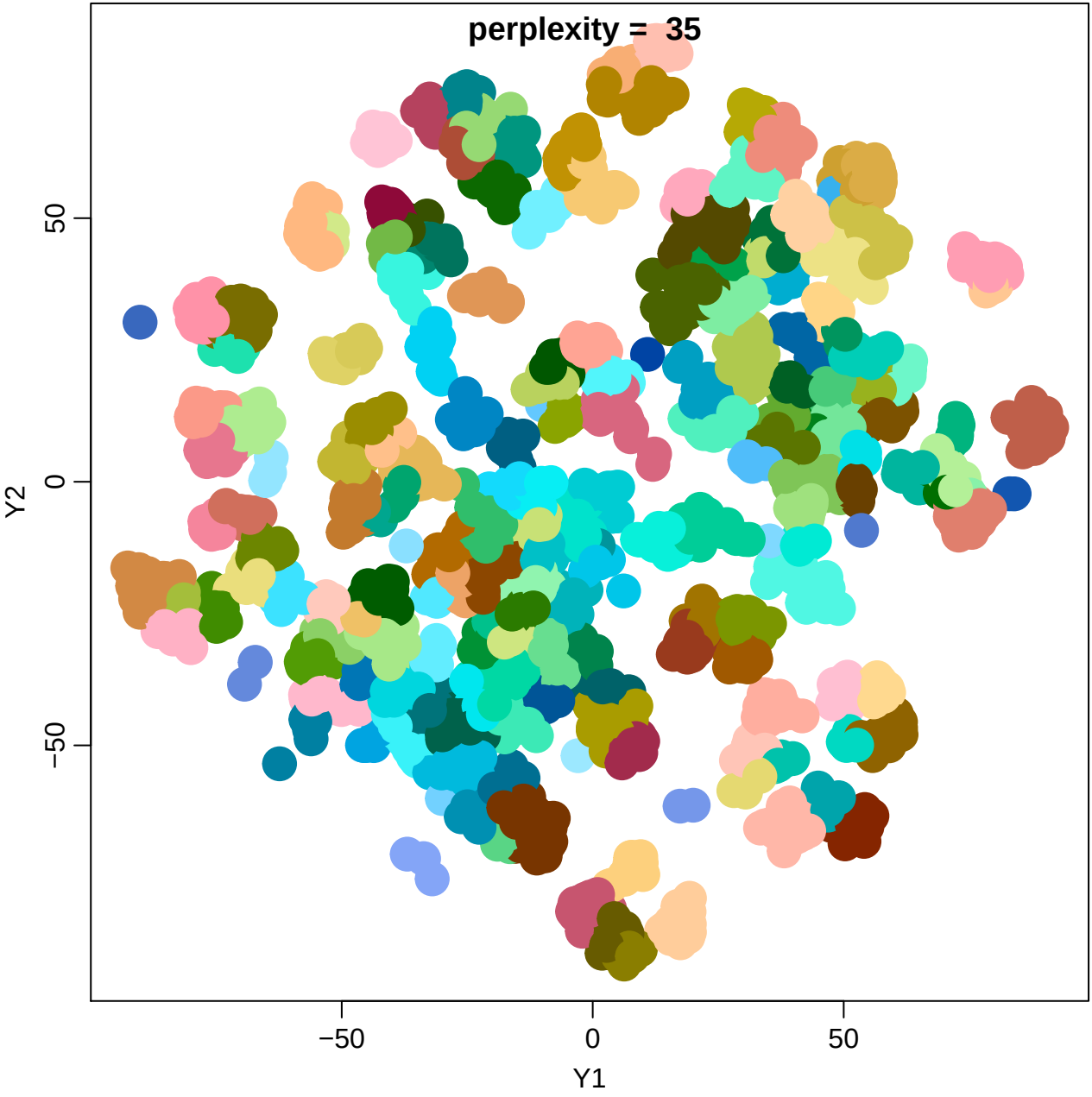

perplexity 415

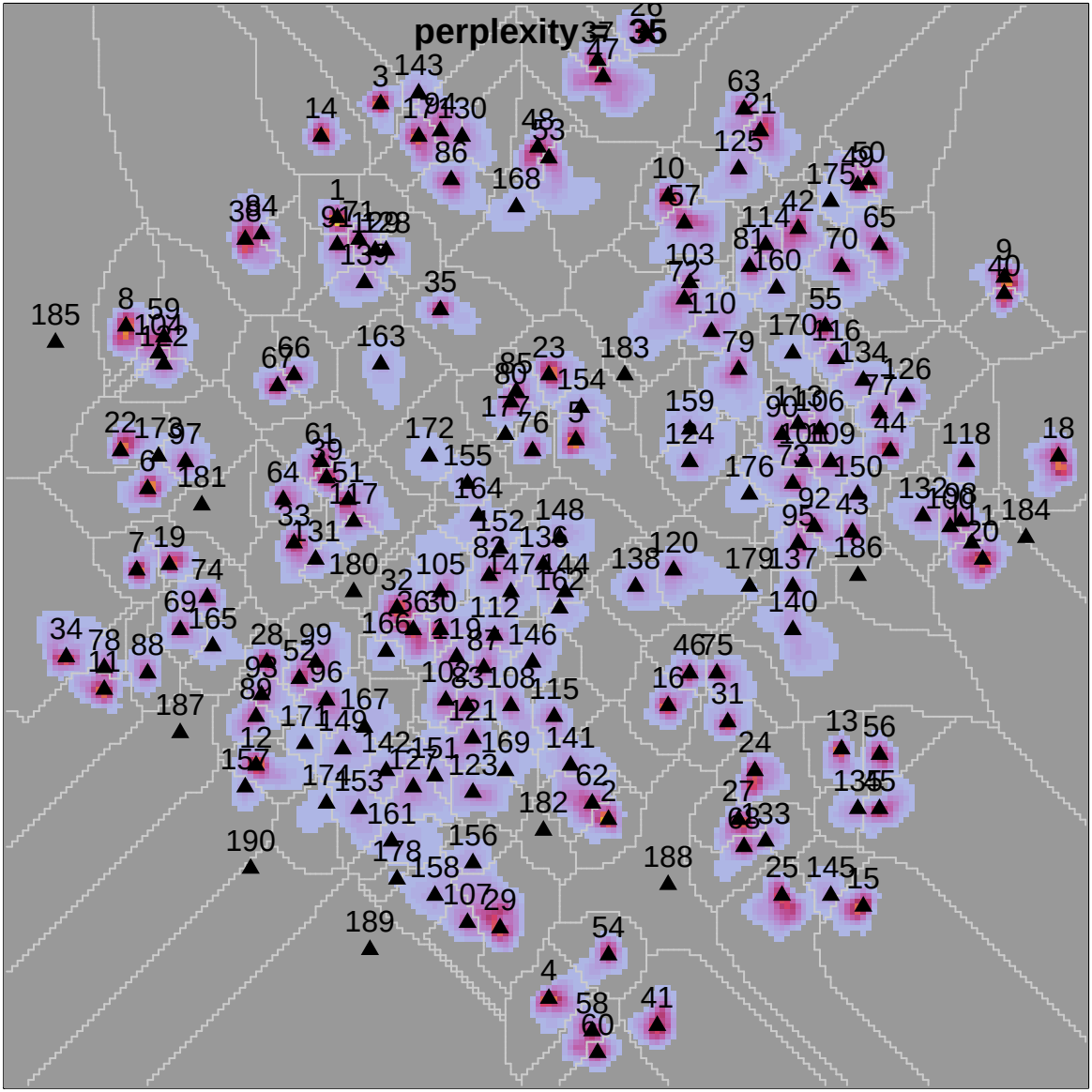

perplexity = 25
Cost
0.9
0.8
0.7
0.6
0.5
Size
250
200
150
100
50
0
0
50
100
150
200
epochs

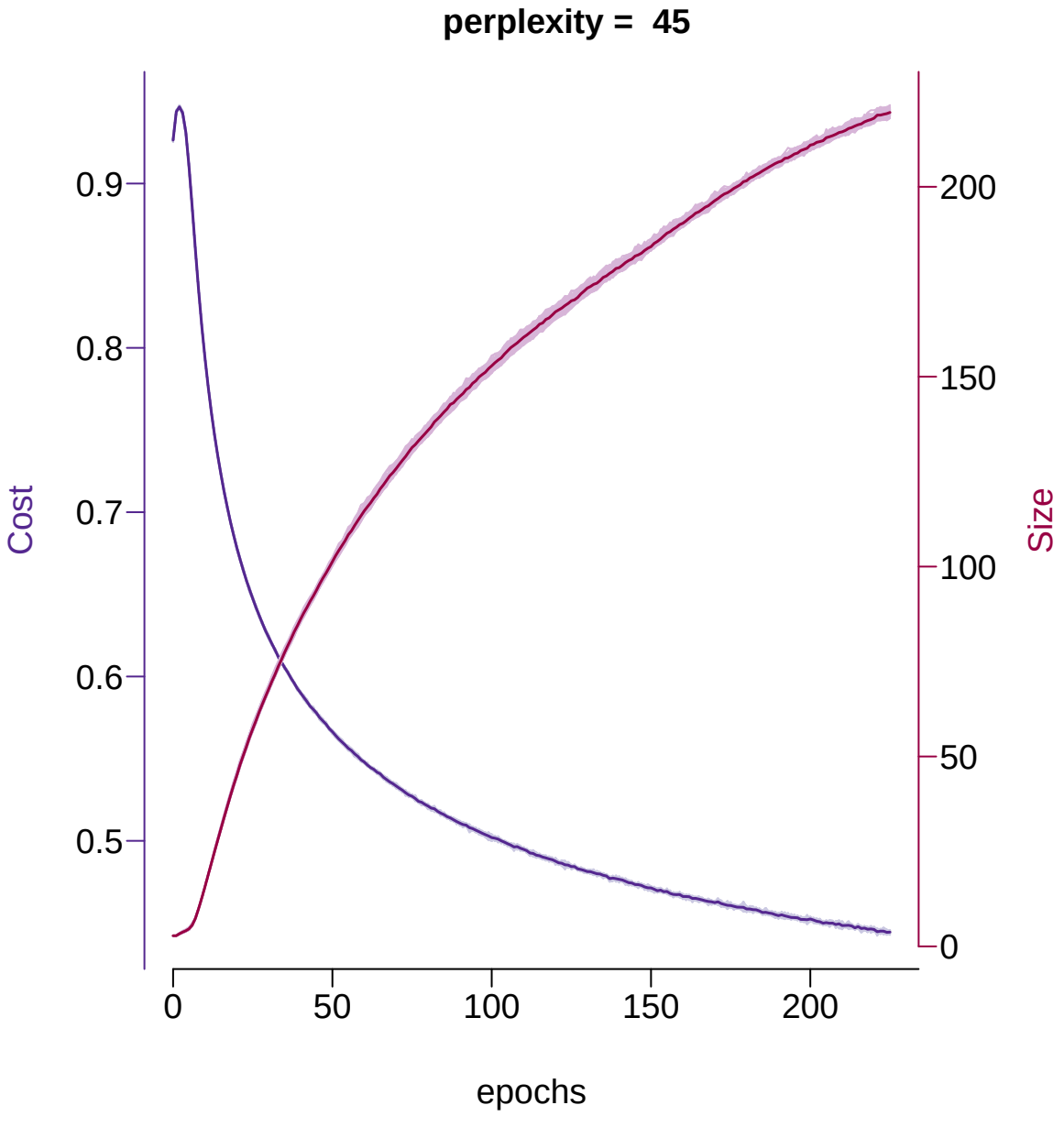

perplexity = 25
Y2
100
50
0
−50
−100
Y1
−100
−50
0
50

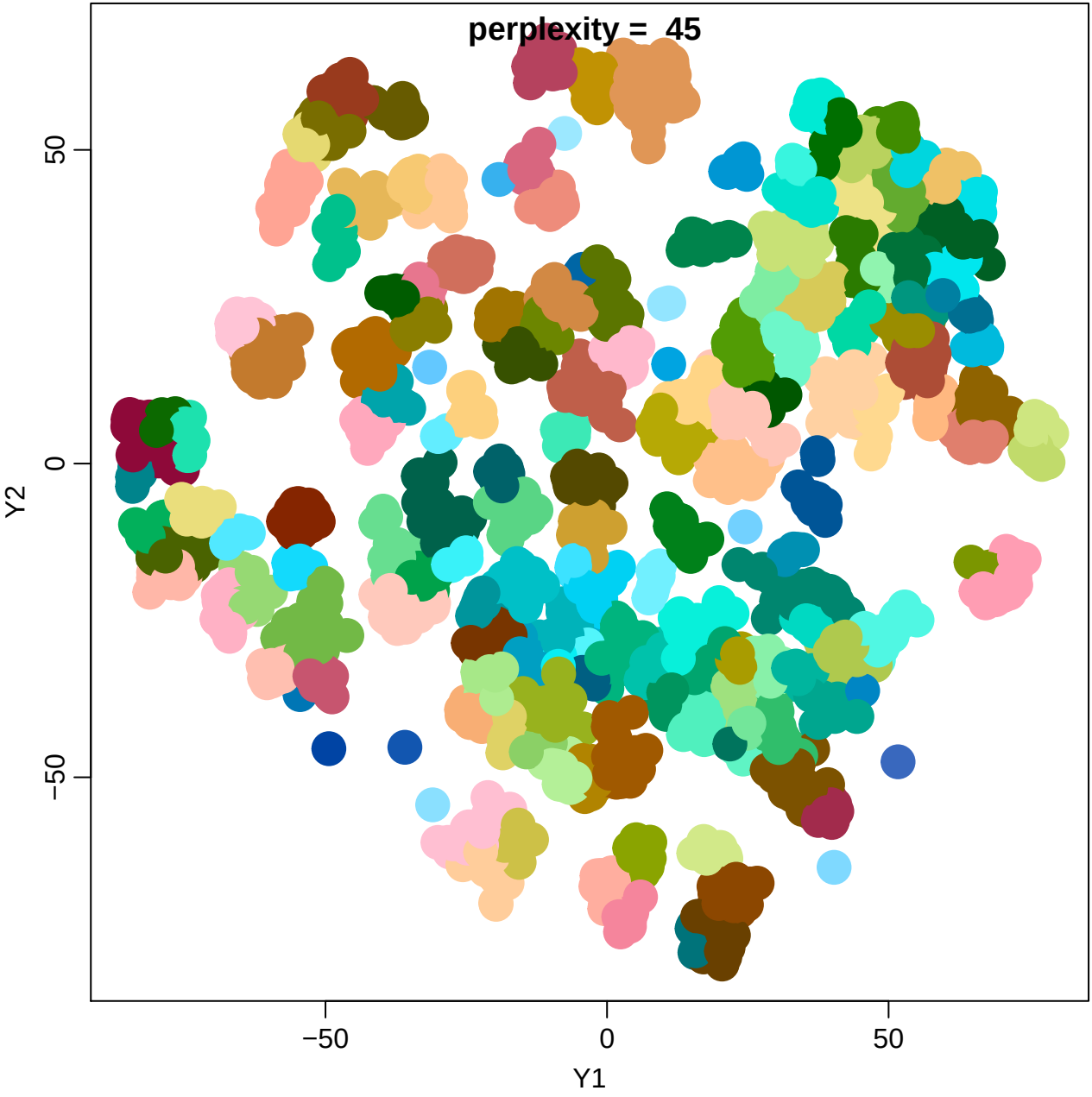

perplexity = 25

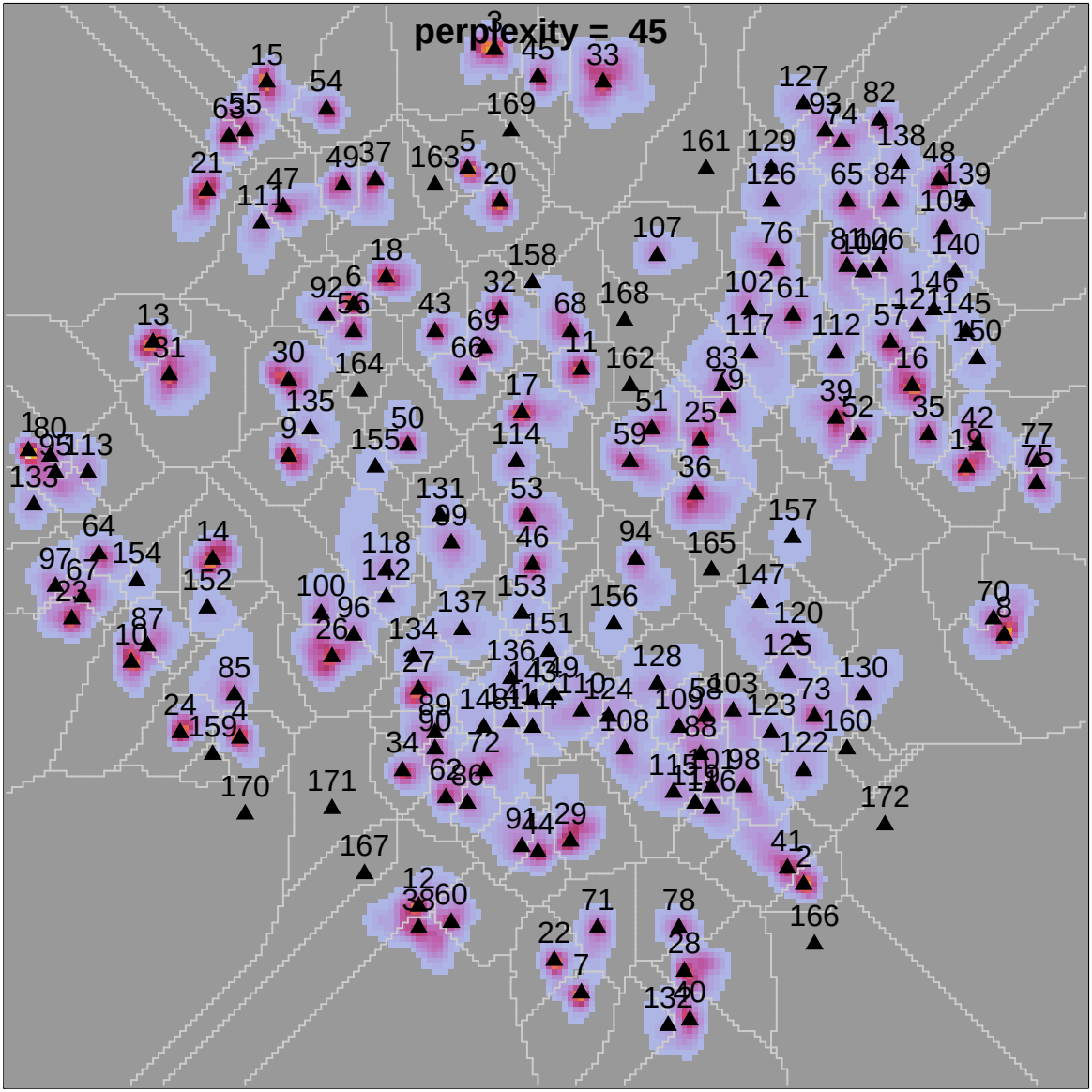

perplexity = 35
Cost
0.9
0.8
0.7
0.6
0.5
Size
250
200
150
100
50
0
0
50
100
150
200
epochs

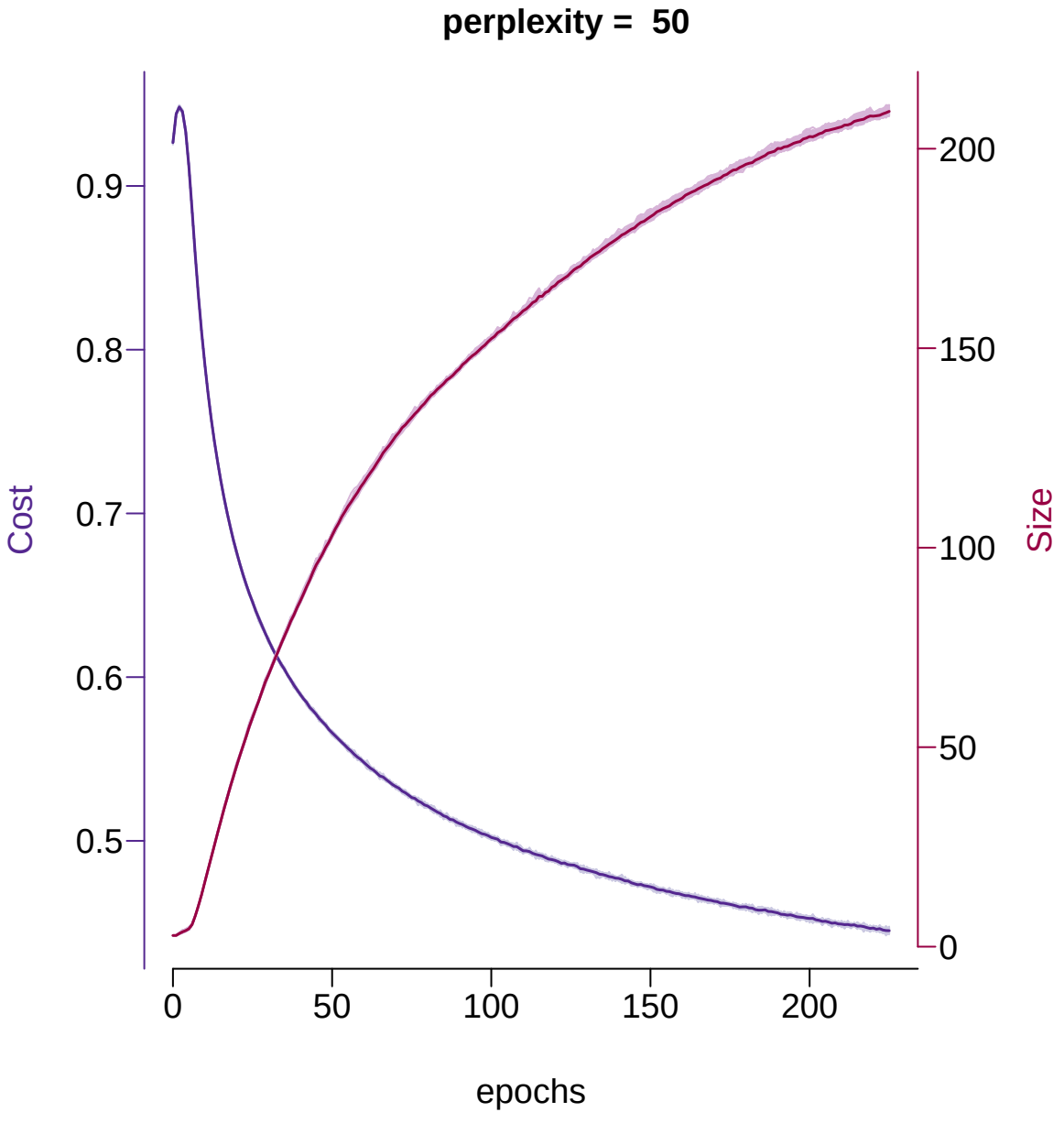

perplexity = 35
50
0
−50
Y2
−50
0
50
Y1

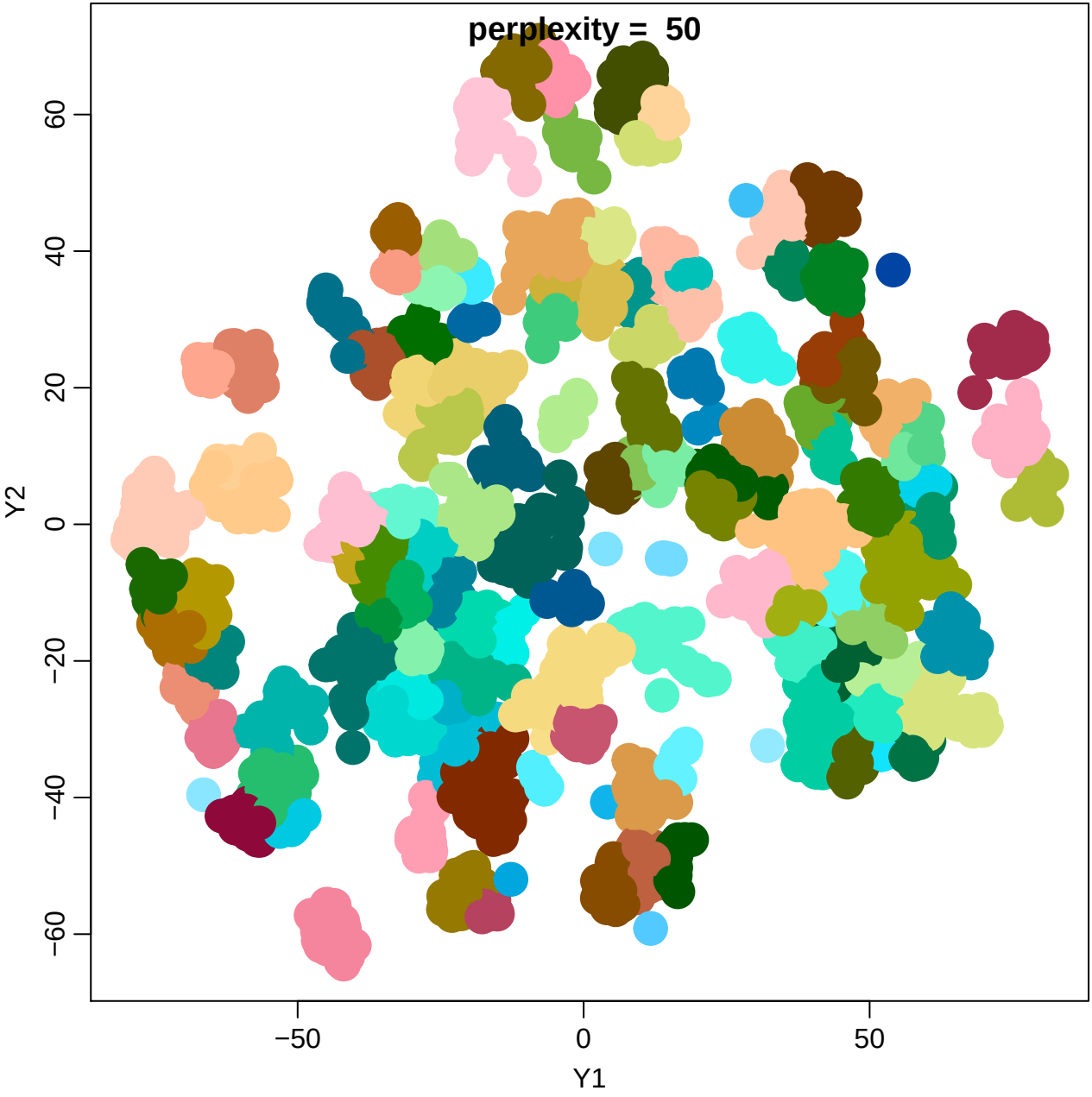

perplexity 35

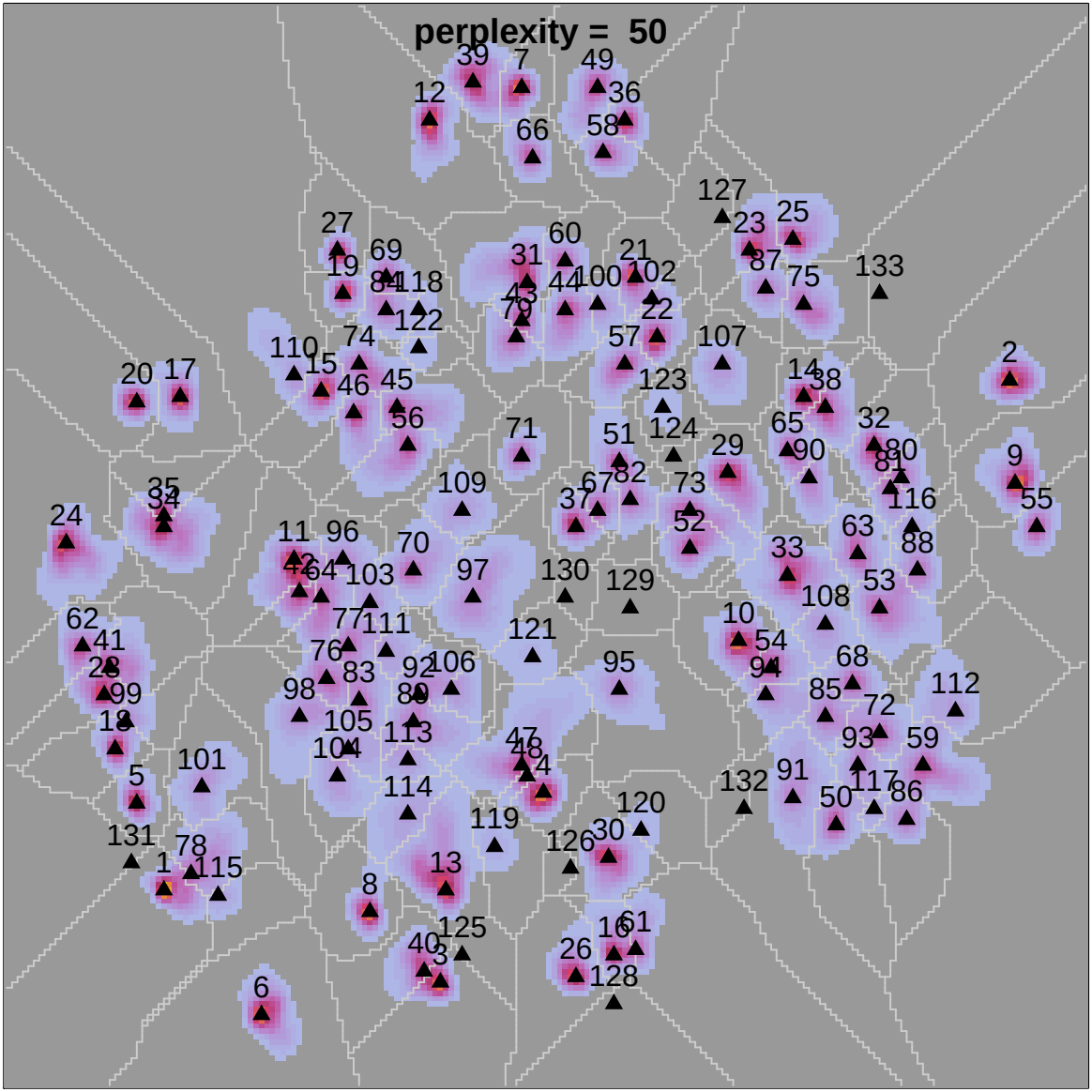

perplexity = 45
Cost
0.9
0.8
0.7
0.6
0.5
Size
200
150
100
50
0
0
50
100
150
200
epochs

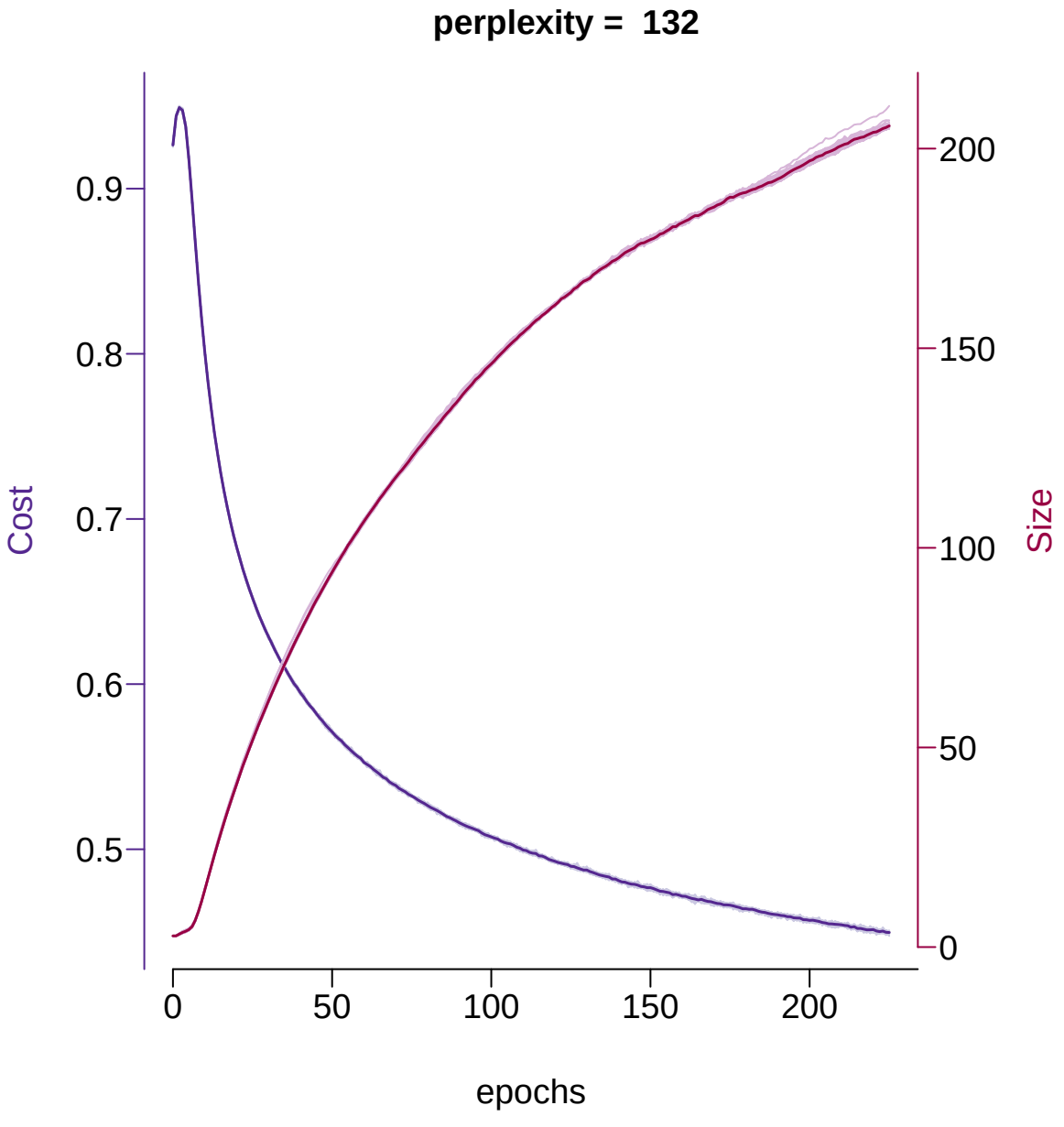

perplexity = 45
50
0
−50
Y2
−50
0
50
Y1

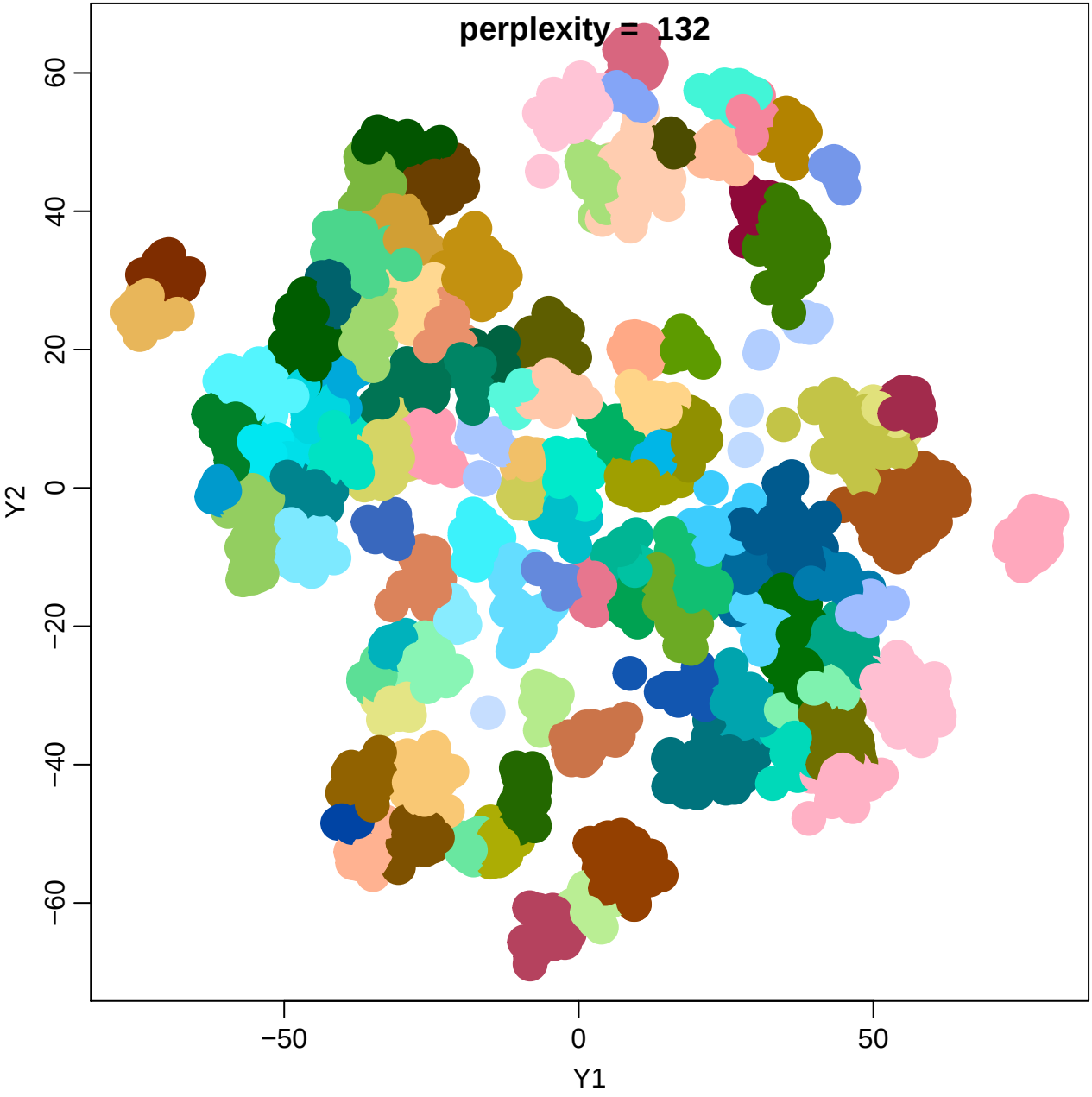

perplexity = 45

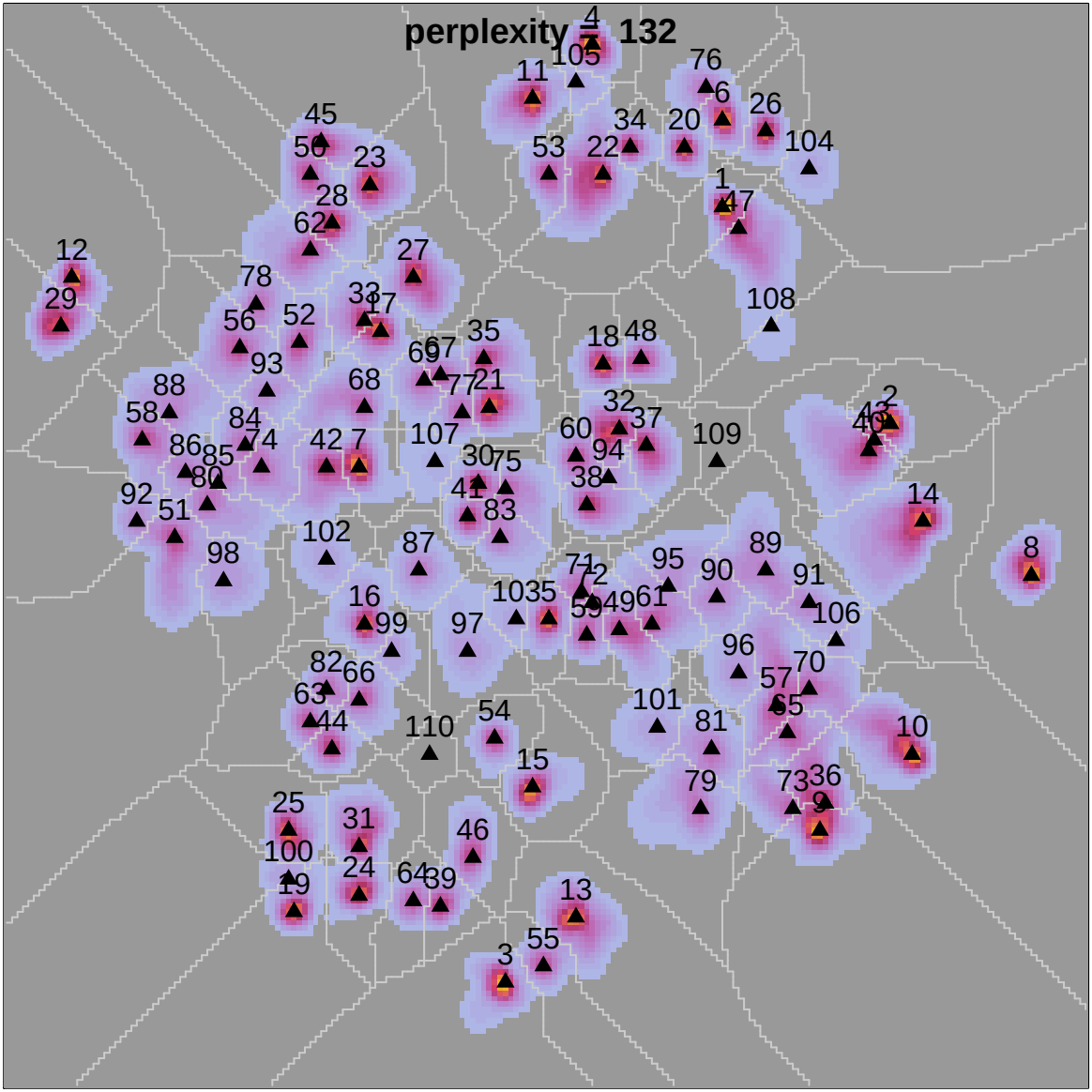

perplexity = 50
Cost
0.9
0.8
0.7
0.6
0.5
Size
200
150
100
50
0
0
50
100
150
200
epochs

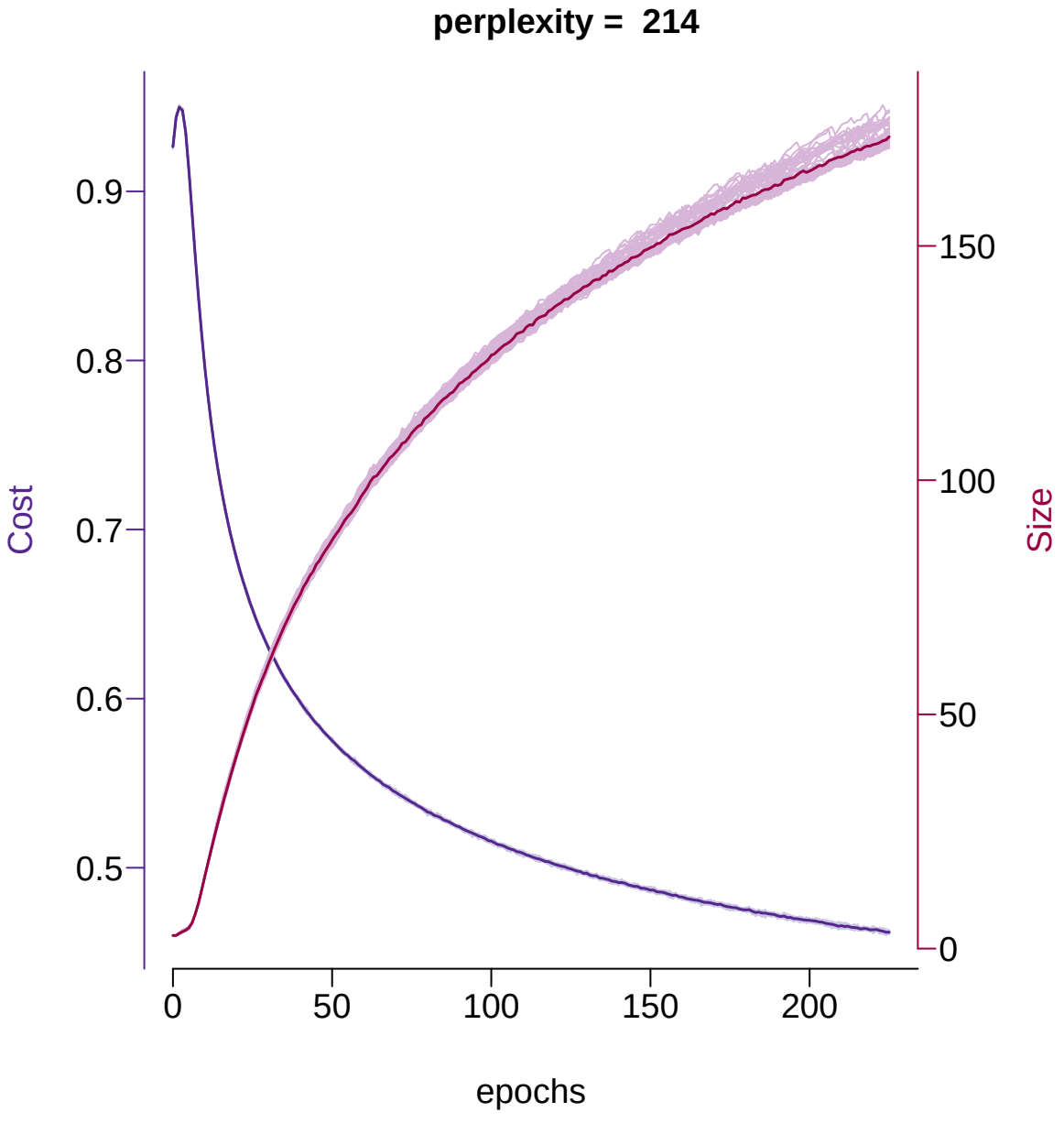

perplexity = 50
Y2
60
40
20
0
−20
−40
−60
−50
0
50
Y1

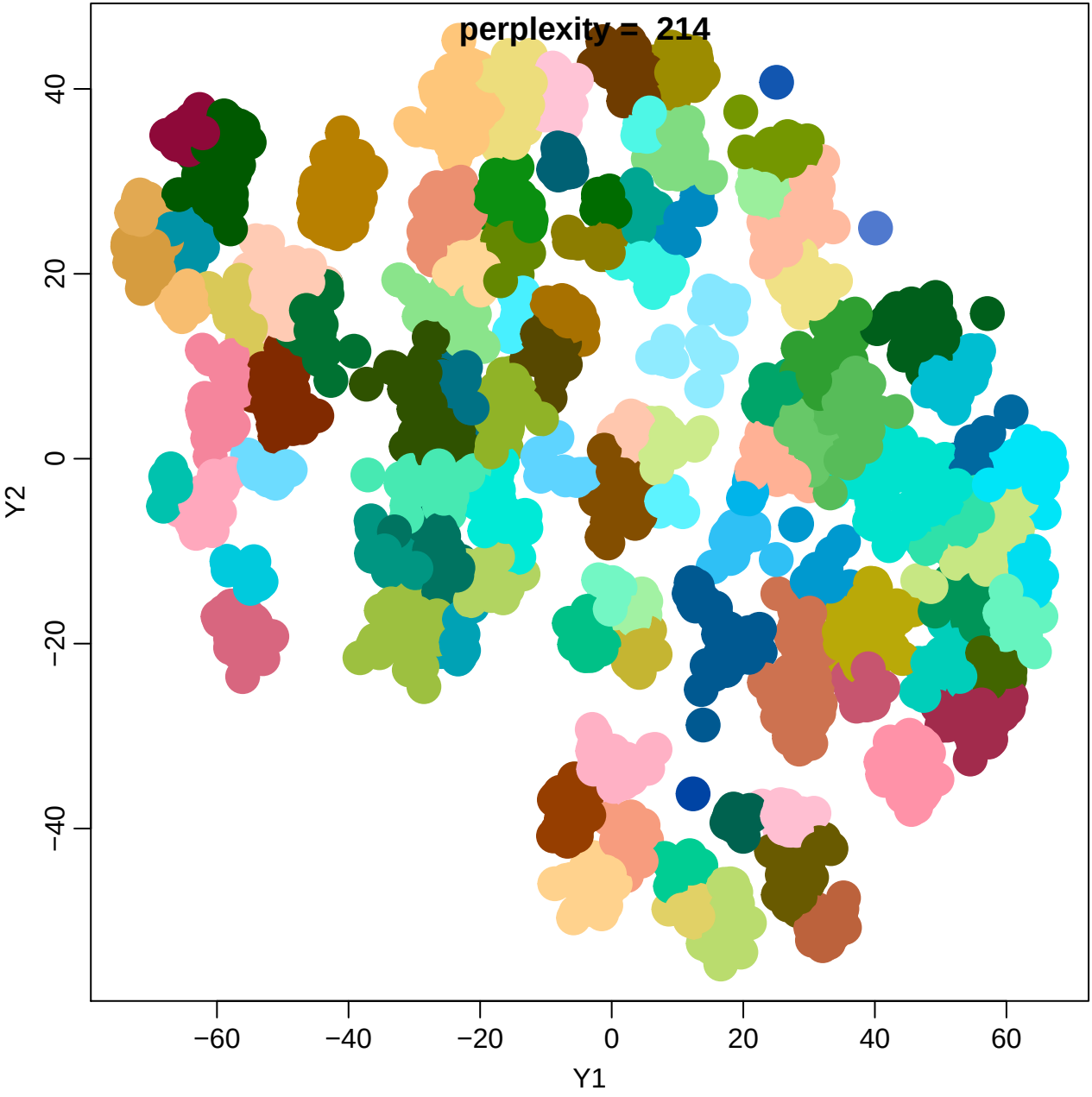

perplexity = 50

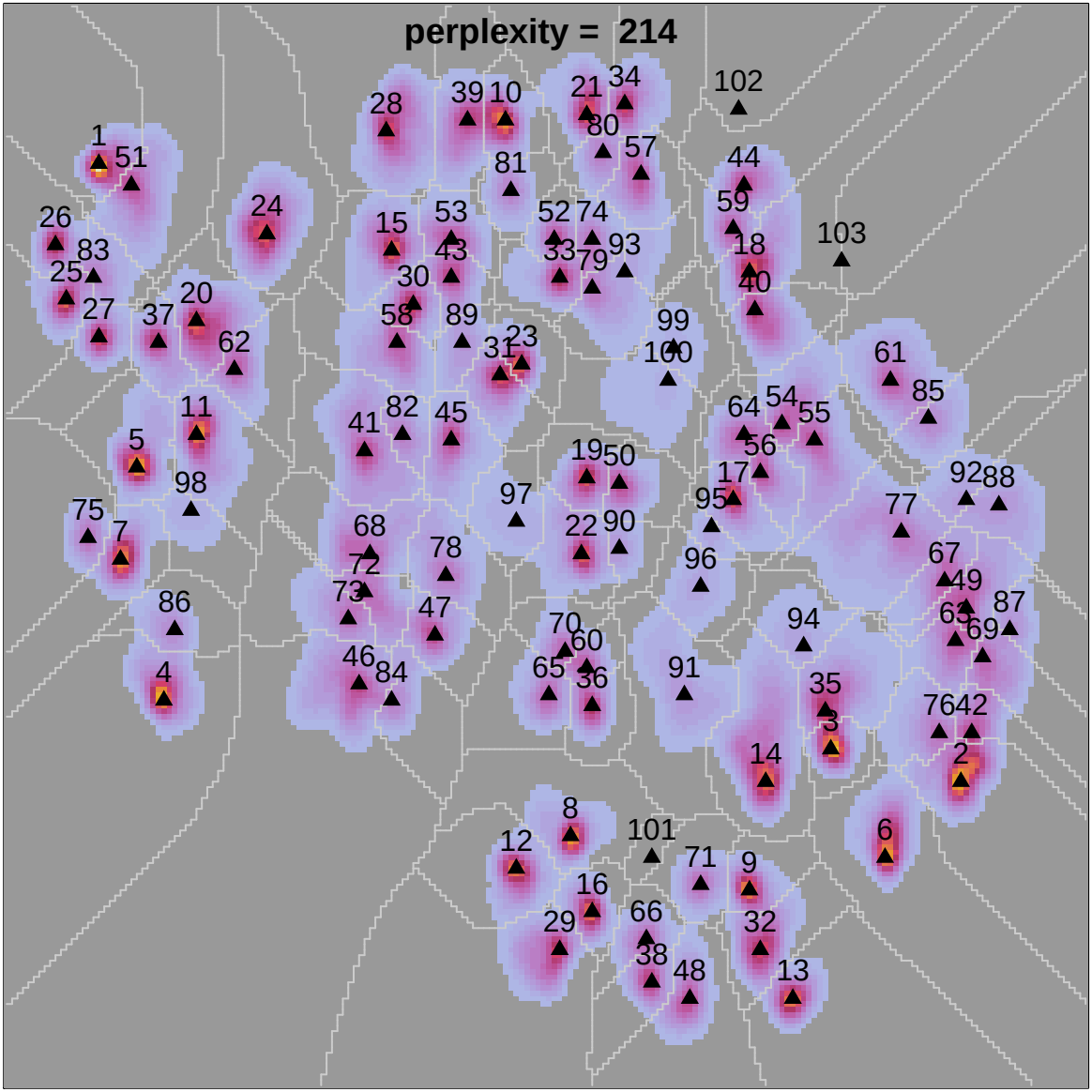

perplexity = 132
Cost
0.9
0.8
0.7
0.6
0.5
Size
200
150
100
50
0
0
50
100
150
200
epochs

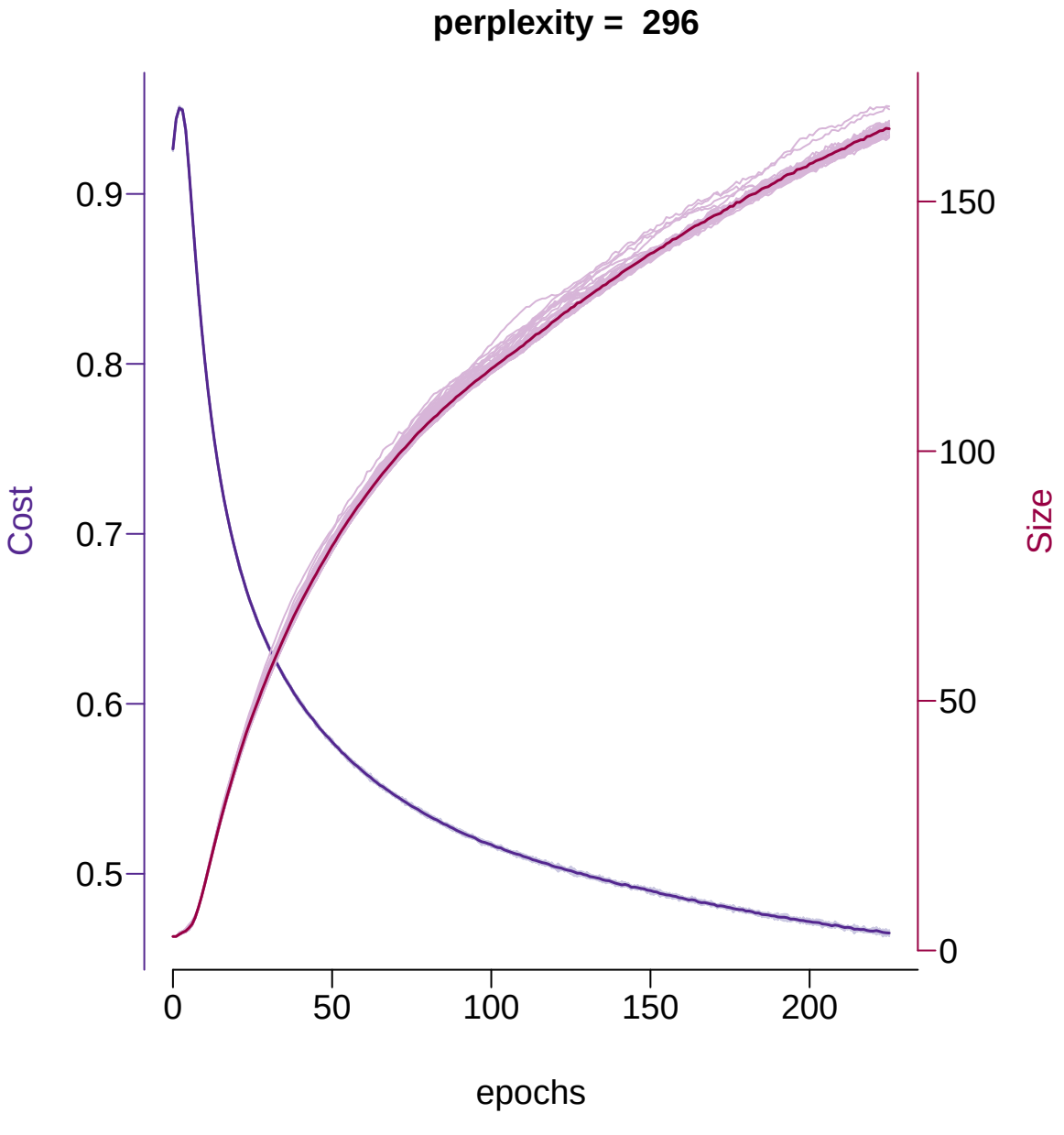

perplexity = 132
60
40
20
0
−20
−40
−60
Y2
−50
0
50
Y1

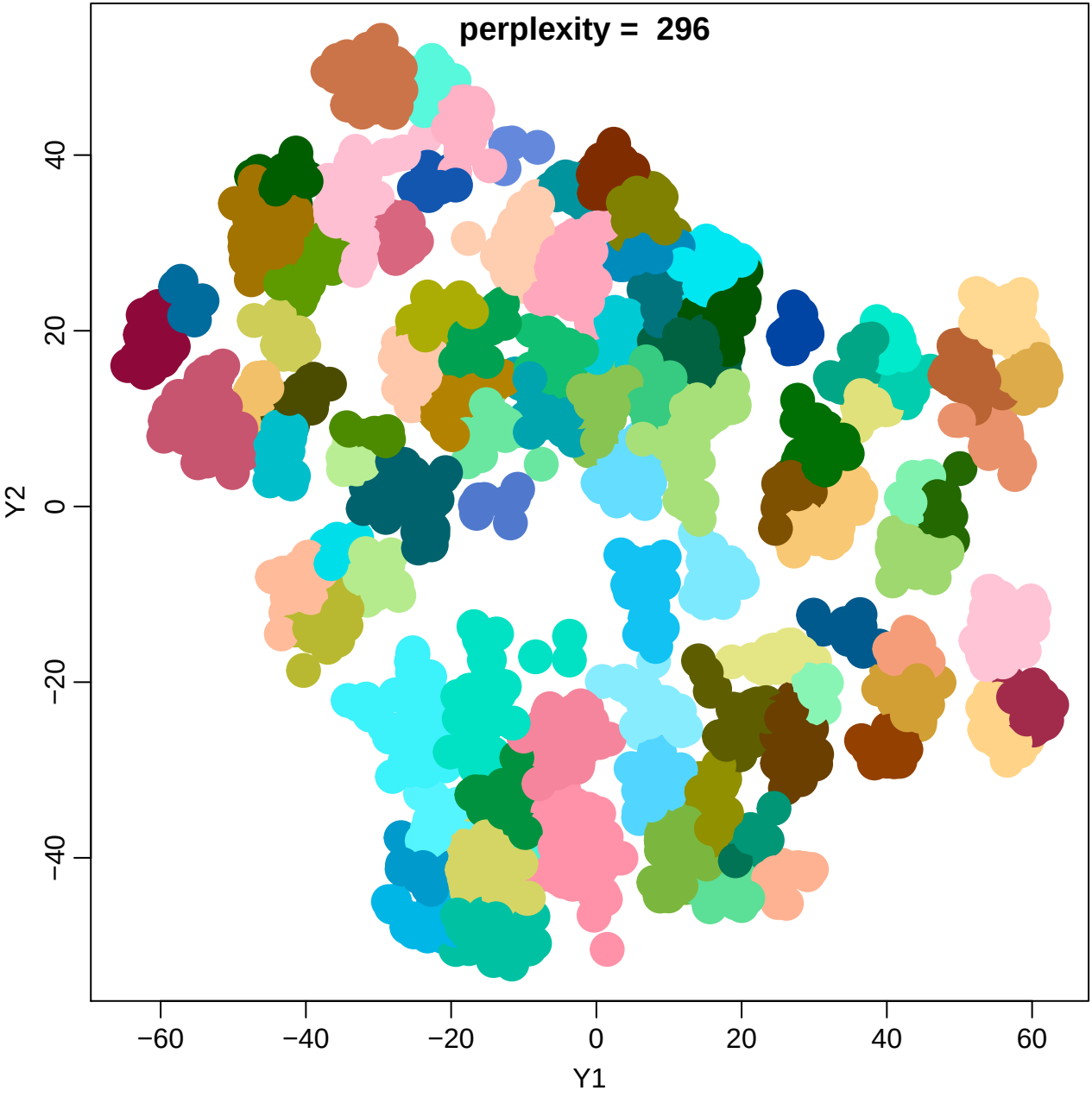

perplexity = 132
4
105
11
76
6
26
45
34
20
104
50
23
53
22
1
28
47
62
12
27
78
108
29
33
17
56
52
35
18
48
69
67
93
88
68
77
21
32
2
58
84
37
43
49
86
74
42
7
107
60
109
85
30
75
94
80
38
92
41
14
51
83
102
87
89
8
98
95
90
91
71
72
16
103
35
59
49
61
106
99
97
96
82
66
57
70
63
101
65
44
110
54
81
10
15
79
73
36
9
25
31
46
100
24
64
39
19
13
3
55

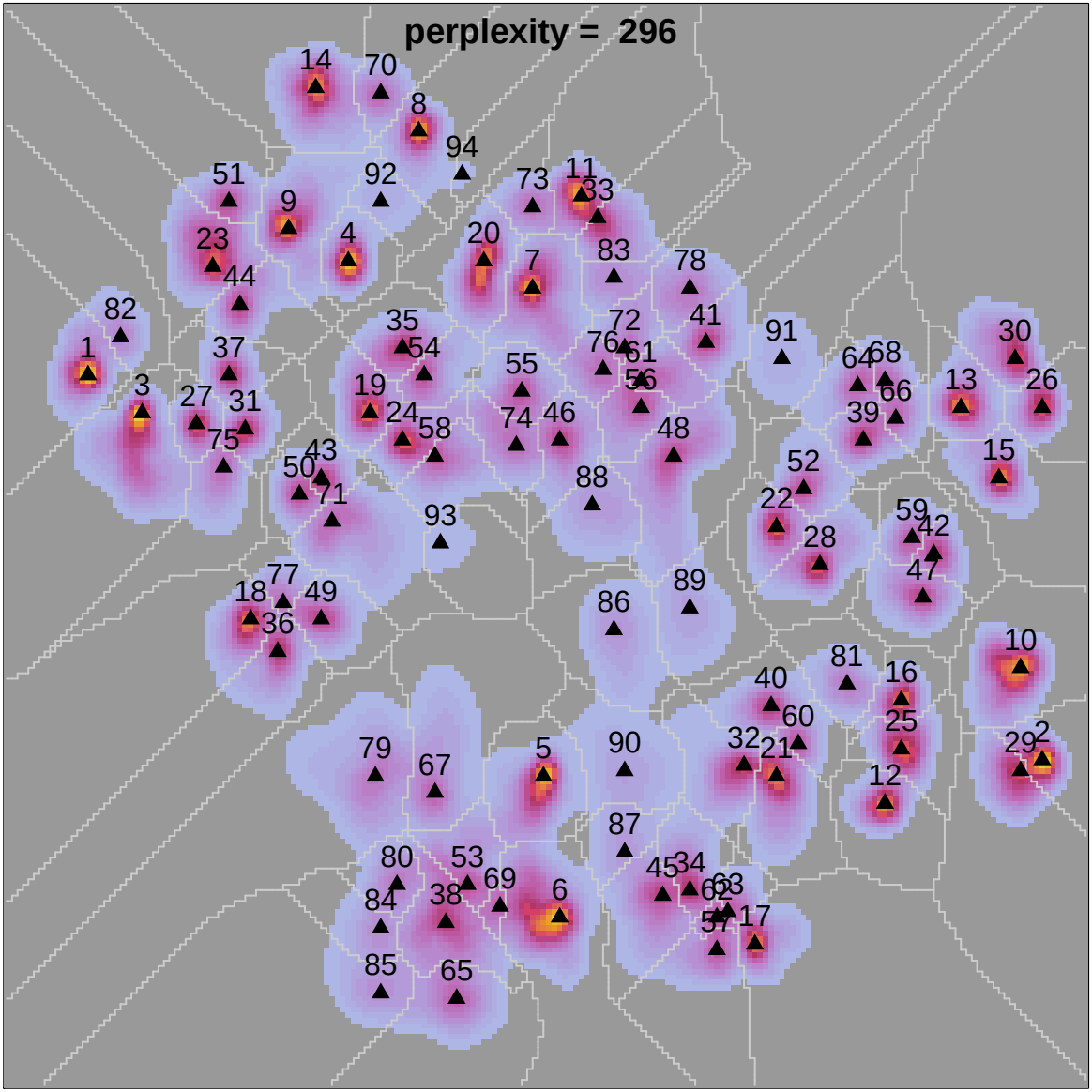

perplexity = 214
Cost
0.9
0.8
0.7
0.6
0.5
Size
150
100
50
0
0
50
100
150
200
epochs

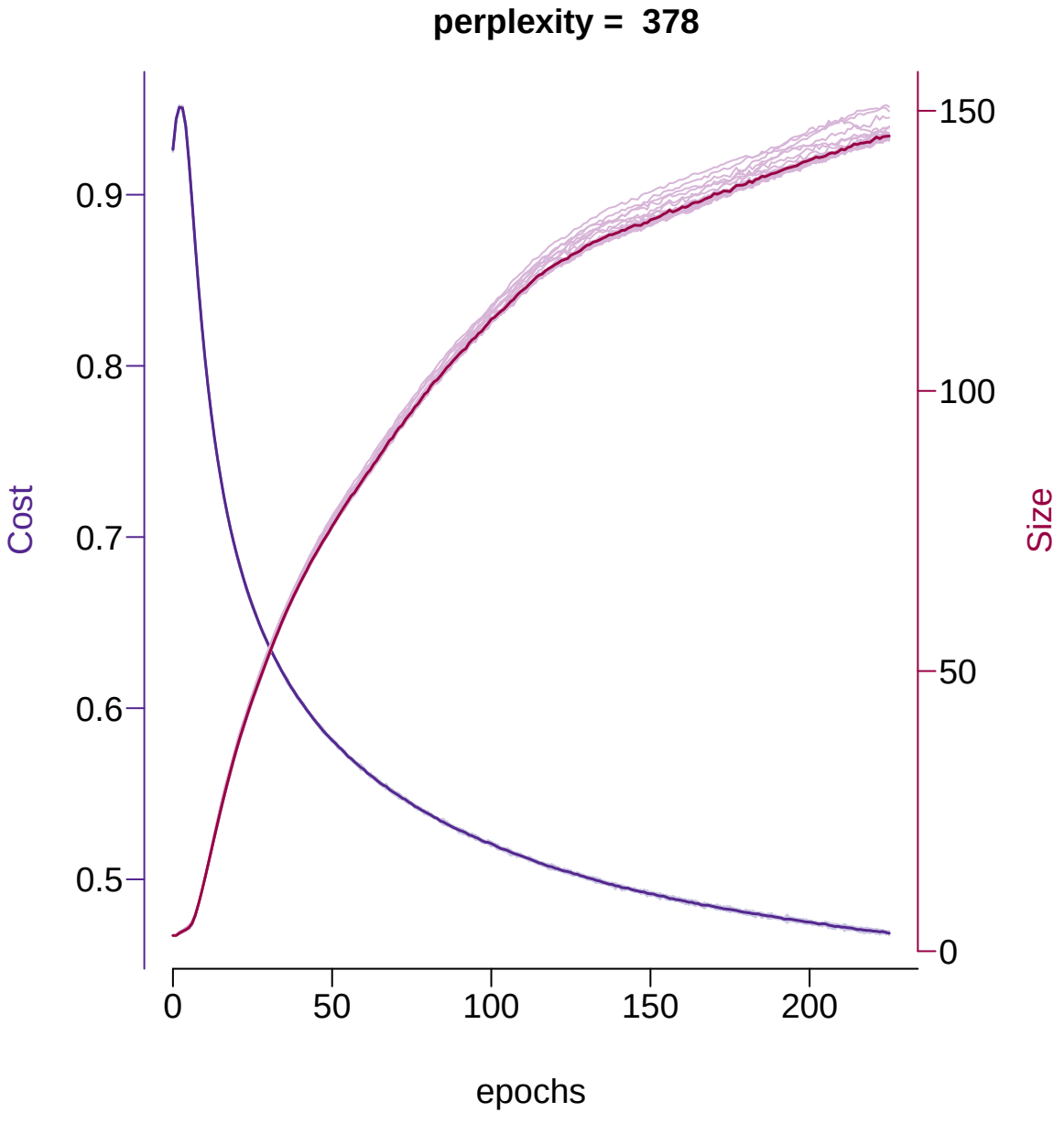

perplexity = 214
Y2
Y1
40
20
0
−20
−40
−60
−40
−20
0
20
40
60

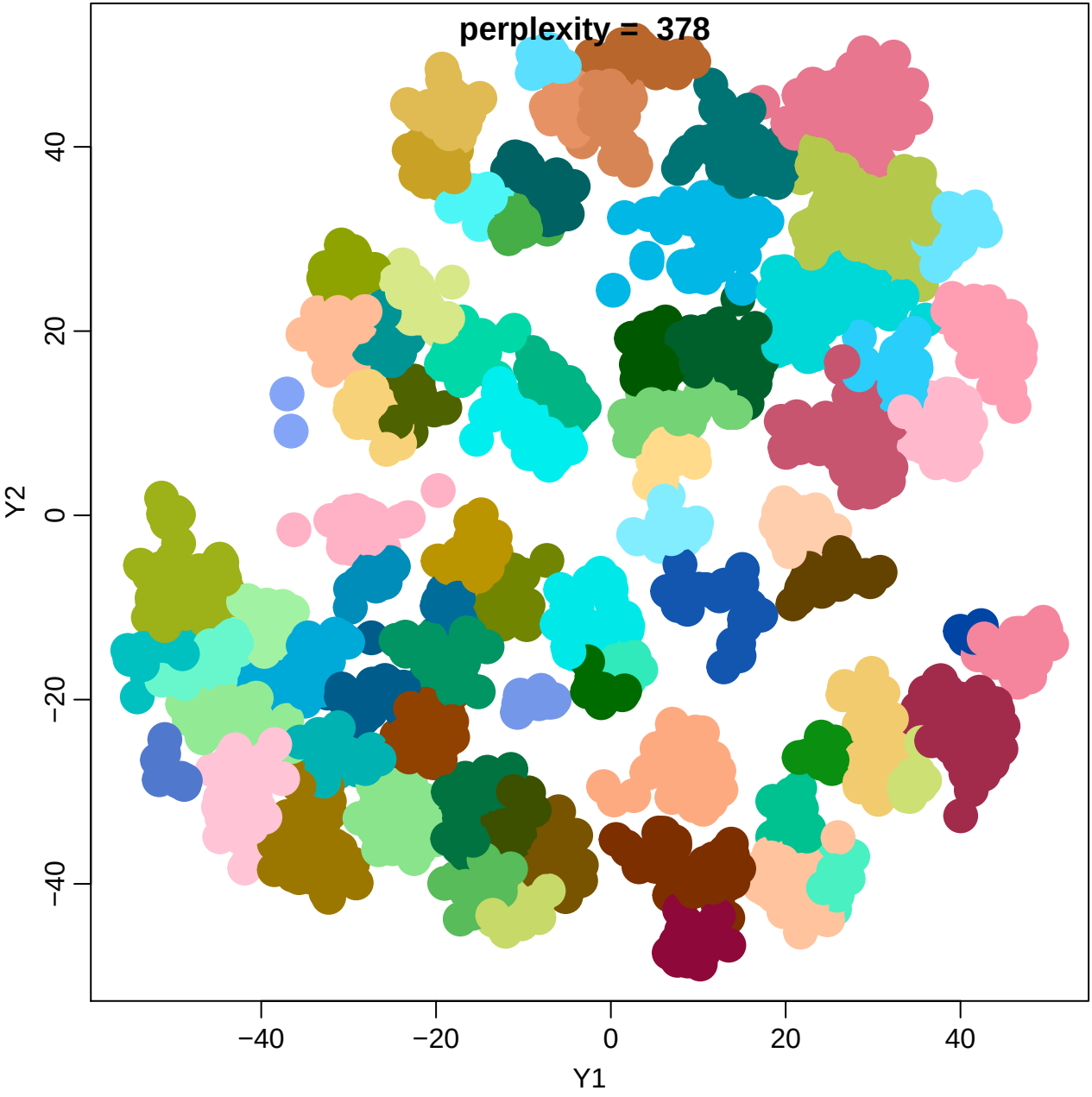

perplexity = 214
1
51
26
83
25
27
37
20
62
24
28
39
10
21
34
80
57
102
81
44
59
15
53
43
52
74
33
79
93
18
103
40
30
58
89
99
100
23
31
61
85
11
41
82
45
64
54
55
56
5
19
50
17
92
88
98
97
95
77
75
7
68
22
90
78
72
96
67
49
73
86
47
63
87
69
70
60
94
46
84
65
36
91
35
4
3
76
42
14
2
8
101
12
71
9
6
16
66
29
32
38
48
13

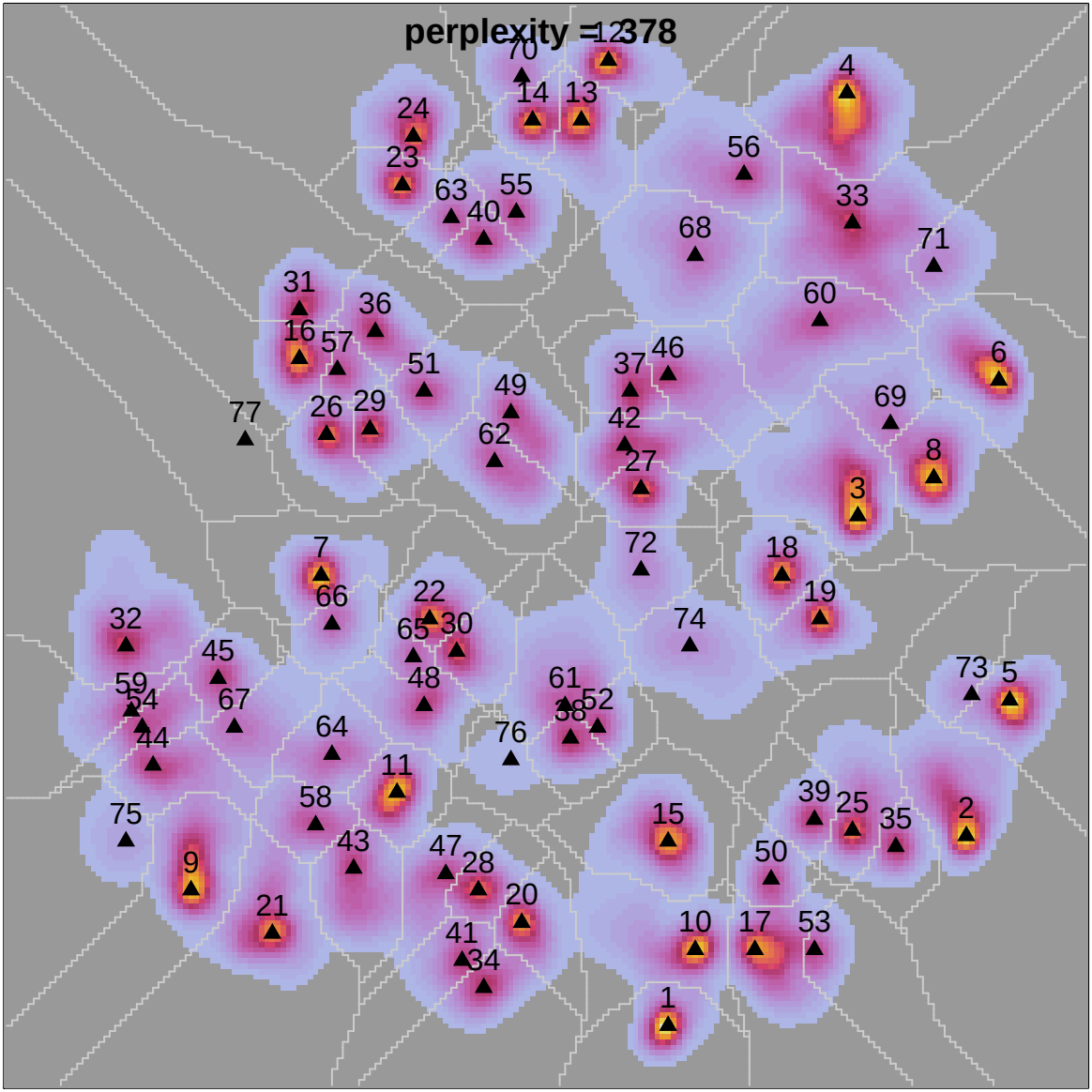

perplexity = 296
Cost
0.9
0.8
0.7
0.6
0.5
Size
150
100
50
0
0
50
100
150
200
epochs

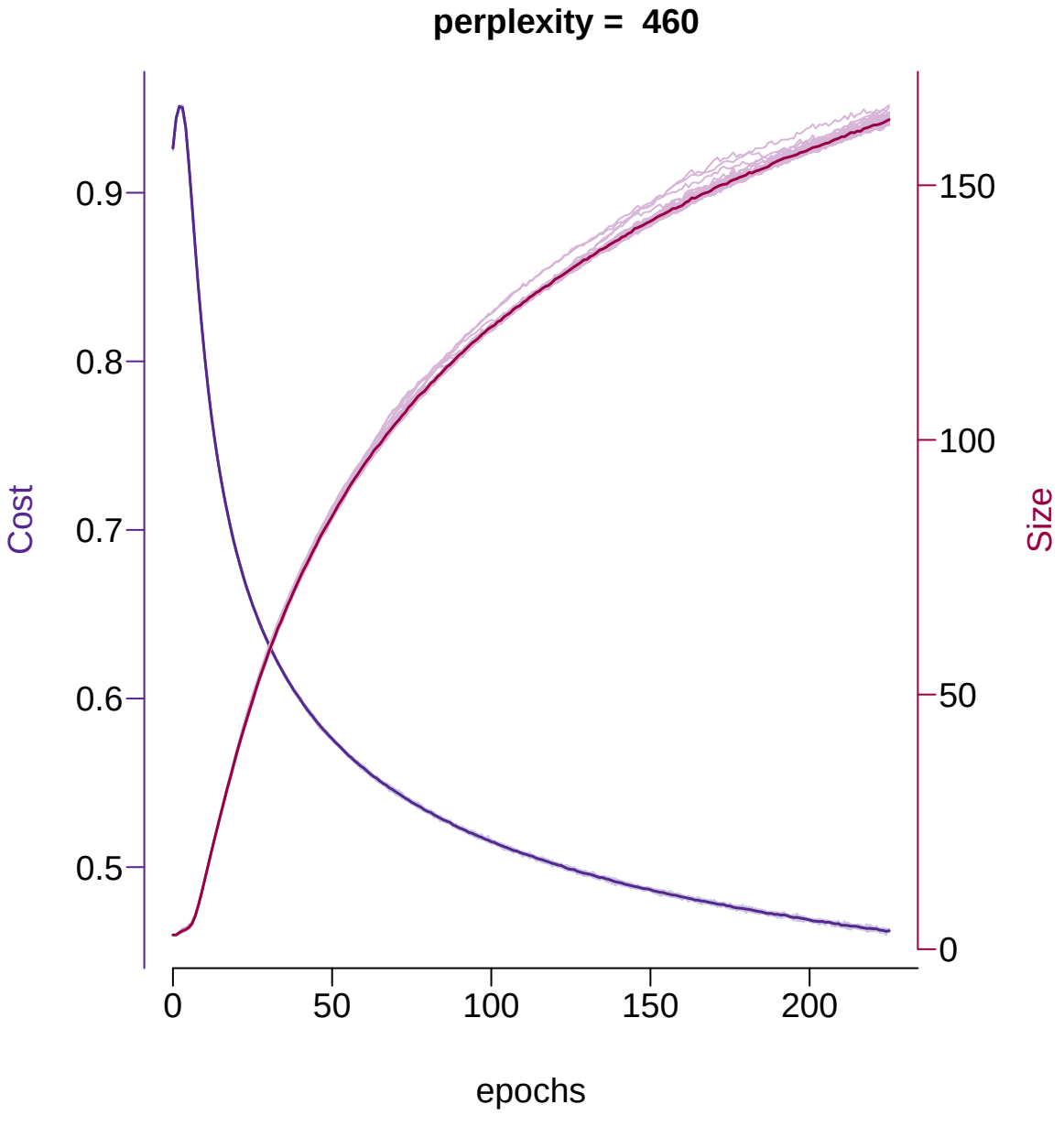

perplexity = 296
Y2
40
20
0
−20
−40
−60
−40
−20
0
20
40
60
Y1

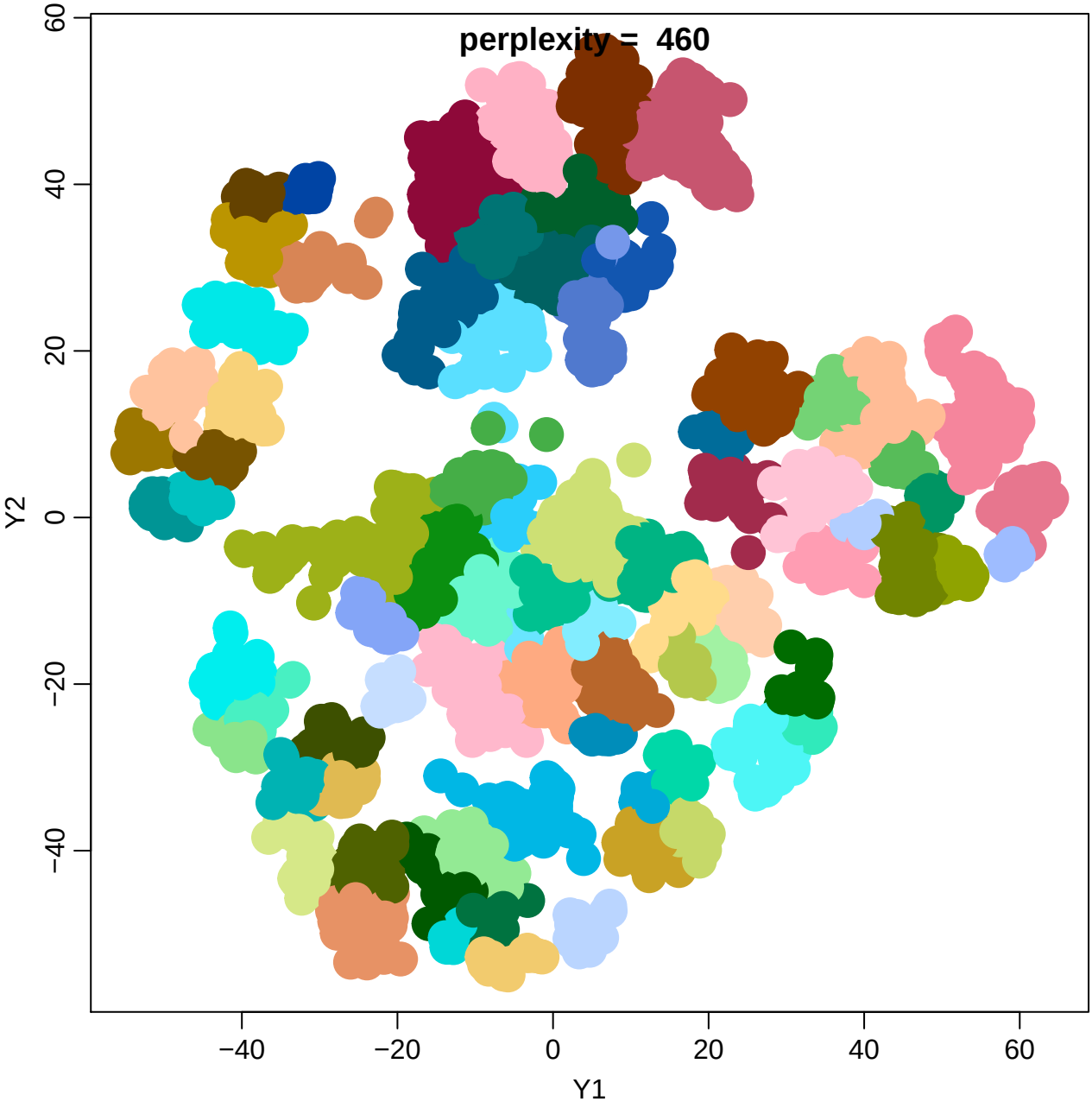

perplexity = 296
14
70
8
94
51
92
73
11
33
9
23
4
20
7
83
78
44
82
35
72
41
91
30
1
37
54
76
61
64
68
3
27
31
19
55
56
66
13
26
24
74
46
39
75
58
48
43
50
52
15
71
88
22
59
93
42
28
77
47
18
49
89
86
36
10
81
16
40
60
25
79
67
5
90
32
21
29
2
12
87
80
53
69
45
34
63
84
38
6
62
17
57
85
65

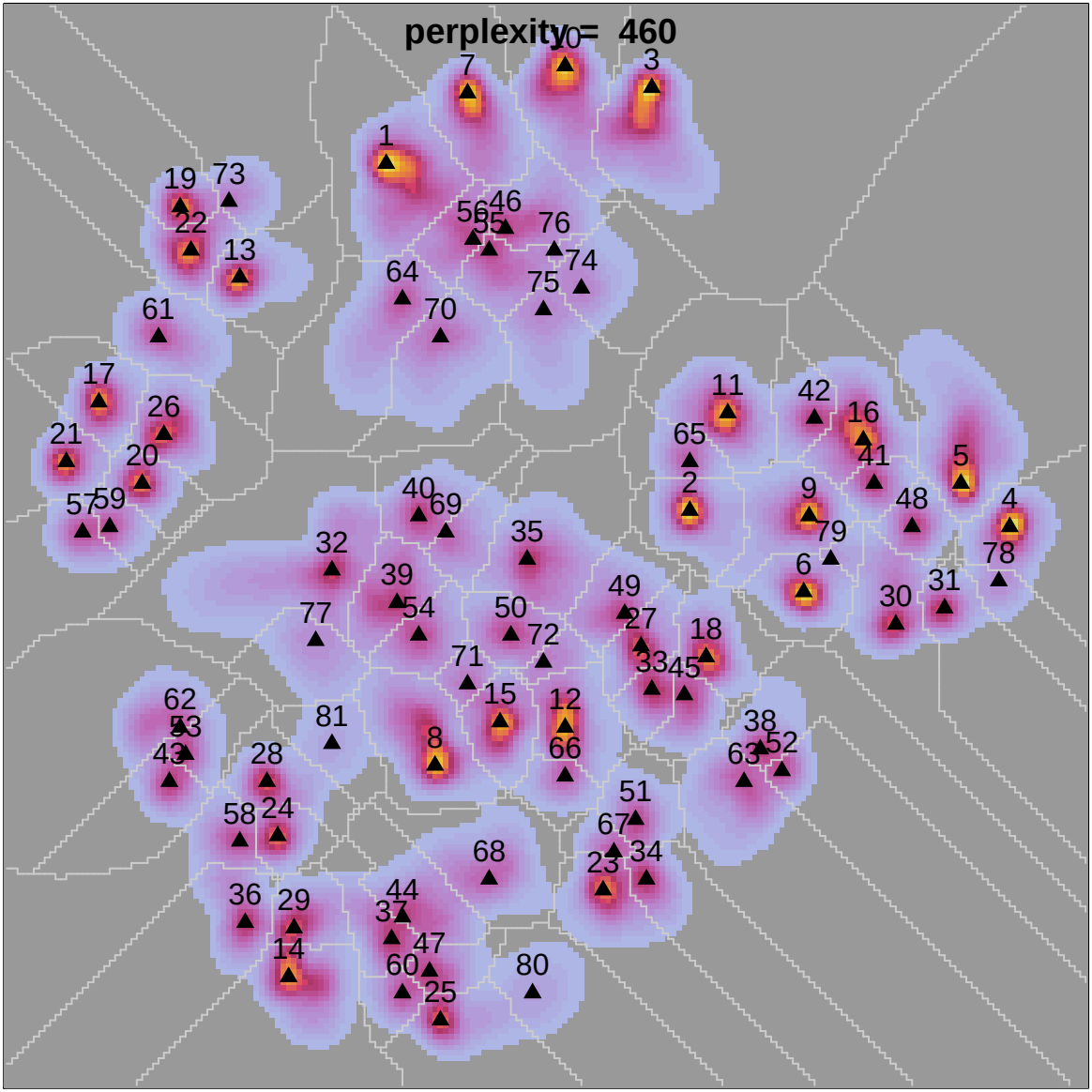

perplexity = 378
Cost
0.9
0.8
0.7
0.6
0.5
Size
150
100
50
0
0
50
100
150
200
epochs

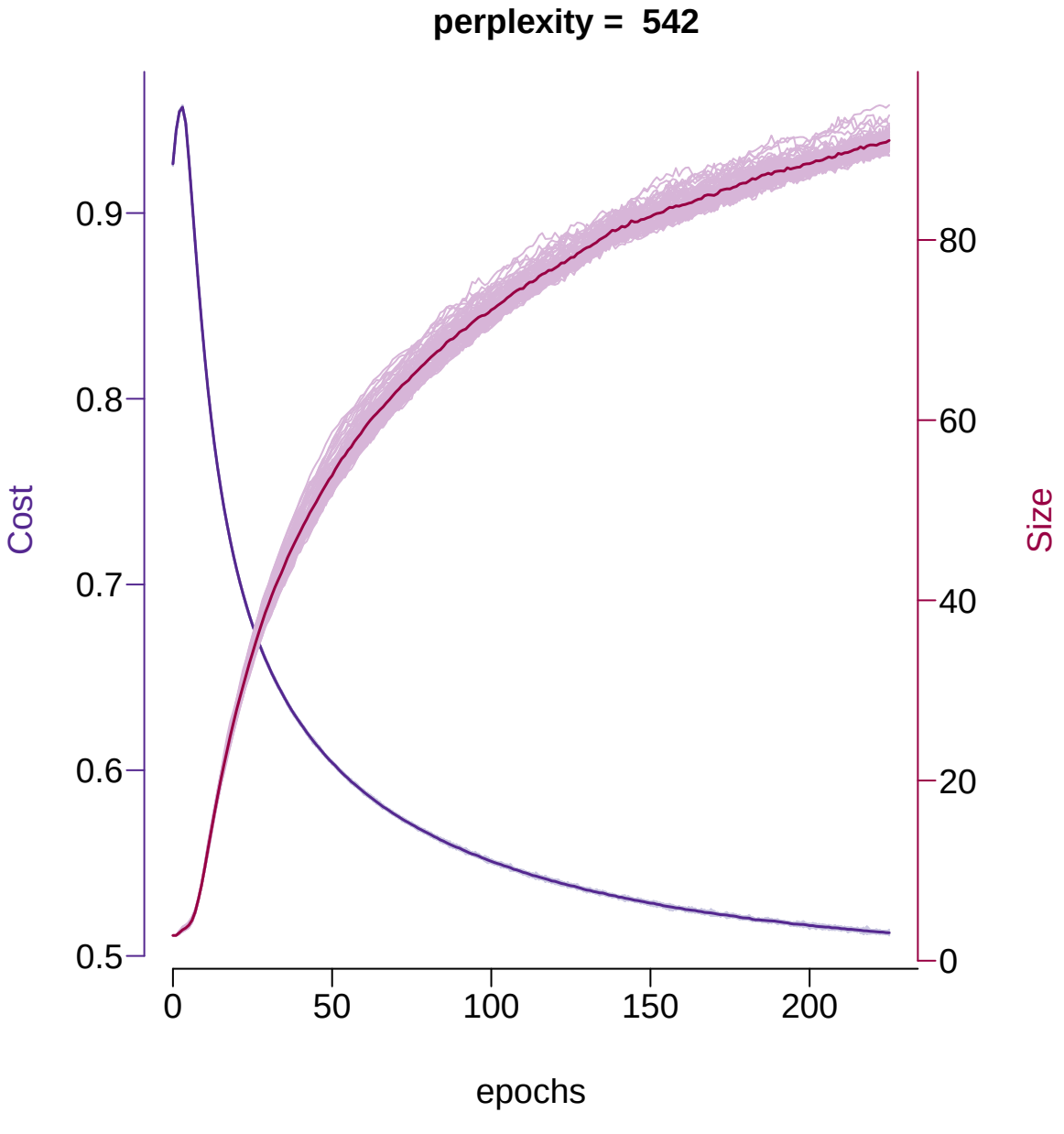

perplexity = 378
Y2
Y1
40
20
0
−20
−40
−40
−20
0
20
40

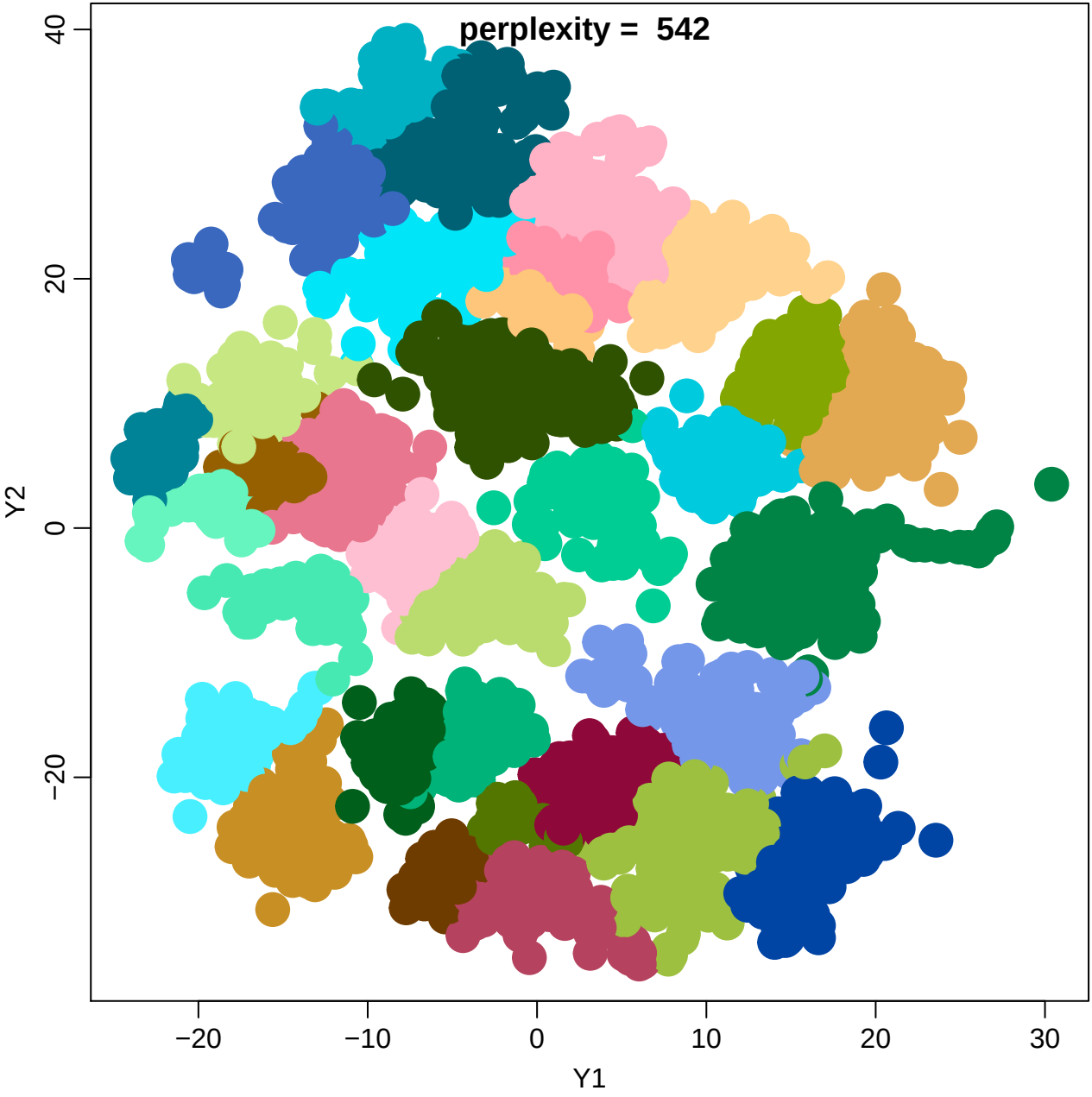

perplexity = 378
12
70
4
14
13
24
56
23
63
55
33
40
68
71
31
36
60
16
57
51
37
46
6
49
69
77
26
29
42
62
8
27
3
7
72
18
66
22
19
32
65
30
74
45
73
5
59
48
61
54
67
38
52
64
76
44
11
39
25
35
2
75
58
15
43
47
50
9
28
20
21
10
17
53
41
34
1

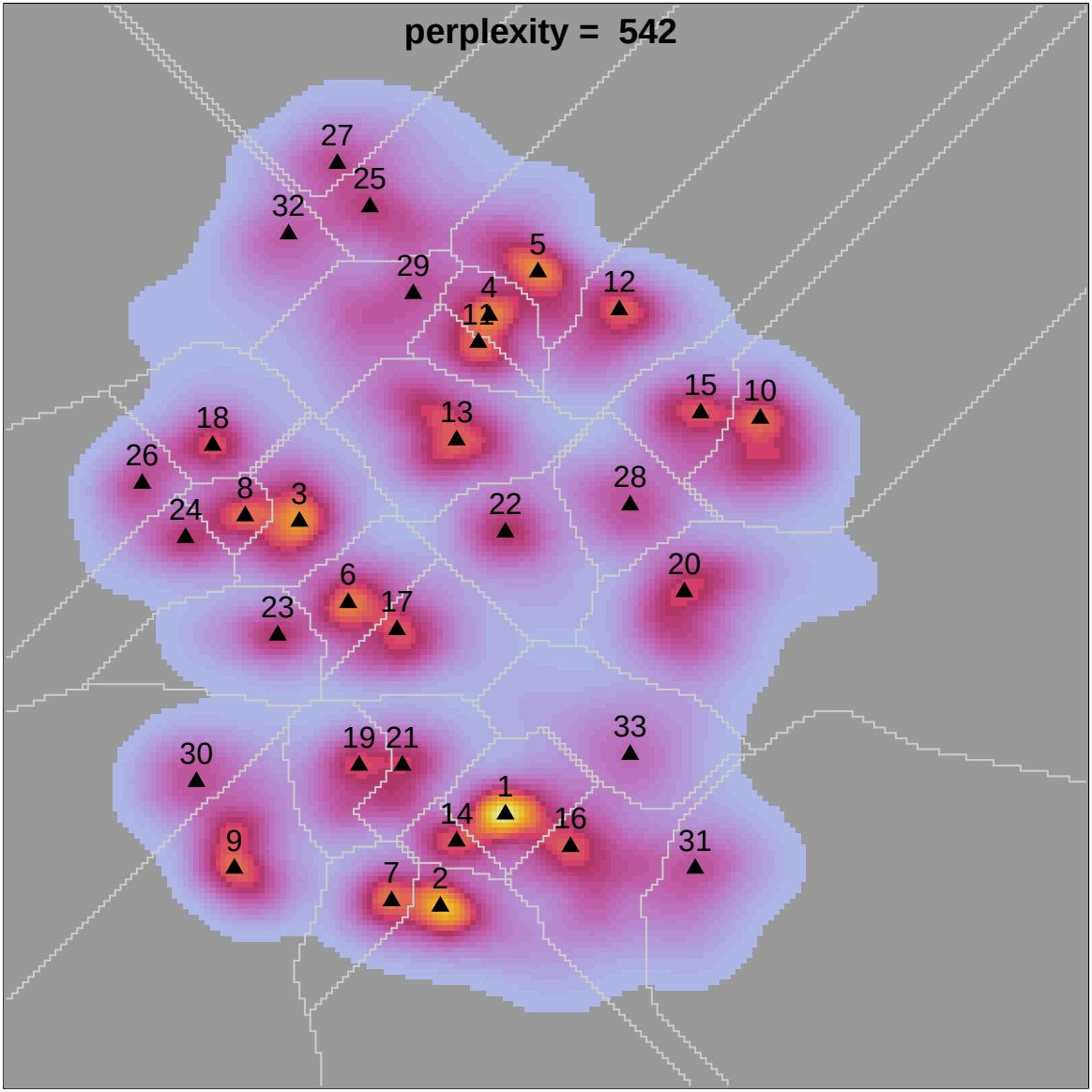

perplexity = 460
Cost
0.9
0.8
0.7
0.6
0.5
Size
150
100
50
0
0
50
100
150
200
epochs

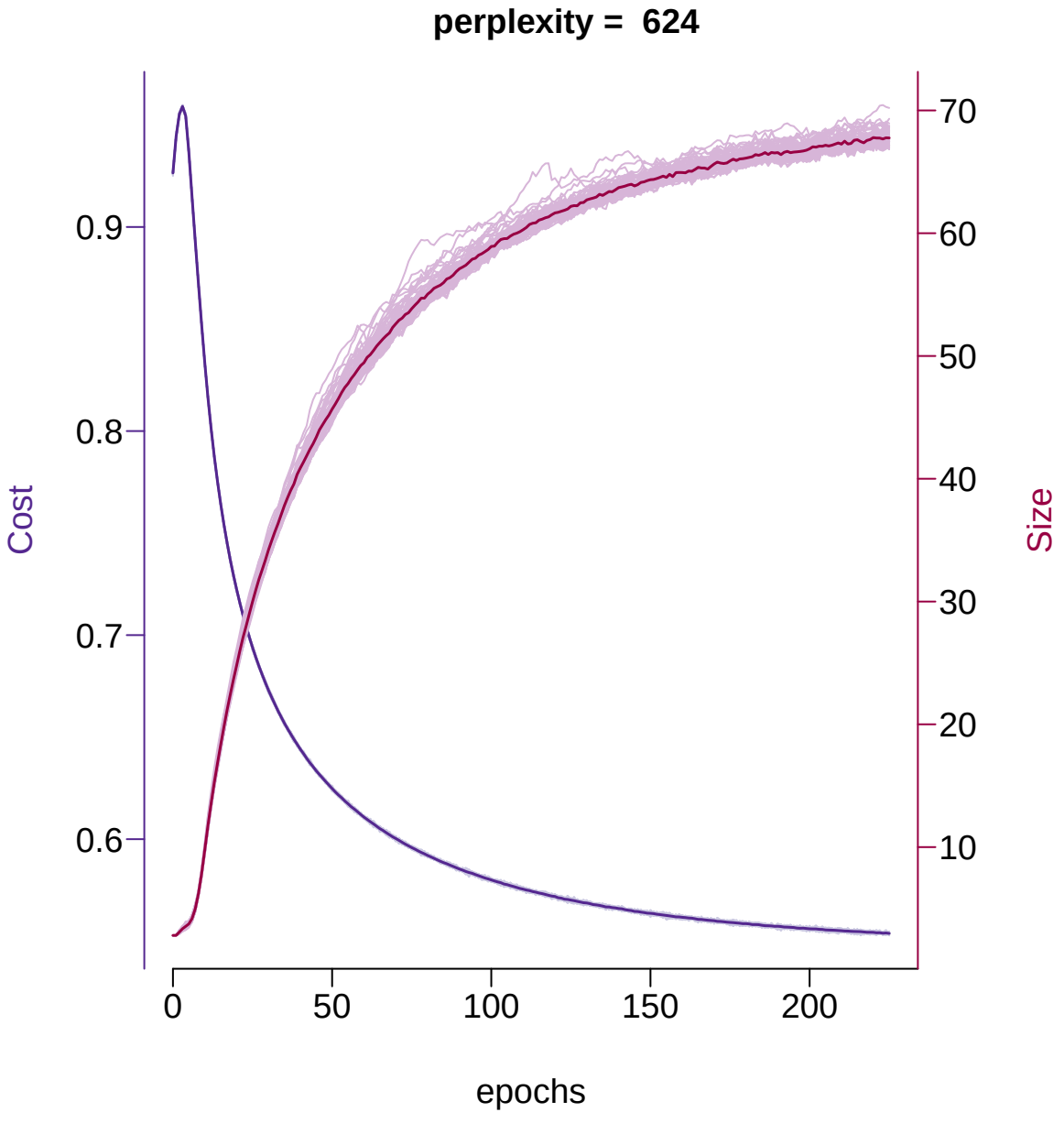

perplexity = 460
Y2
Y1
60
40
20
0
−20
−40
−40
−20
0
20
40
60

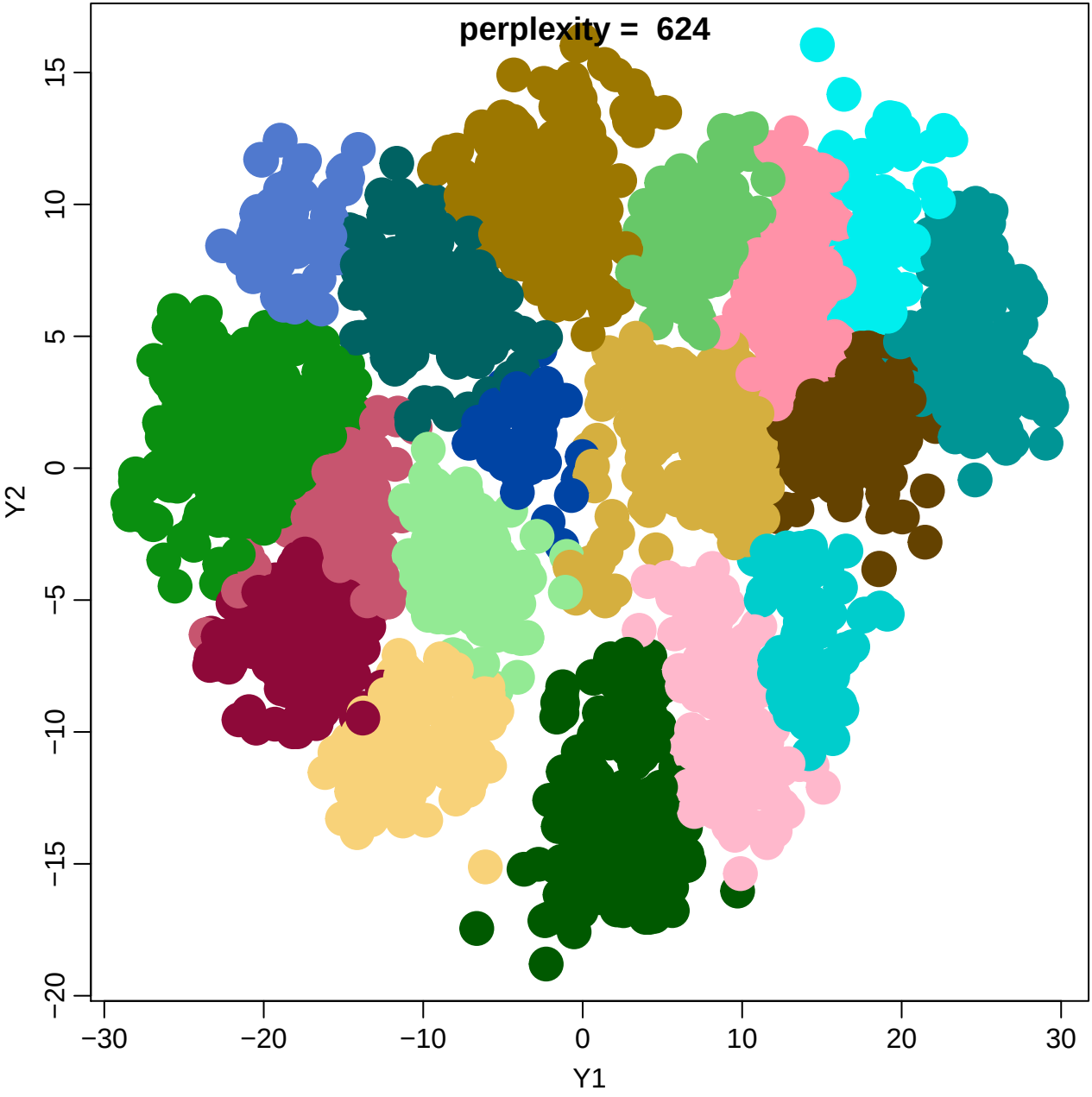

perplexity= 460
10
7
3
1
19
73
22
13
56
35
46
76
74
75
64
70
61
17
26
21
20
57
59
11
42
16
65
41
5
2
9
48
4
40
69
79
32
35
6
78
39
49
31
30
77
54
50
27
72
18
71
33
45
62
15
12
81
38
53
8
66
63
52
43
28
51
58
24
67
68
34
23
36
29
44
37
14
47
60
80
25

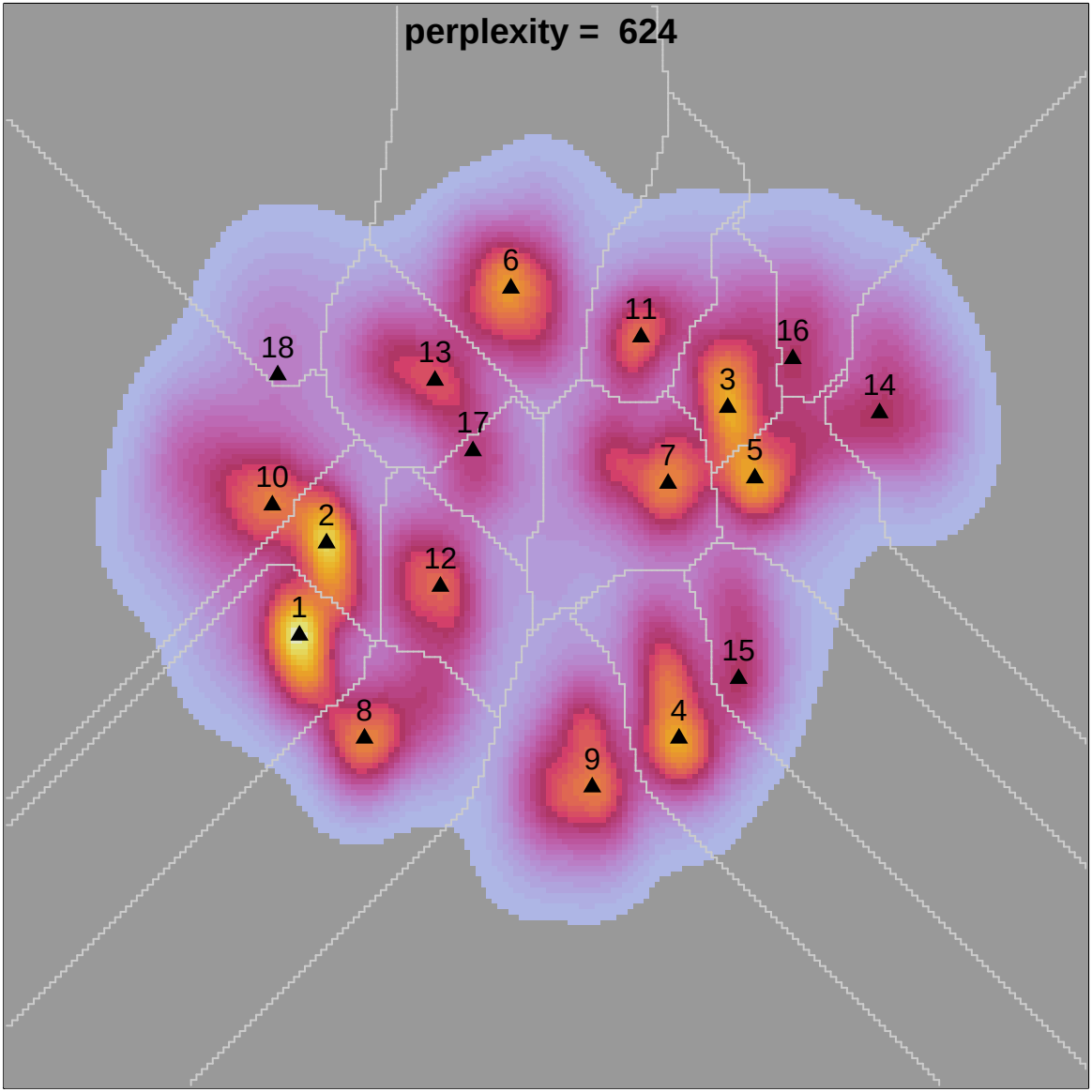

perplexity = 542
Cost
Size
0.9
0.8
0.7
0.6
0.5
80
60
40
20
0
0
50
100
150
200
epochs

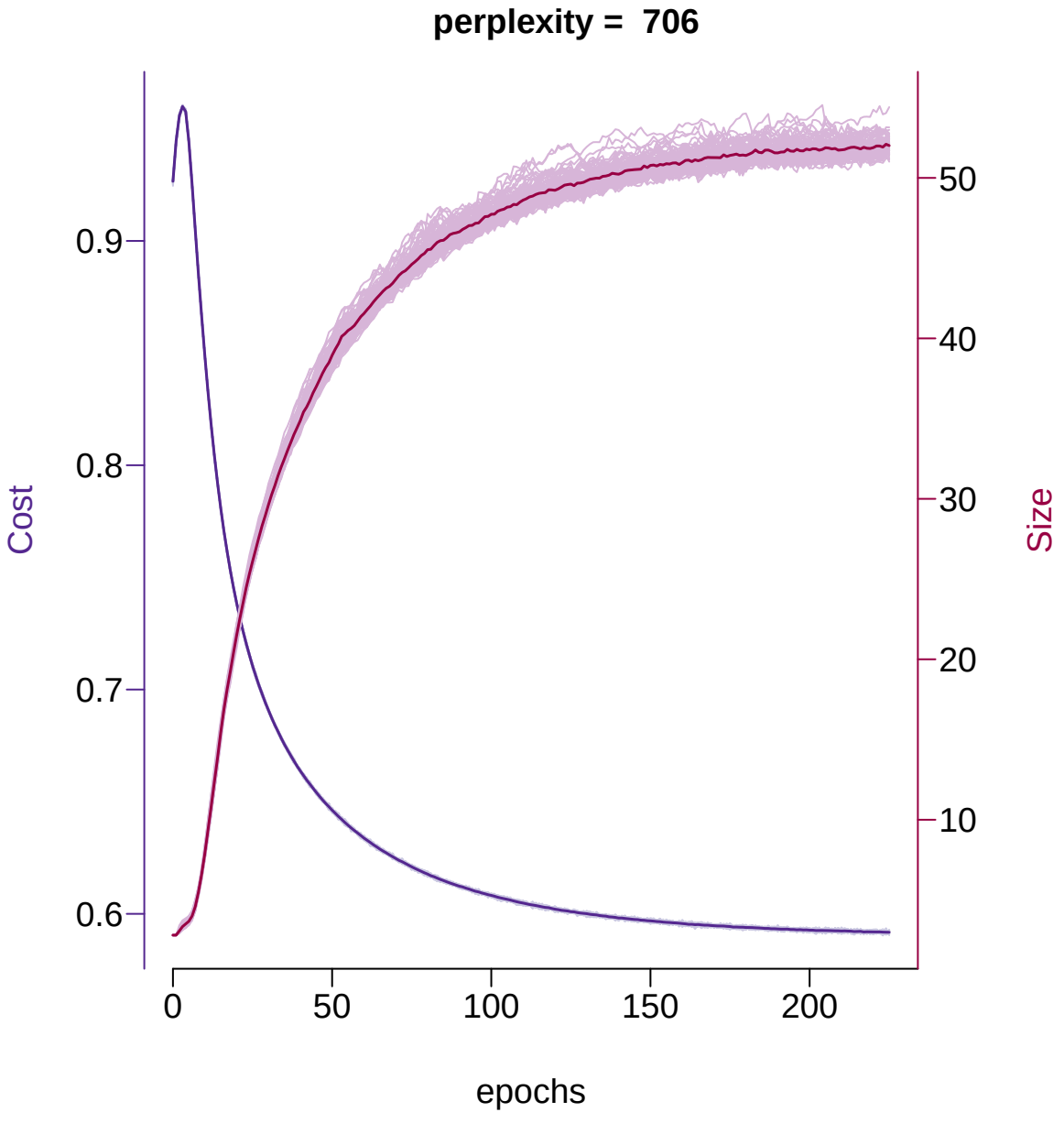

perplexity = 542
Y2
40
20
0
−20
−20
−10
0
10
20
30
Y1

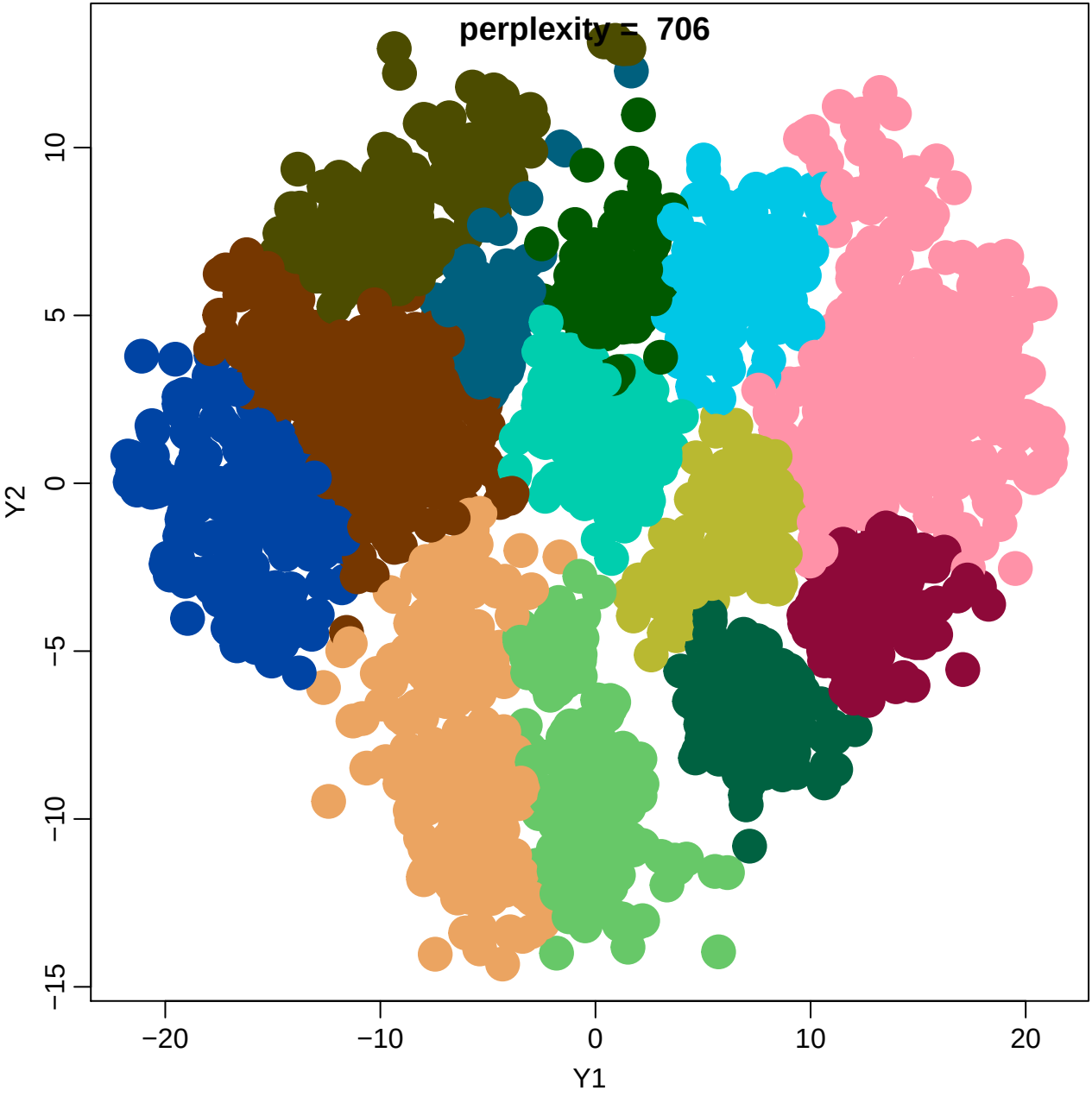

perplexity = 542
27
25
32
5
29
4
12
11
15
10
13
18
26
28
8
3
24
22
20
6
23
17
33
19
21
30
1
14
16
9
31
7
2

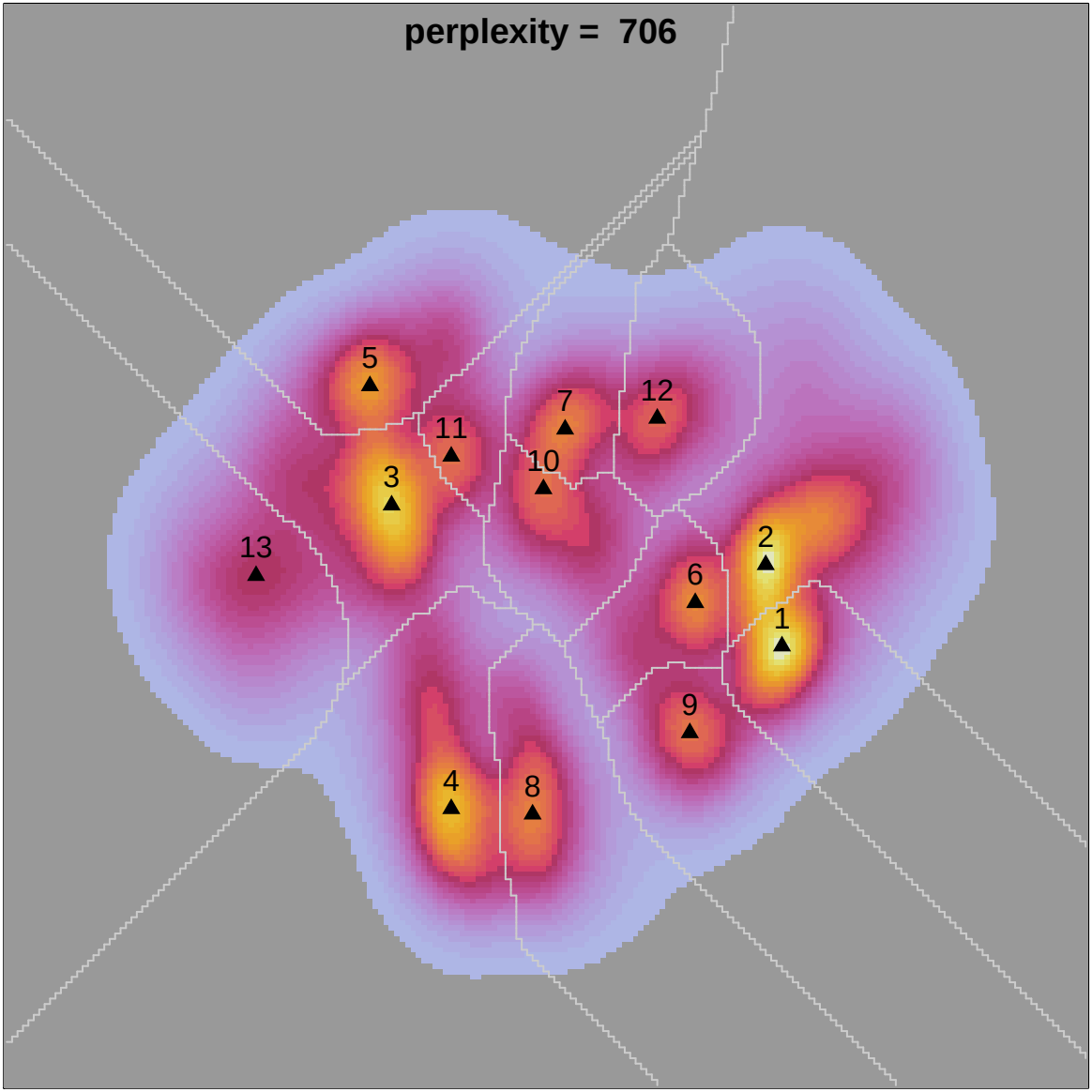

perplexity = 624
Cost
Size
0.9
0.8
0.7
0.6
70
60
50
40
30
20
10
0
50
100
150
200
epochs

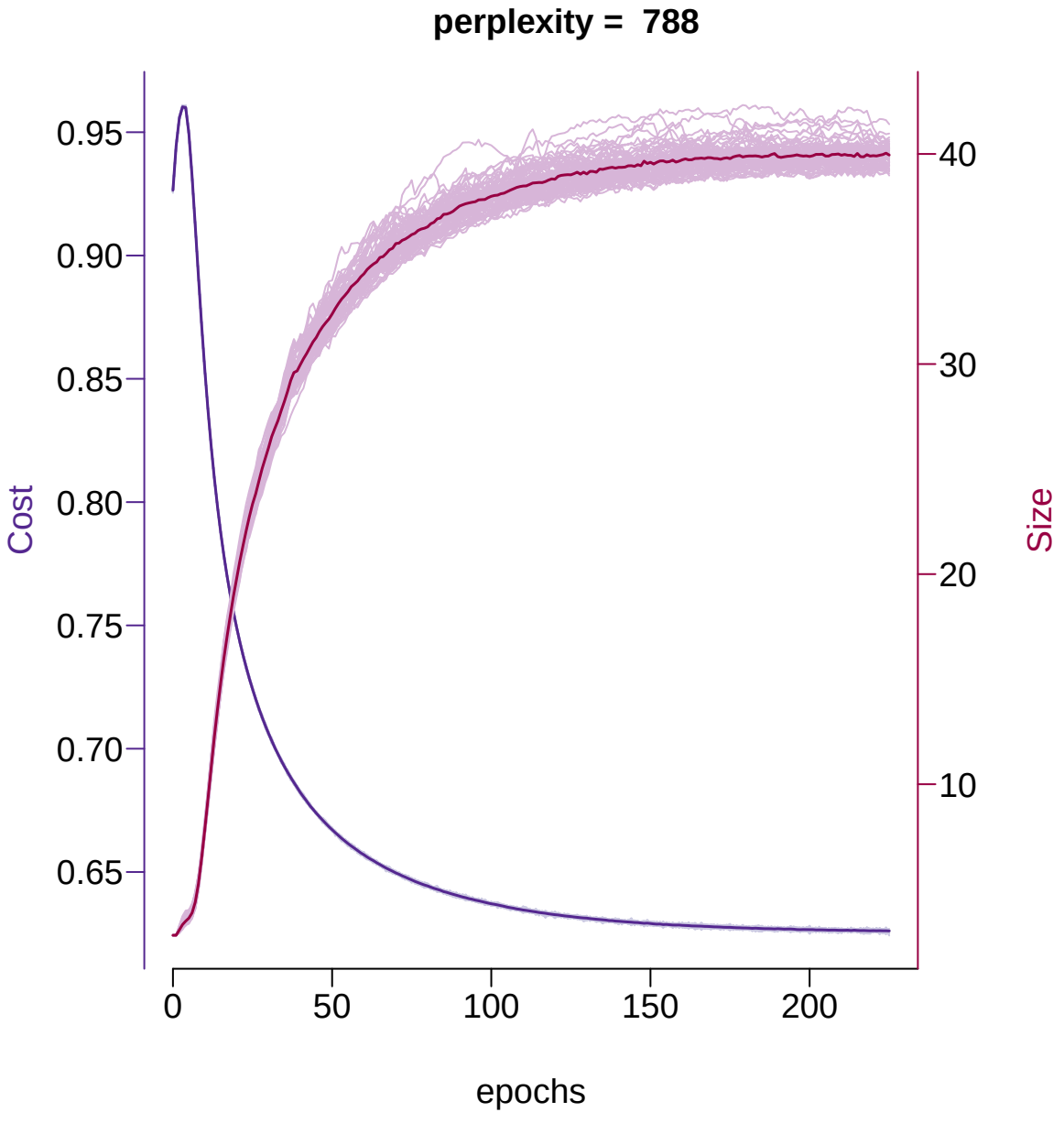

perplexity = 624
Y2
Y1
15
10
5
0
−5
−10
−15
−20
−30
−20
−10
0
10
20
30

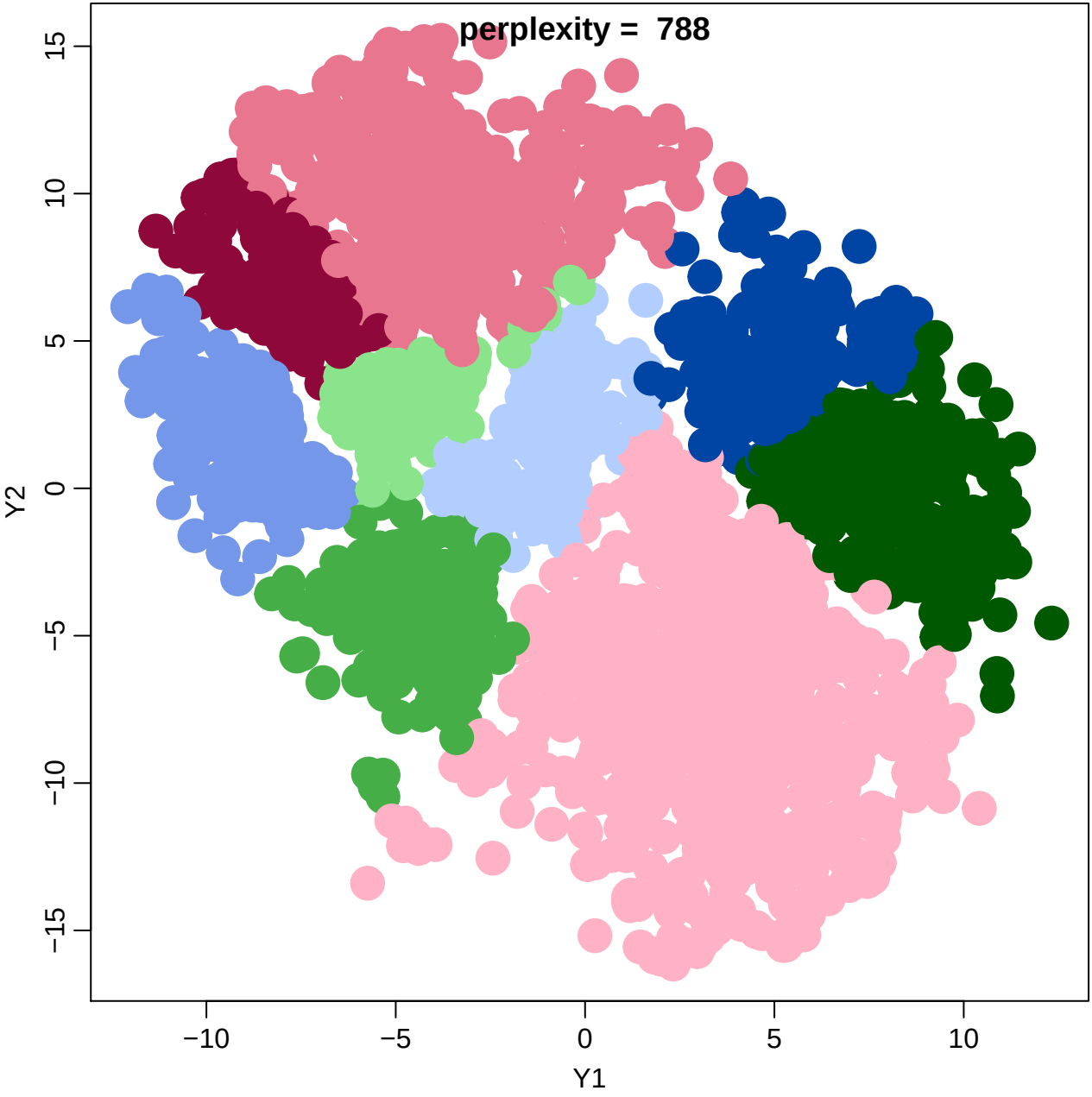

perplexity = 624
6
11
16
18
13
3
14
17
7
5
10
2
12
1
15
8
4
9

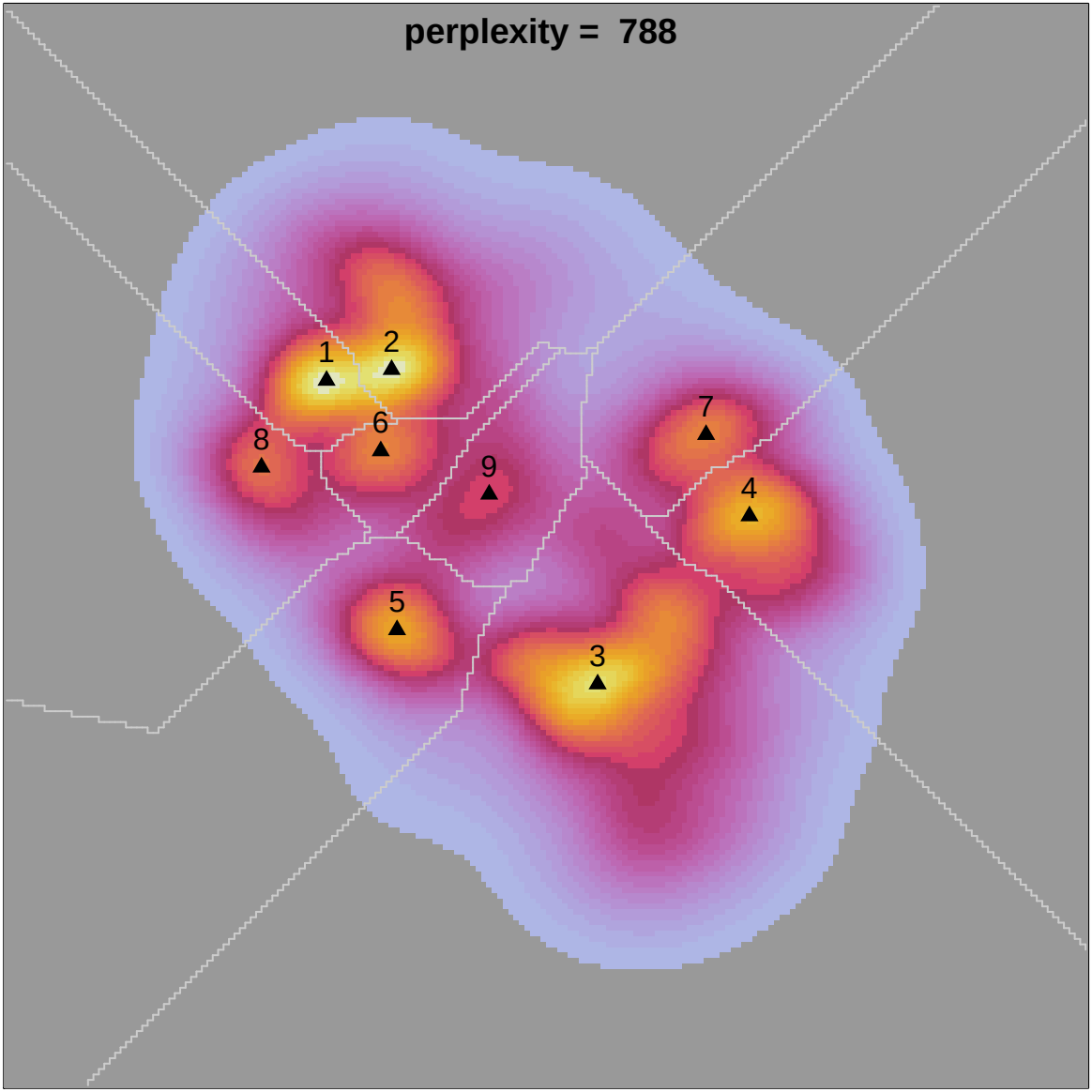

perplexity = 706
Cost
0.9
0.8
0.7
0.6
Size
50
40
30
20
10
0
50
100
150
200
epochs

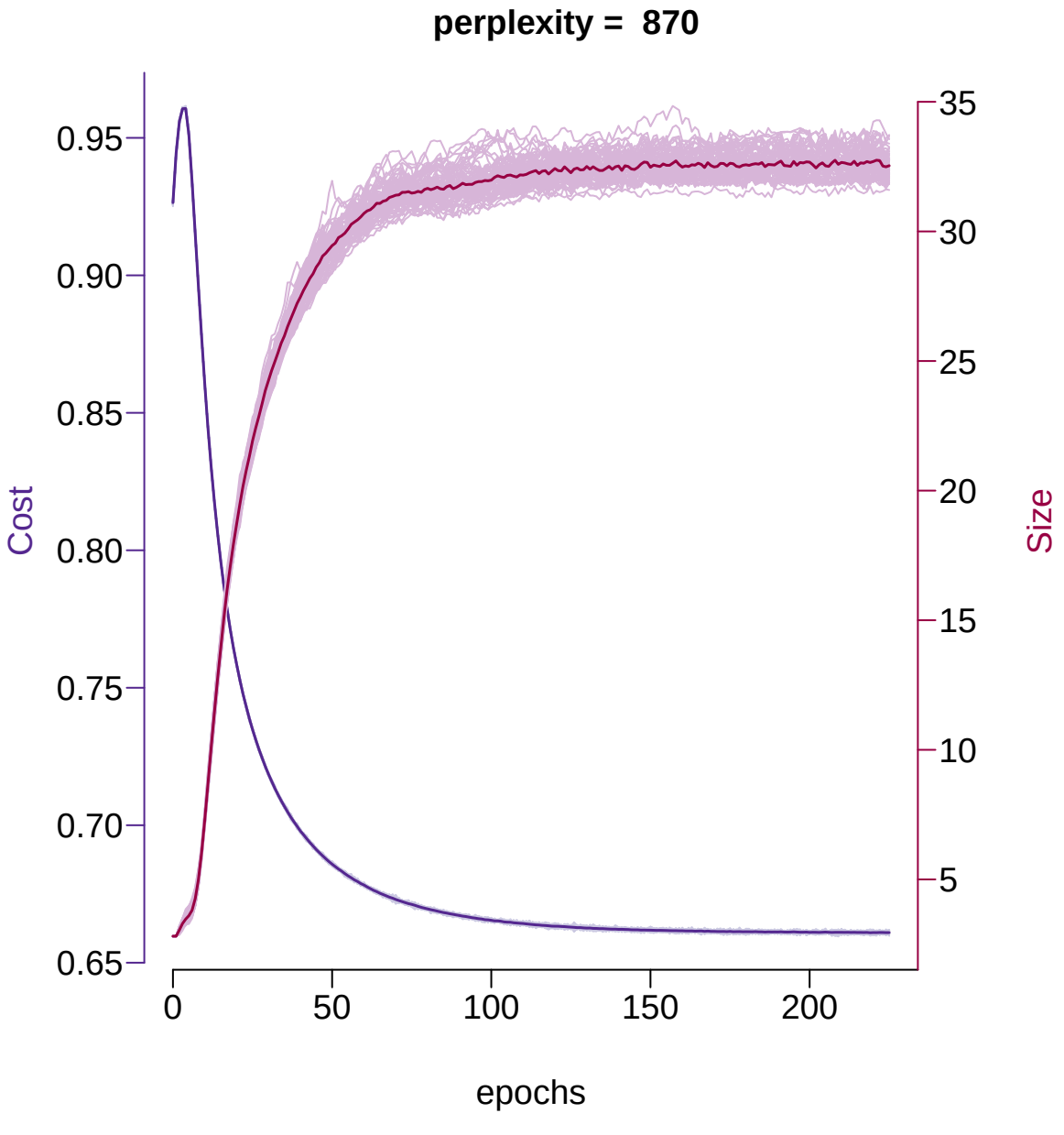

perplexity = 706
Y2
Y1
10
5
0
−5
−10
−15
−20
−10
0
10
20

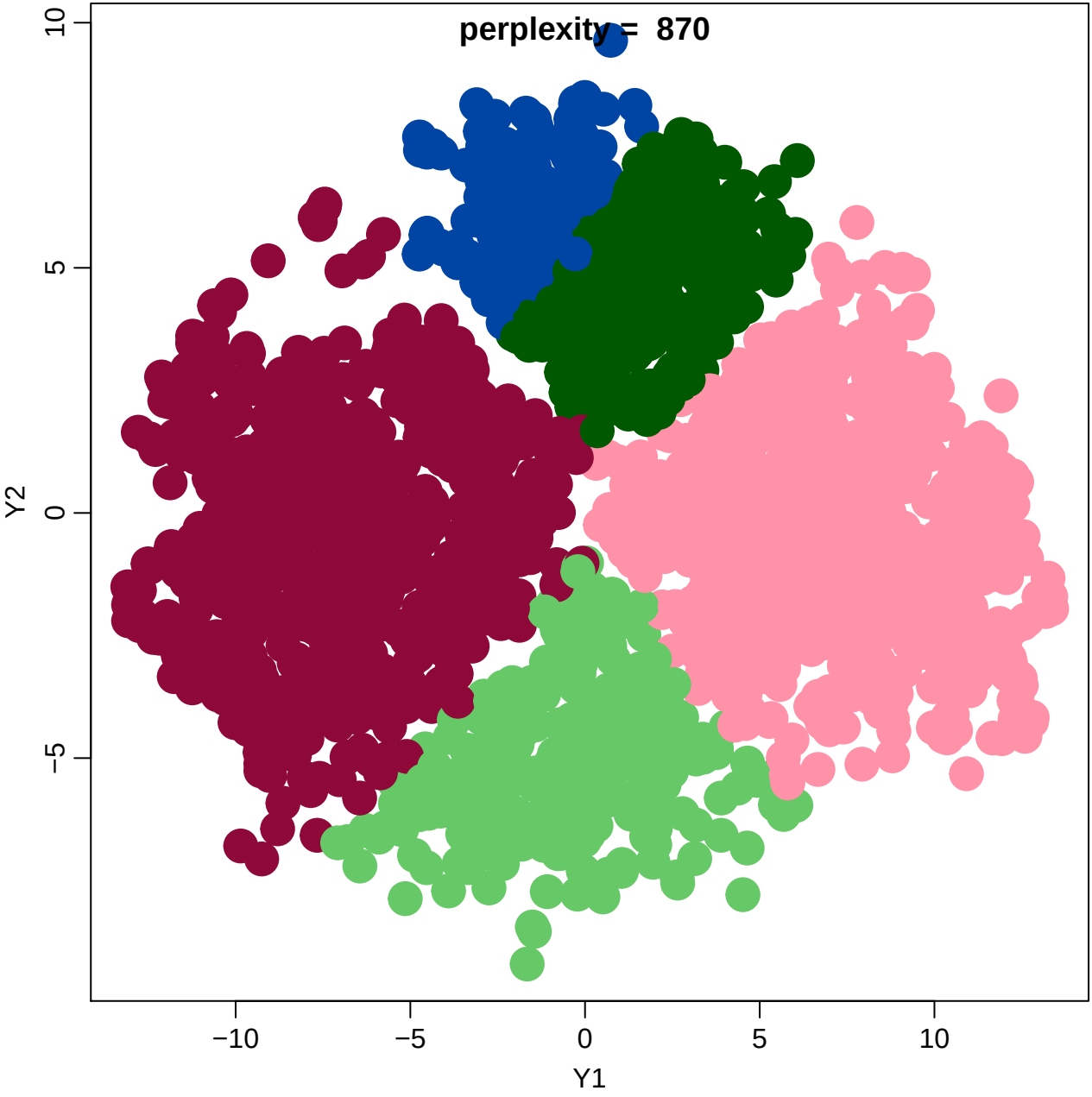

perplexity = 706
5
7
12
11
10
3
13
2
6
1
9
4
8

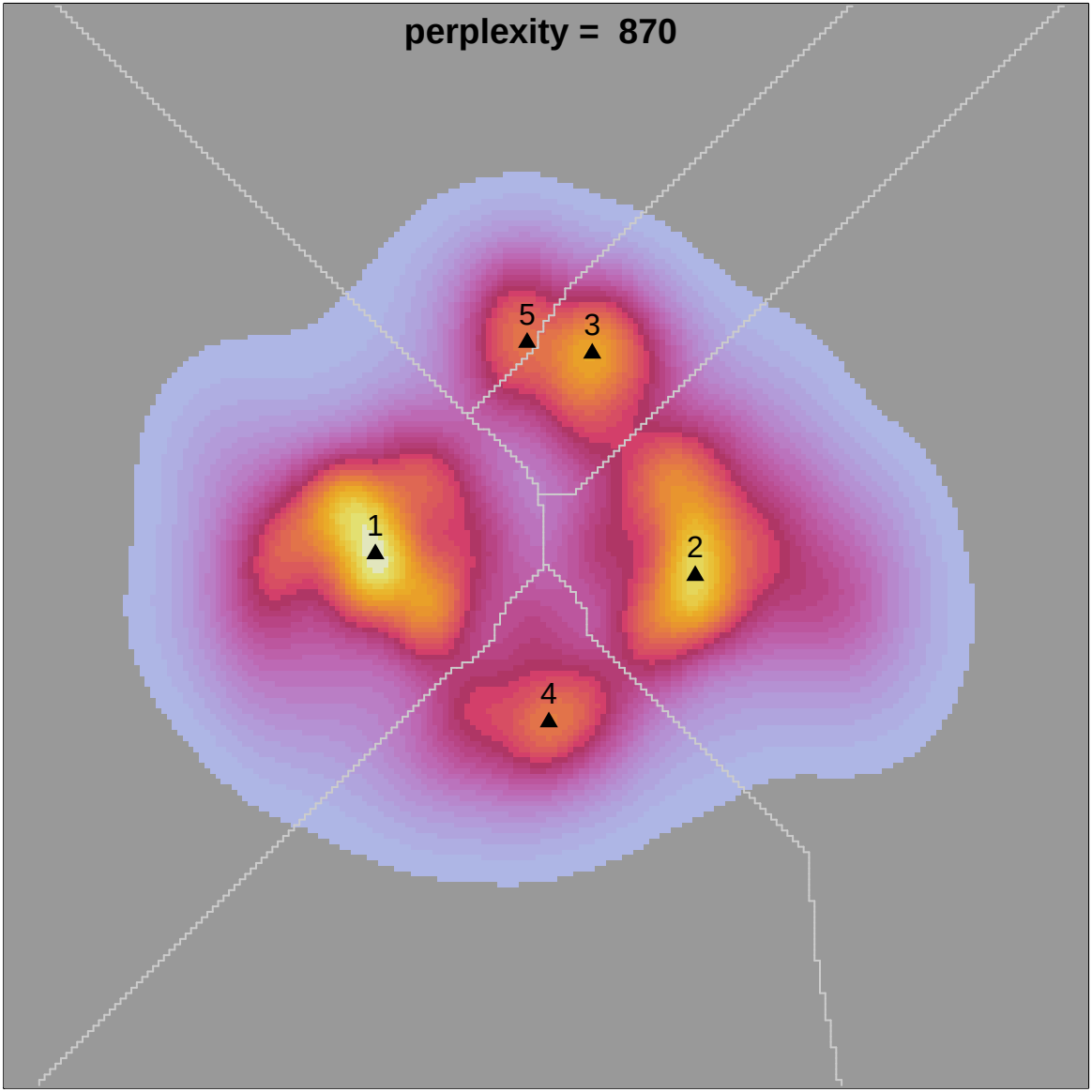

perplexity = 788
Cost
Size
0.95
0.90
0.85
0.80
0.75
0.70
0.65
40
30
20
10
0
50
100
150
200
epochs

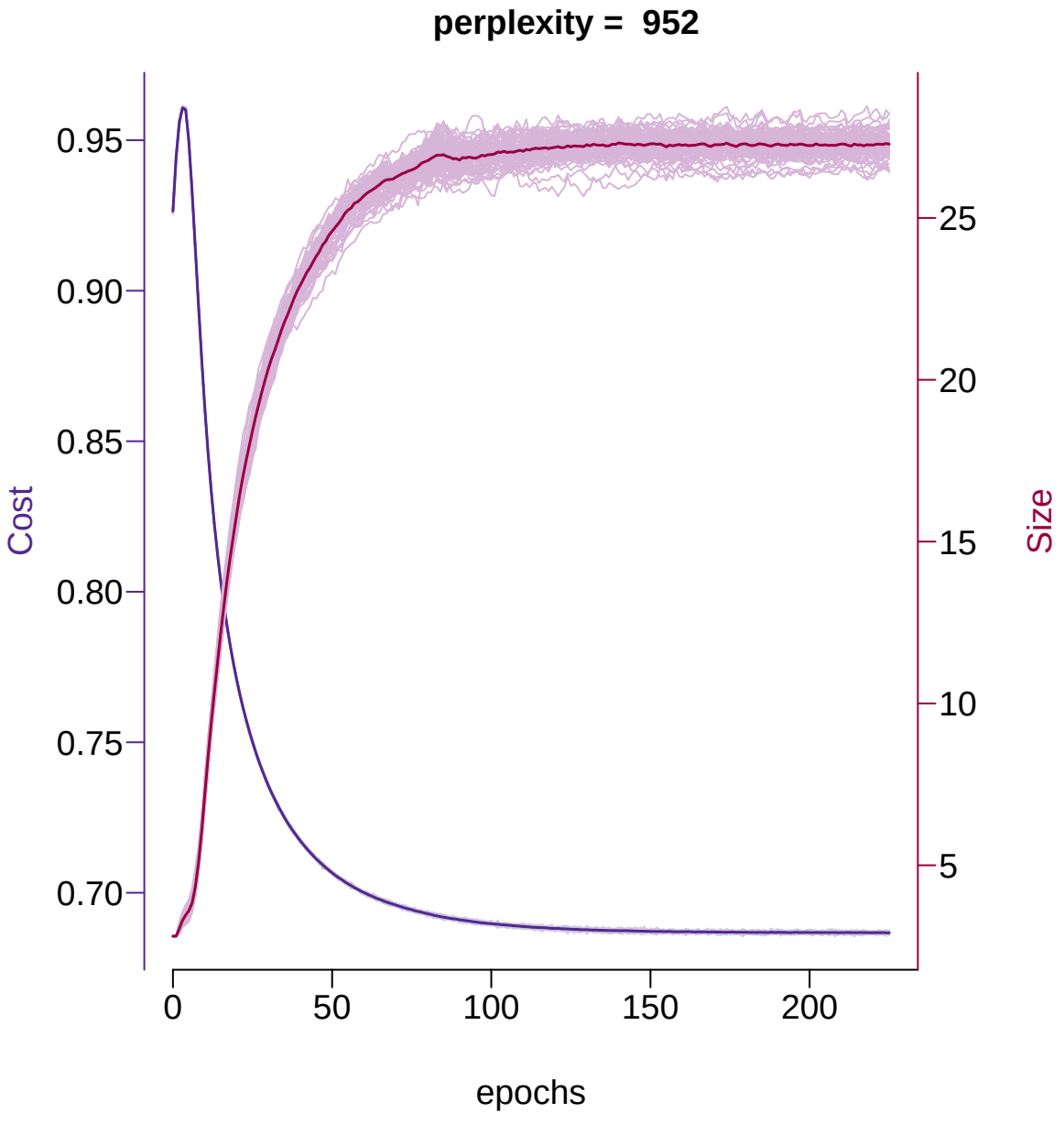

perplexity = 788
15
10
5
0
−5
−10
−15
Y2
−10
−5
0
5
10
Y1

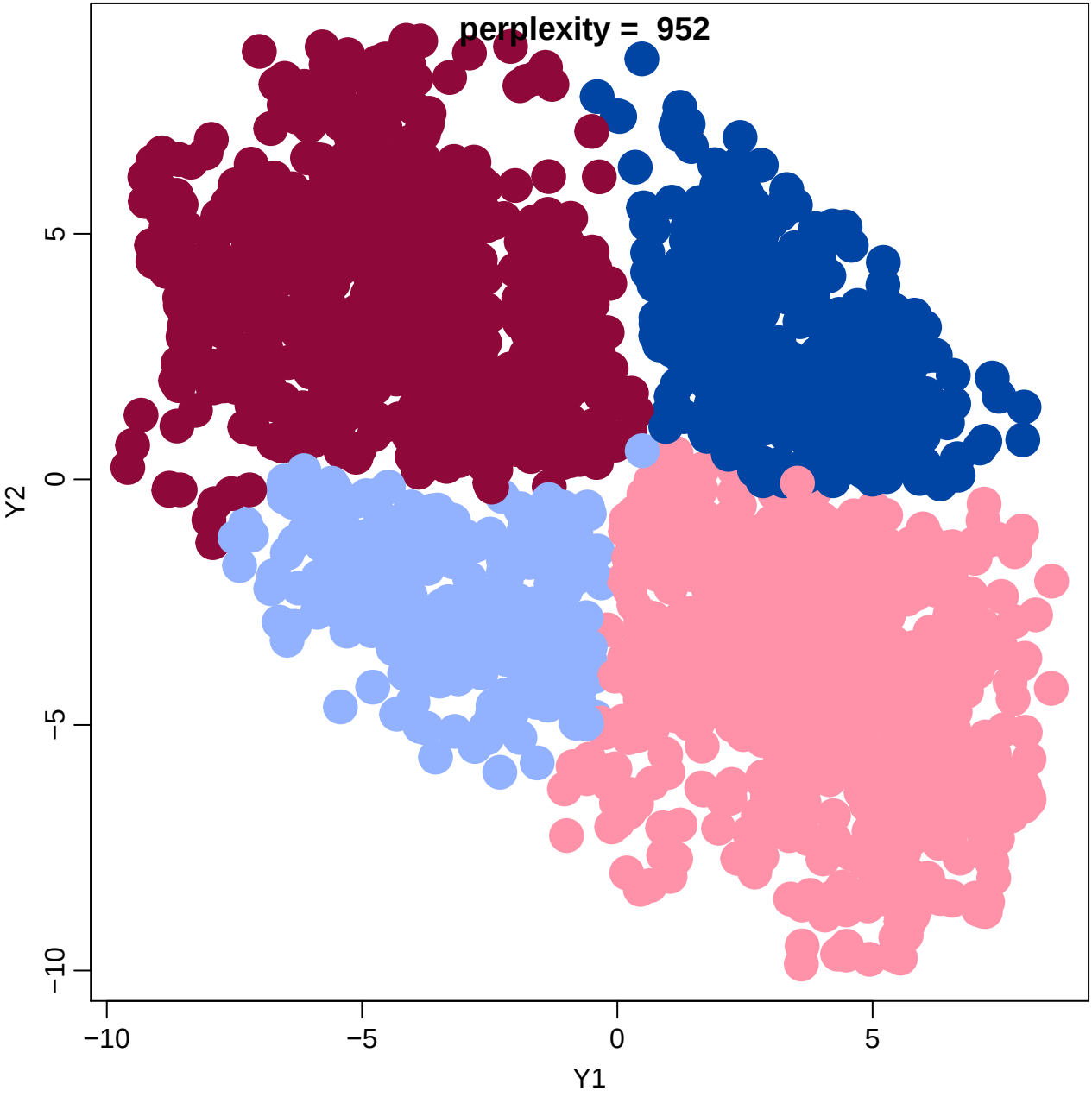

perplexity = 788
1
2
6
8
9
7
4
5
3

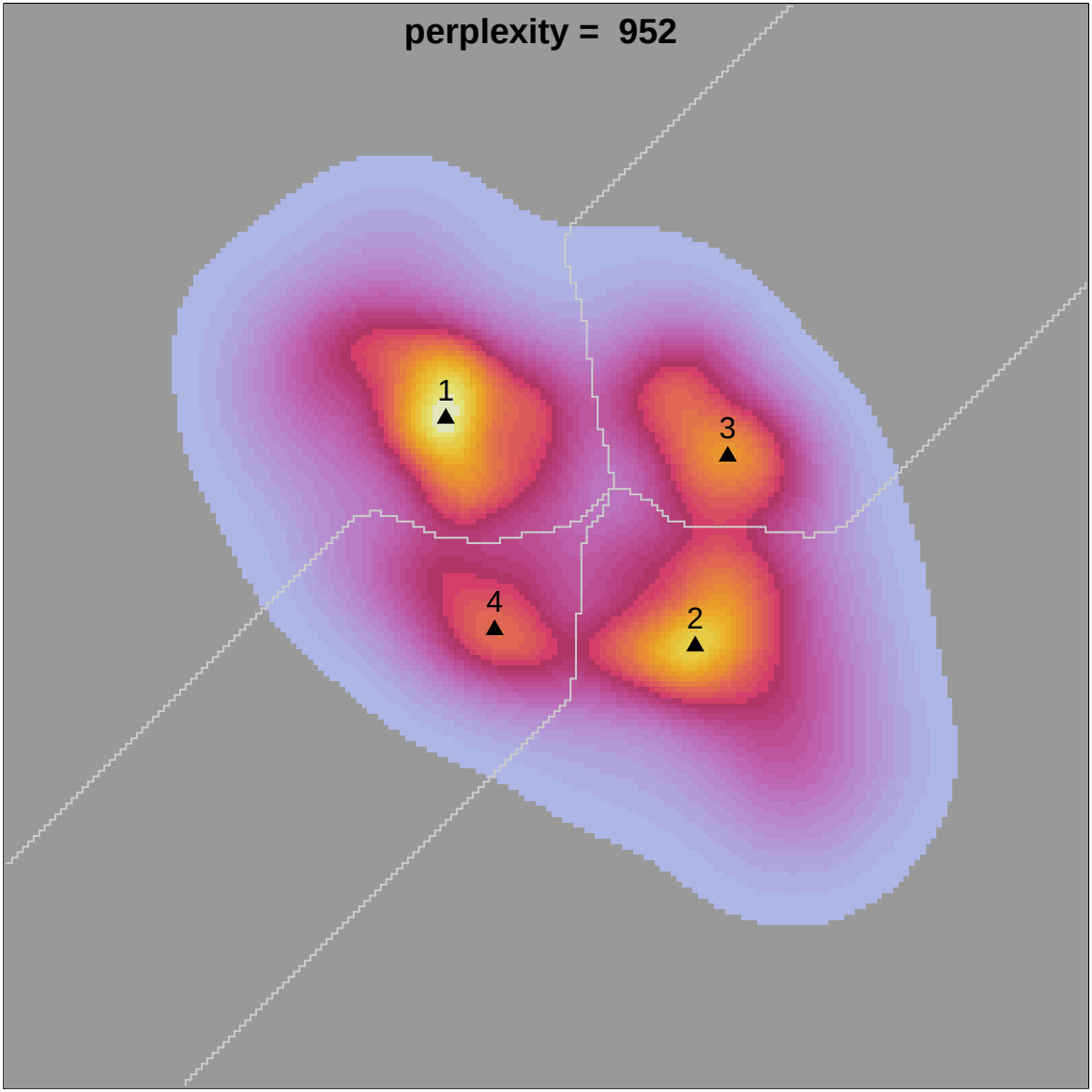

perplexity = 870
Cost
Size
epochs
0.95
0.90
0.85
0.80
0.75
0.70
0.65
35
30
25
20
15
10
5
0
50
100
150
200

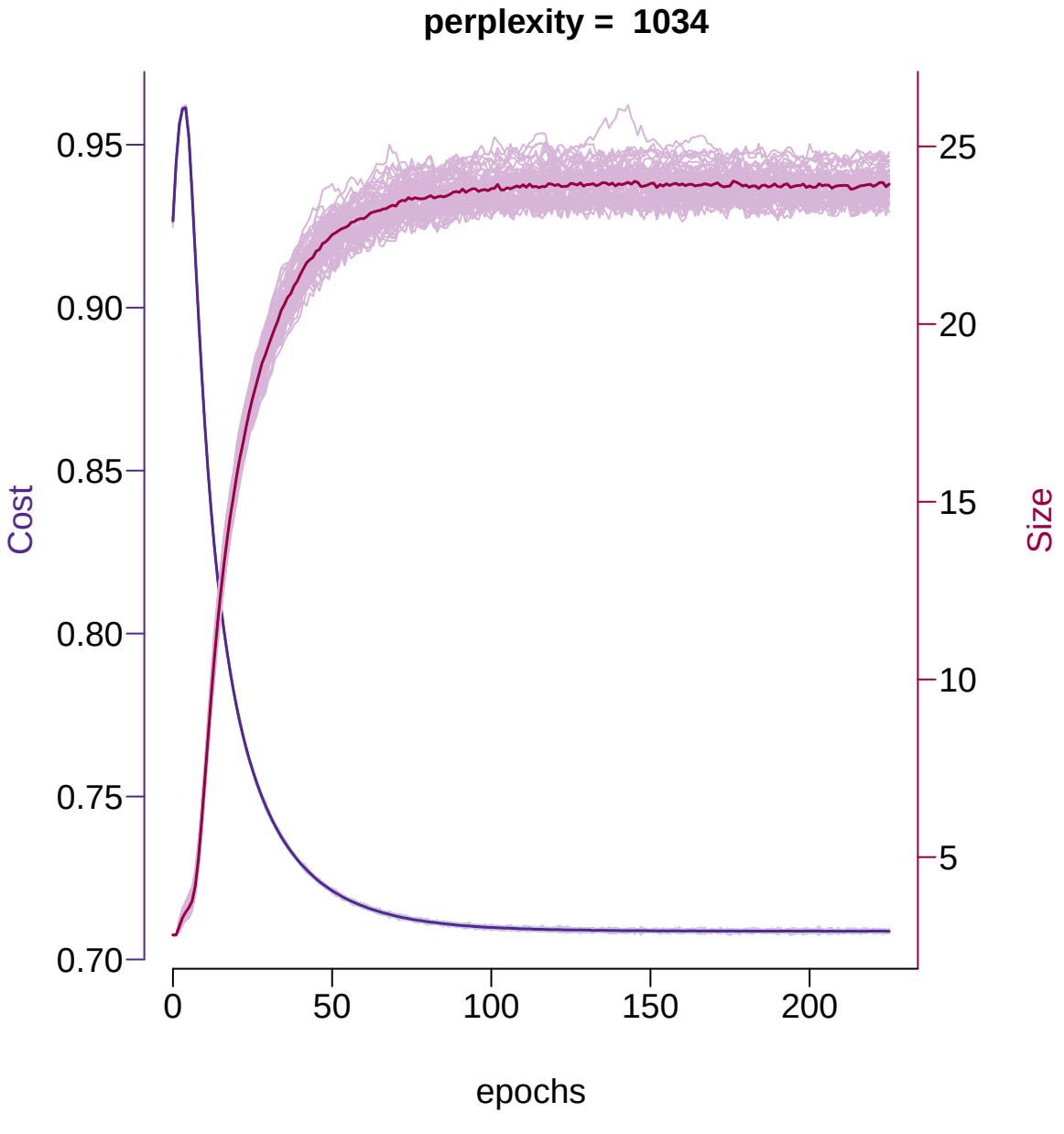

perplexity = 870
Y2
Y1
10
5
0
−5
−10
−5
0
5
10

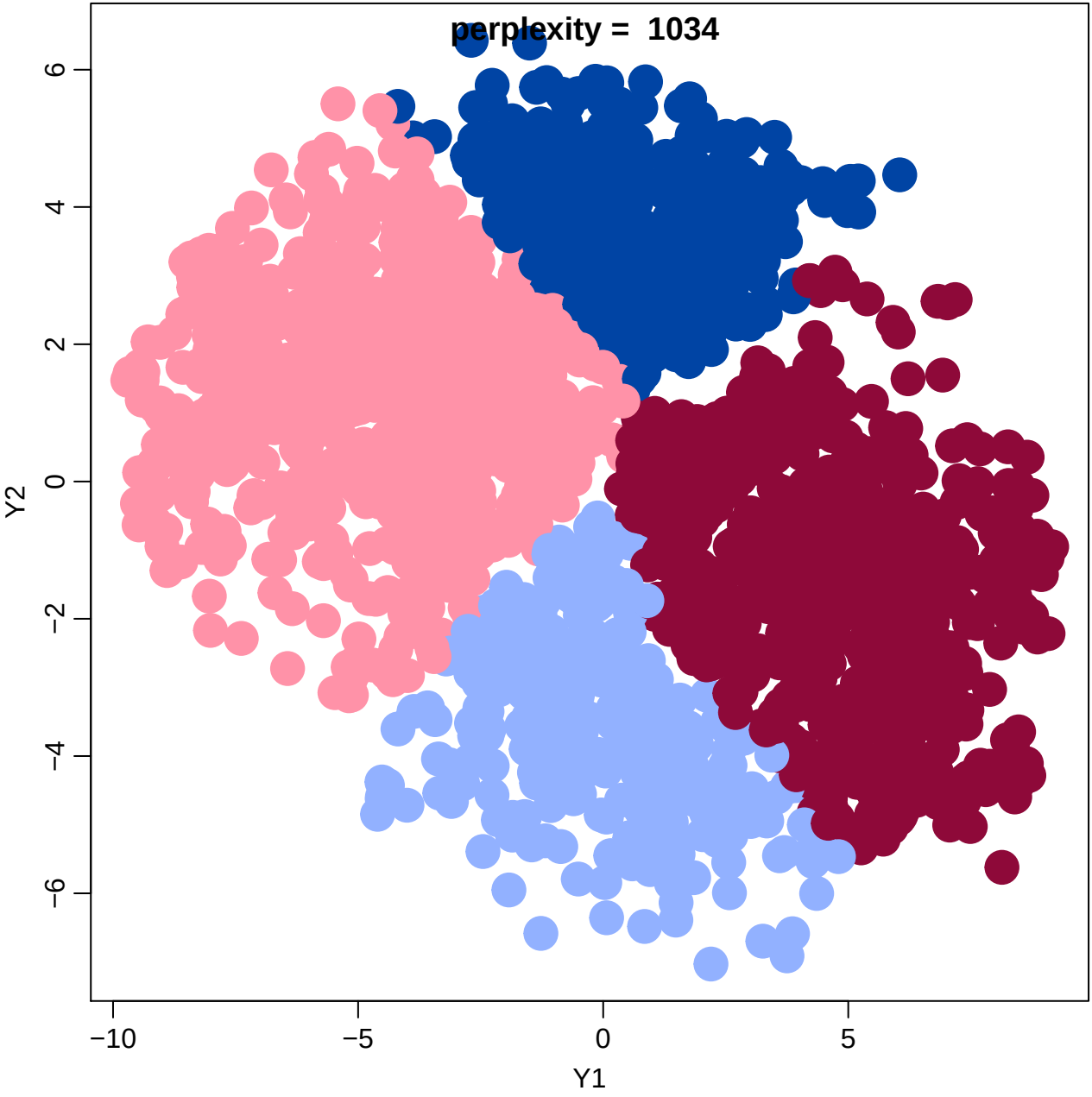

perplexity = 870
5
3
1
2
4

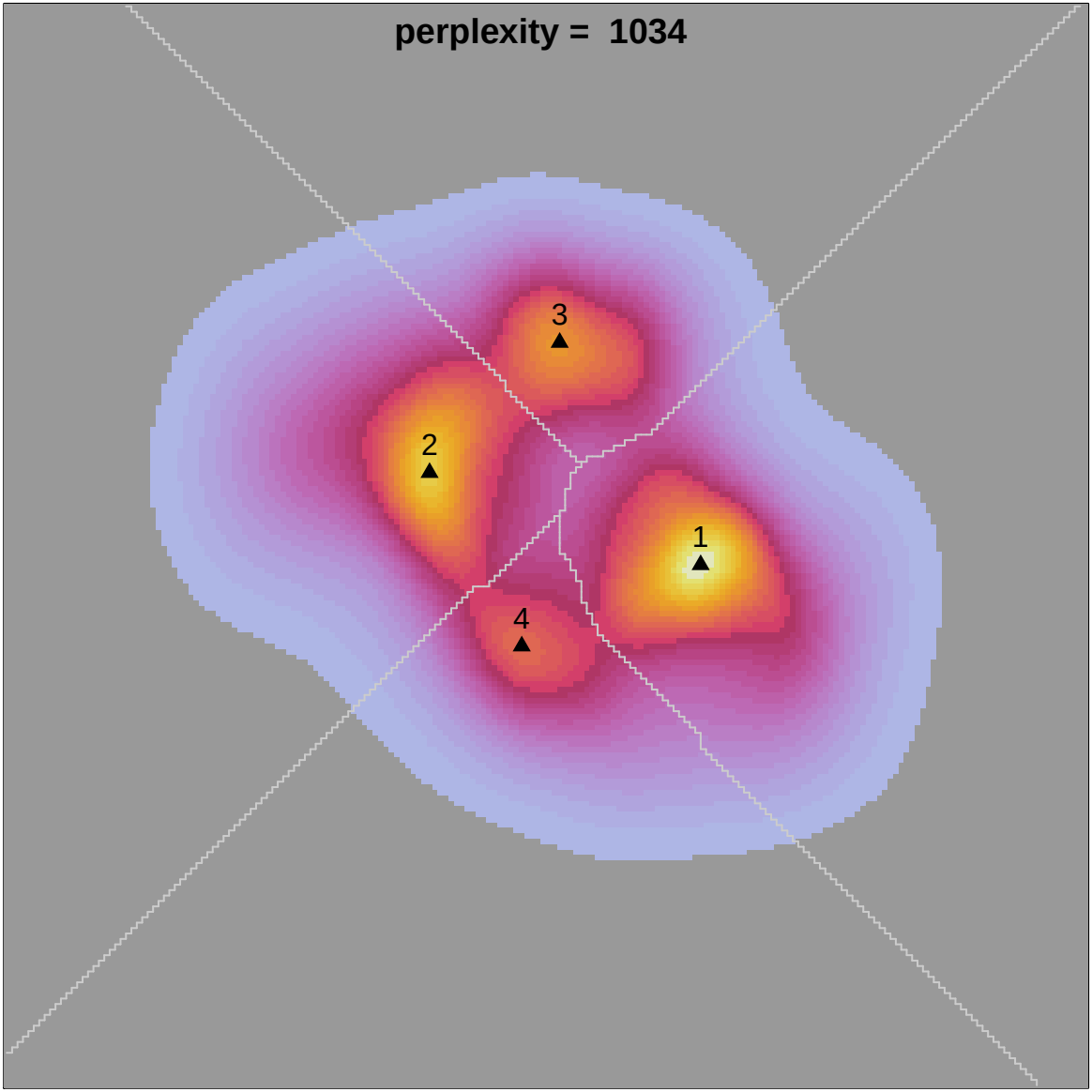

perplexity = 952
Cost
0.95
0.90
0.85
0.80
0.75
0.70
Size
25
20
15
10
5
0
50
100
150
200
epochs

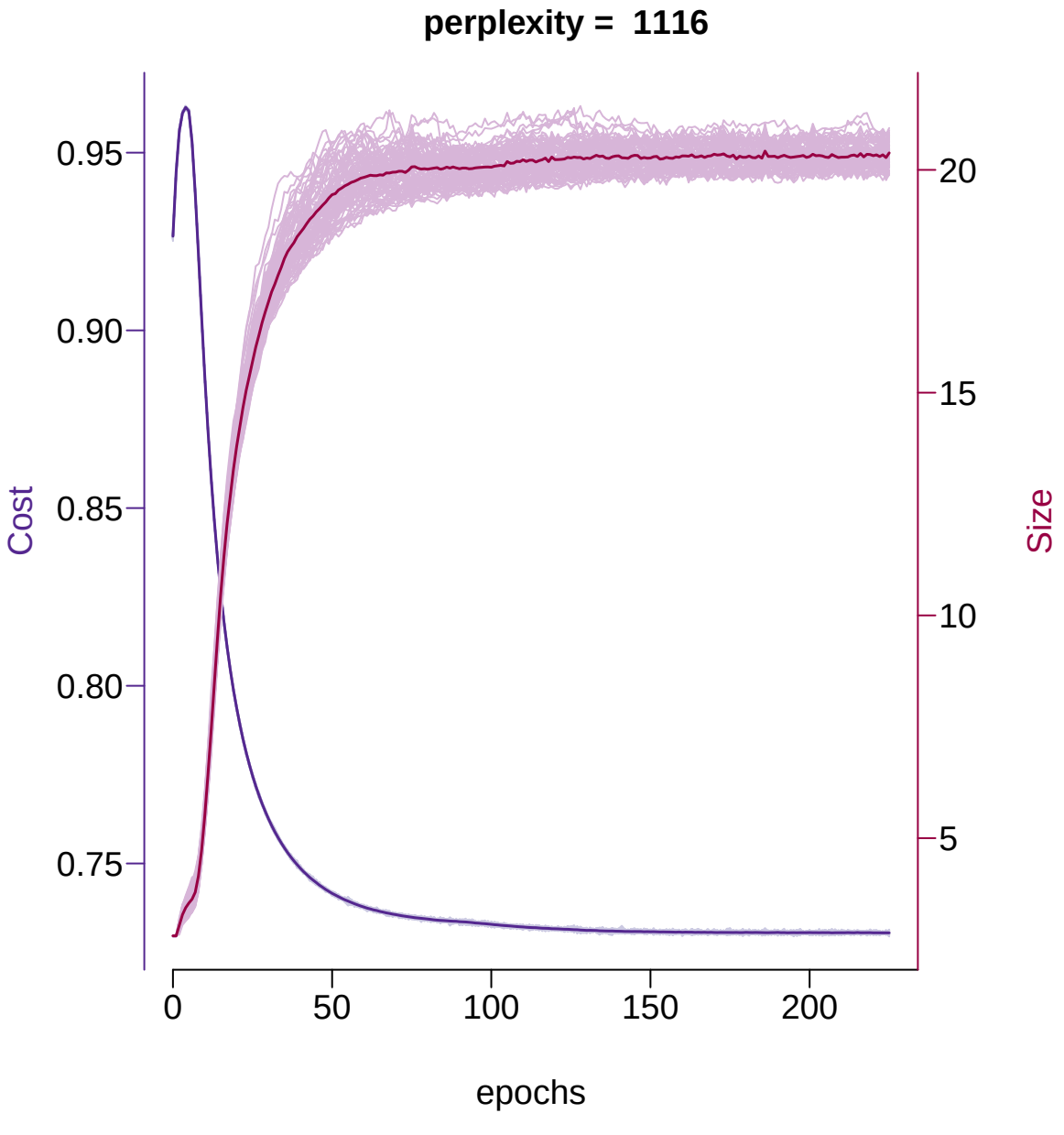

perplexity = 952
Y2
Y1
5
0
−5
−10
−10
−5
0
5

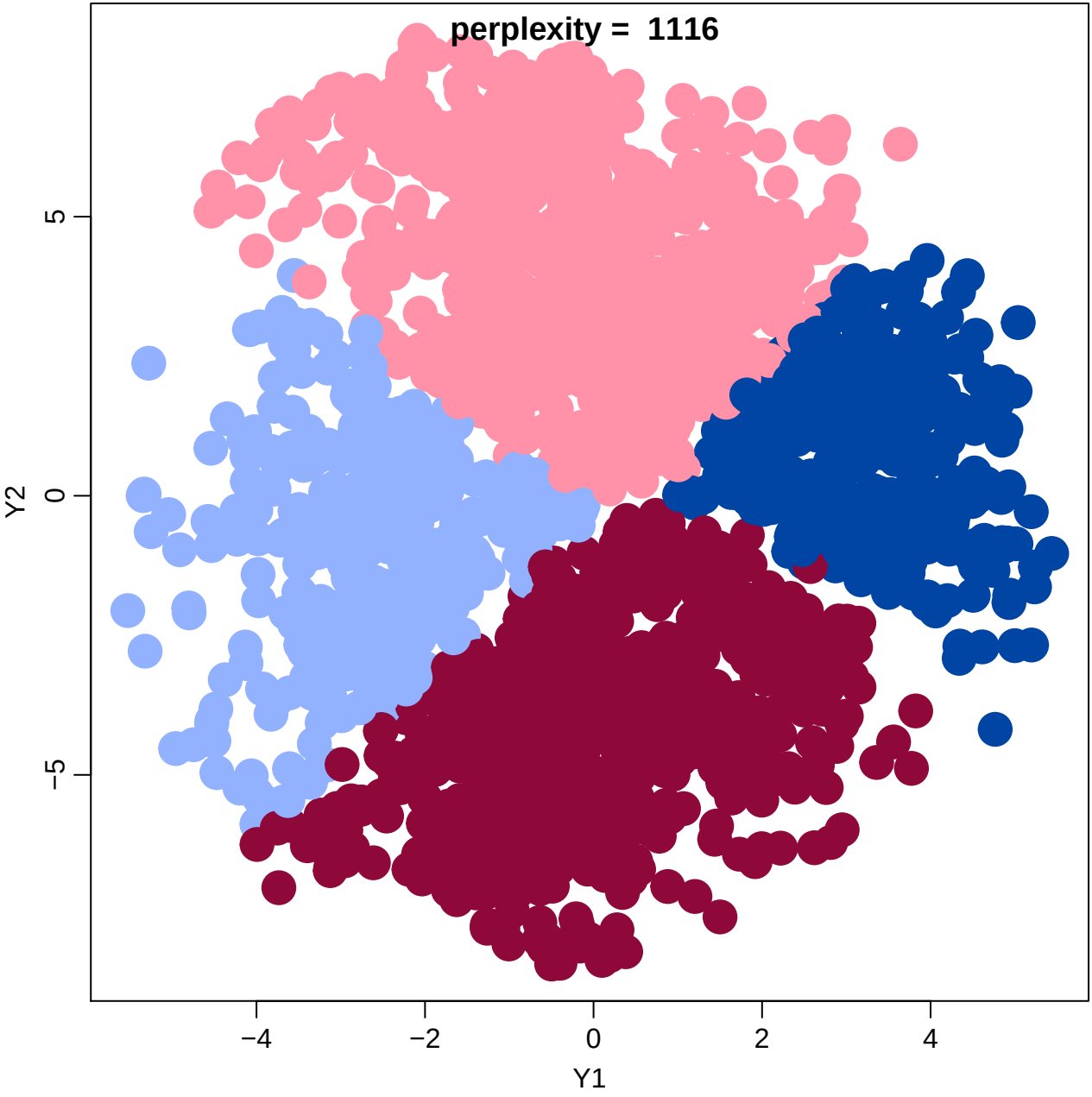

perplexity = 952
1
3
4
2

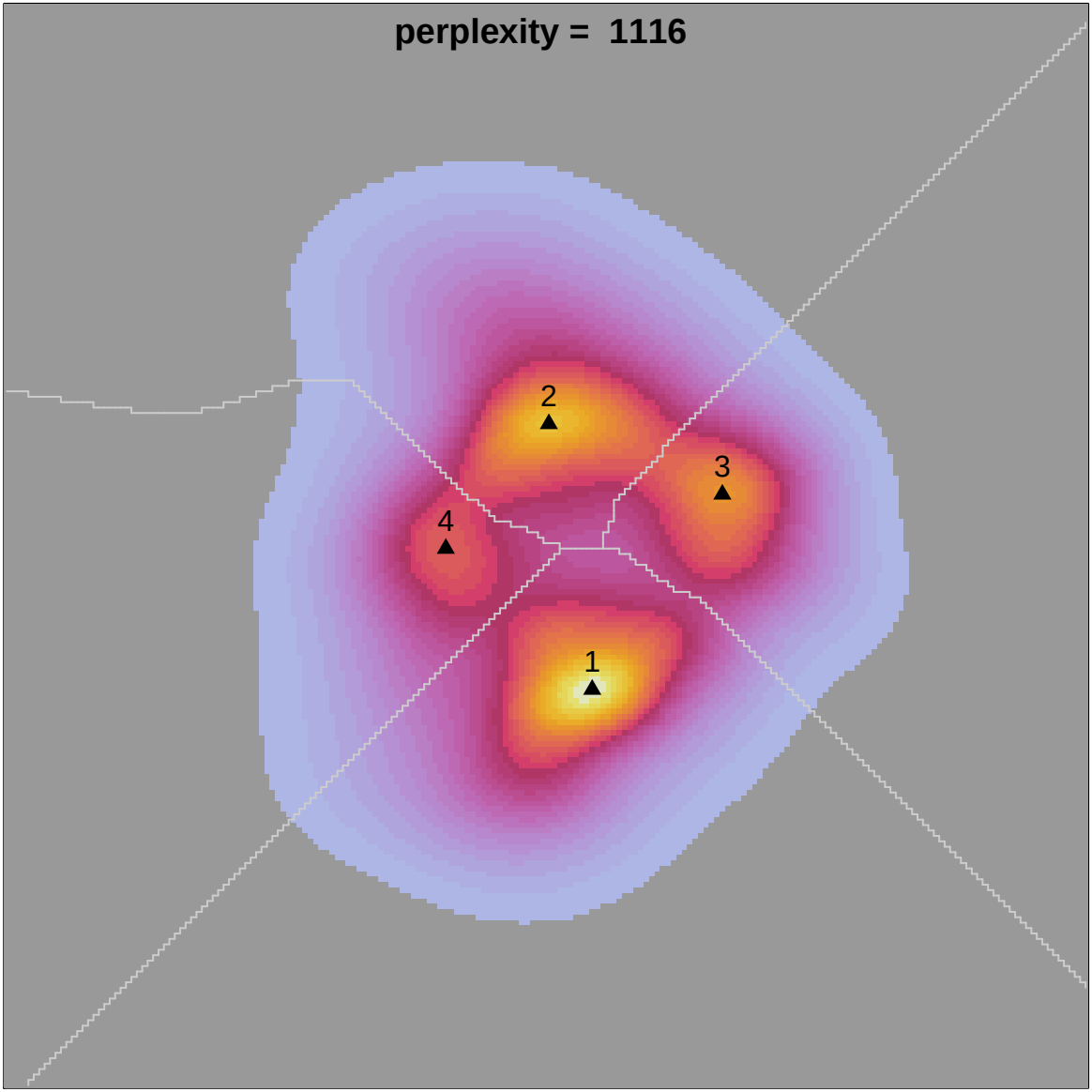

perplexity = 1034
Cost
Size
epochs
0.95
0.90
0.85
0.80
0.75
0.70
25
20
15
10
5
0
50
100
150
200

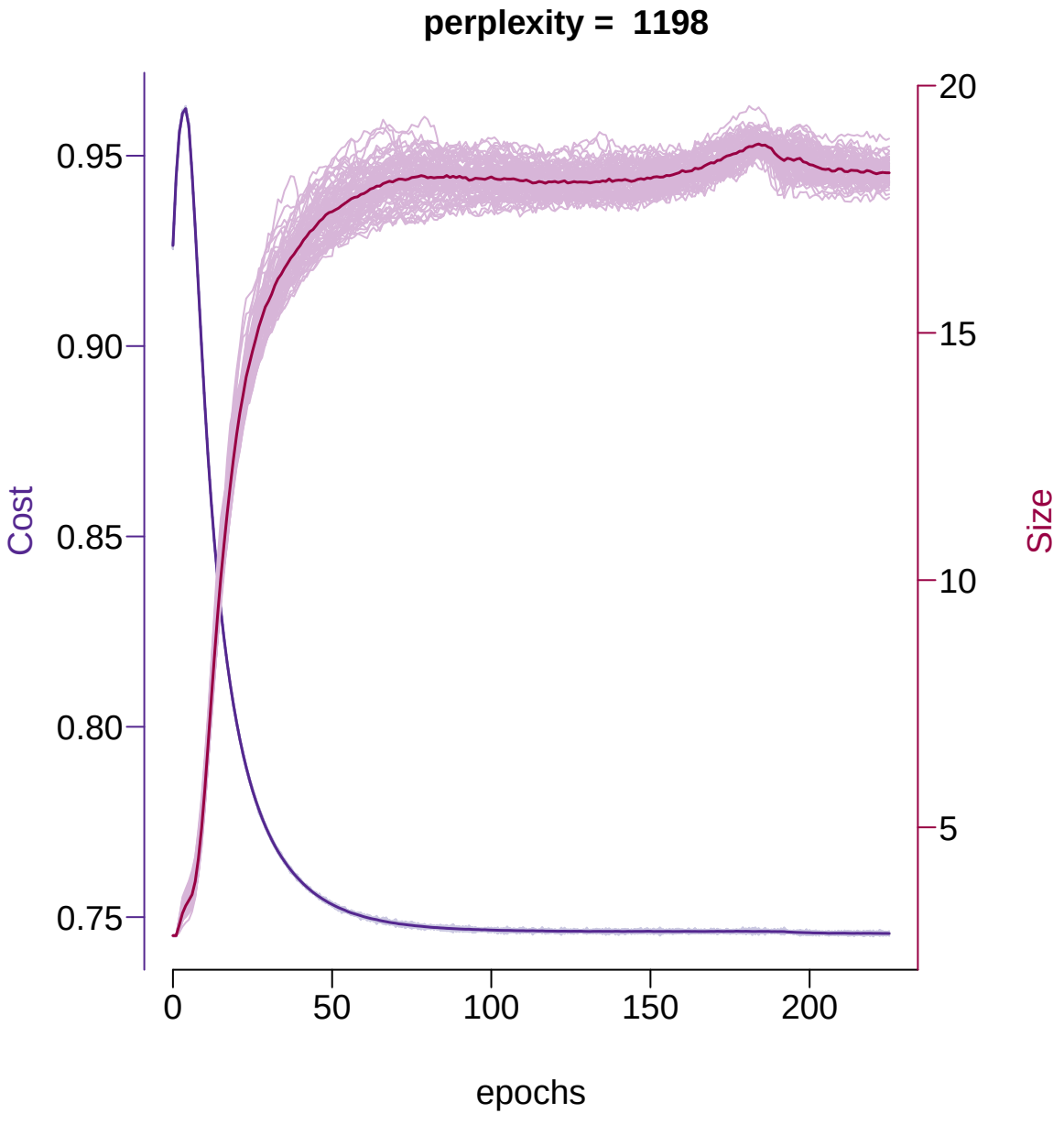

perplexity = 1034
Y2
Y1
6
4
2
0
−2
−4
−6
−10
−5
0
5

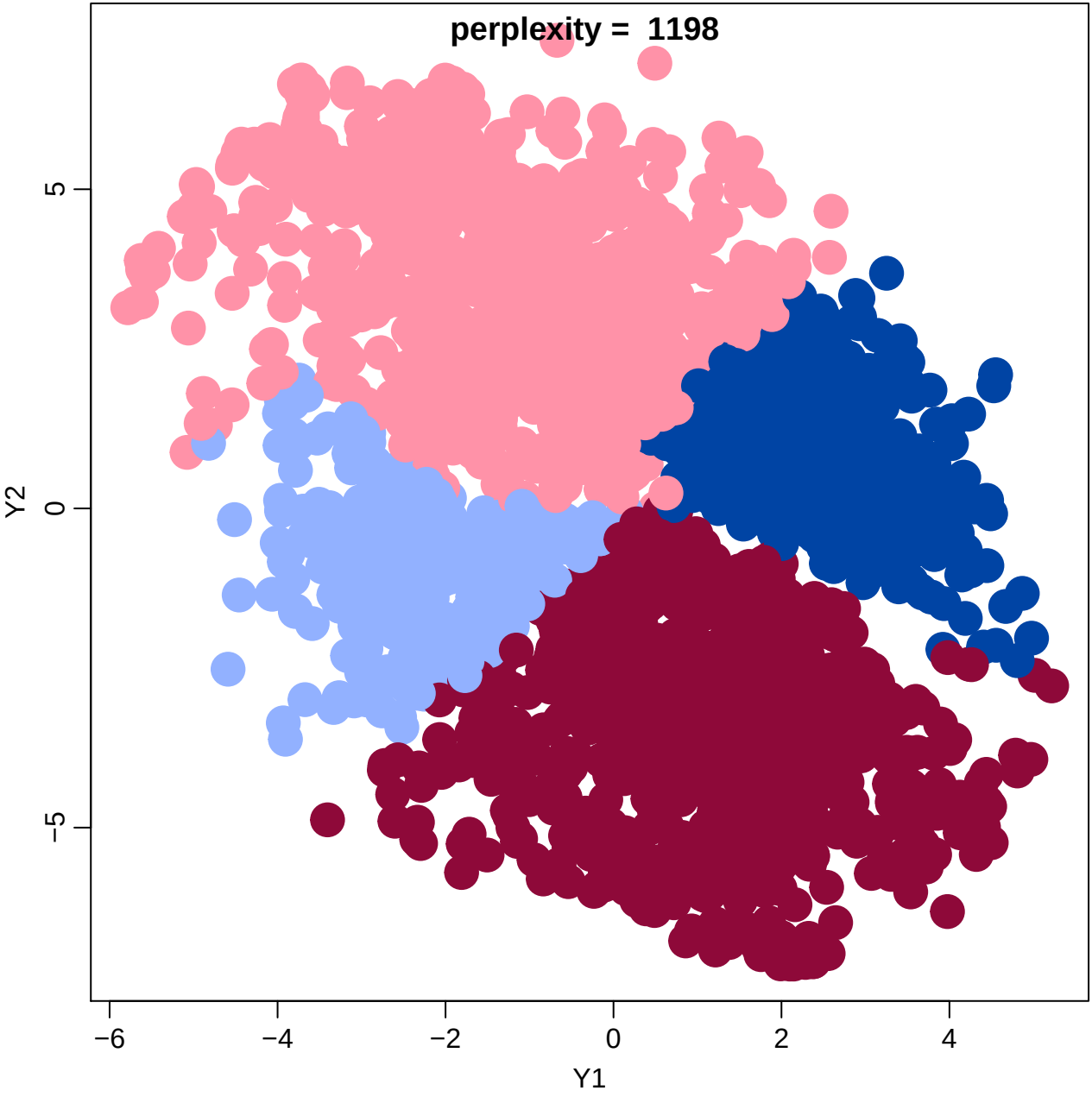

perplexity = 1034
3
2
1
4

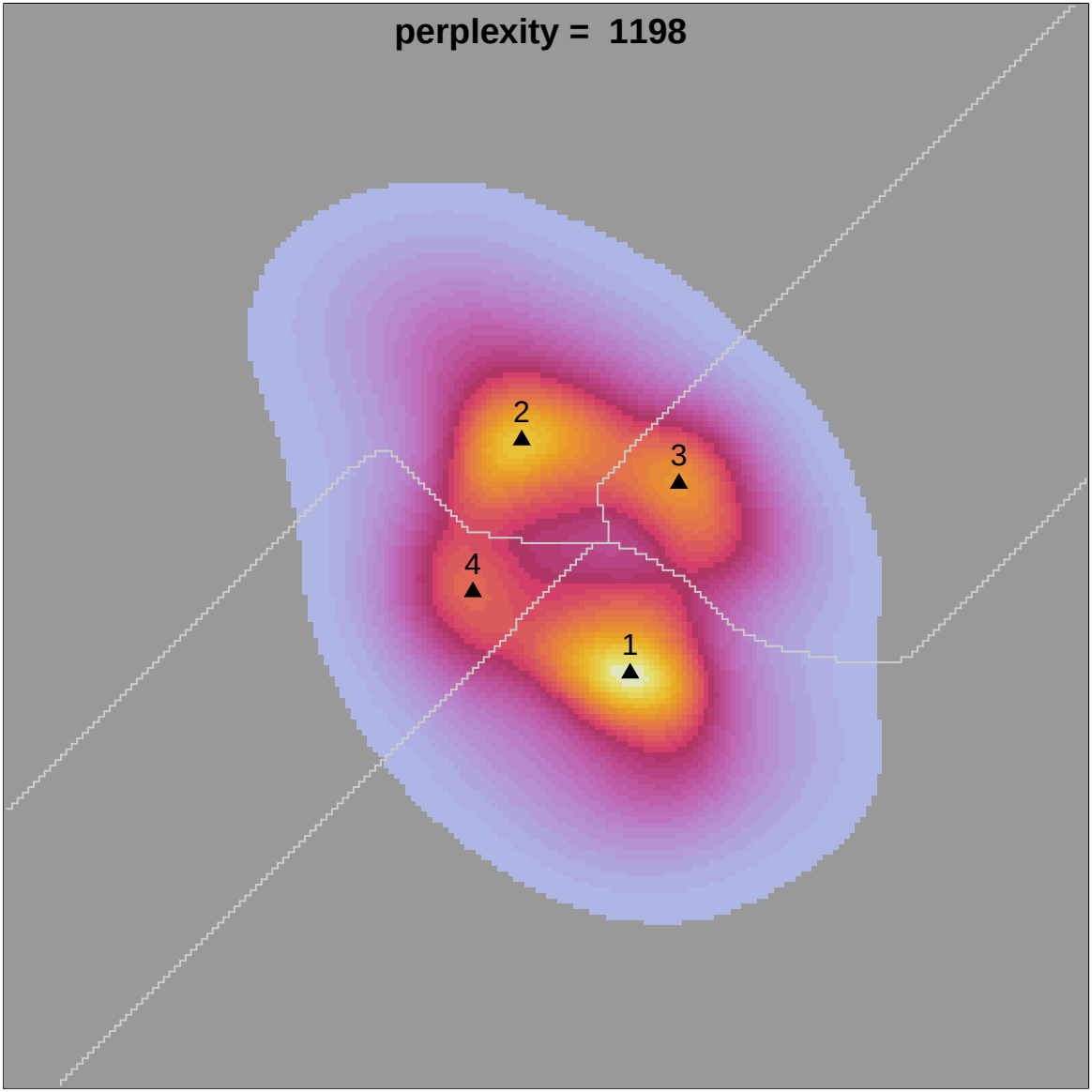

perplexity = 1116
Cost
0.95
0.90
0.85
0.80
0.75
Size
20
15
10
5
0
50
100
150
200
epochs

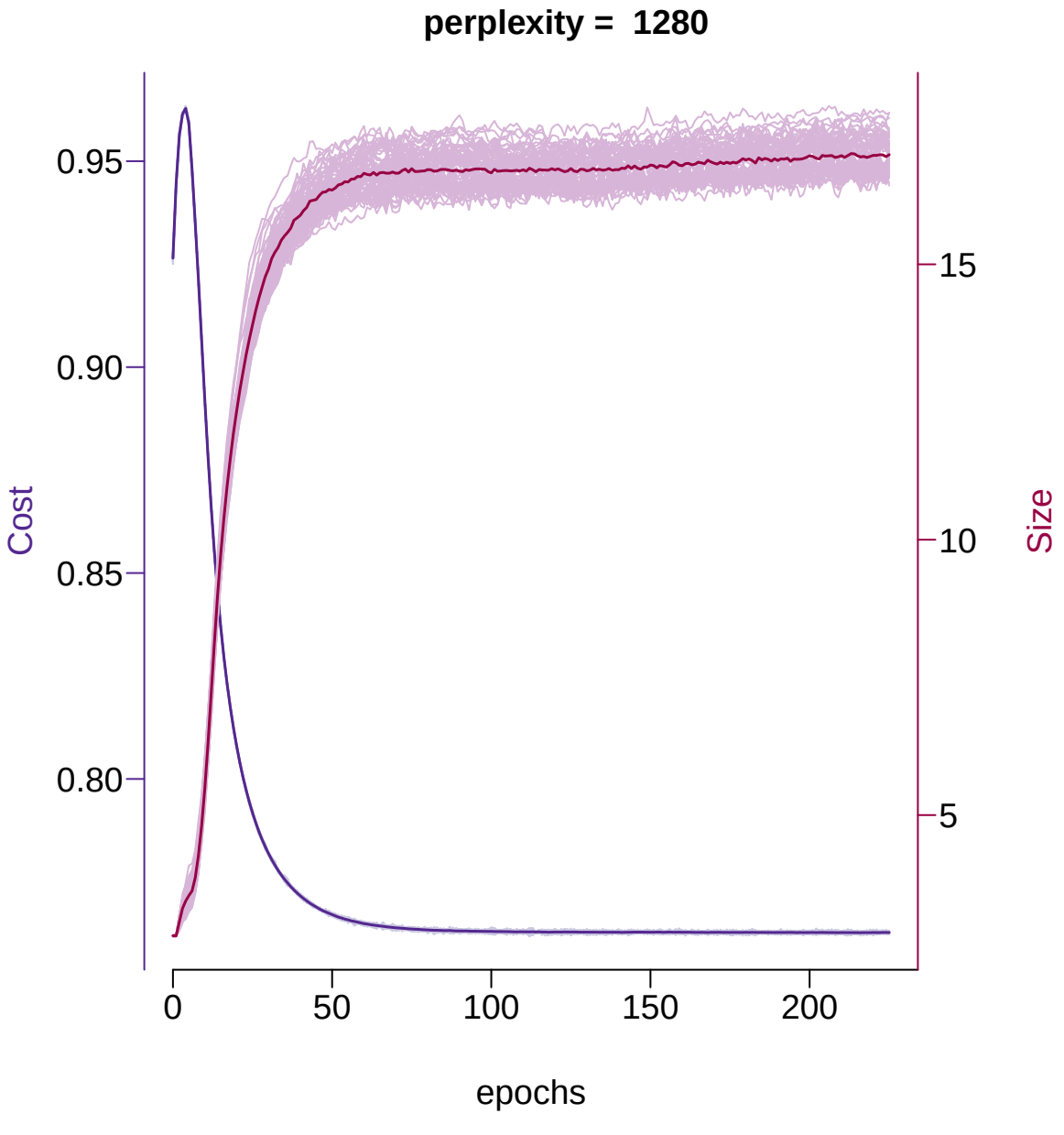

perplexity = 1116
Y2
5
0
−5
−4
−2
0
2
4
Y1

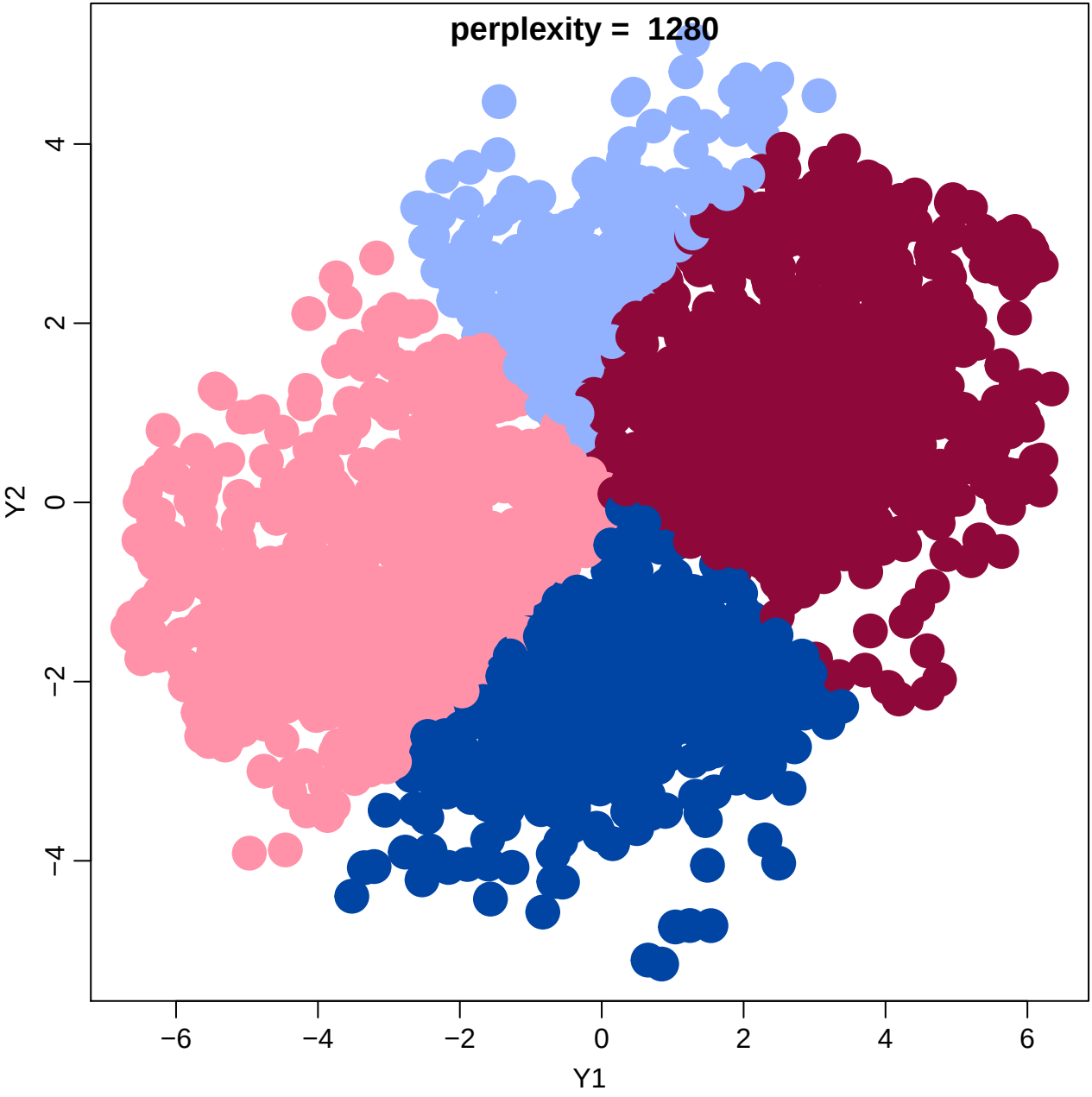

perplexity = 1116
2
3
4
1

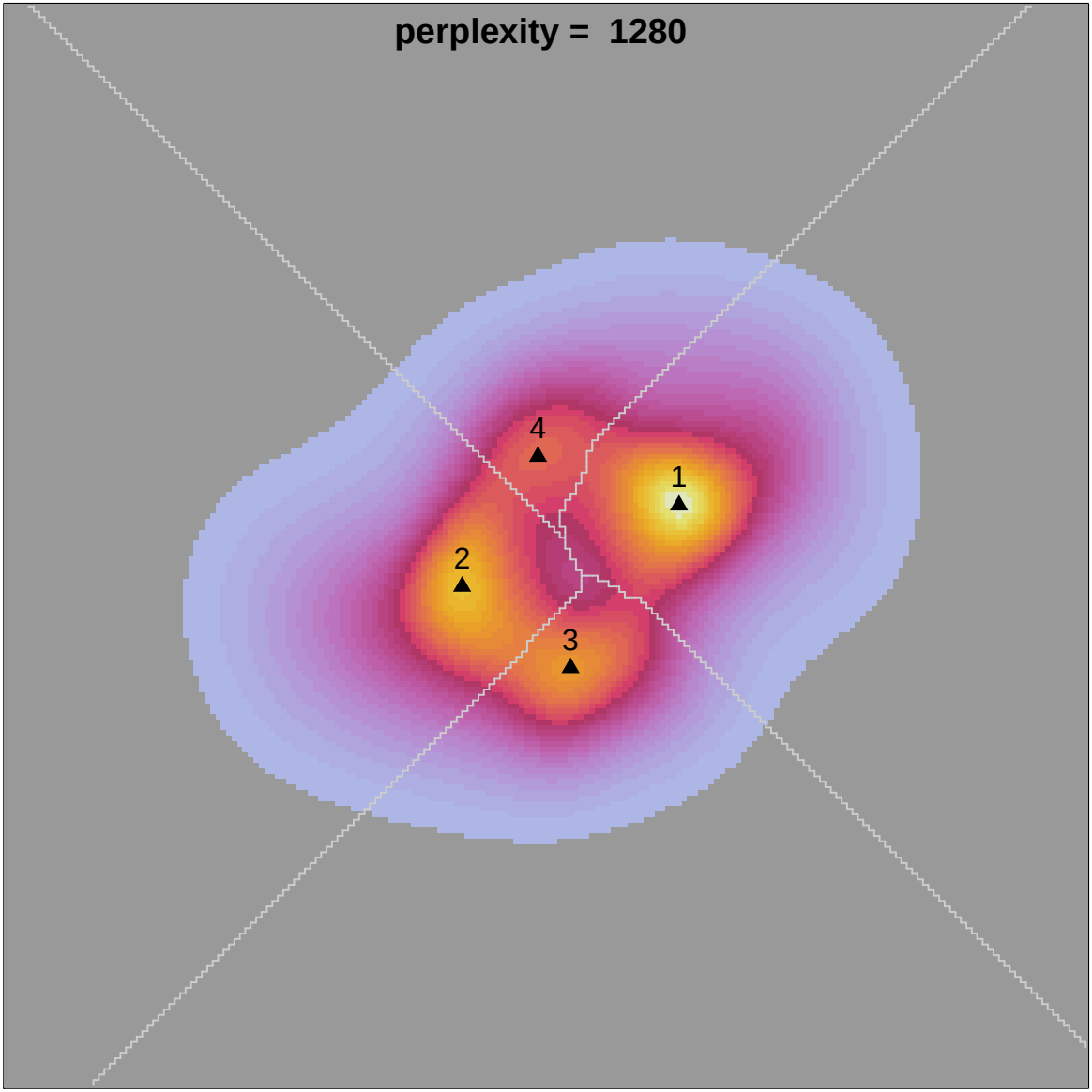

perplexity = 1198
Cost
0.95
0.90
0.85
0.80
0.75
Size
20
15
10
5
0
50
100
150
200
epochs

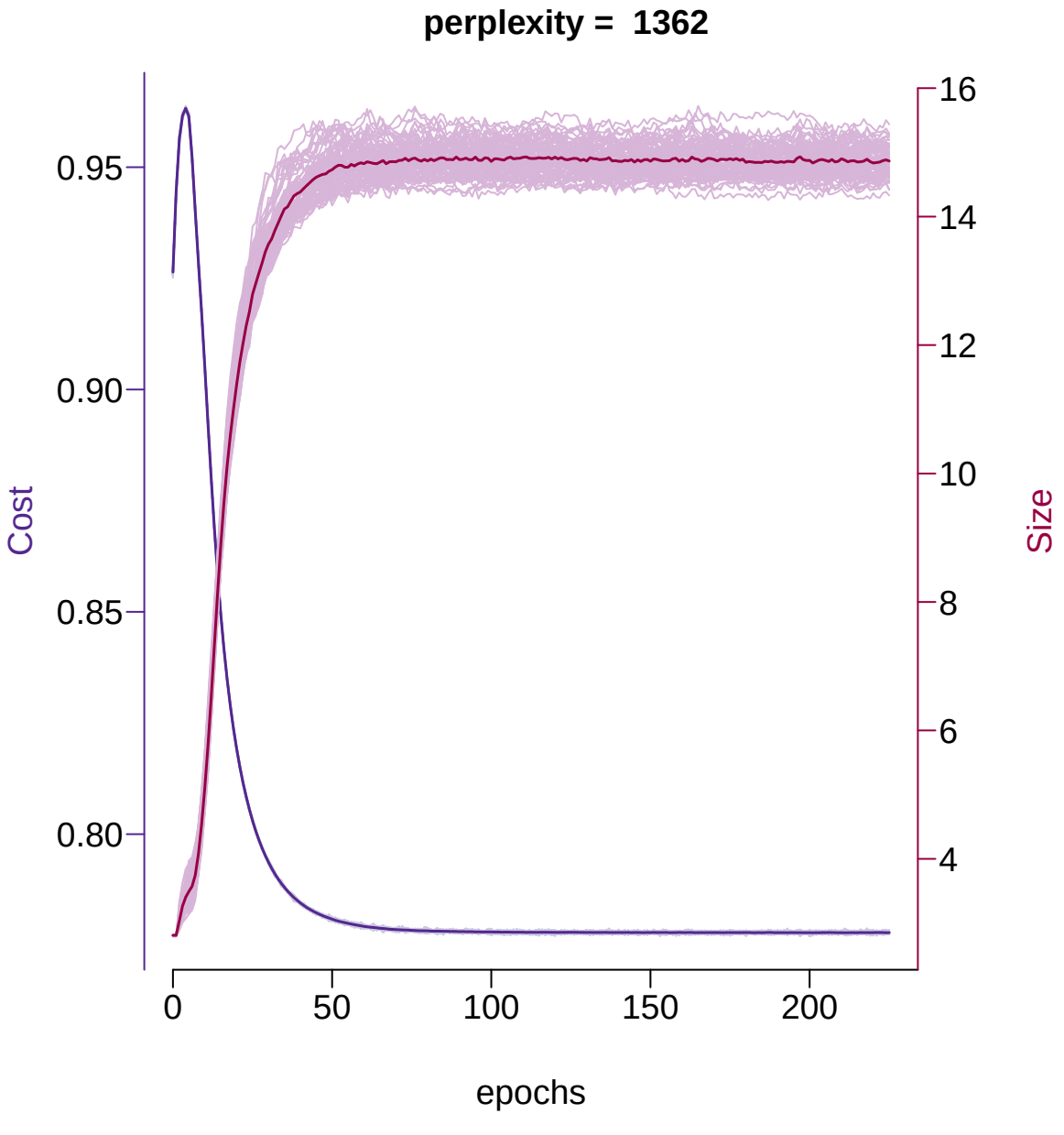

perplexity = 1198
Y2
5
0
−5
−6
−4
−2
0
2
4
Y1

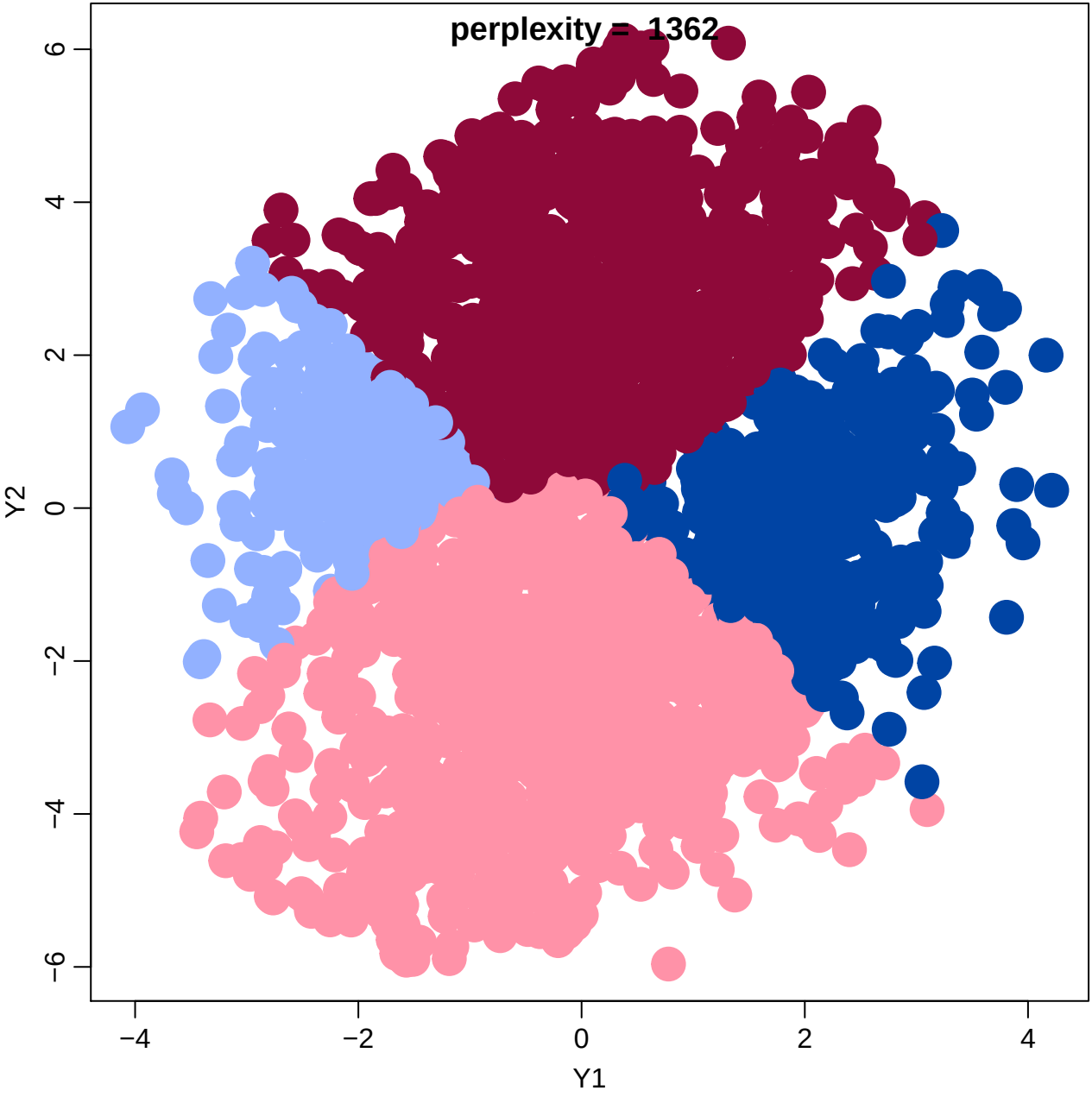

perplexity = 1198
2
3
4
1

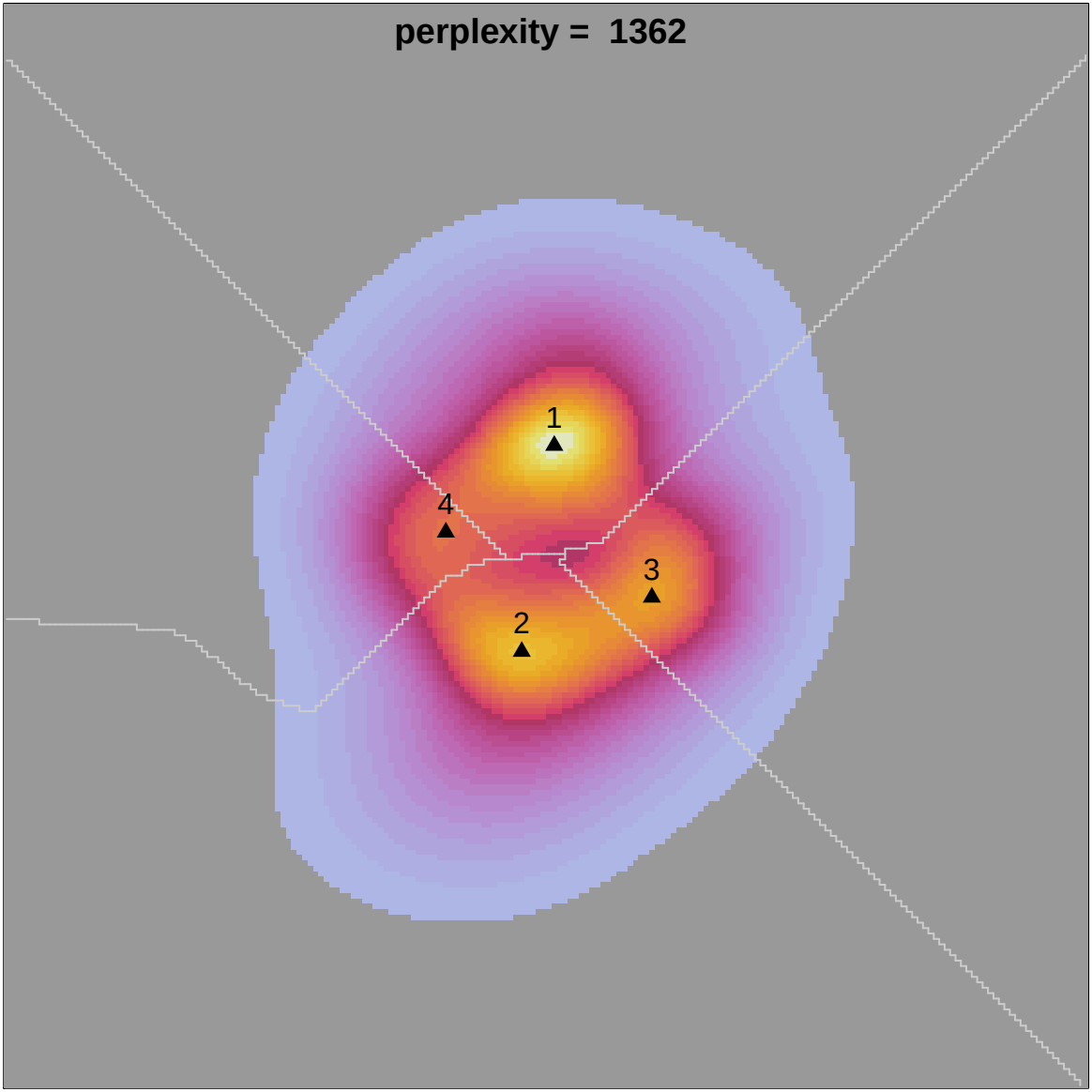

perplexity = 1280
Cost
0.95
0.90
0.85
0.80
Size
15
10
5
0
50
100
150
200
epochs

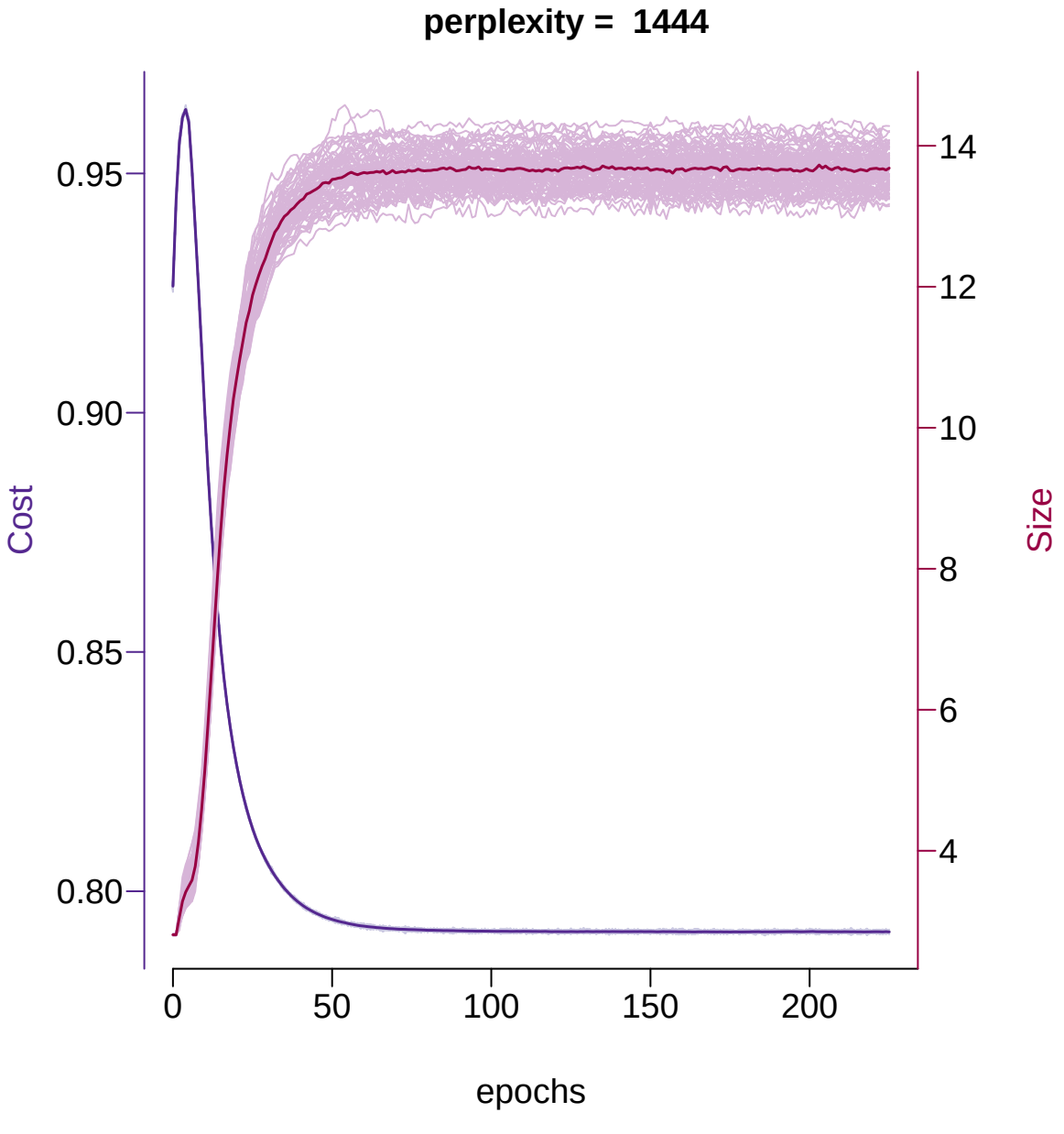

perplexity = 1280
Y2
Y1
4
2
0
−2
−4
−6
−4
−2
0
2
4
6

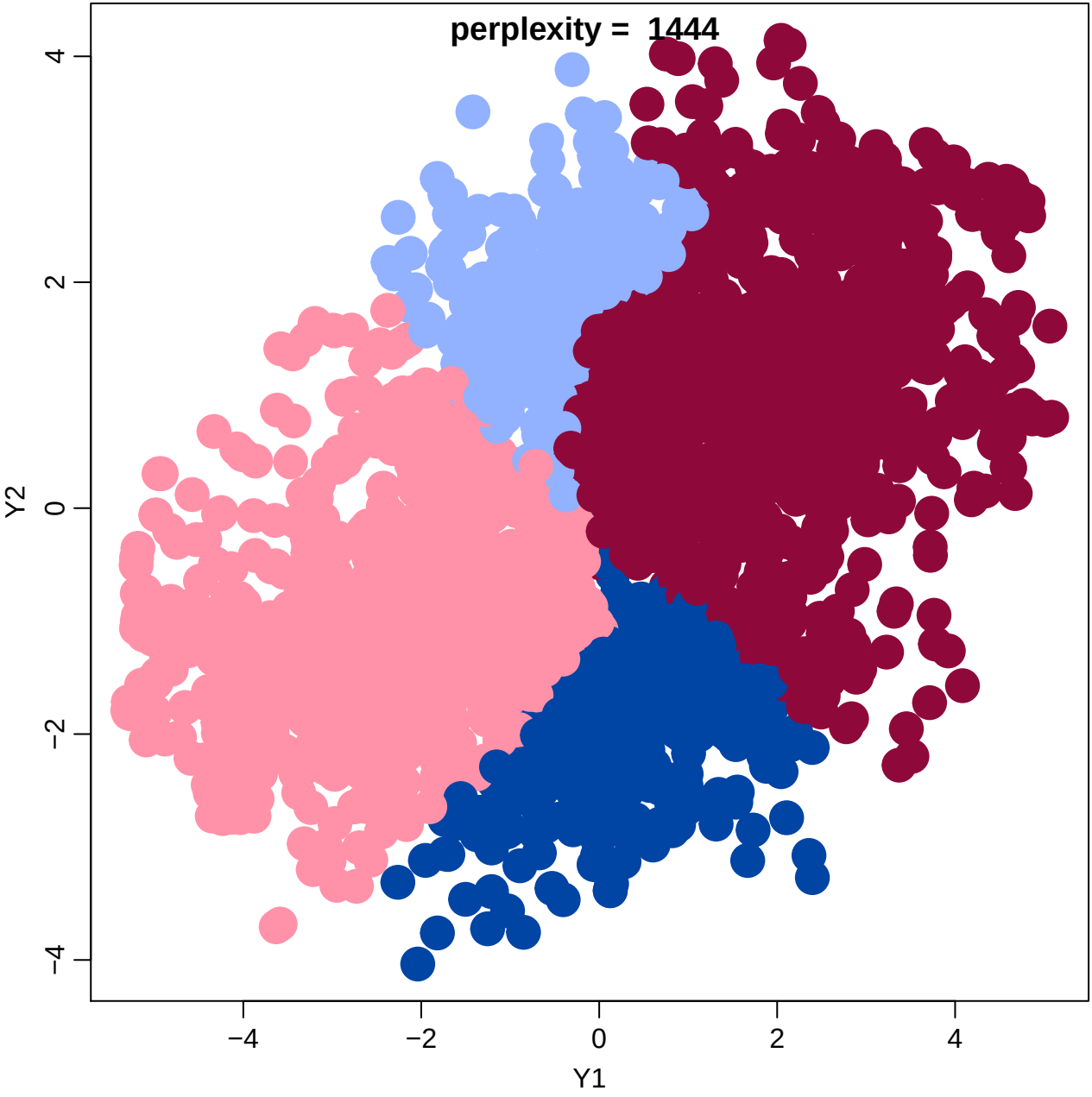

perplexity = 1280
4
1
2
3

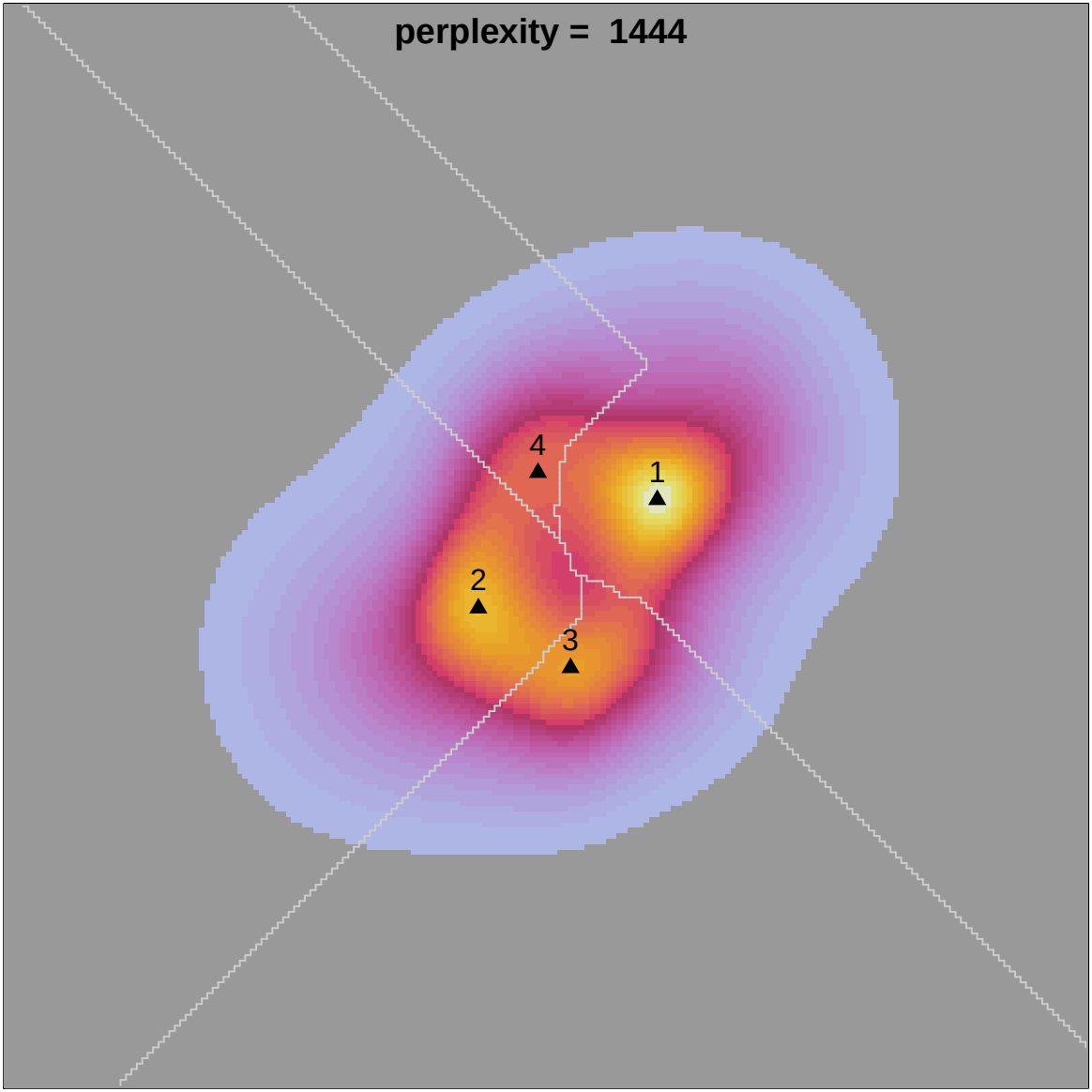

perplexity = 1362
Cost
0.95
0.90
0.85
0.80
Size
16
14
12
10
8
6
4
0
50
100
150
200
epochs

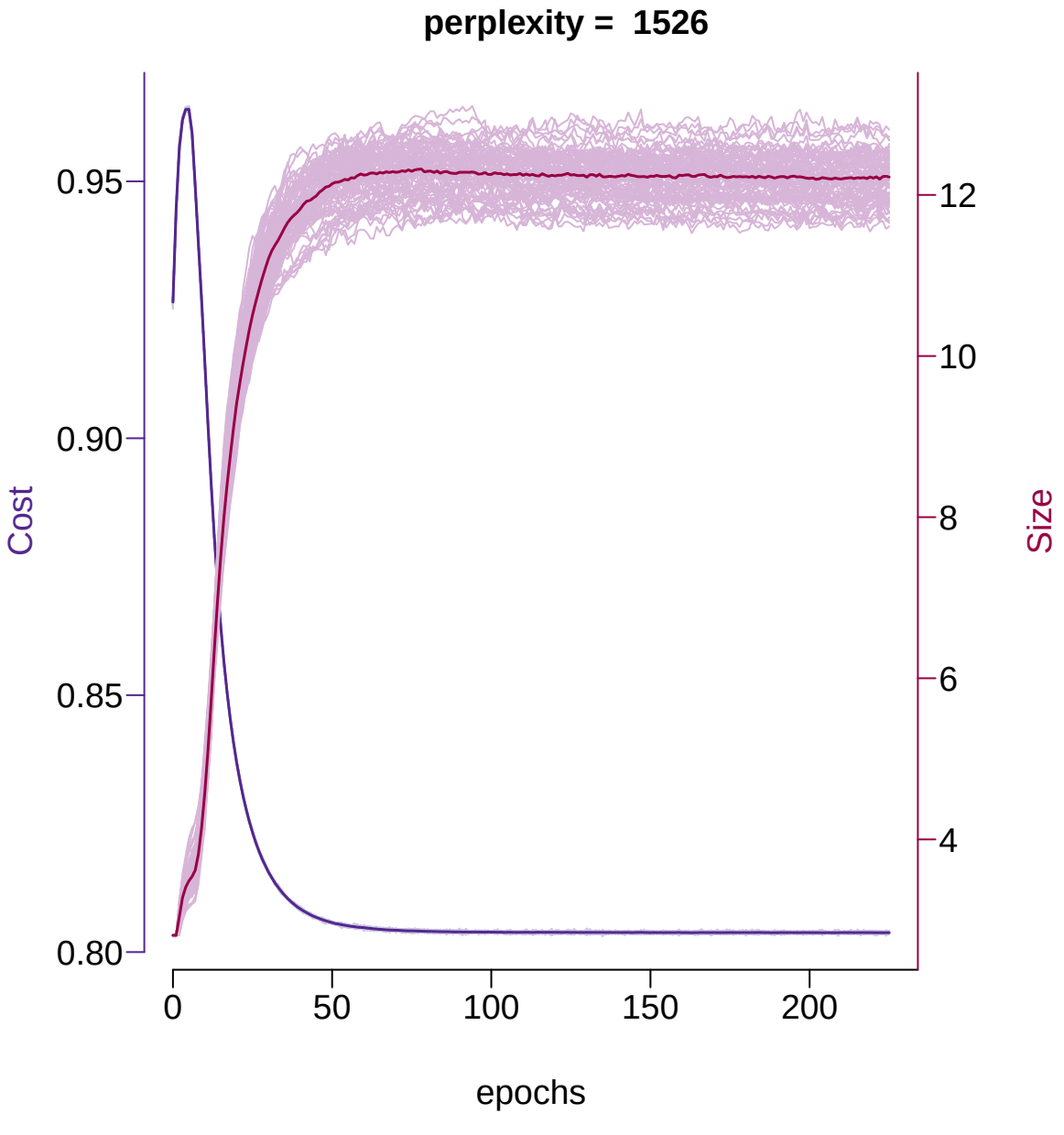

perplexity = 1362
Y2
Y1
6
4
2
0
−2
−4
−6
−4
−2
0
2
4

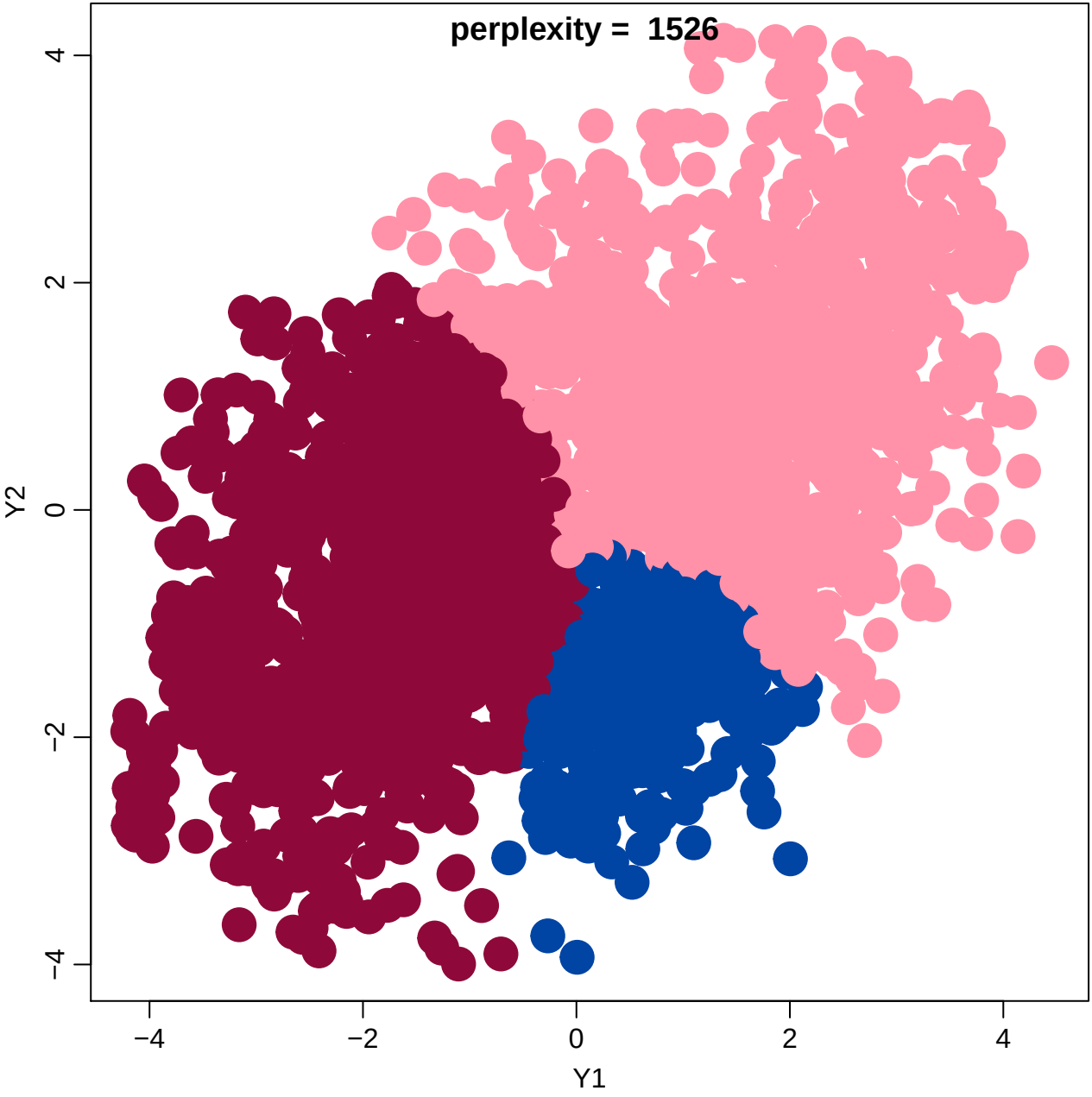

perplexity = 1362
1
2
3
4

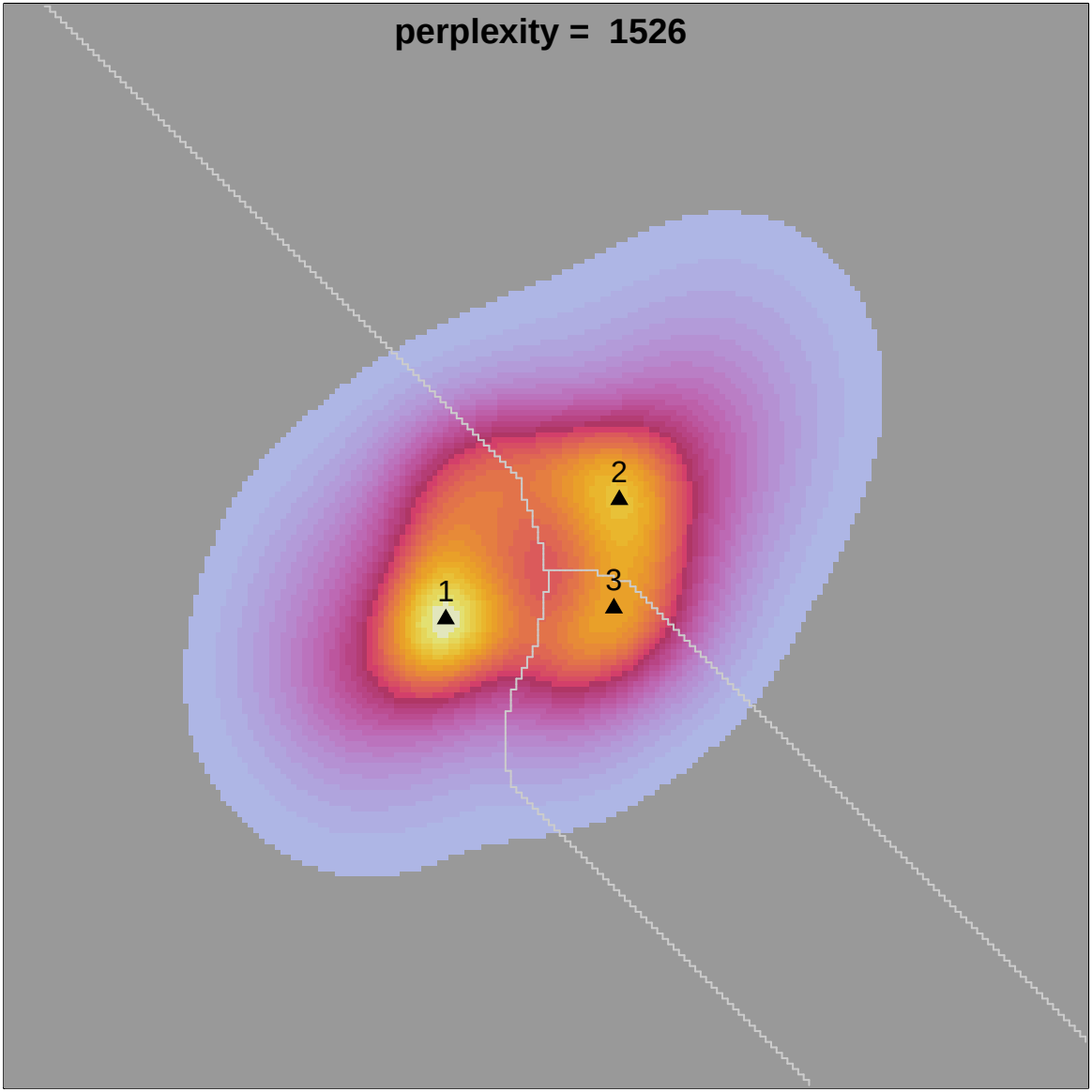

perplexity = 1444
Cost
0.95
0.90
0.85
0.80
Size
14
12
10
8
6
4
0
50
100
150
200
epochs

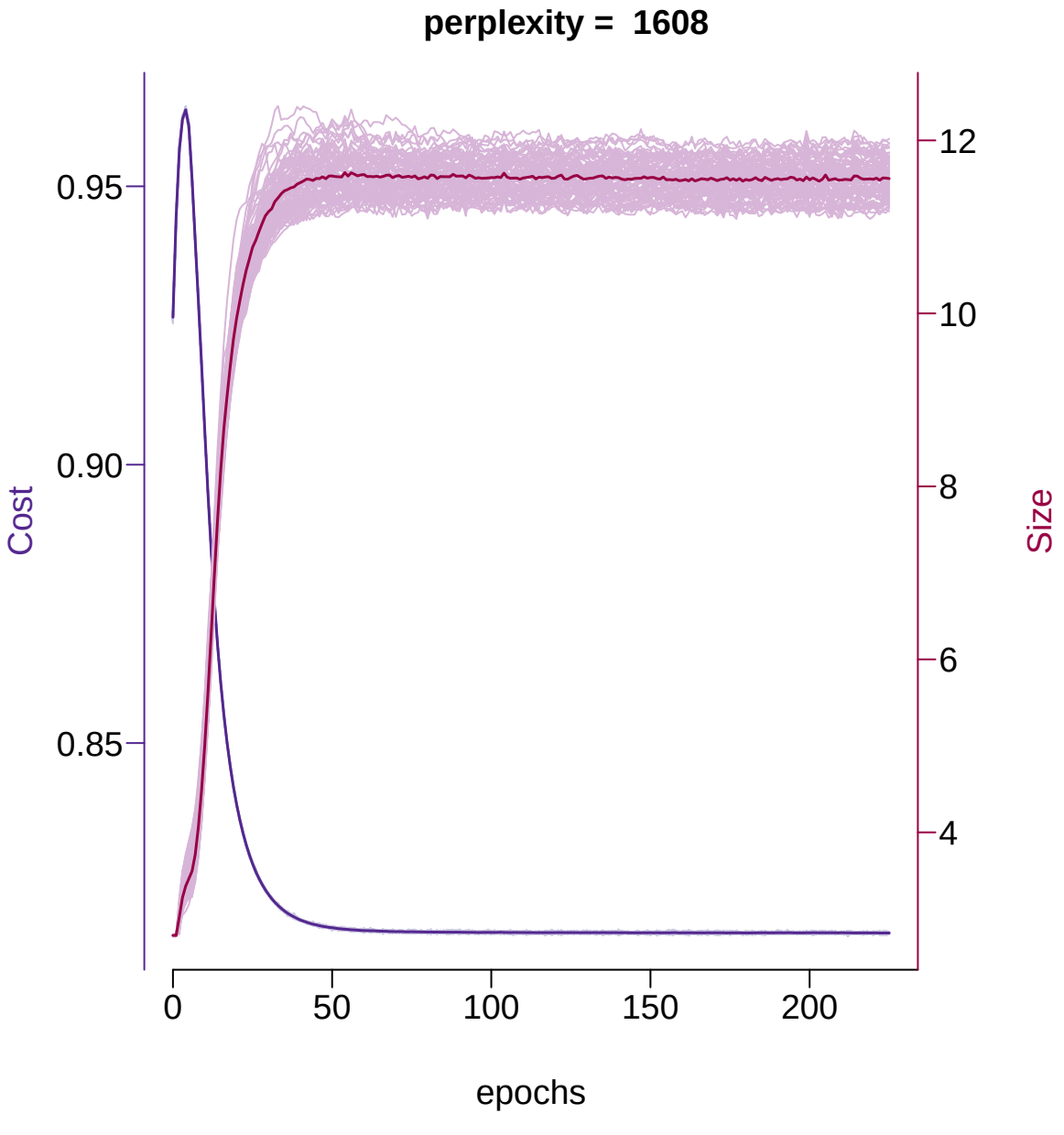

perplexity = 1444
Y2
Y1
4
2
0
−2
−4
−4
−2
0
2
4

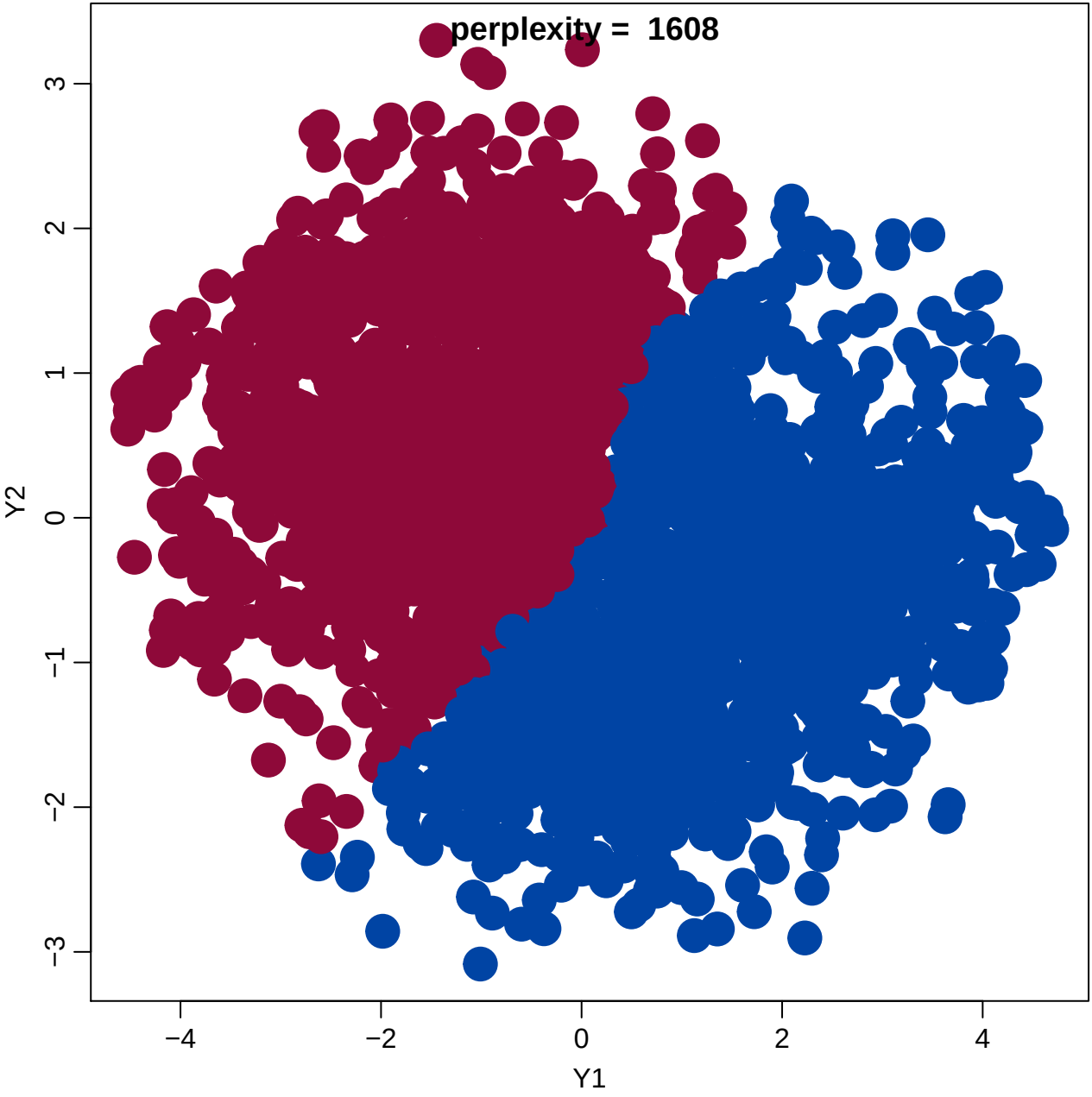

perplexity = 1444
4
1
2
3

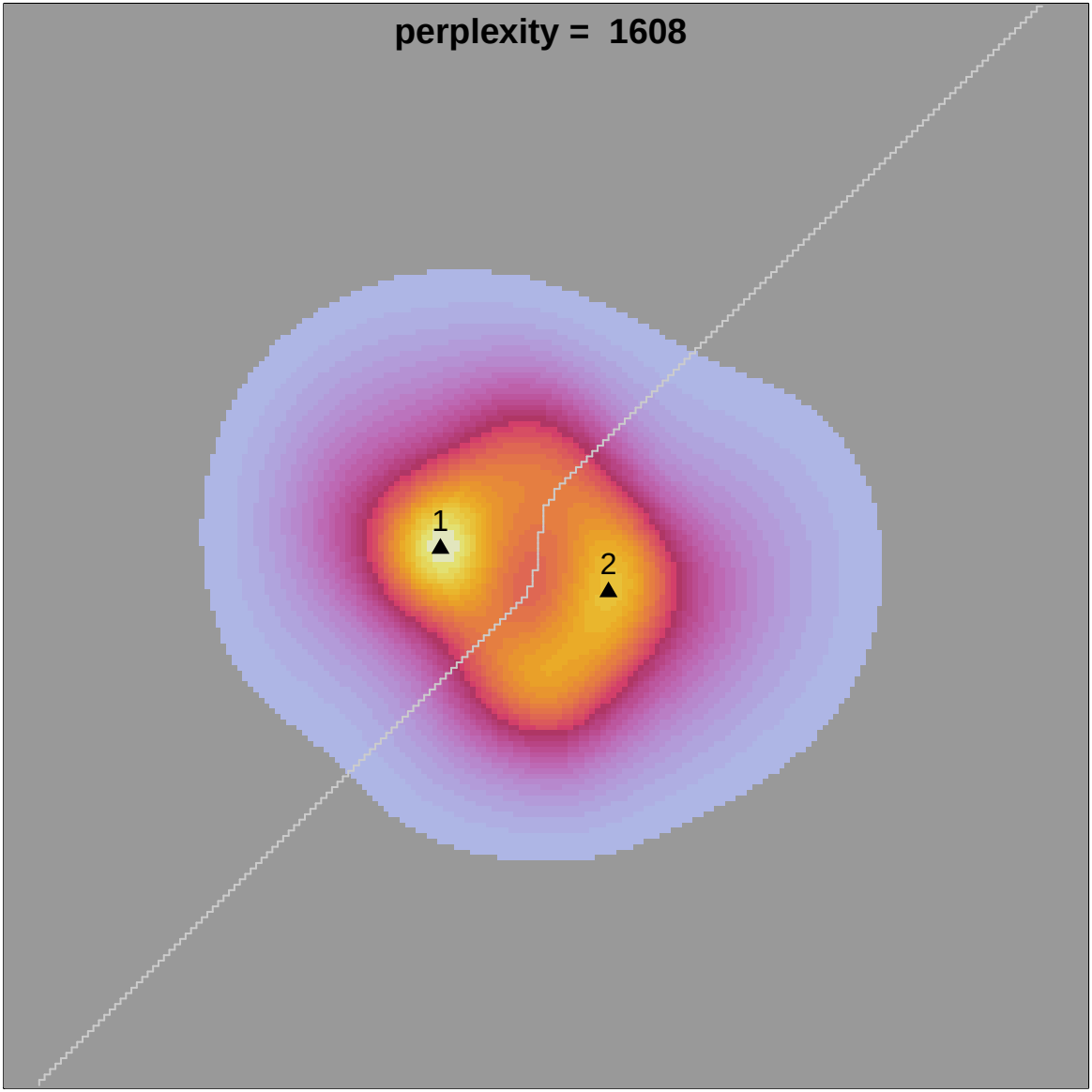

perplexity = 1526
Cost
0.95
0.90
0.85
0.80
Size
12
10
8
6
4
0
50
100
150
200
epochs

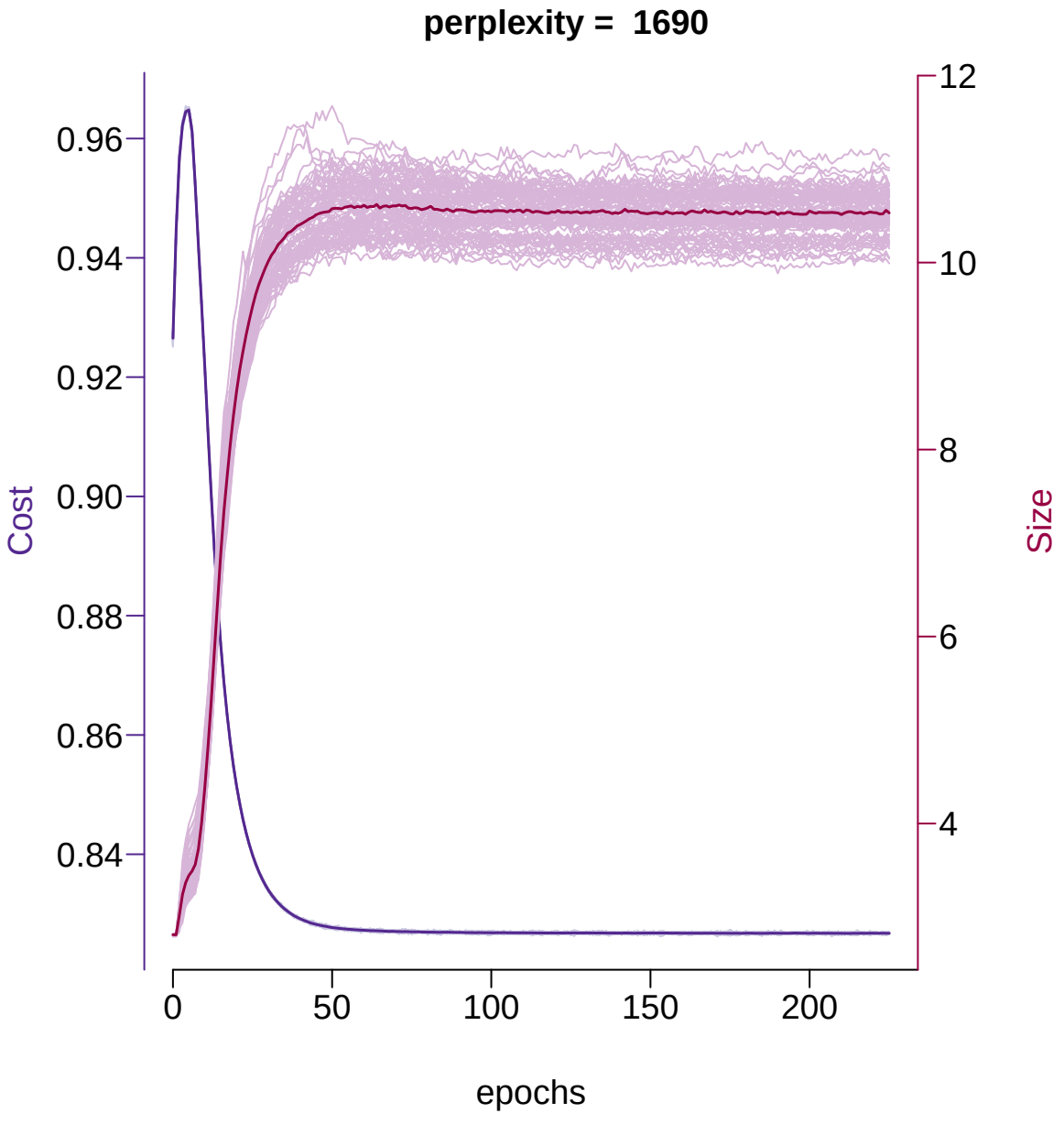

perplexity = 1526
Y2
Y1
4
2
0
−2
−4
−4
−2
0
2
4

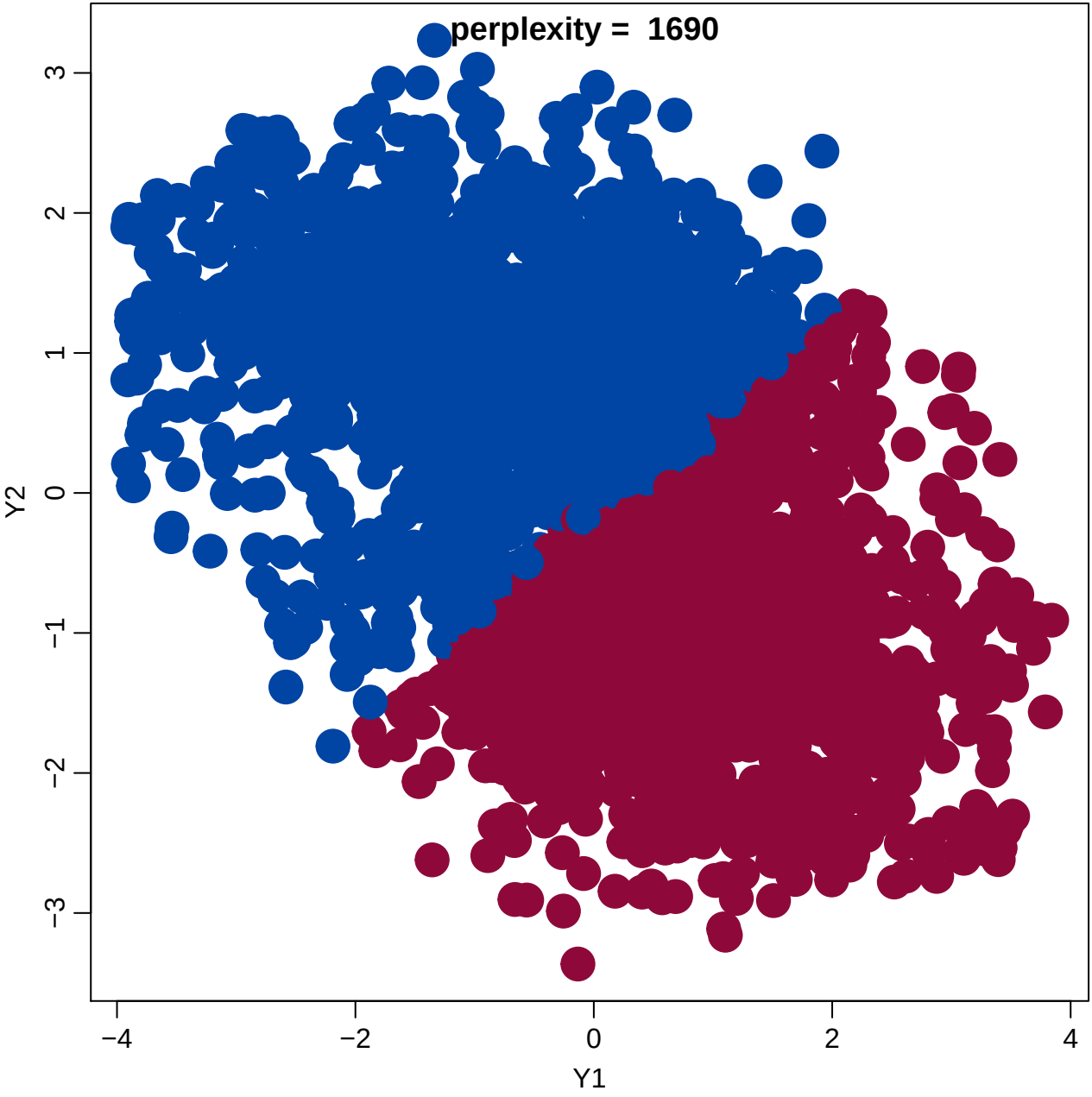

perplexity = 1526
1
2
3

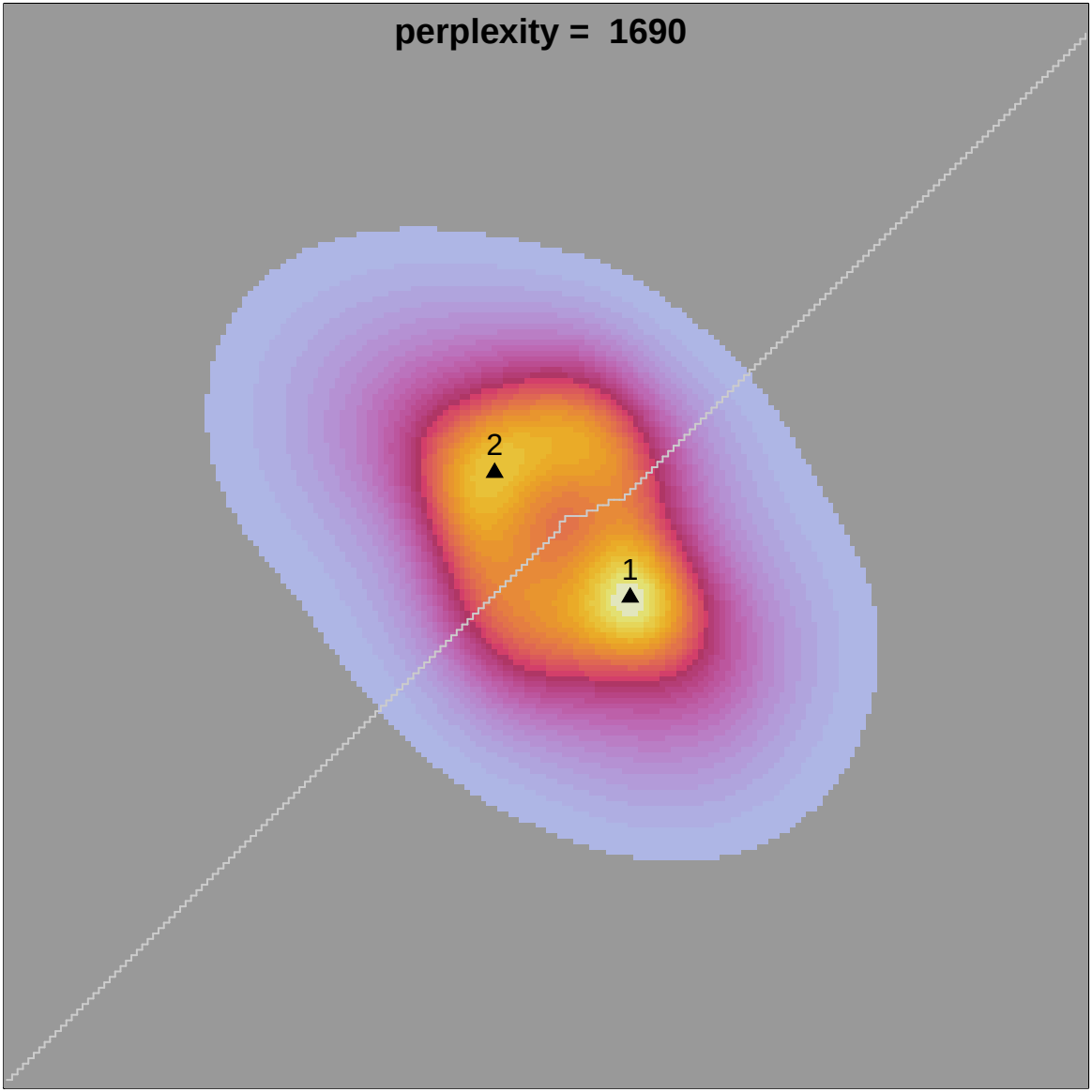

perplexity = 1608
Cost
0.95
0.90
0.85
Size
12
10
8
6
4
0
50
100
150
200
epochs

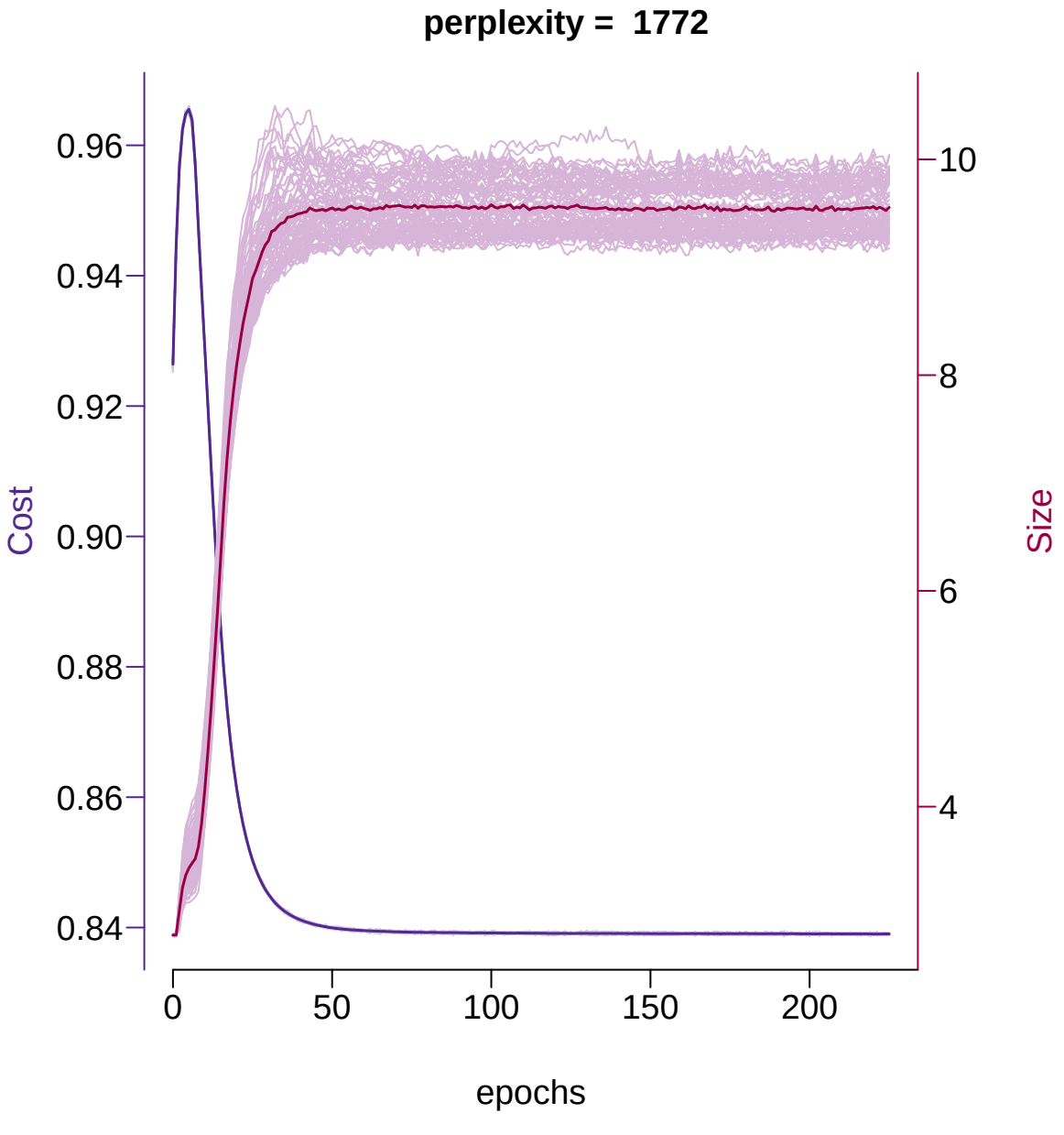

perplexity = 1608
Y2
Y1
3
2
1
0
−1
−2
−3
−4
−2
0
2
4

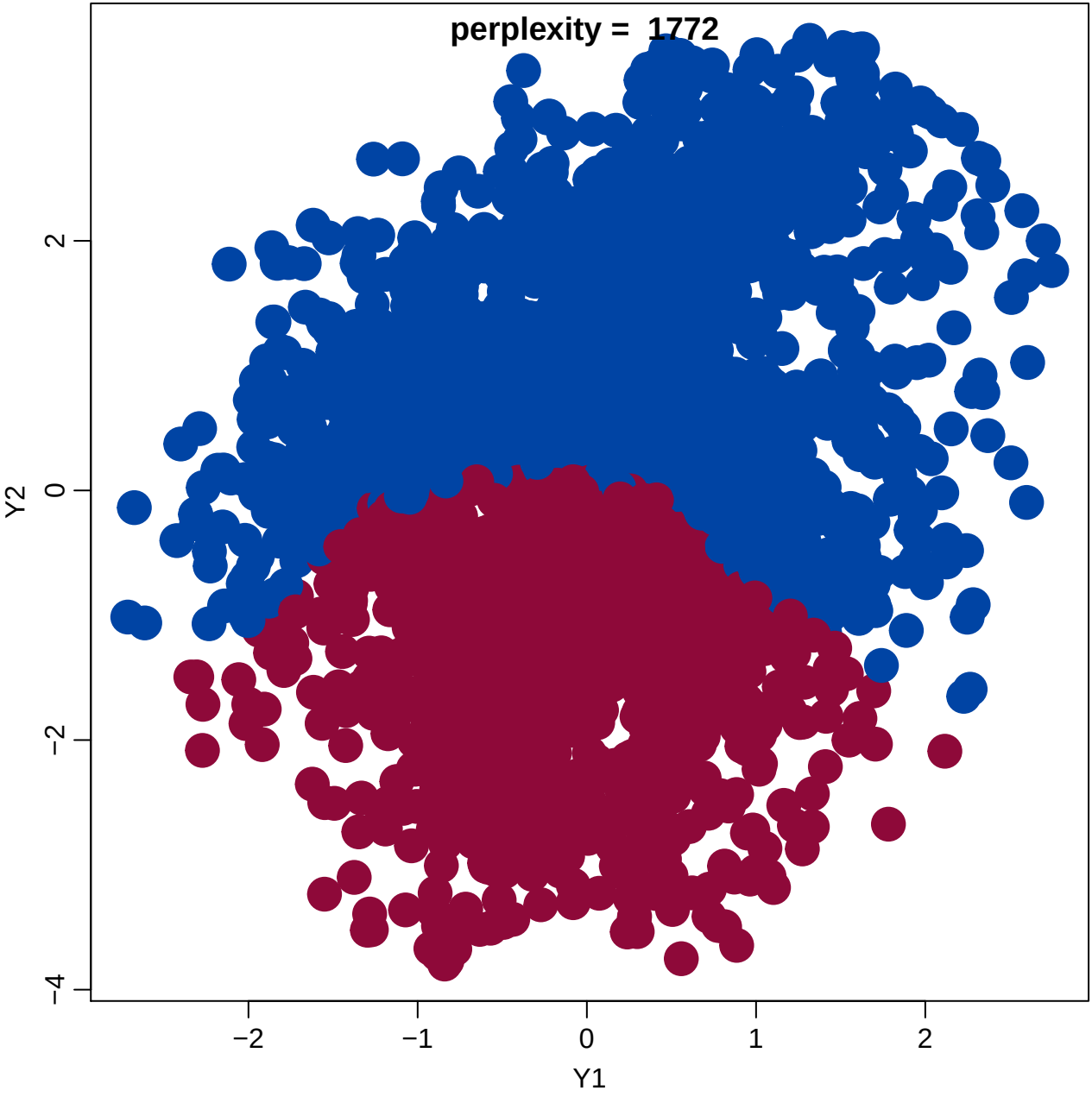

perplexity = 1608
1
2

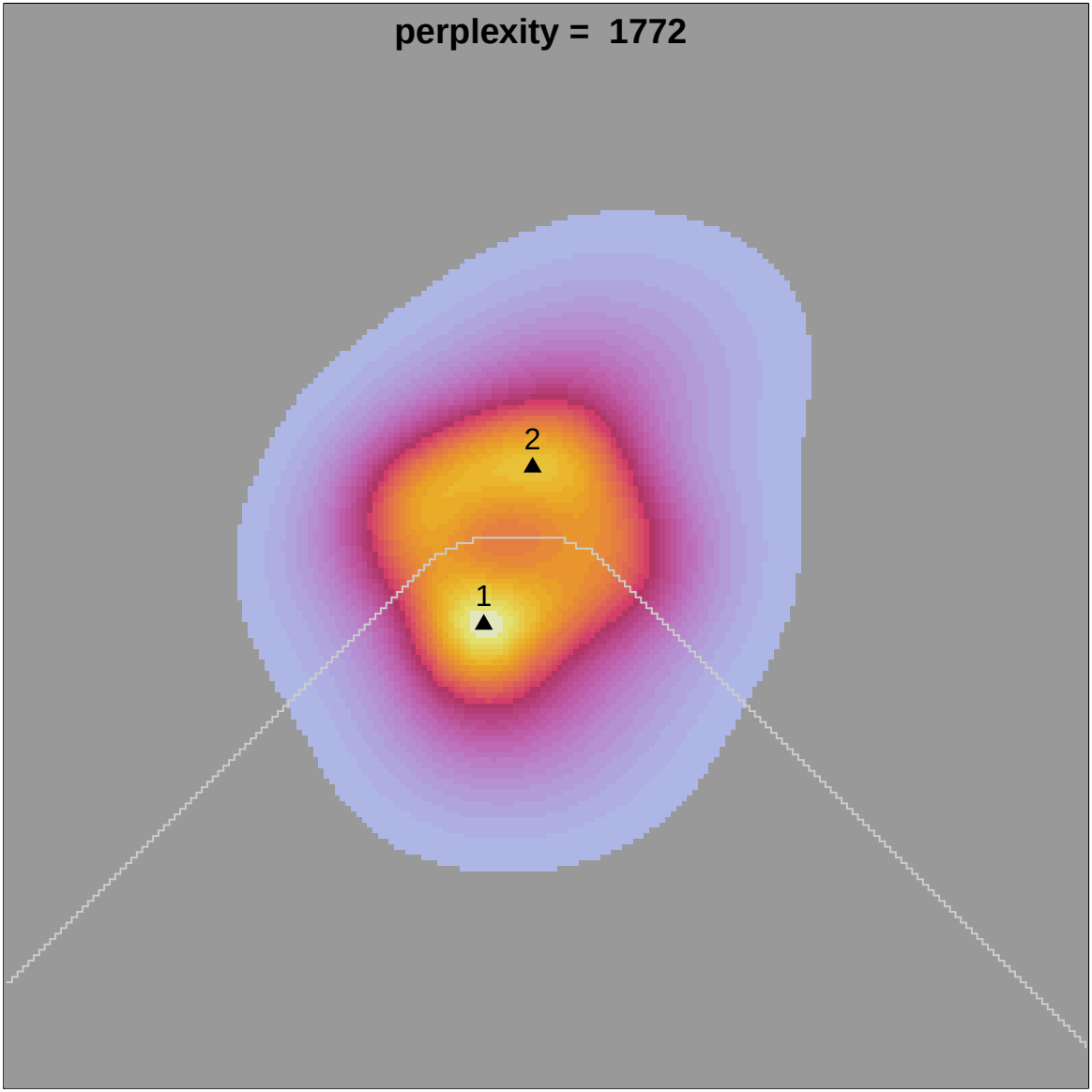

perplexity = 1690
Cost
0.96
0.94
0.92
0.90
0.88
0.86
0.84
Size
12
10
8
6
4
0
50
100
150
200
epochs

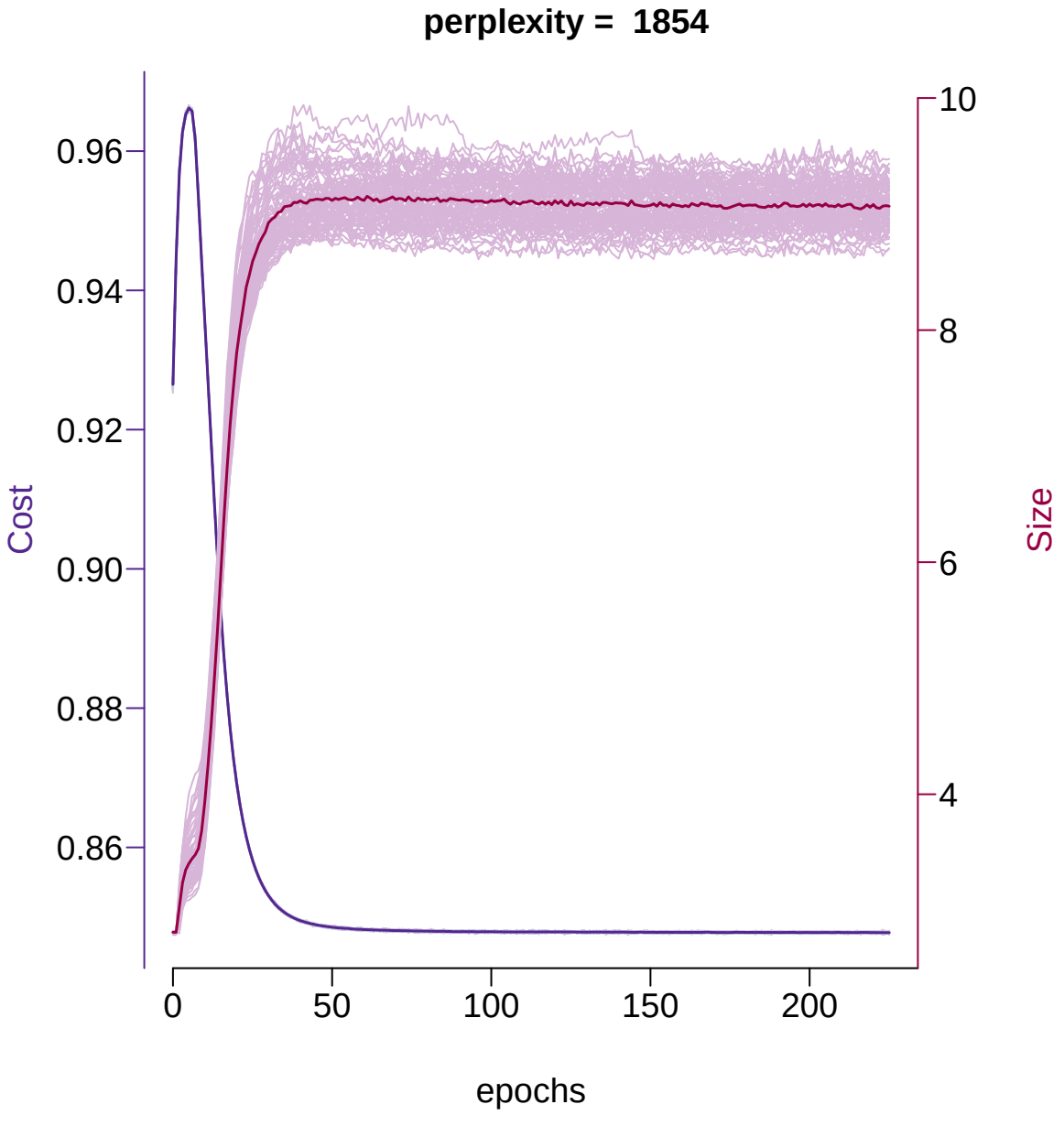

perplexity = 1690
Y2
Y1
3
2
1
0
−1
−2
−3
−4
−2
0
2
4

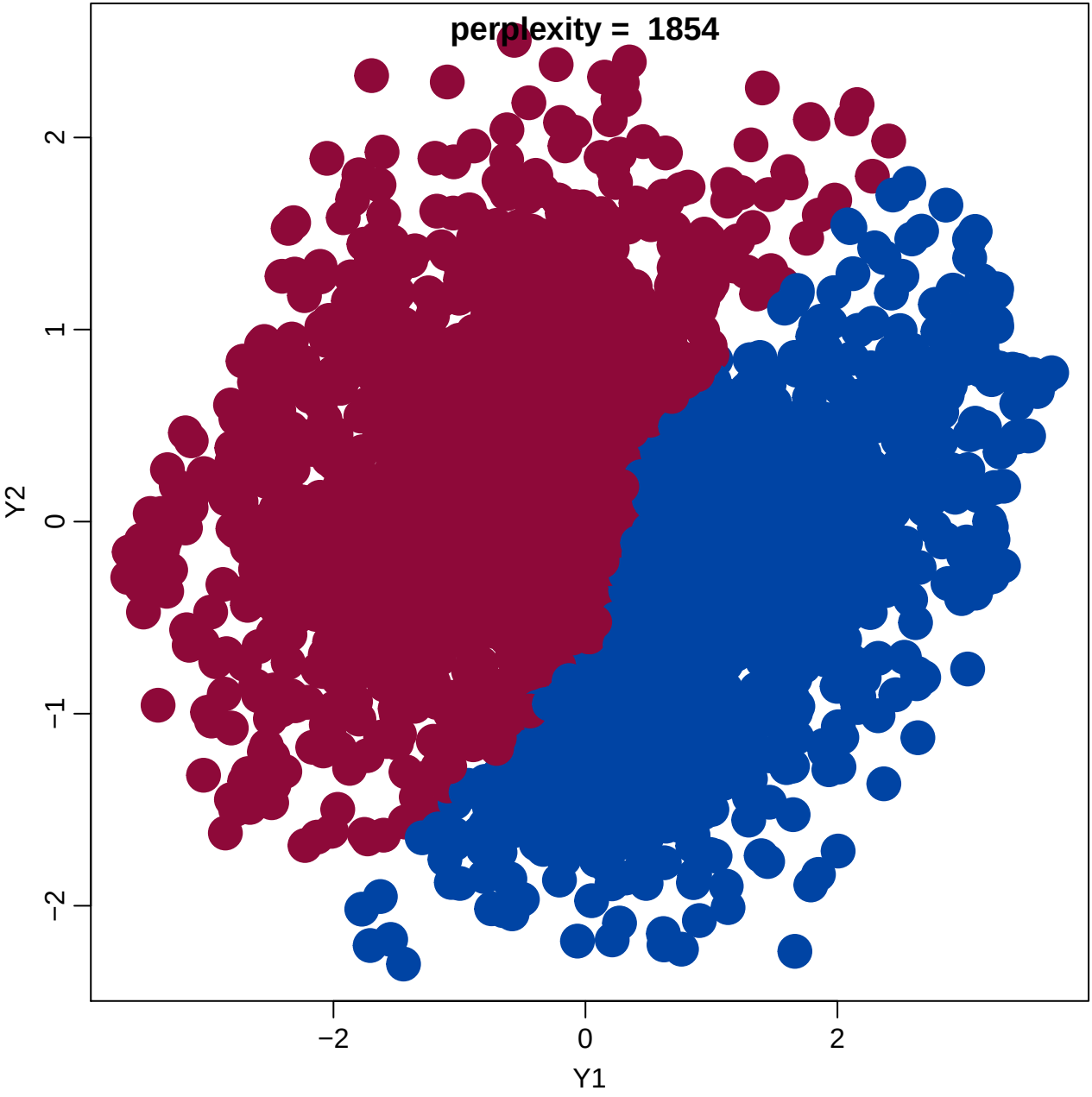

perplexity = 1690
2
1

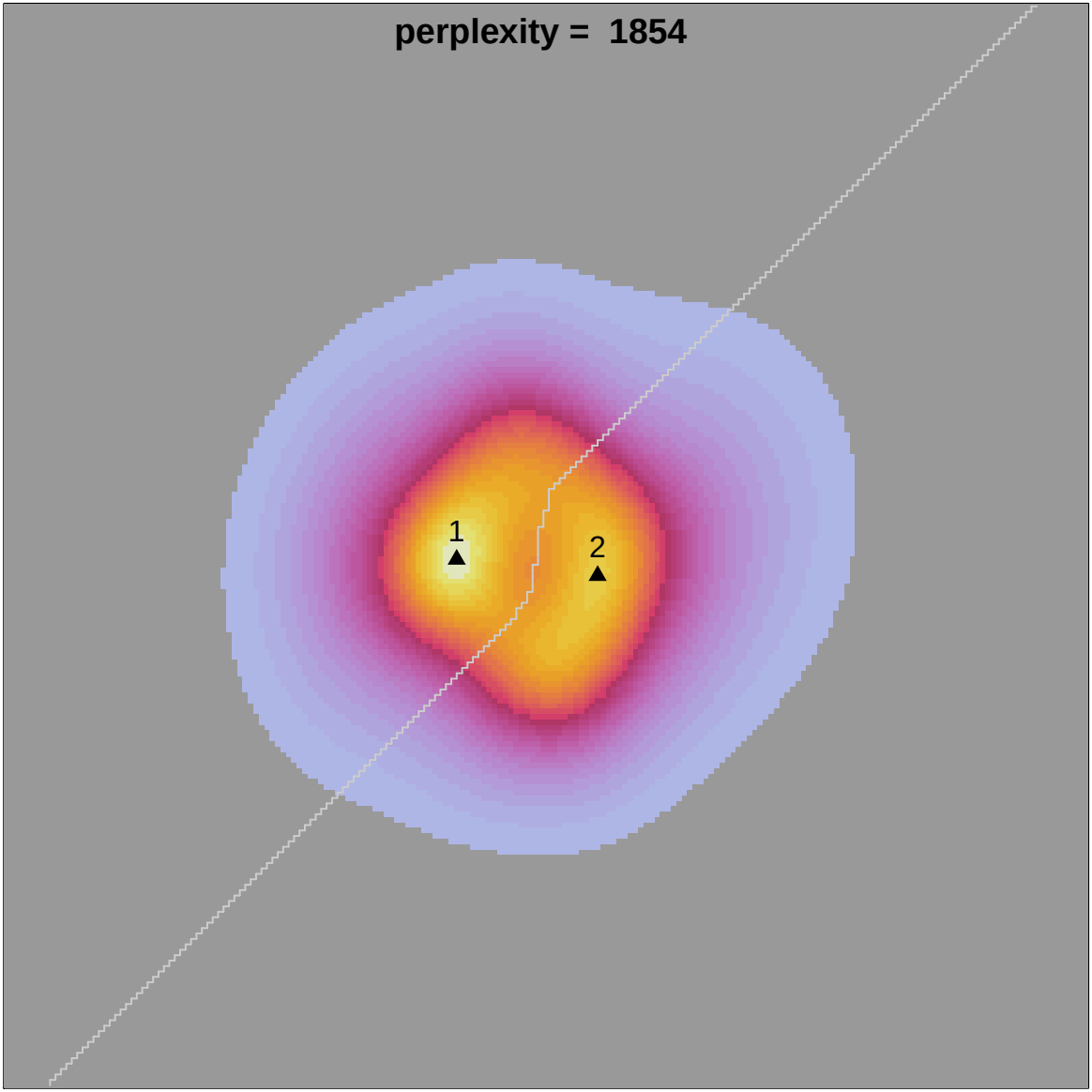

perplexity = 1772
Cost
Size
epochs
0.96
0.94
0.92
0.90
0.88
0.86
0.84
10
8
6
4
0
50
100
150
200

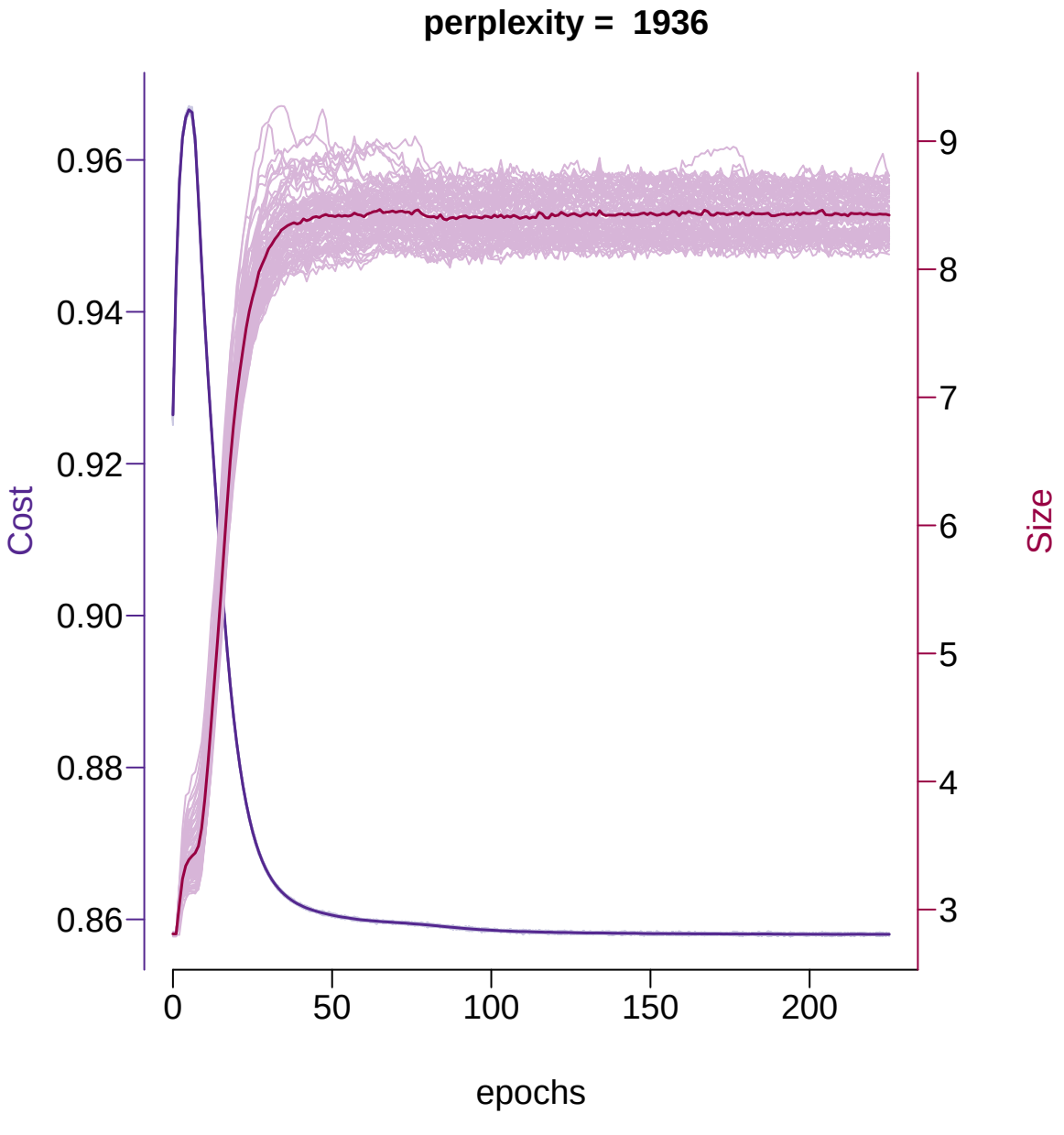

perplexity = 1772
Y2
Y1
2
0
−2
−4
−2
−1
0
1
2

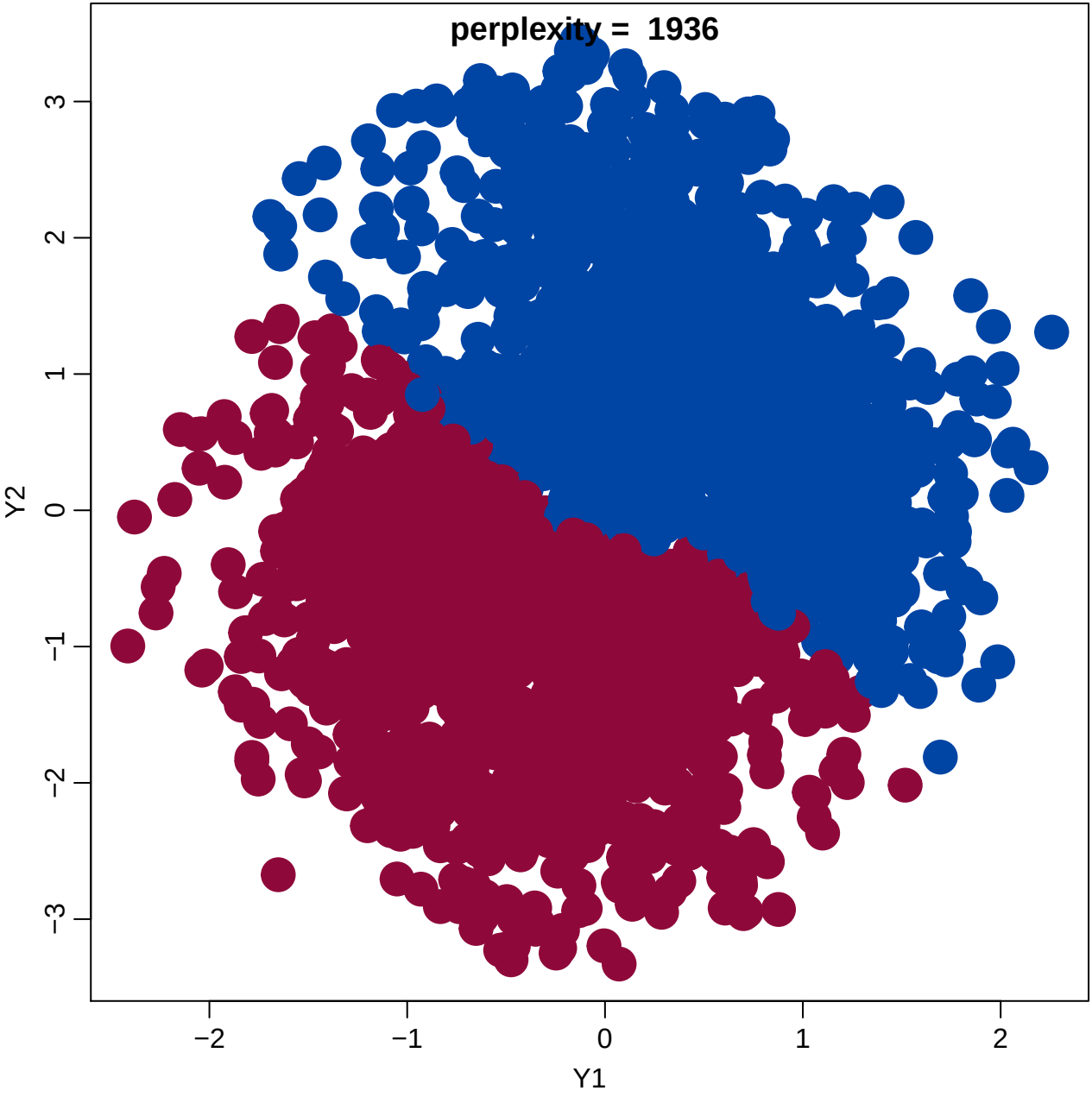

perplexity = 1772
2
1

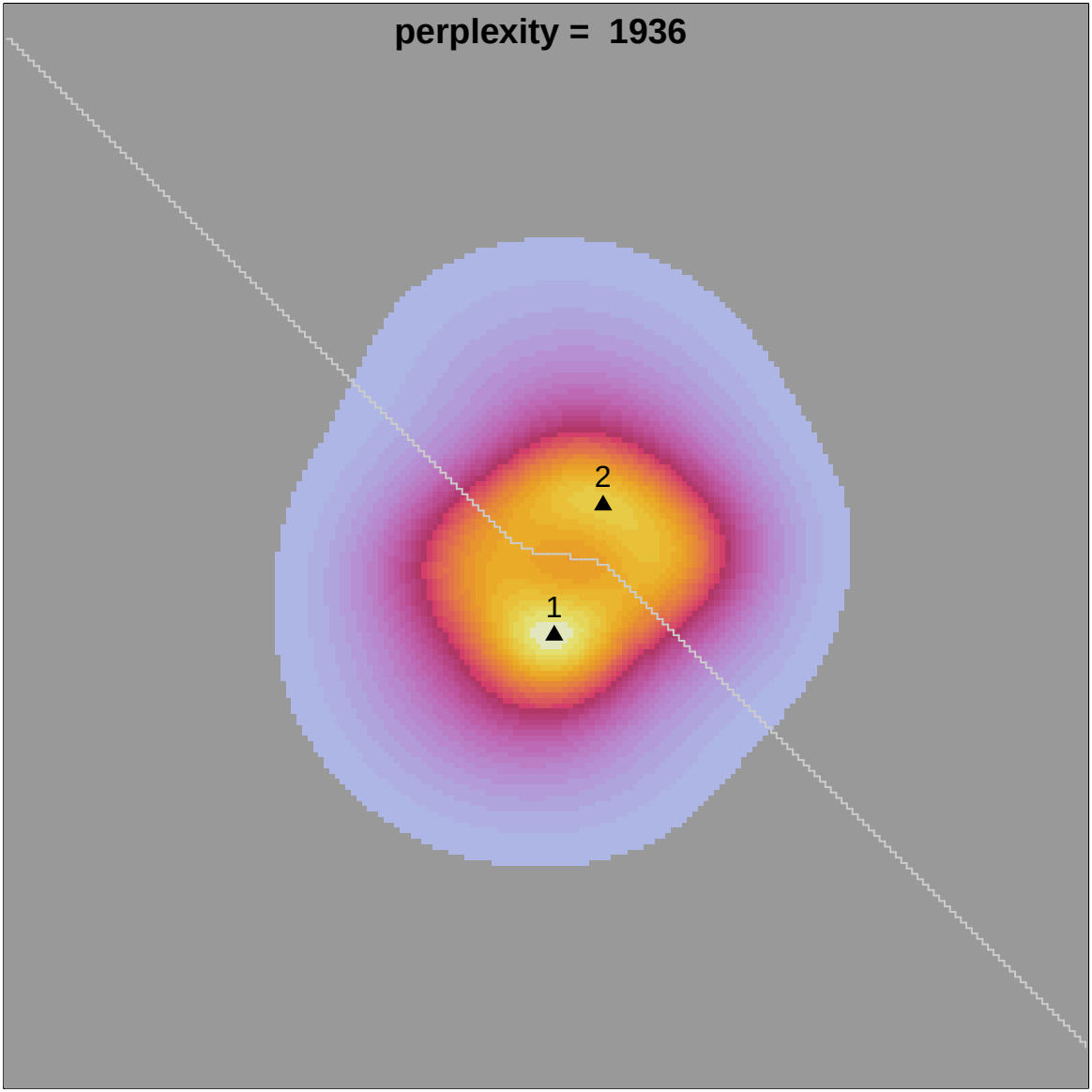

perplexity = 1854
Cost
0.96
0.94
0.92
0.90
0.88
0.86
Size
10
8
6
4
0
50
100
150
200
epochs

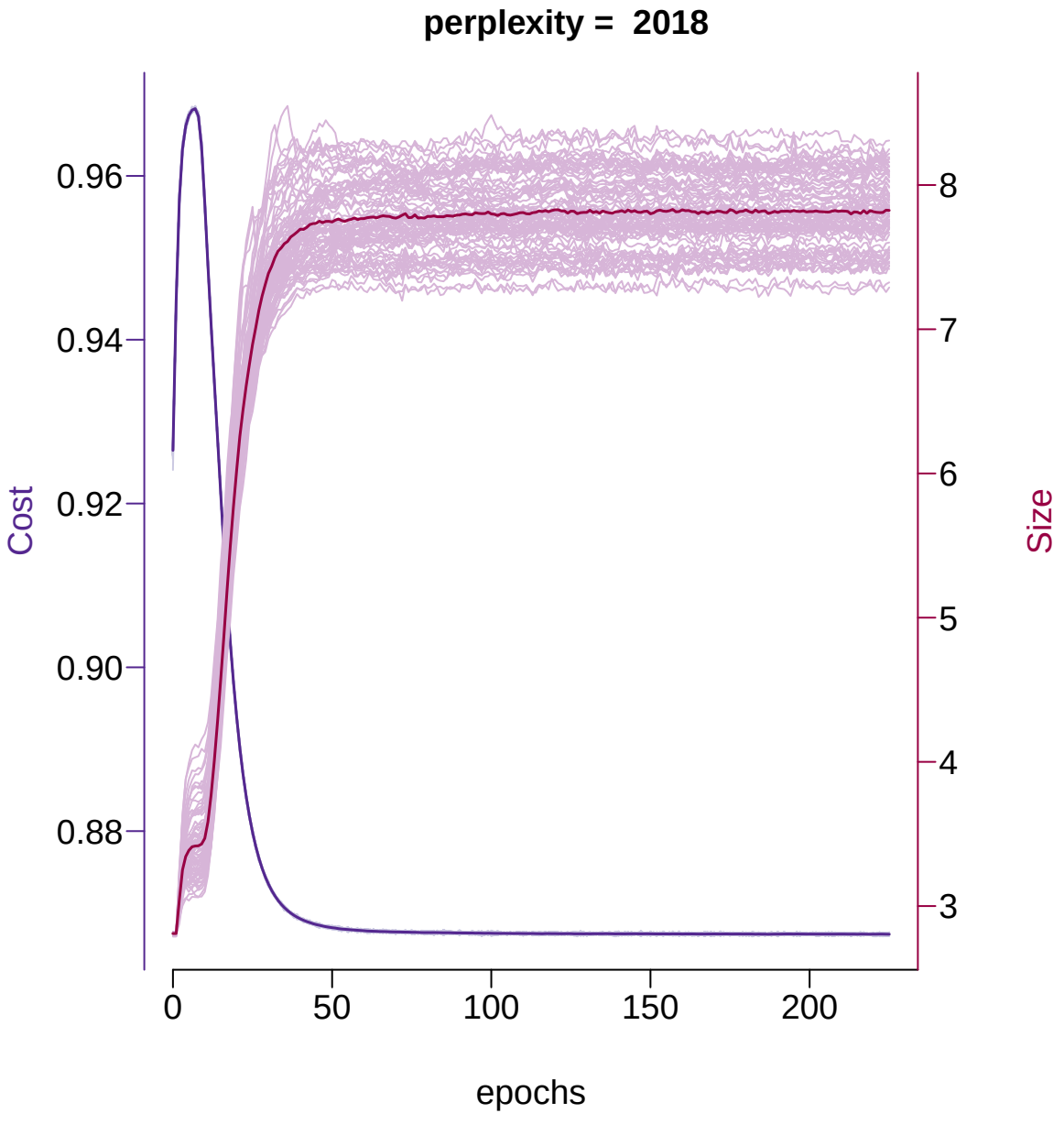

perplexity = 1854
Y2
Y1
2
1
0
−1
−2
−2
0
2

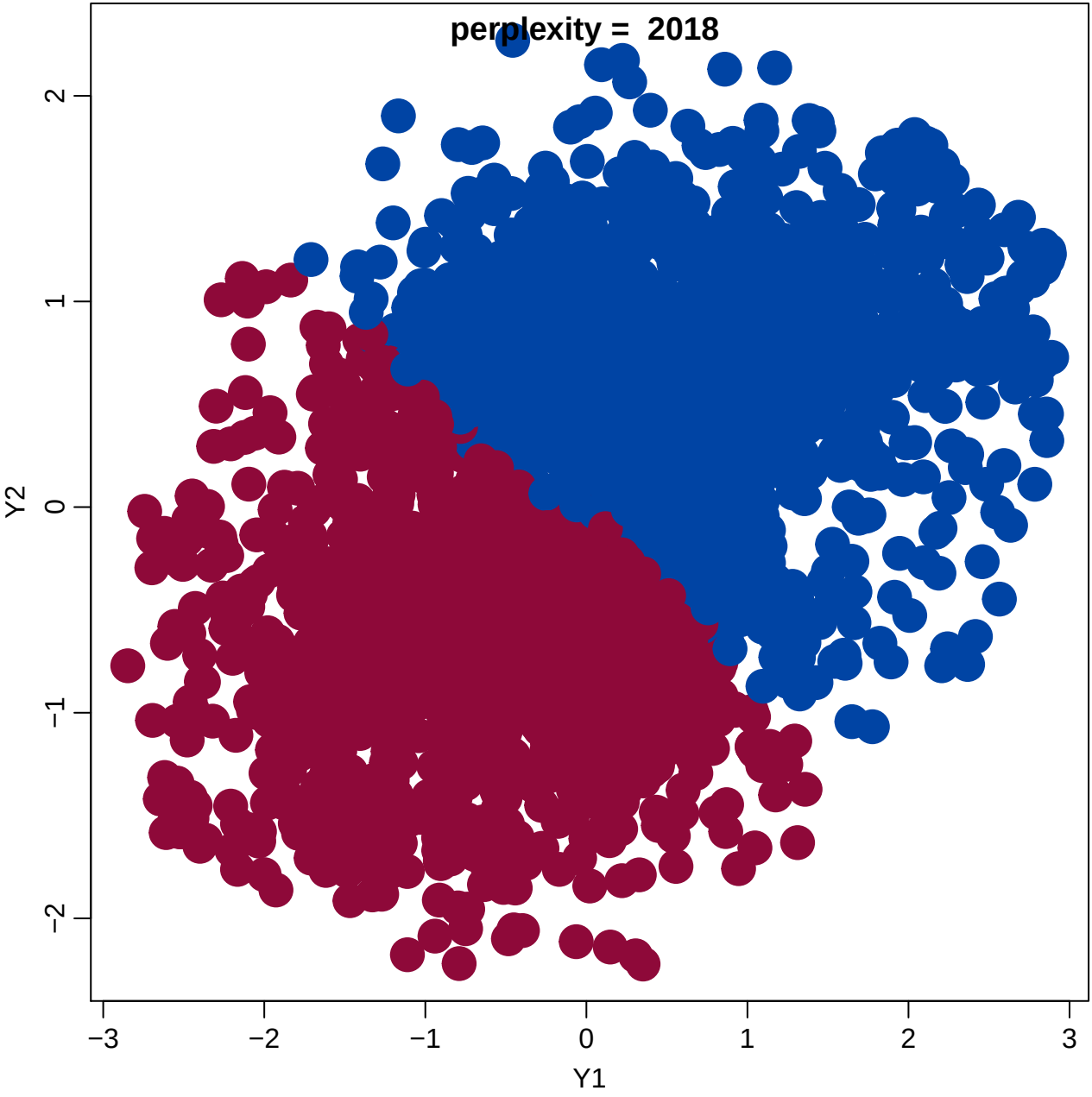

perplexity = 1854
1
2

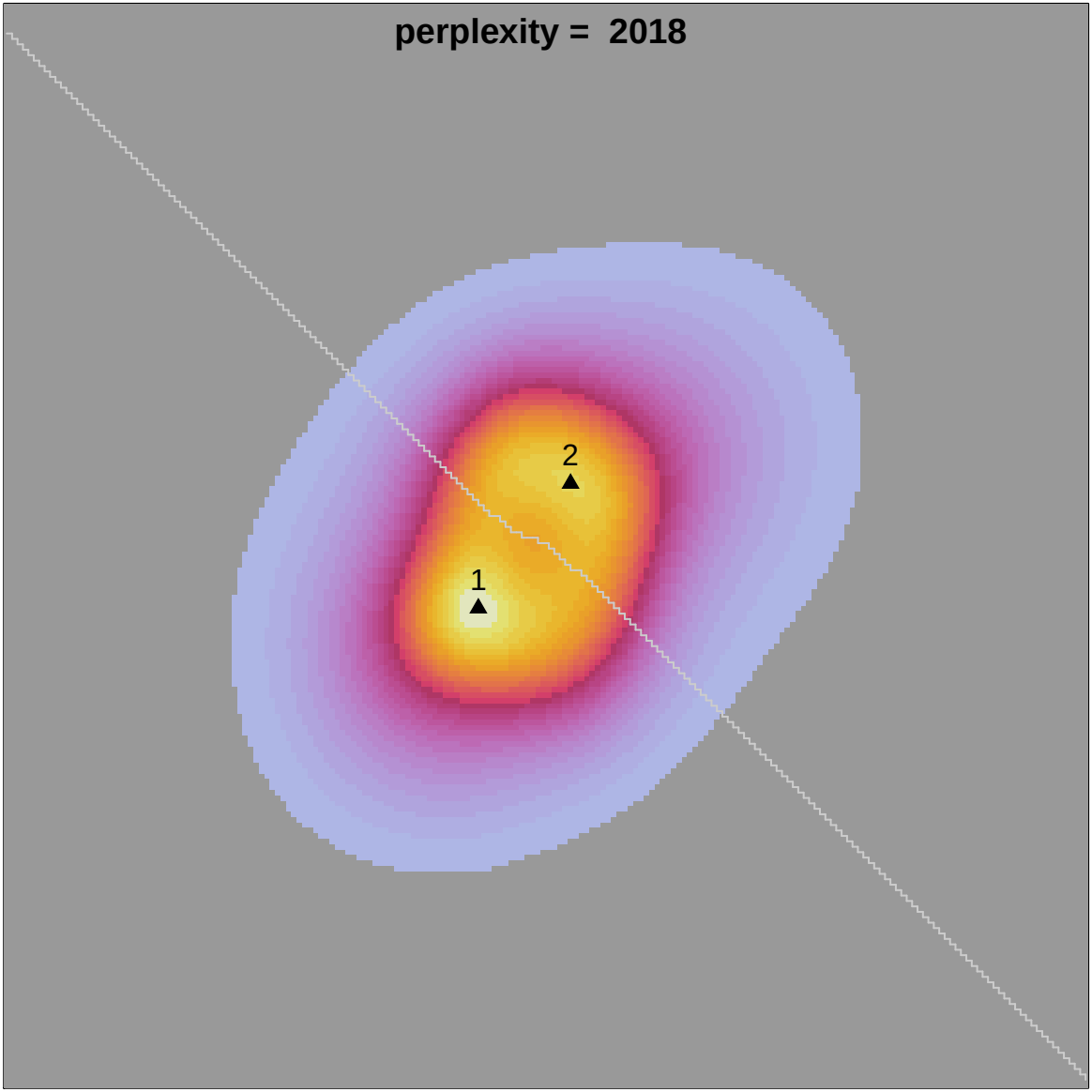

perplexity = 1936
Cost
0.96
0.94
0.92
0.90
0.88
0.86
Size
9
8
7
6
5
4
3
0
50
100
150
200
epochs

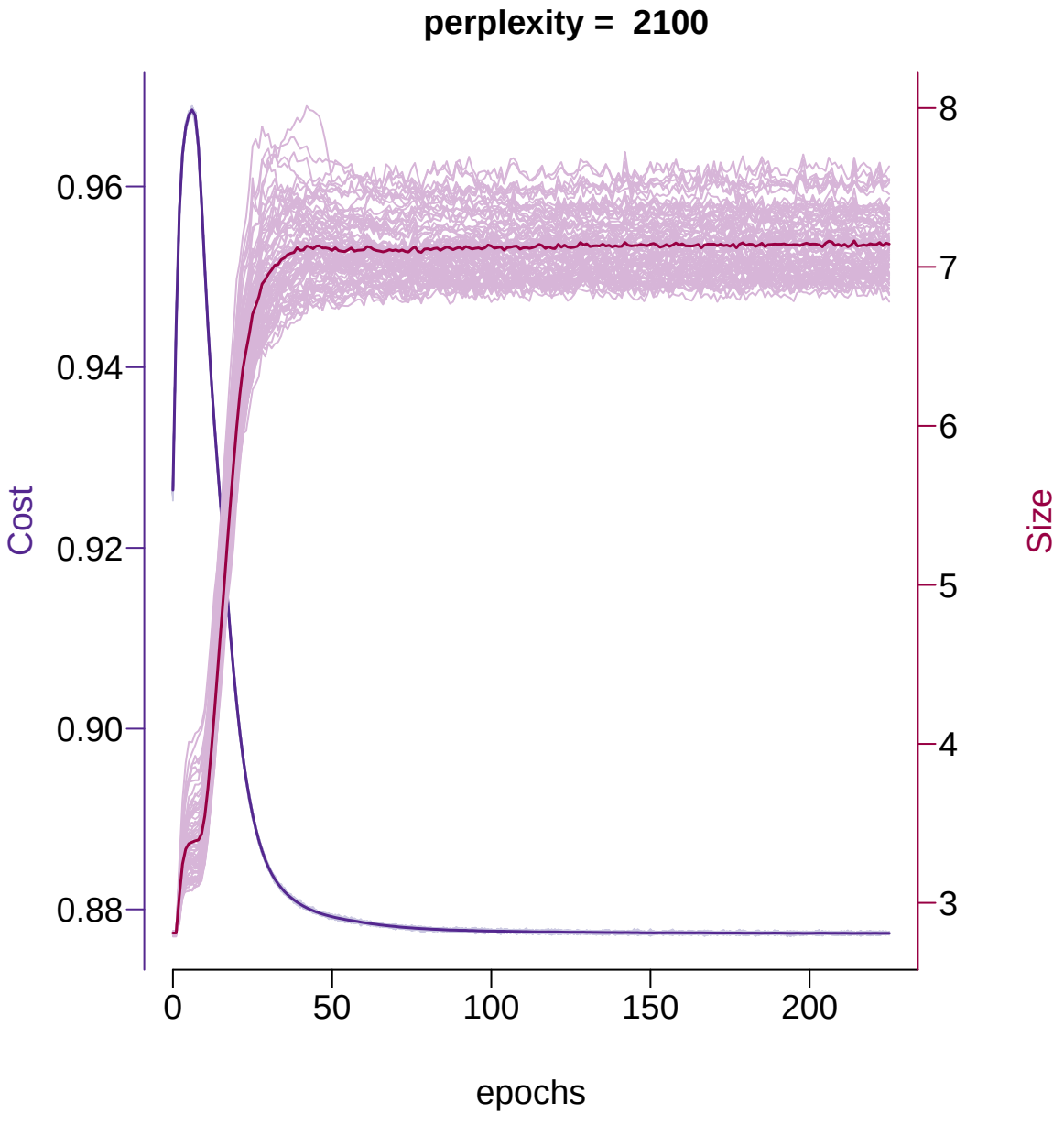

perplexity = 1936
Y2
Y1
3
2
1
0
−1
−2
−3
−2
−1
0
1
2

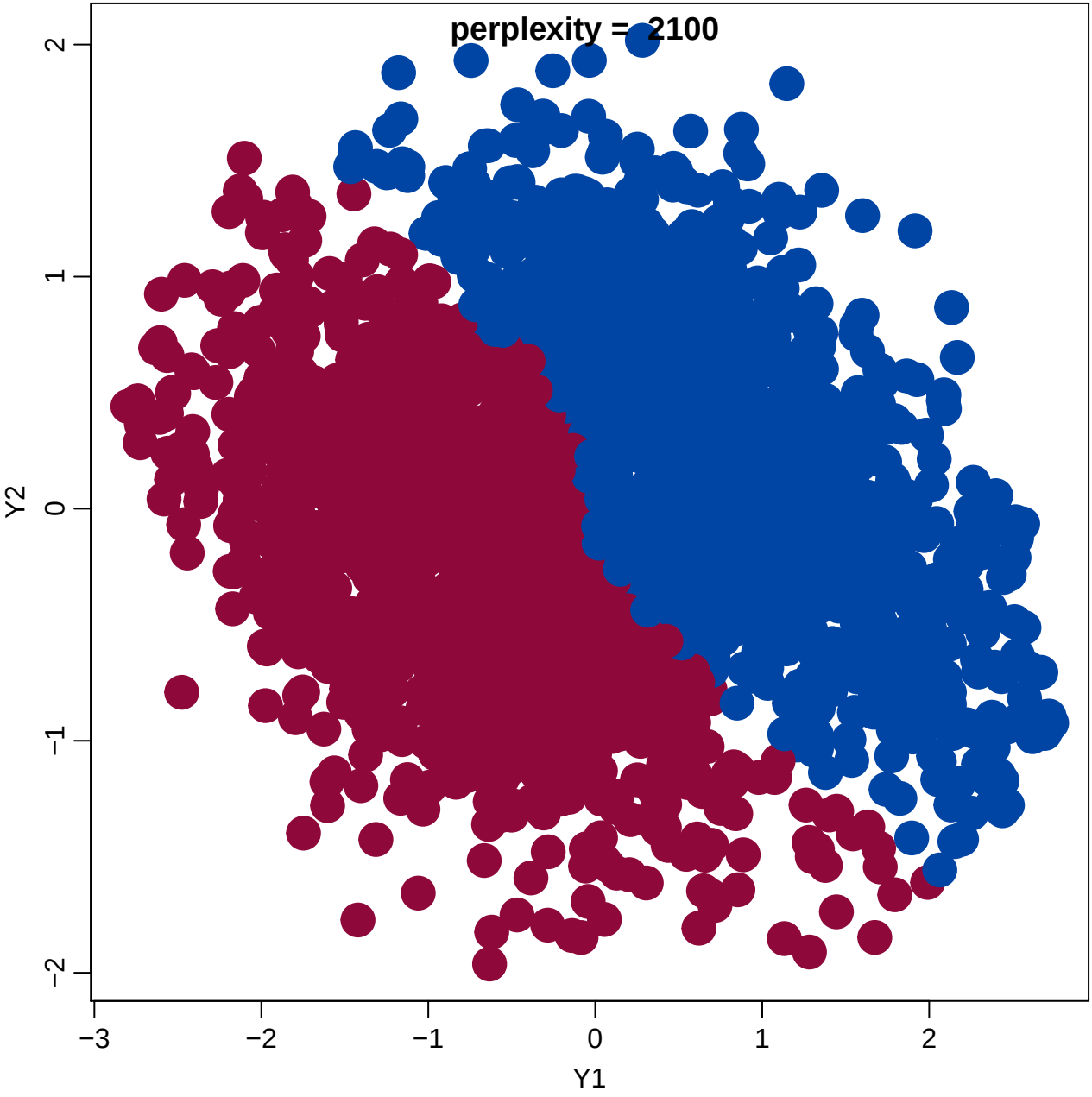

perplexity = 1936
2
1

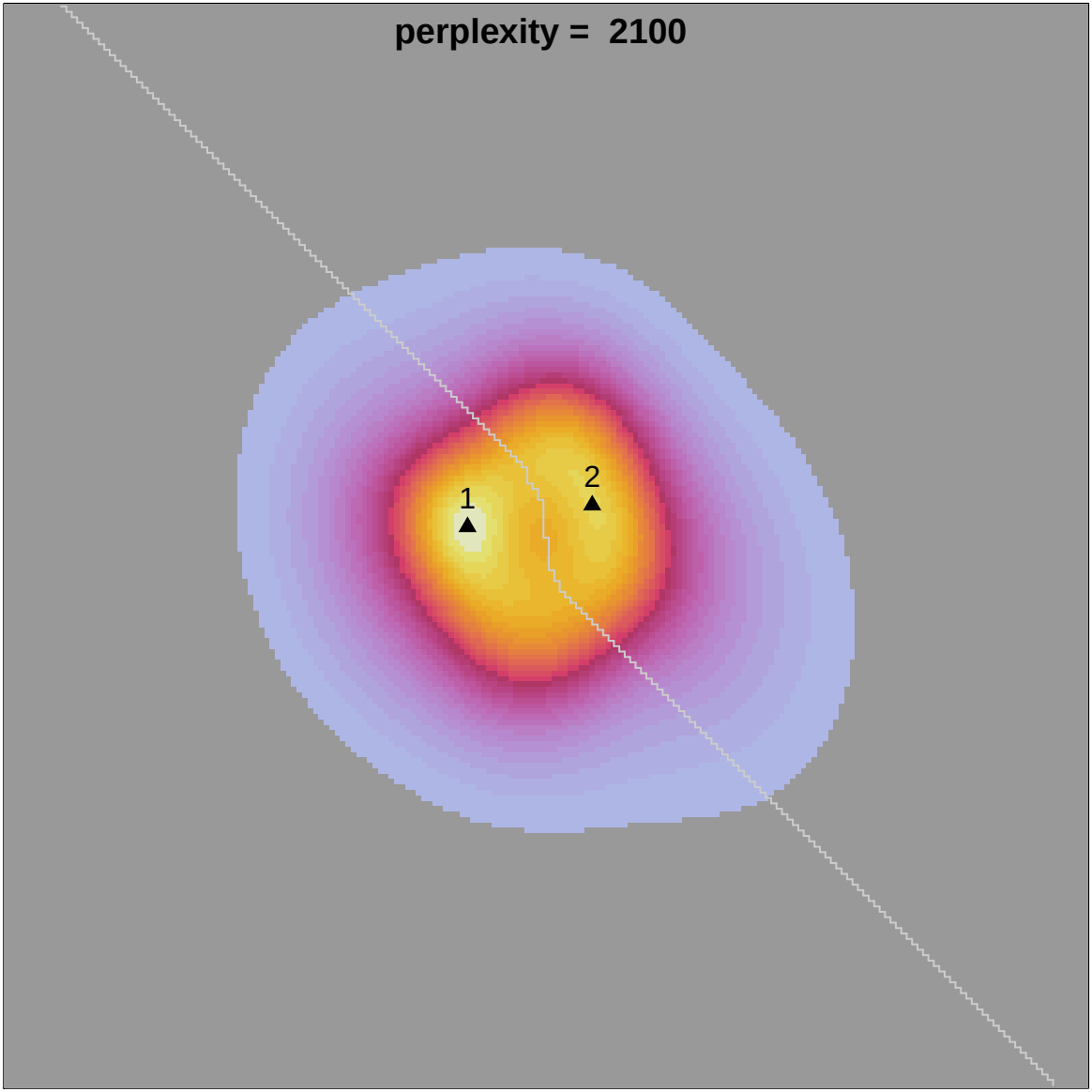

perplexity = 2018
Cost
0.96
0.94
0.92
0.90
0.88
Size
8
7
6
5
4
3
0
50
100
150
200
epochs

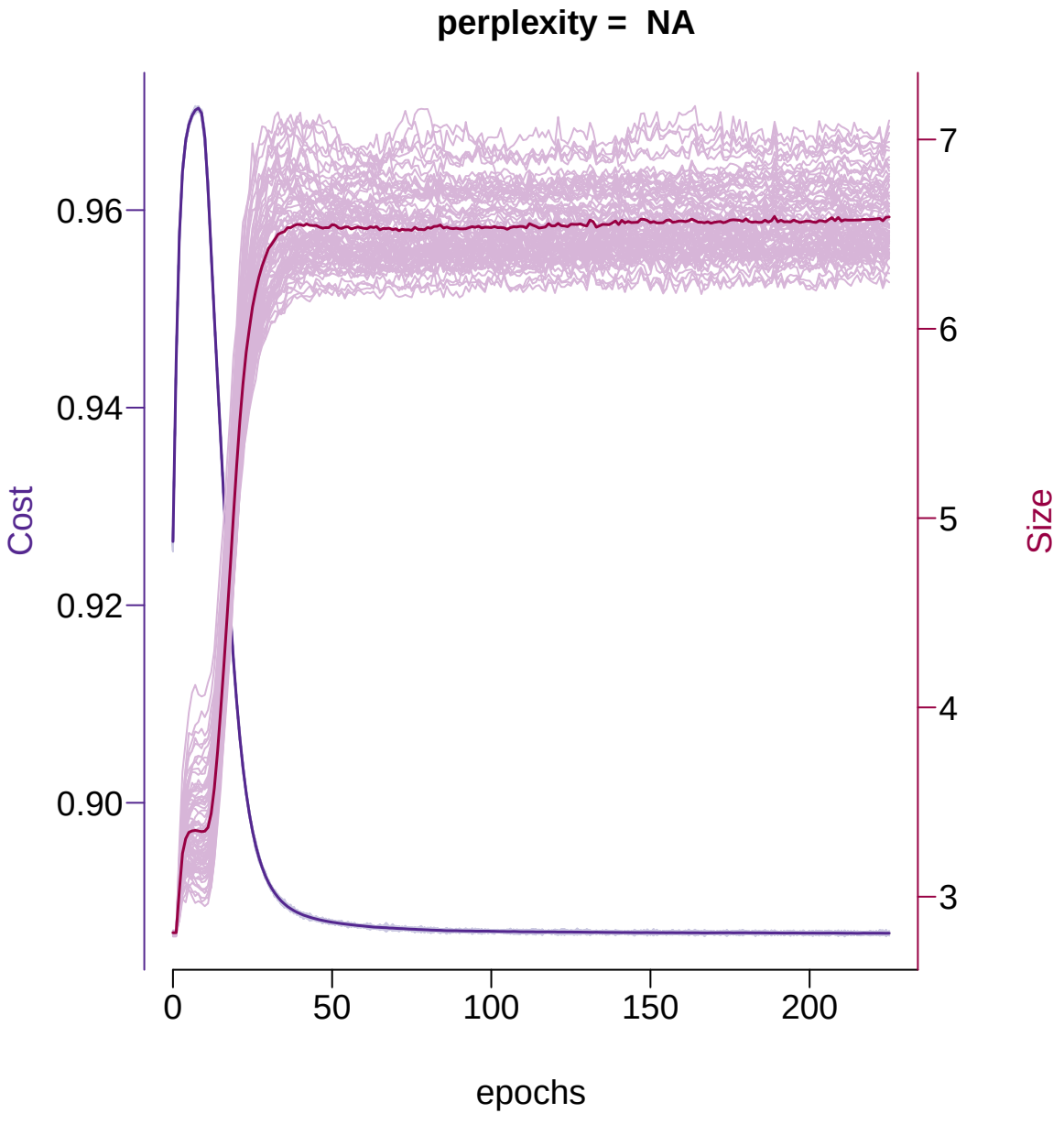

perplexity = 2018
Y2
Y1
2
1
0
−1
−2
−3
−2
−1
0
1
2
3

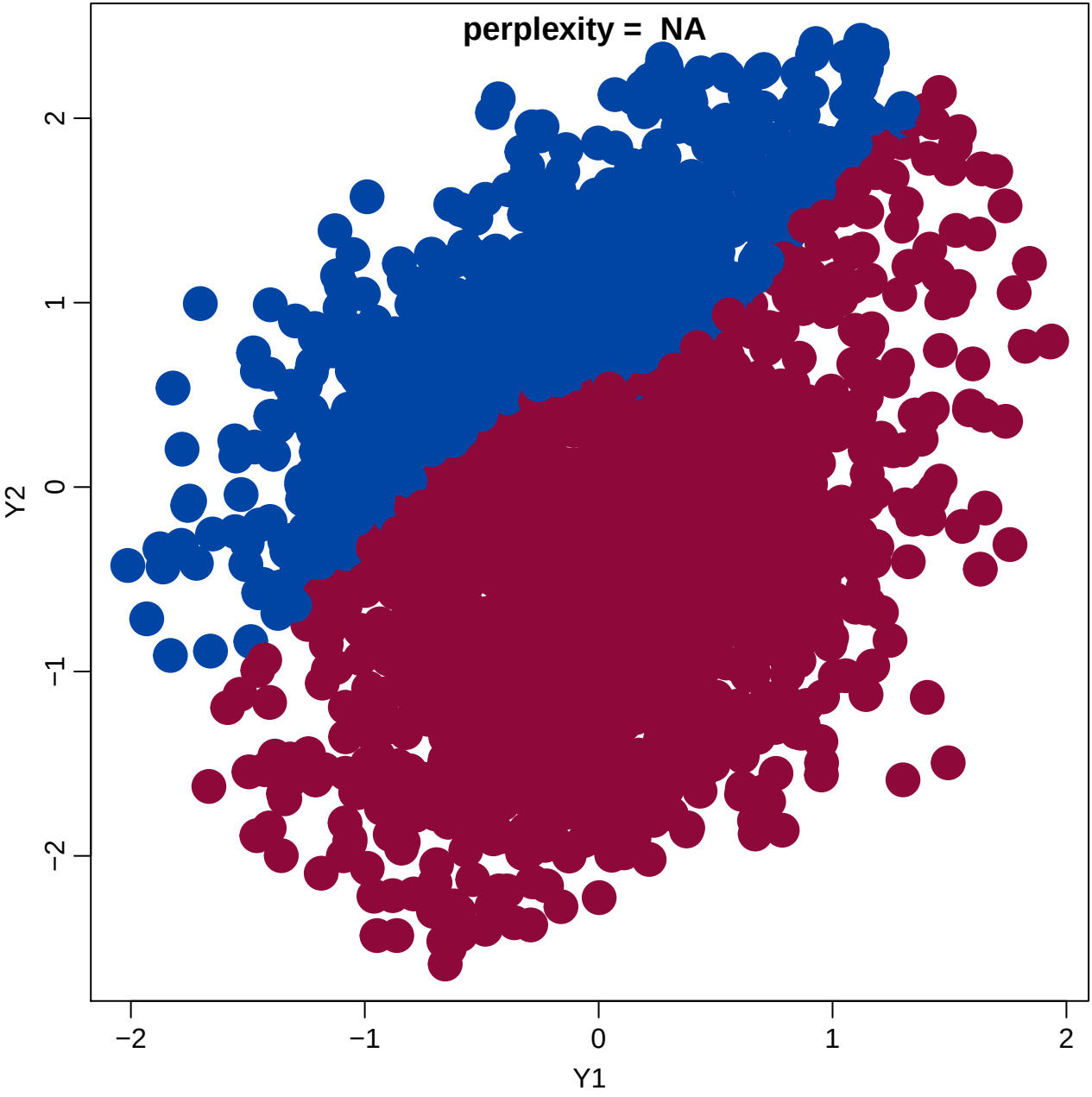

perplexity = 2018
2
1

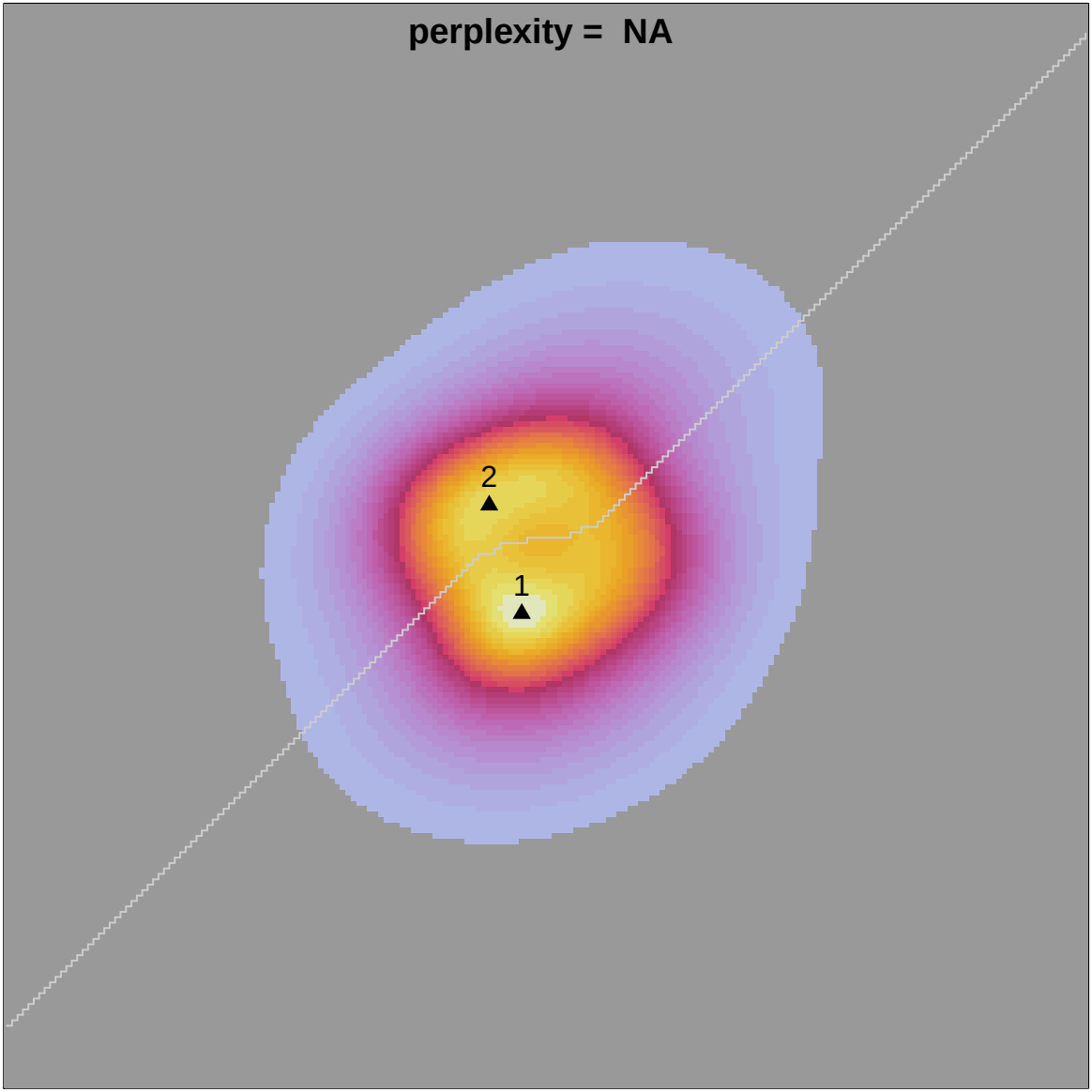

perplexity = 2100
Cost
0.96
0.94
0.92
0.90
0.88
Size
8
7
6
5
4
3
0
50
100
150
200
epochs

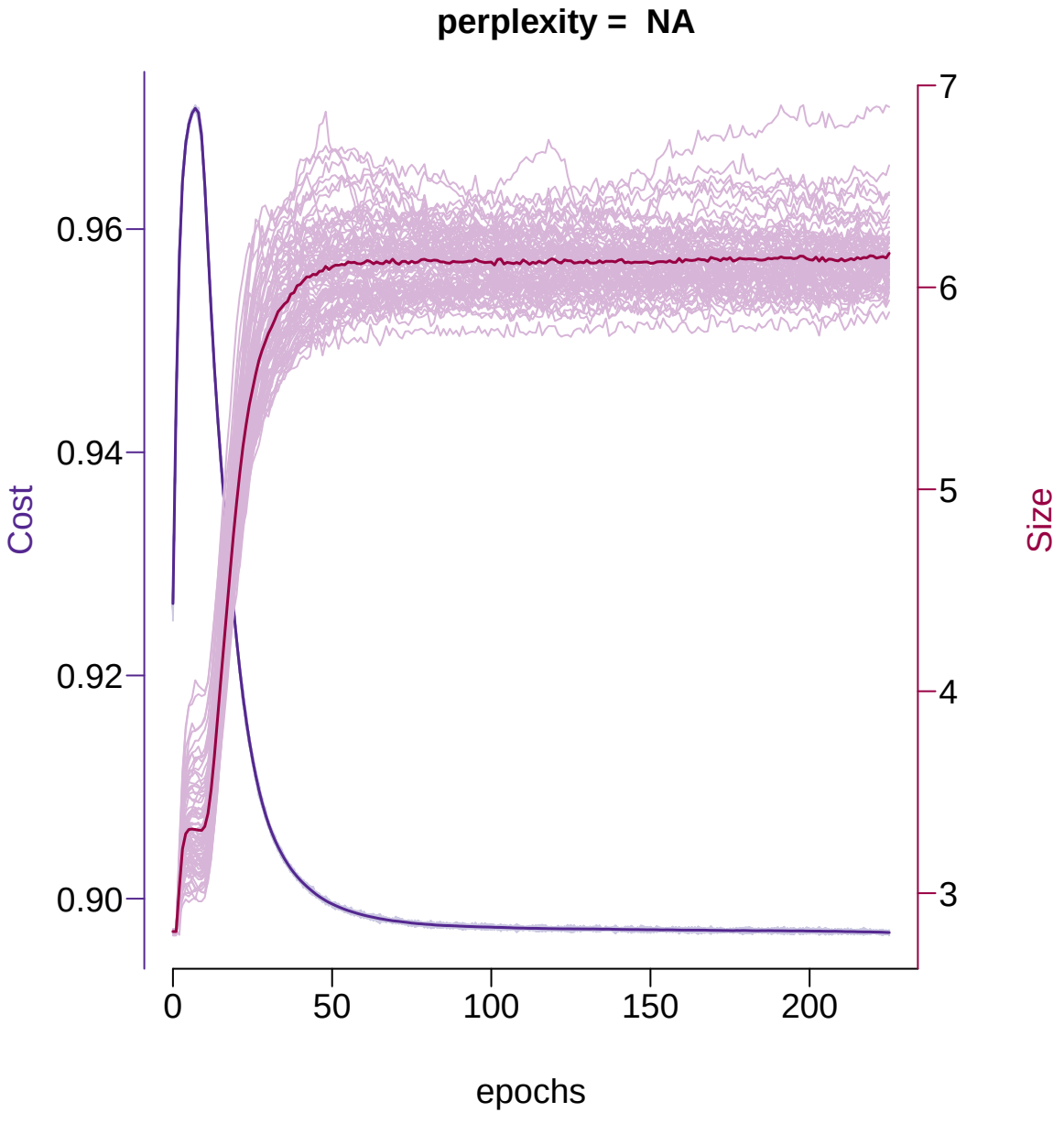

perplexity = 2100
Y2
Y1
2
1
0
−1
−2
−3
−2
−1
0
1
2

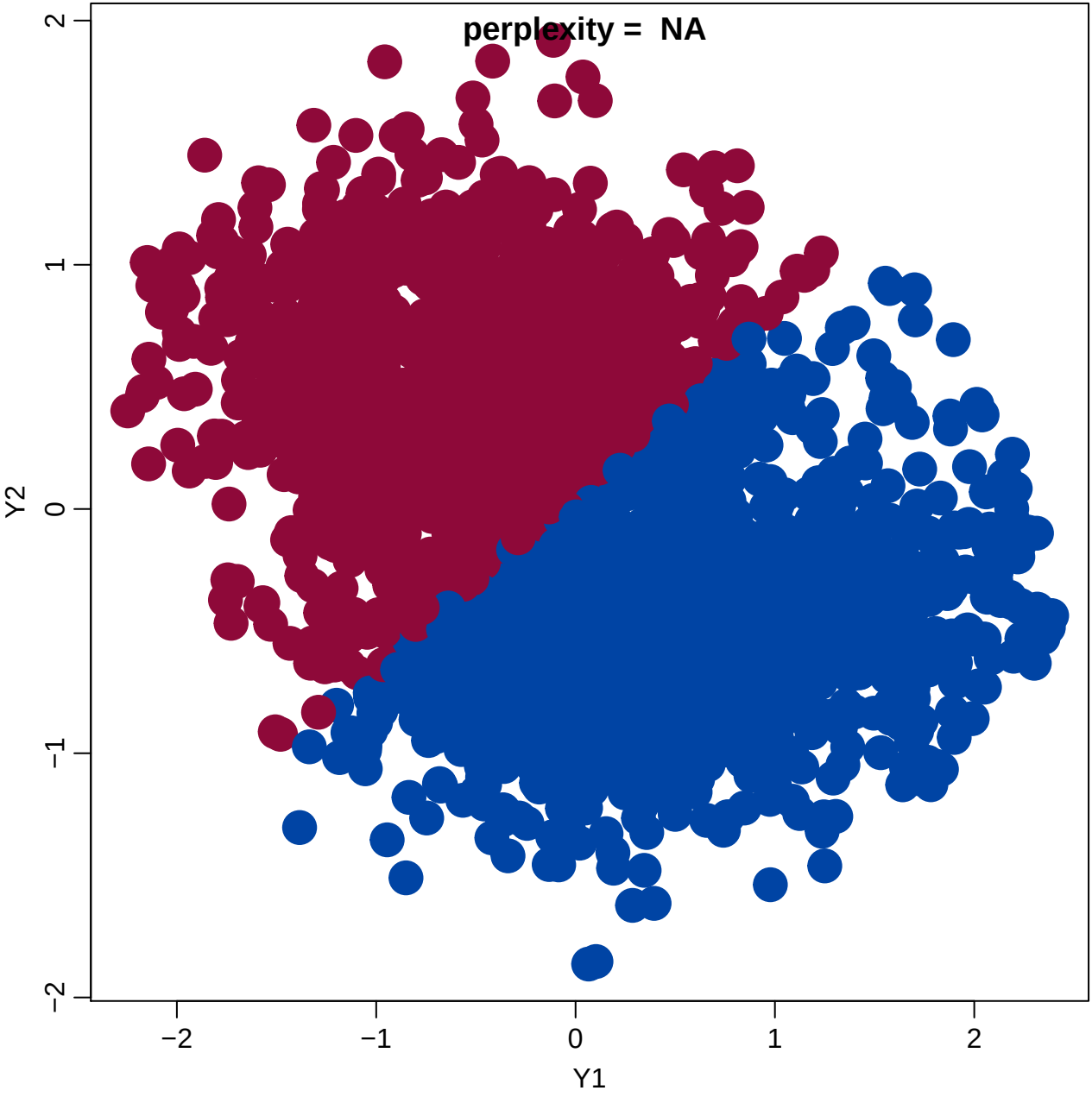

perplexity = 2100
1
2

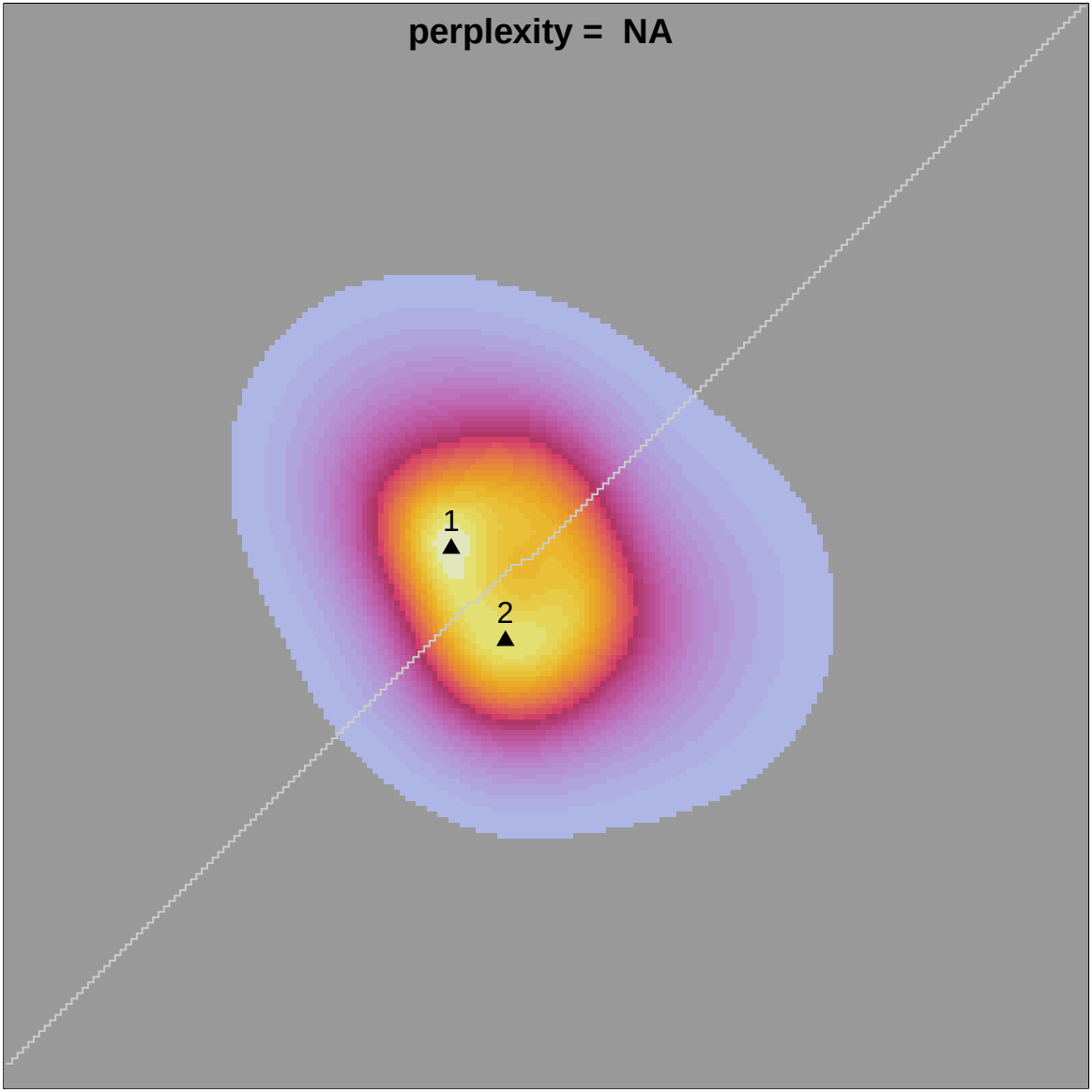

perplexity = NA
Cost
0.96
0.94
0.92
0.90
Size
7
6
5
4
3
0
50
100
150
200
epochs

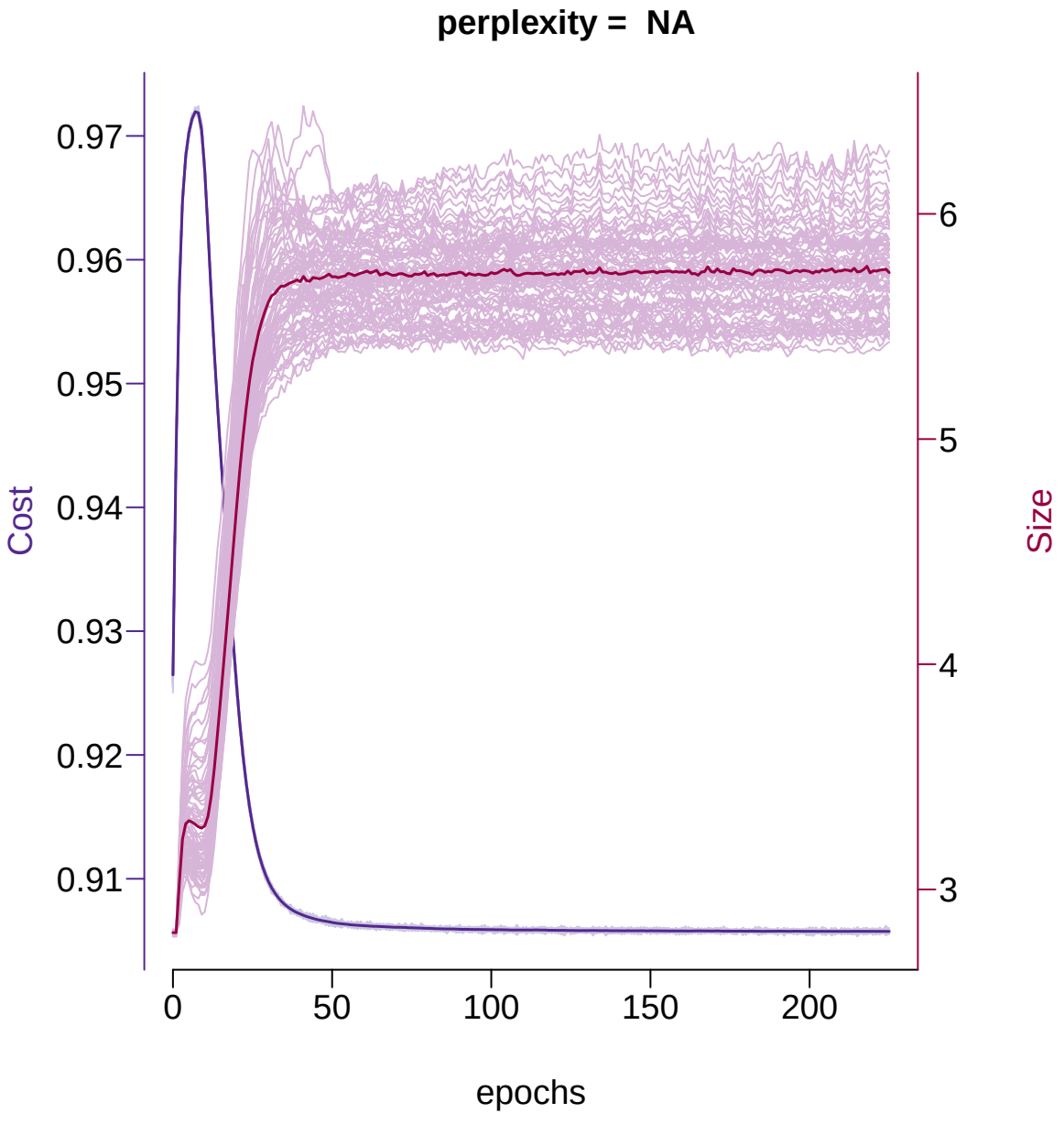

perplexity = NA
Y2
Y1
2
1
0
−1
−2
−2
−1
0
1
2

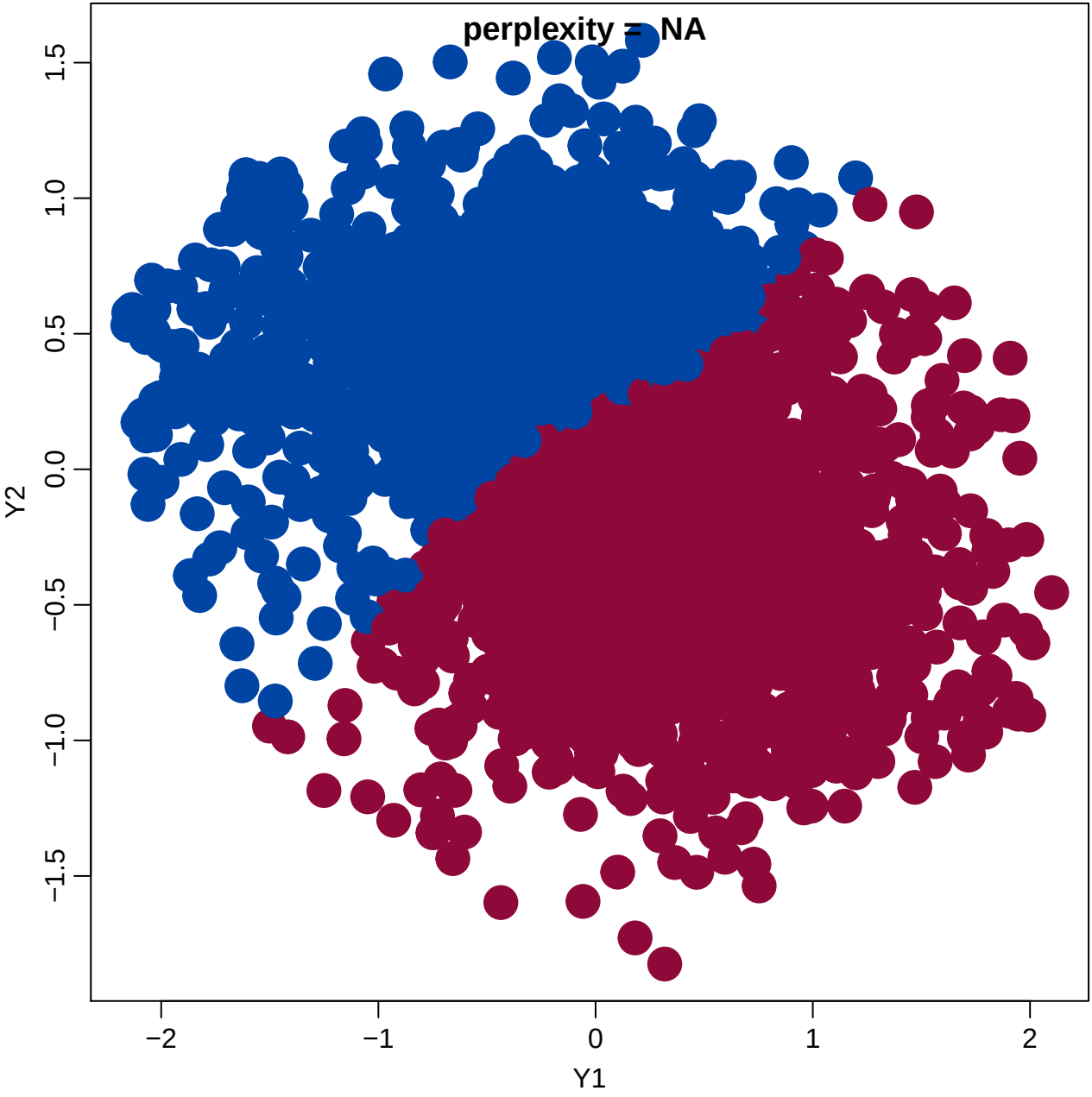

perplexity = NA
2
1

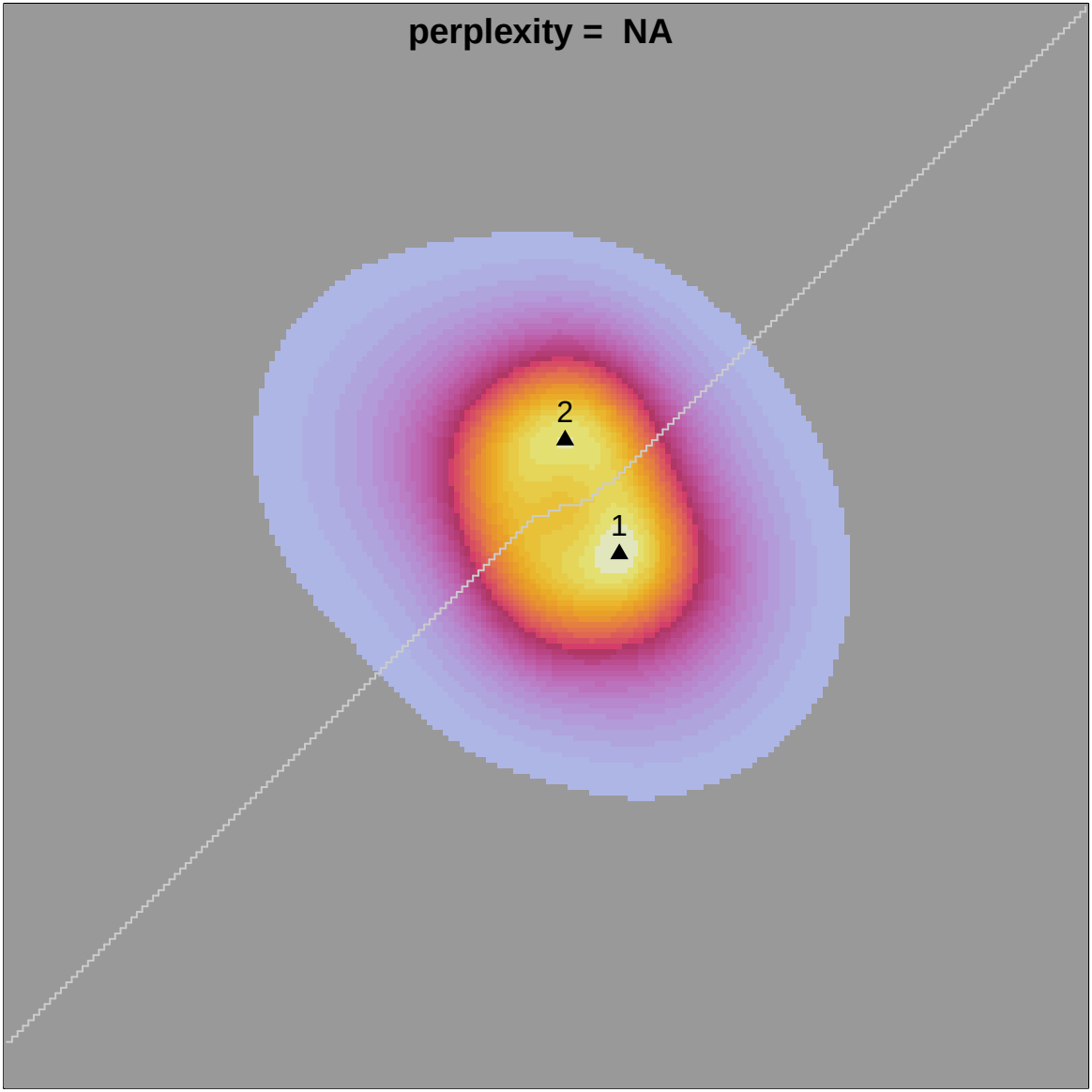

perplexity = NA
Cost
0.96
0.94
0.92
0.90
Size
7
6
5
4
3
0
50
100
150
200
epochs

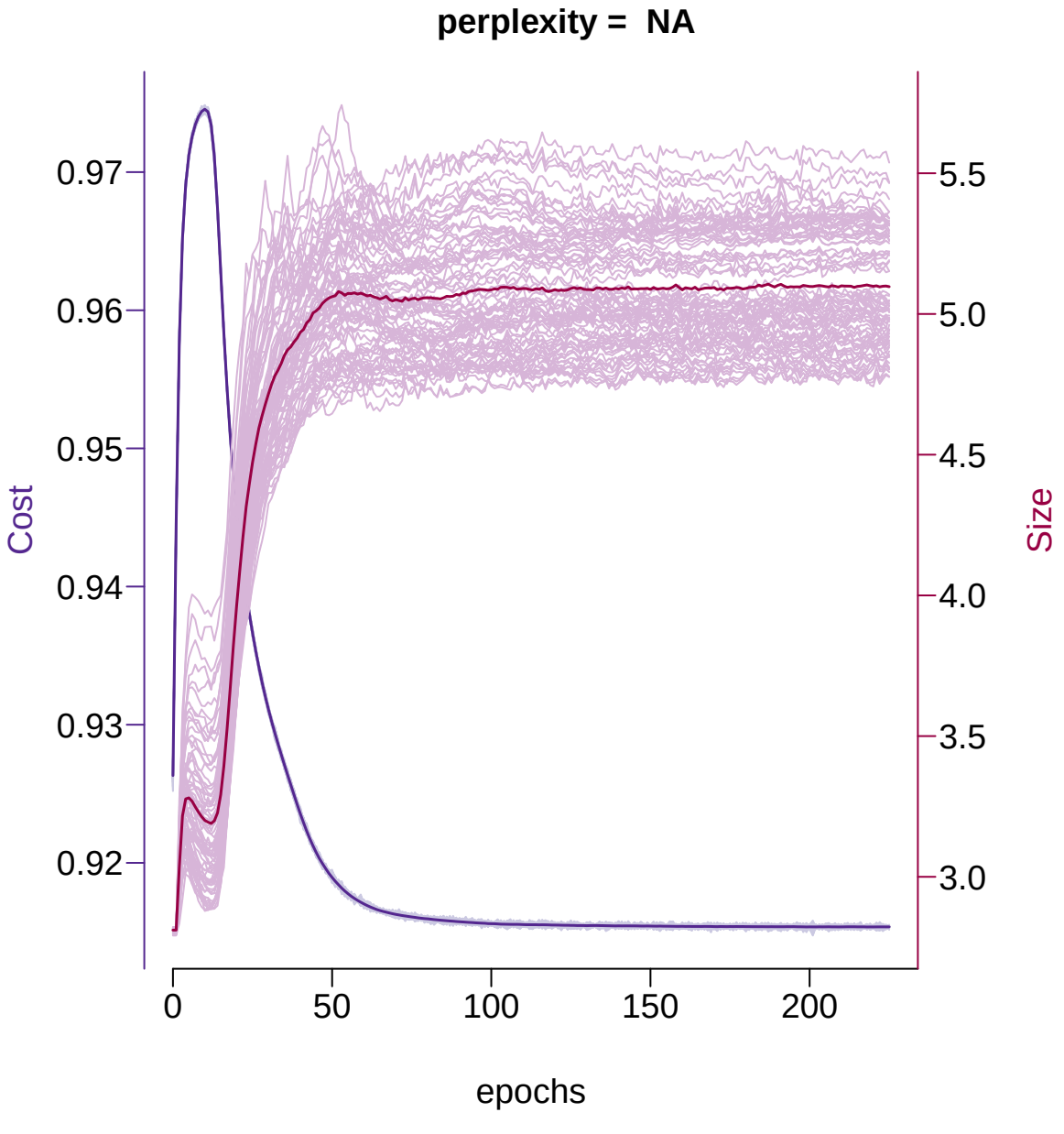

perplexity = NA
Y2
Y1
2
1
0
−1
−2
−2
−1
0
1
2

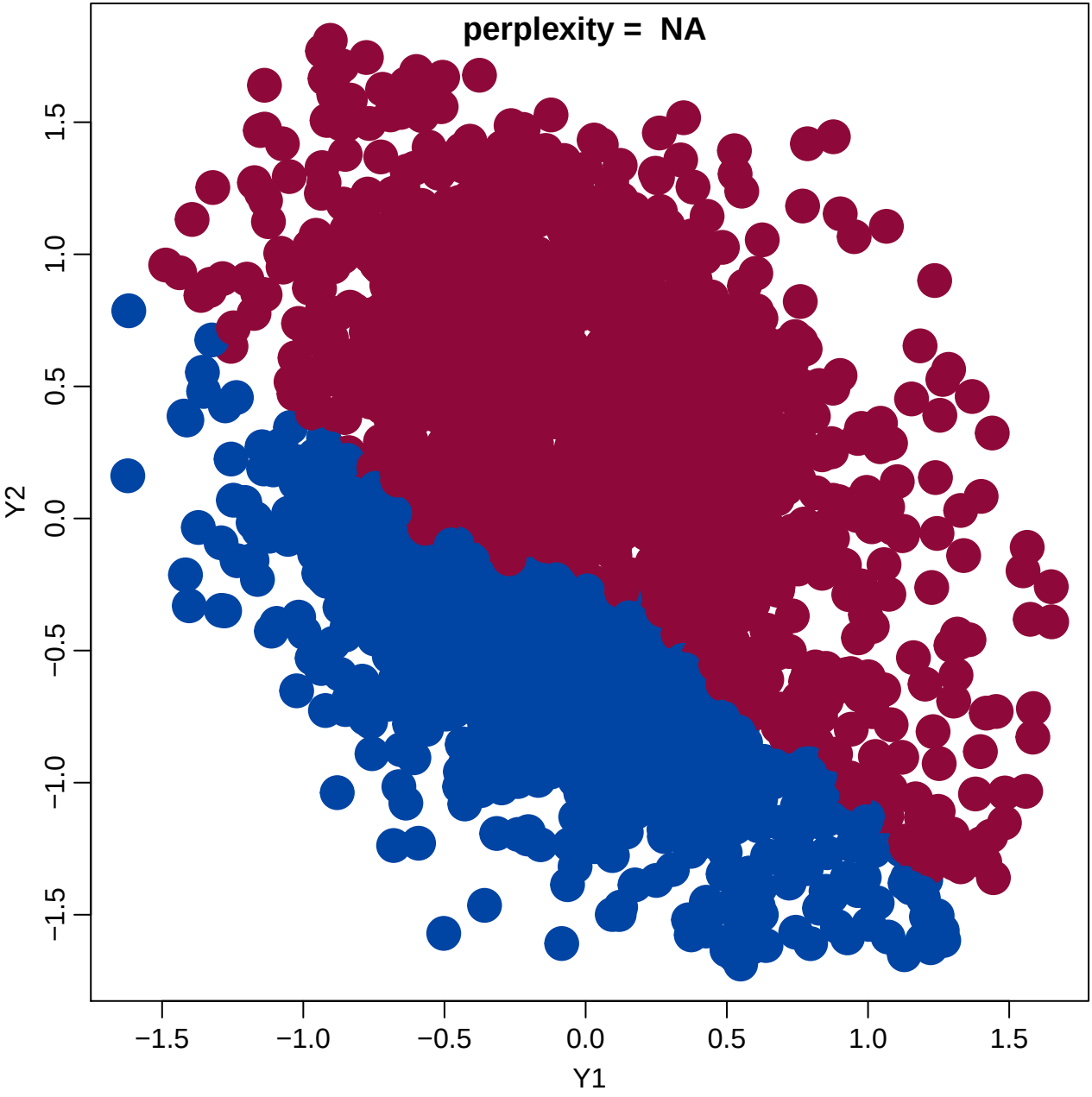

perplexity = NA
1
2

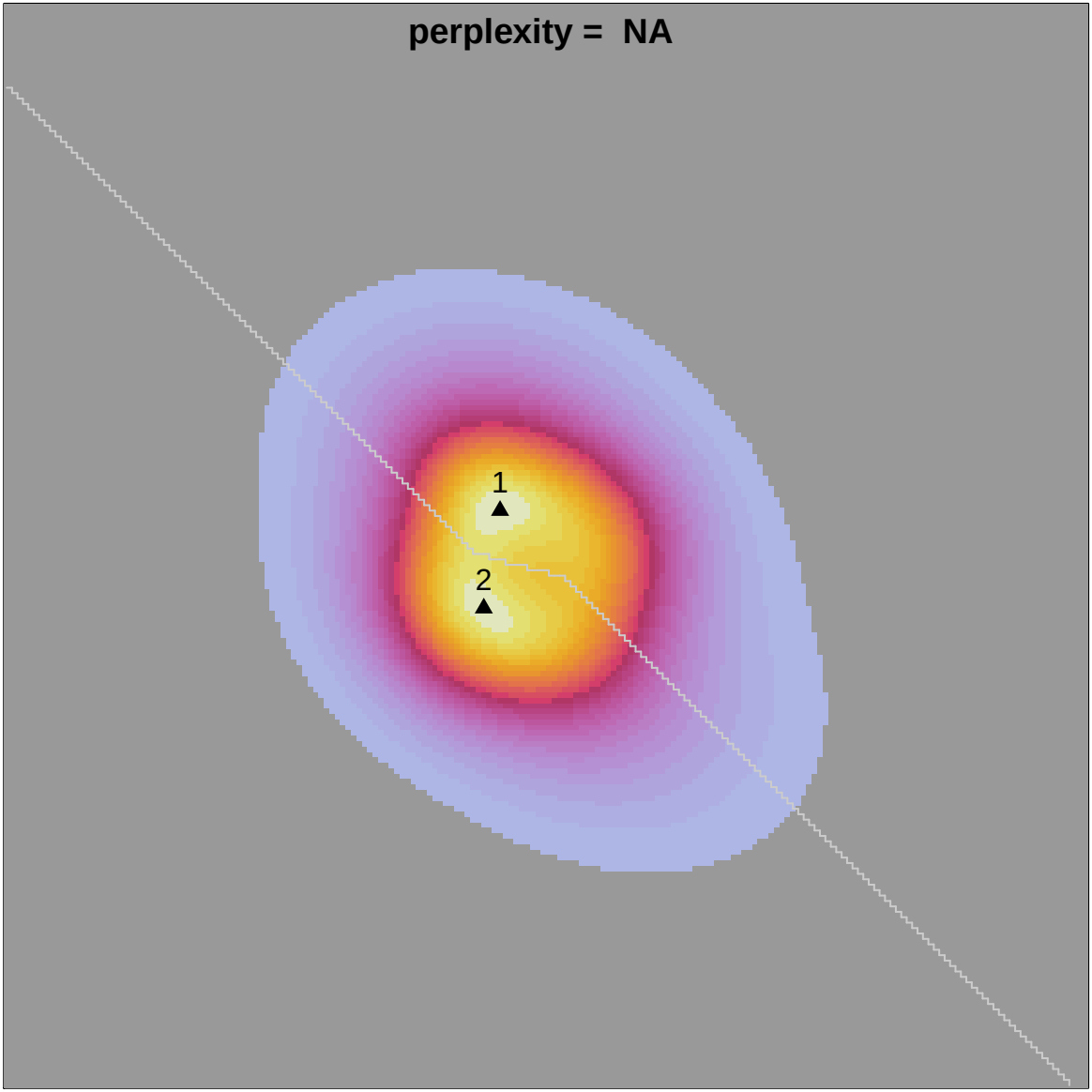

perplexity = NA
Cost
0.97
0.96
0.95
0.94
0.93
0.92
0.91
Size
6
5
4
3
0
50
100
150
200
epochs

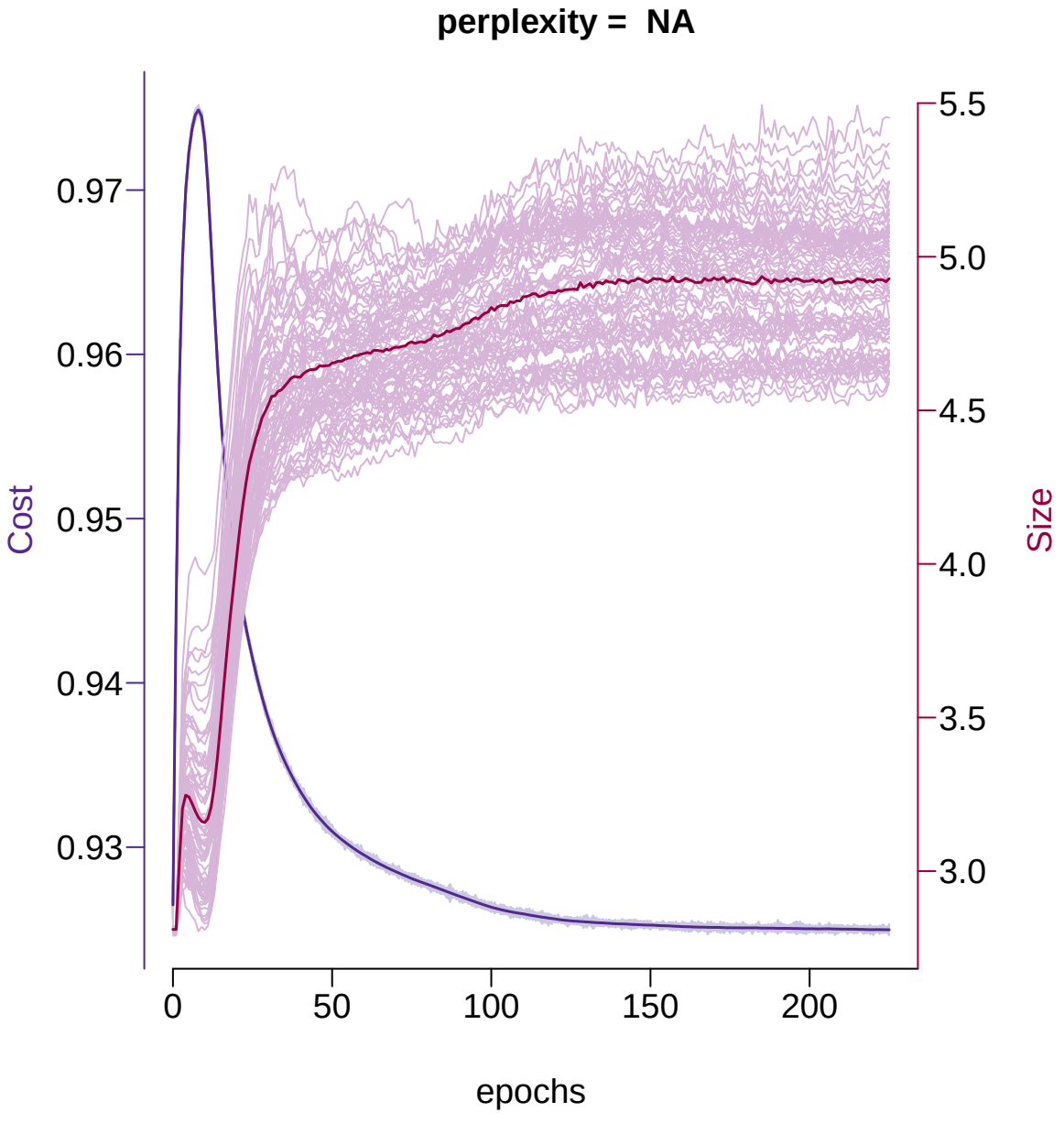

perplexity = NA
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
Y2
−2
−1
0
1
2
Y1

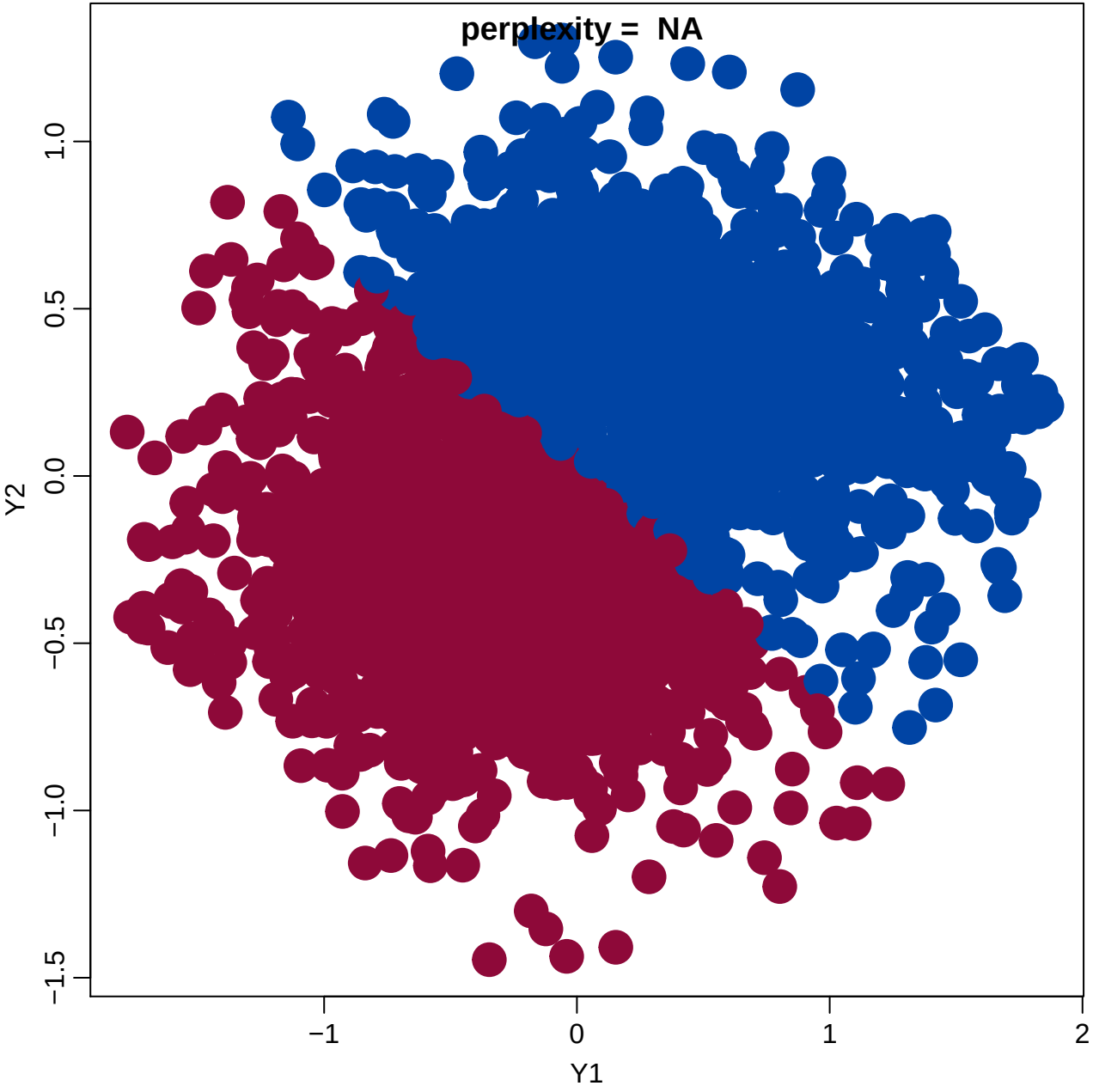

perplexity = NA
2
1

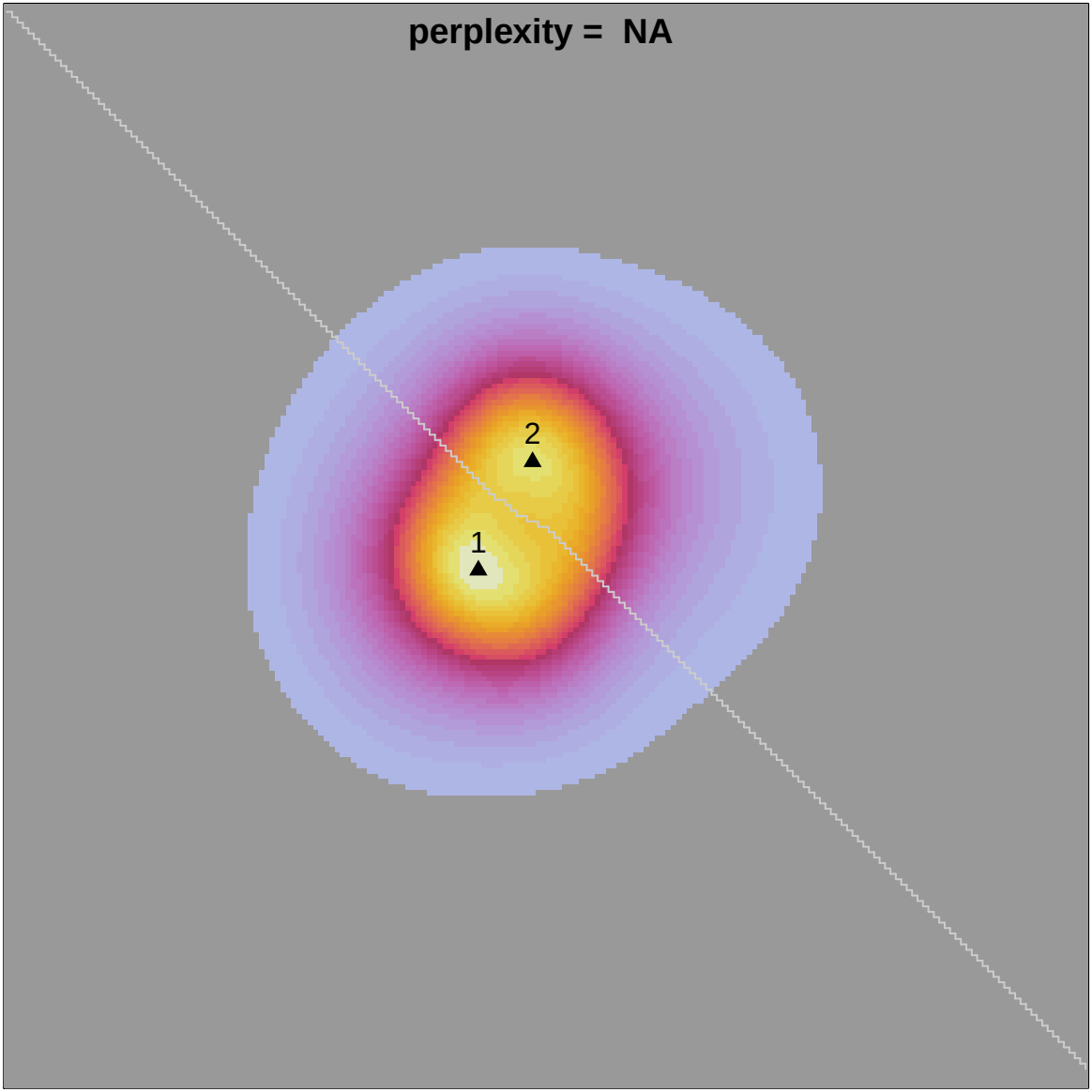

perplexity = NA
Cost
Size
0.97
0.96
0.95
0.94
0.93
0.92
5.5
5.0
4.5
4.0
3.5
3.0
0
50
100
150
200
epochs


perplexity = NA
Y2
Y1
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5


perplexity = NA
1
2


perplexity = NA
Cost
0.97
0.96
0.95
0.94
0.93
Size
5.5
5.0
4.5
4.0
3.5
3.0
0
50
100
150
200
epochs


perplexity = NA
Y2
Y1
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1
0
1
2


perplexity = NA
2
1