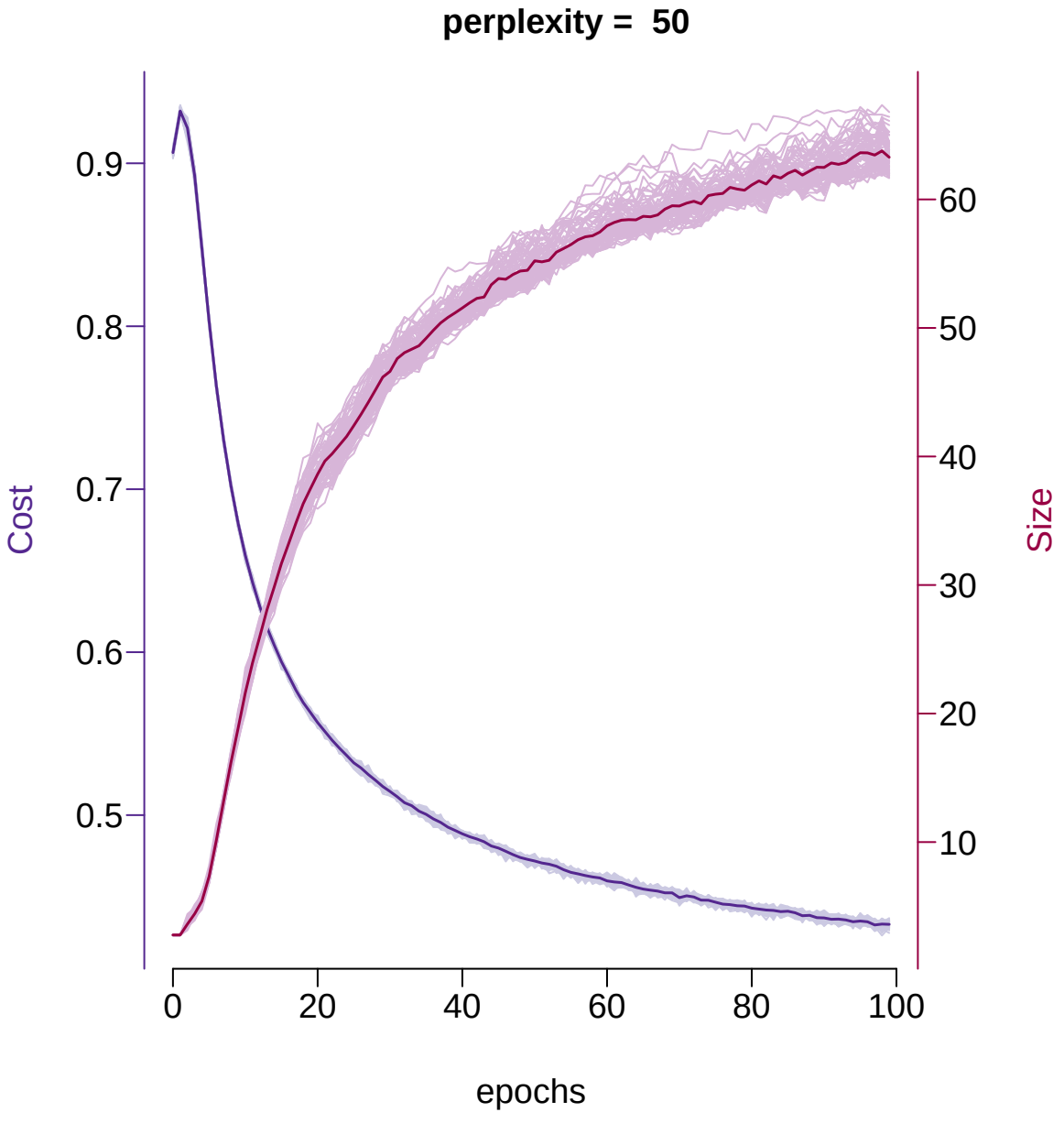
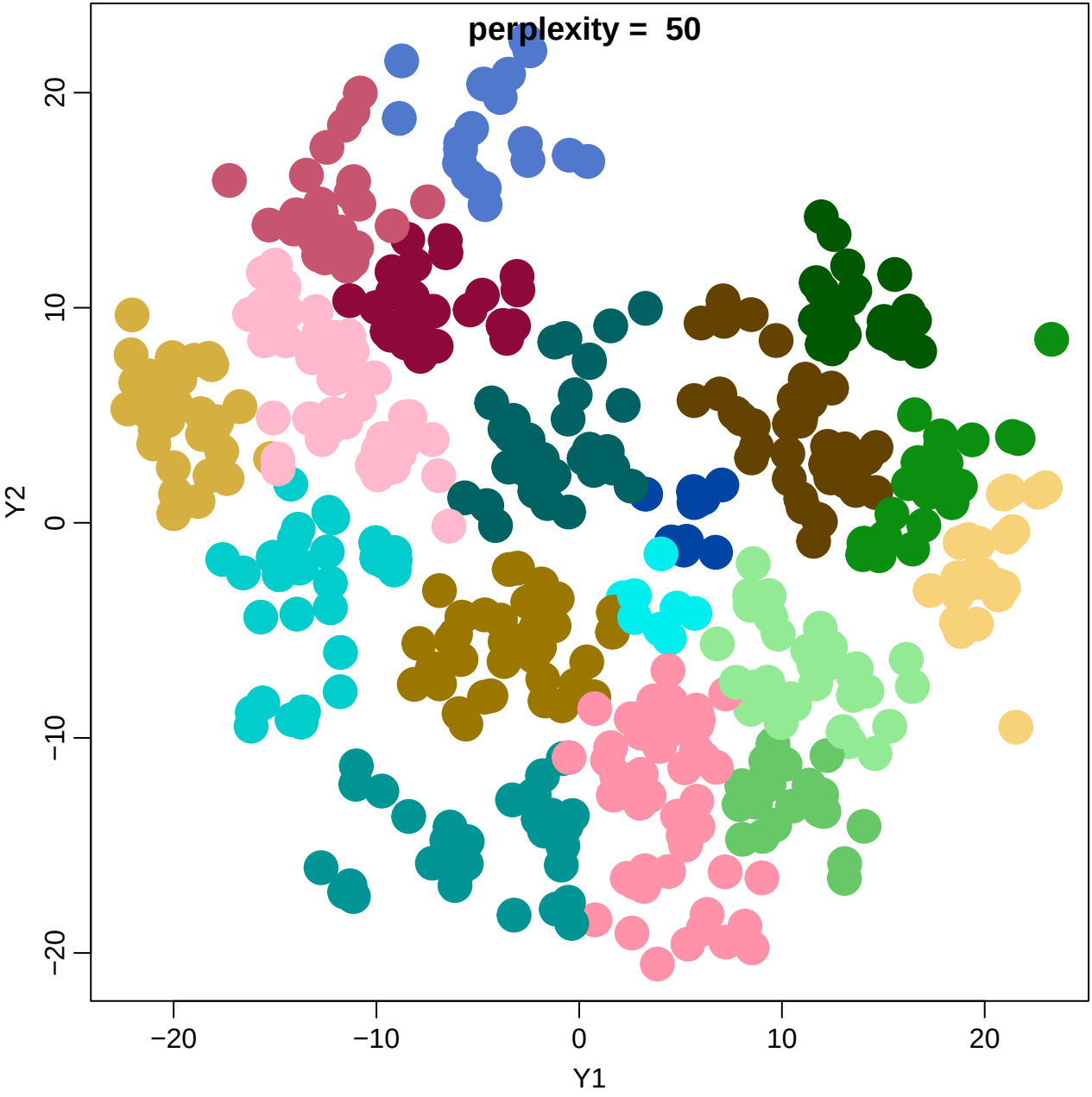
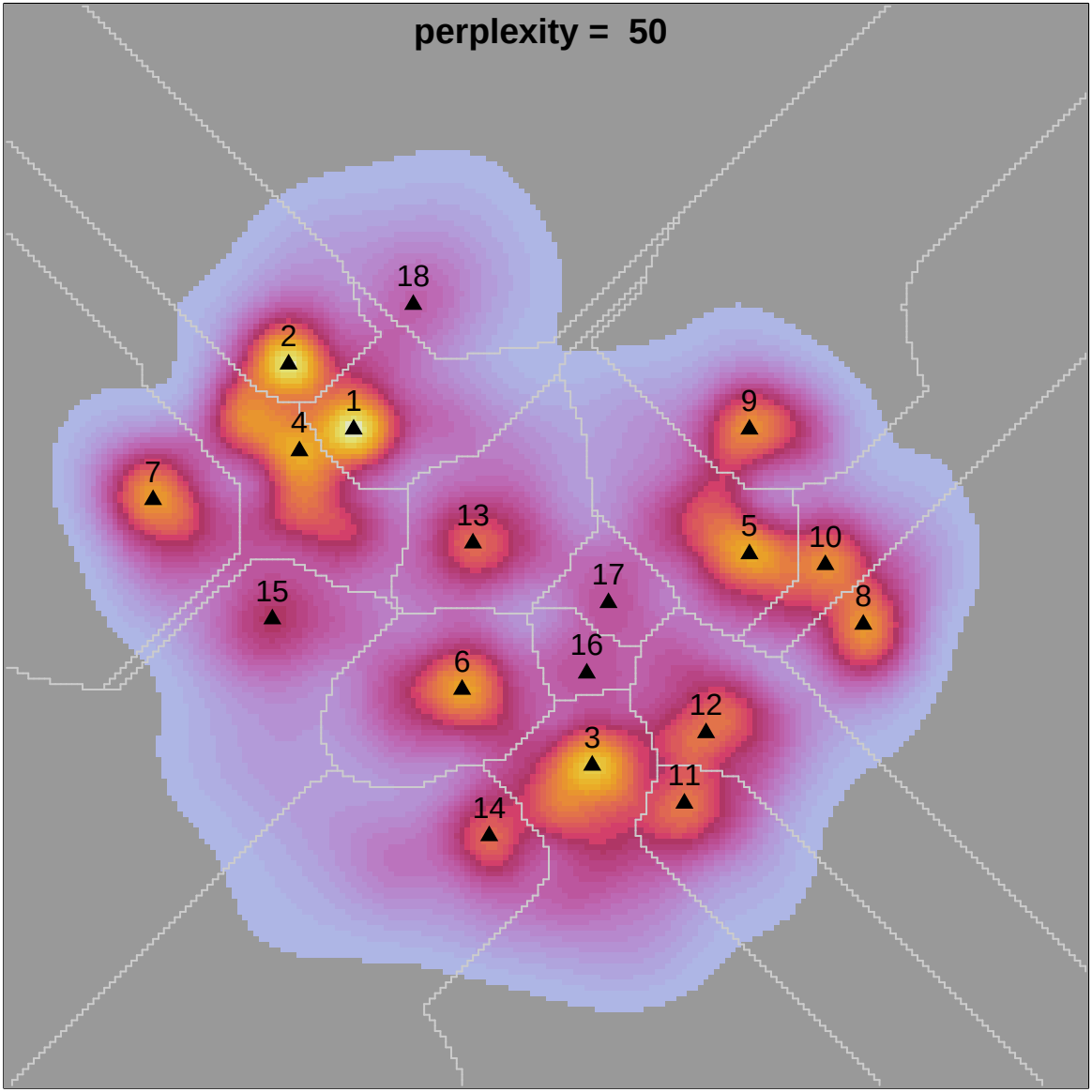
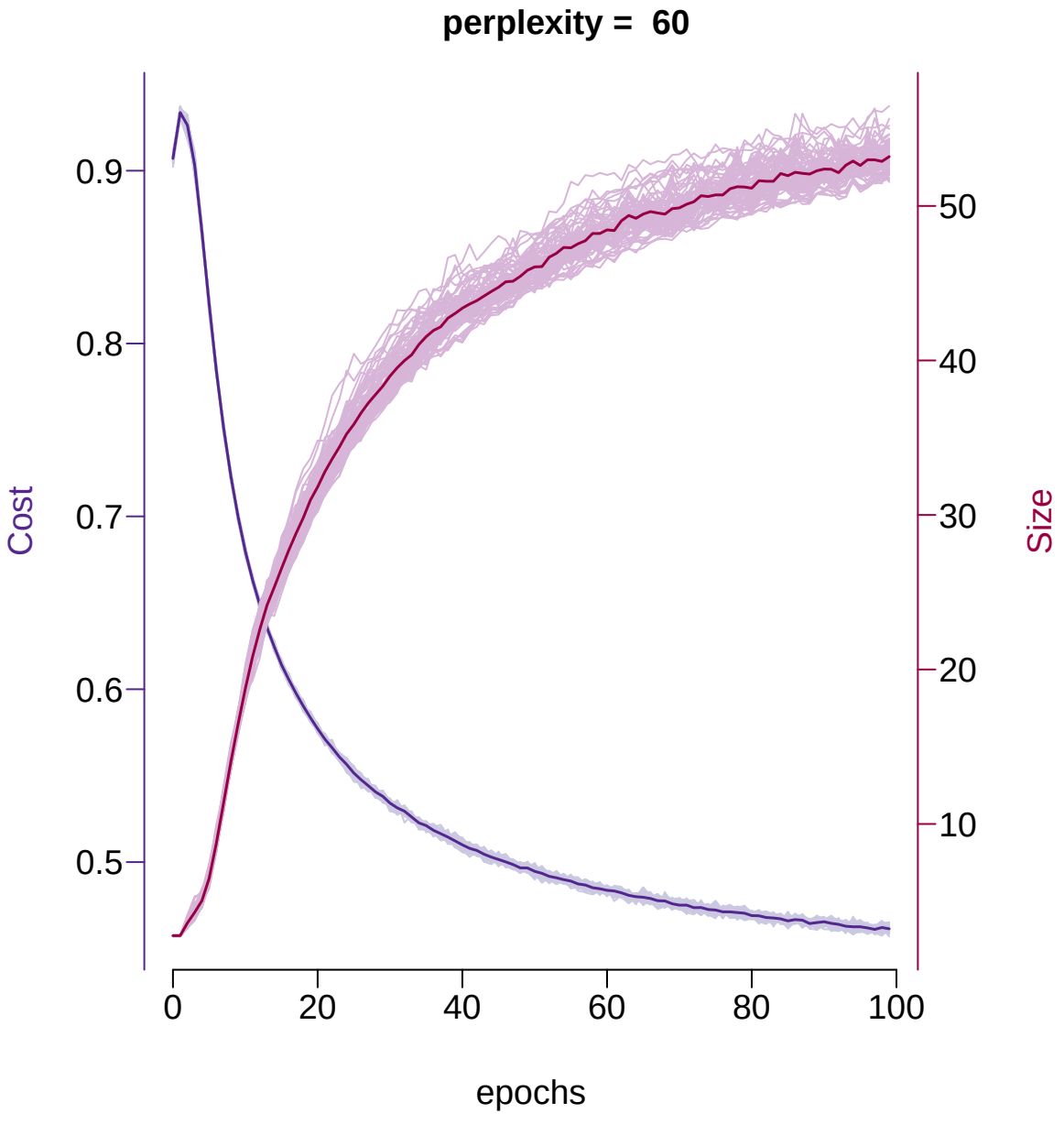
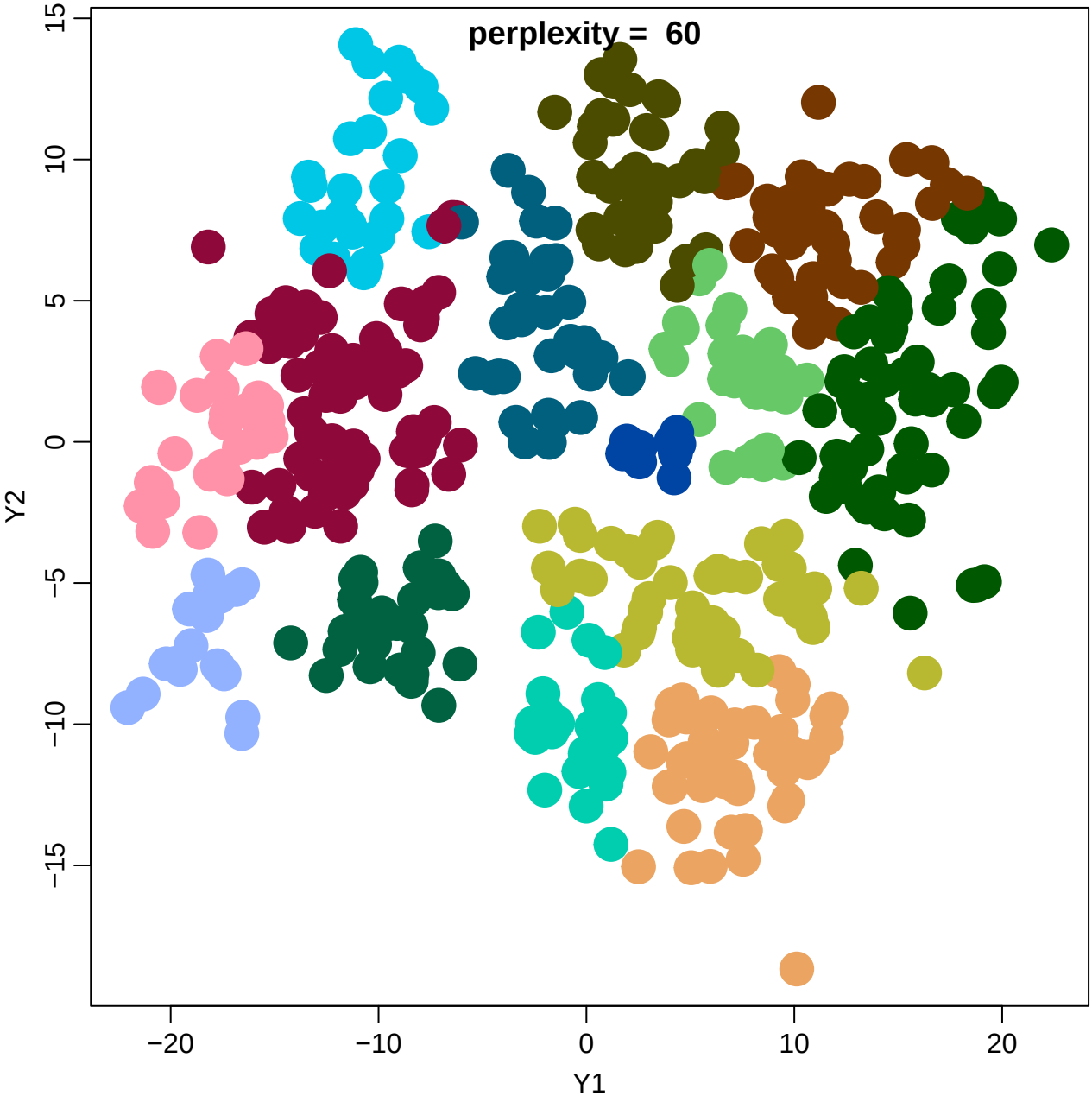
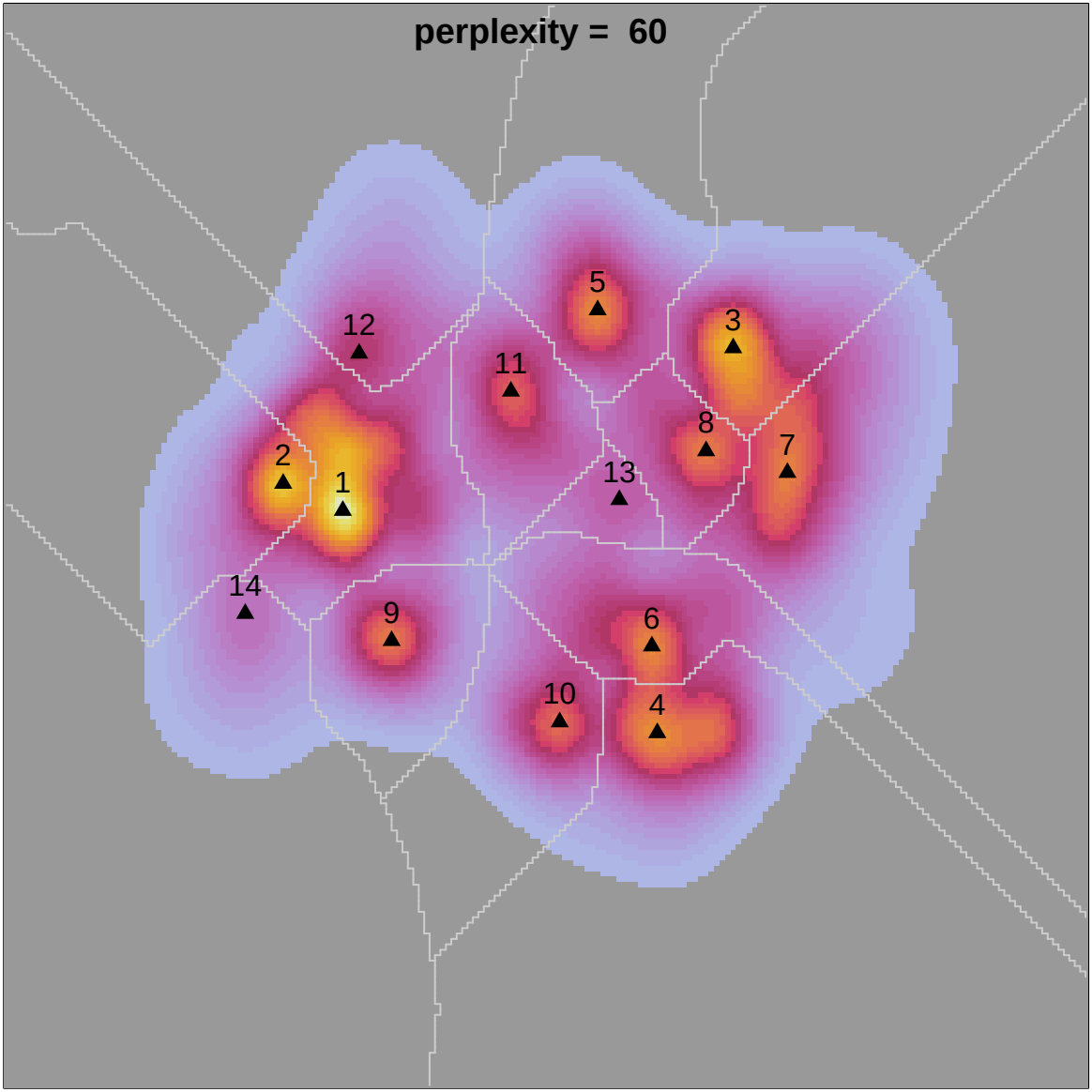

perplexity = 50
Cost
Size
epochs
0.9
0.8
0.7
0.6
0.5
60
50
40
30
20
10
0
20
40
60
80
100

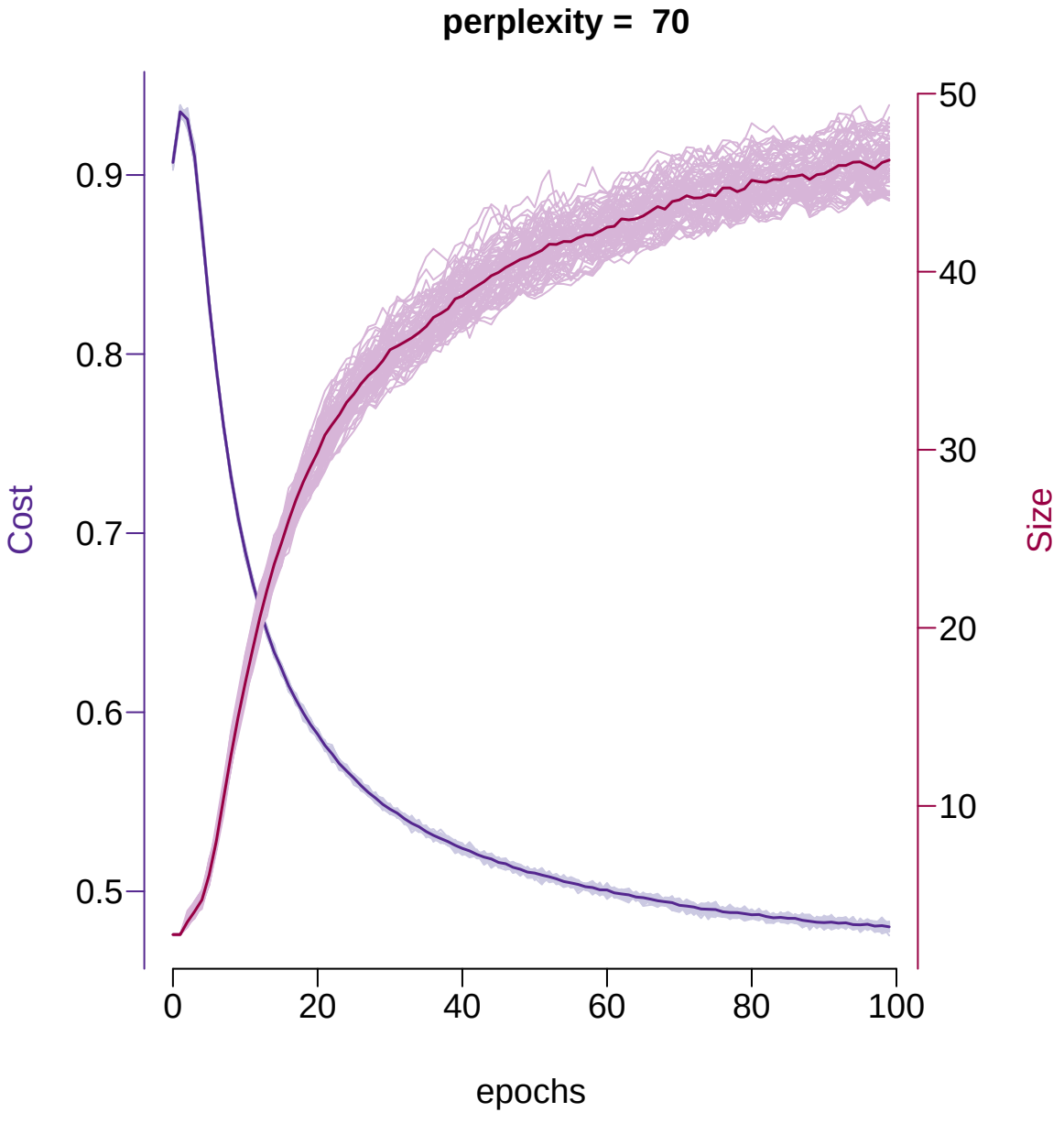

perplexity = 50
Y2
Y1
20
10
0
−10
−20
−20
−10
0
10
20

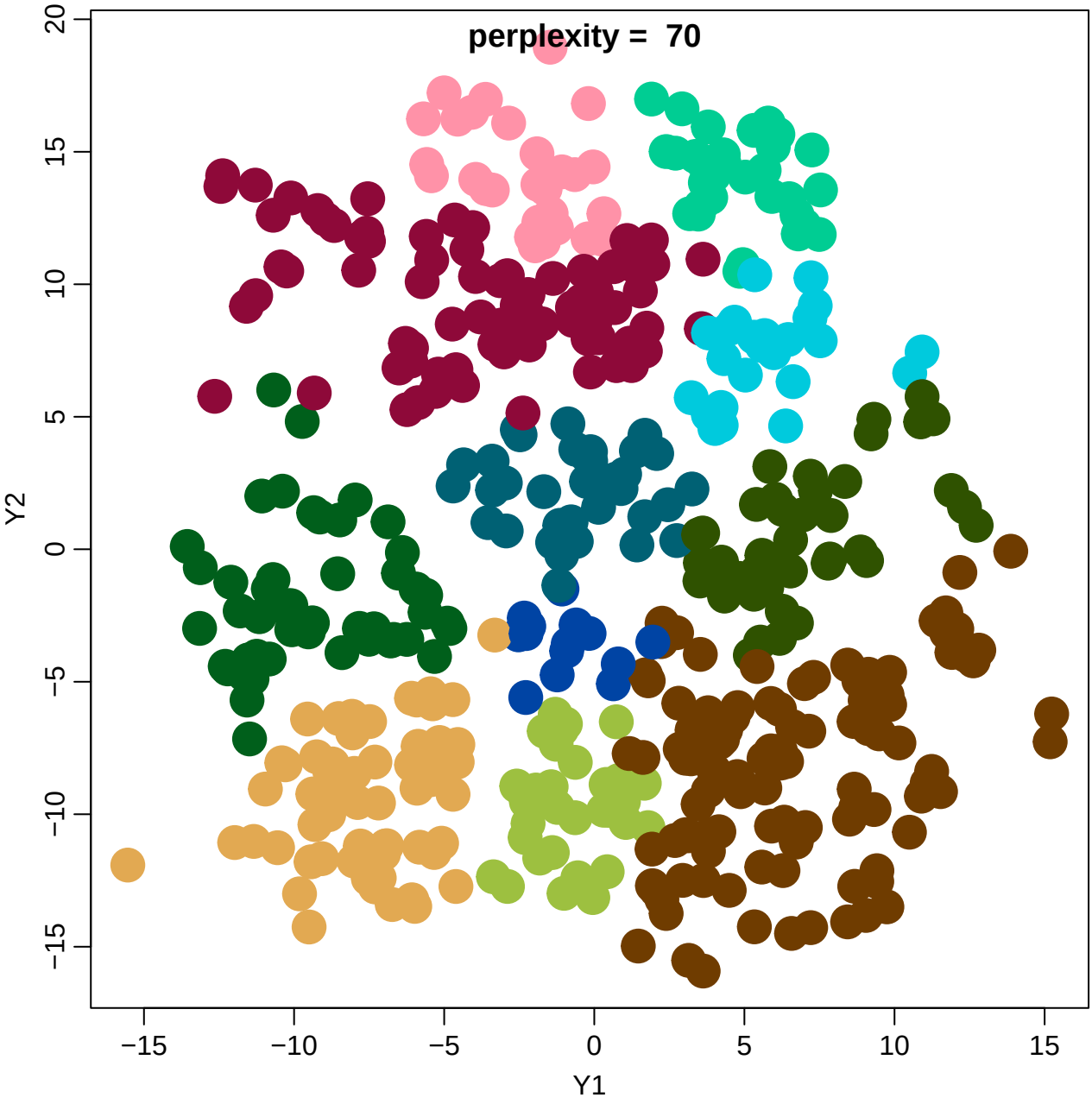

perplexity = 50
18
2
1
4
9
7
13
5
10
17
15
8
16
6
12
3
11
14

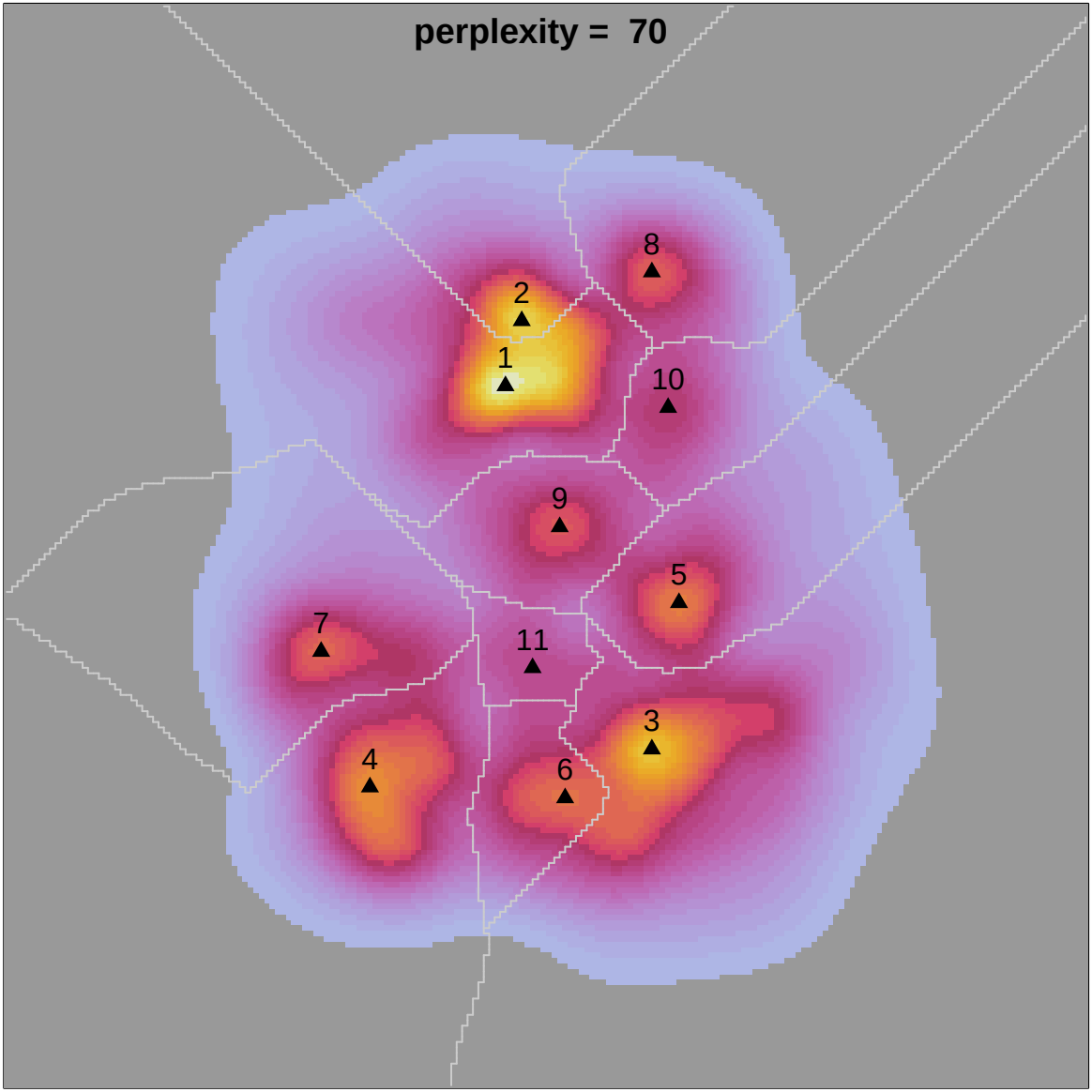

perplexity = 60
Cost
0.9
0.8
0.7
0.6
0.5
Size
50
40
30
20
10
0
20
40
60
80
100
epochs

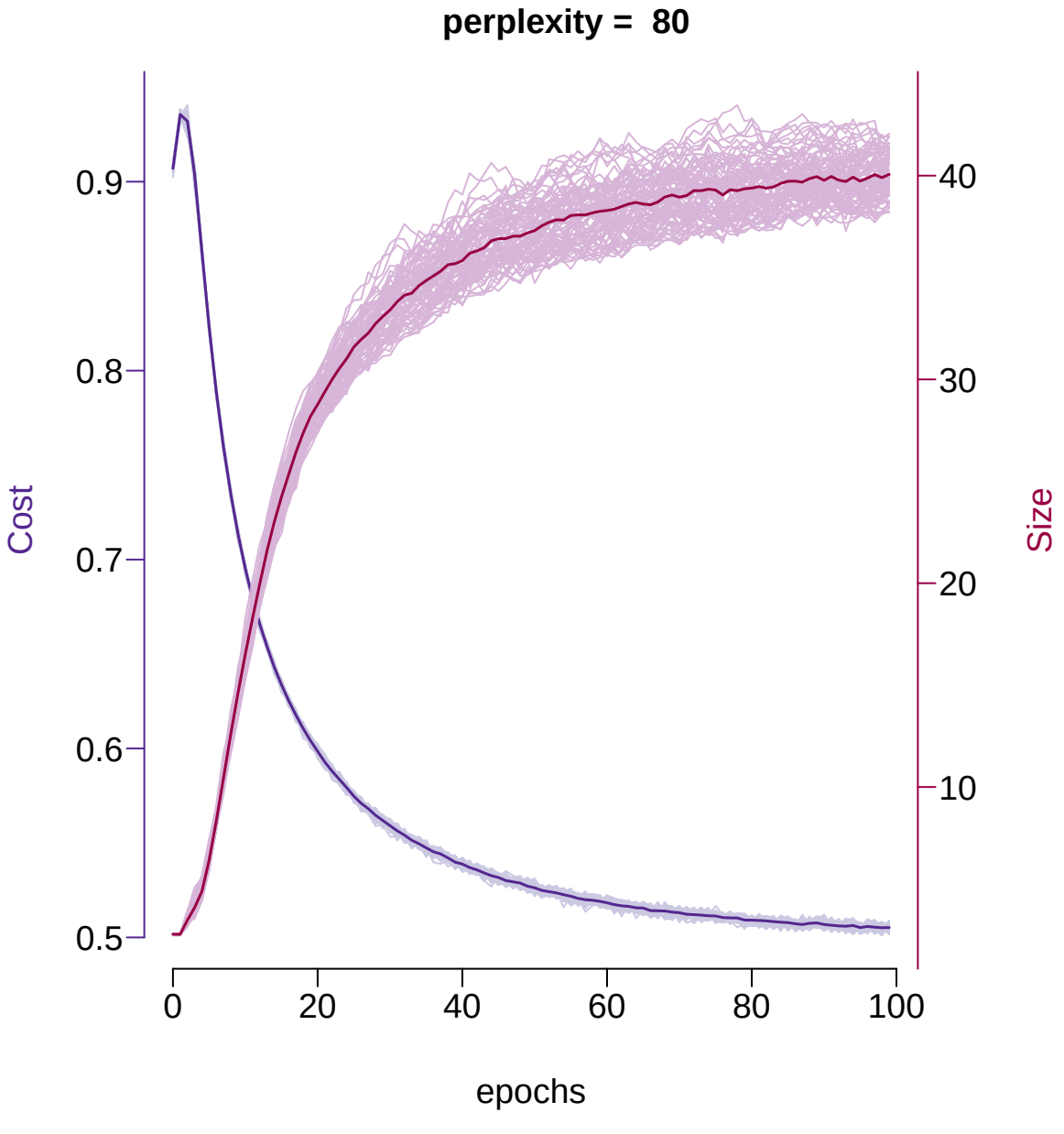

perplexity = 60
Y2
Y1
15
10
5
0
−5
−10
−15
−20
−10
0
10
20

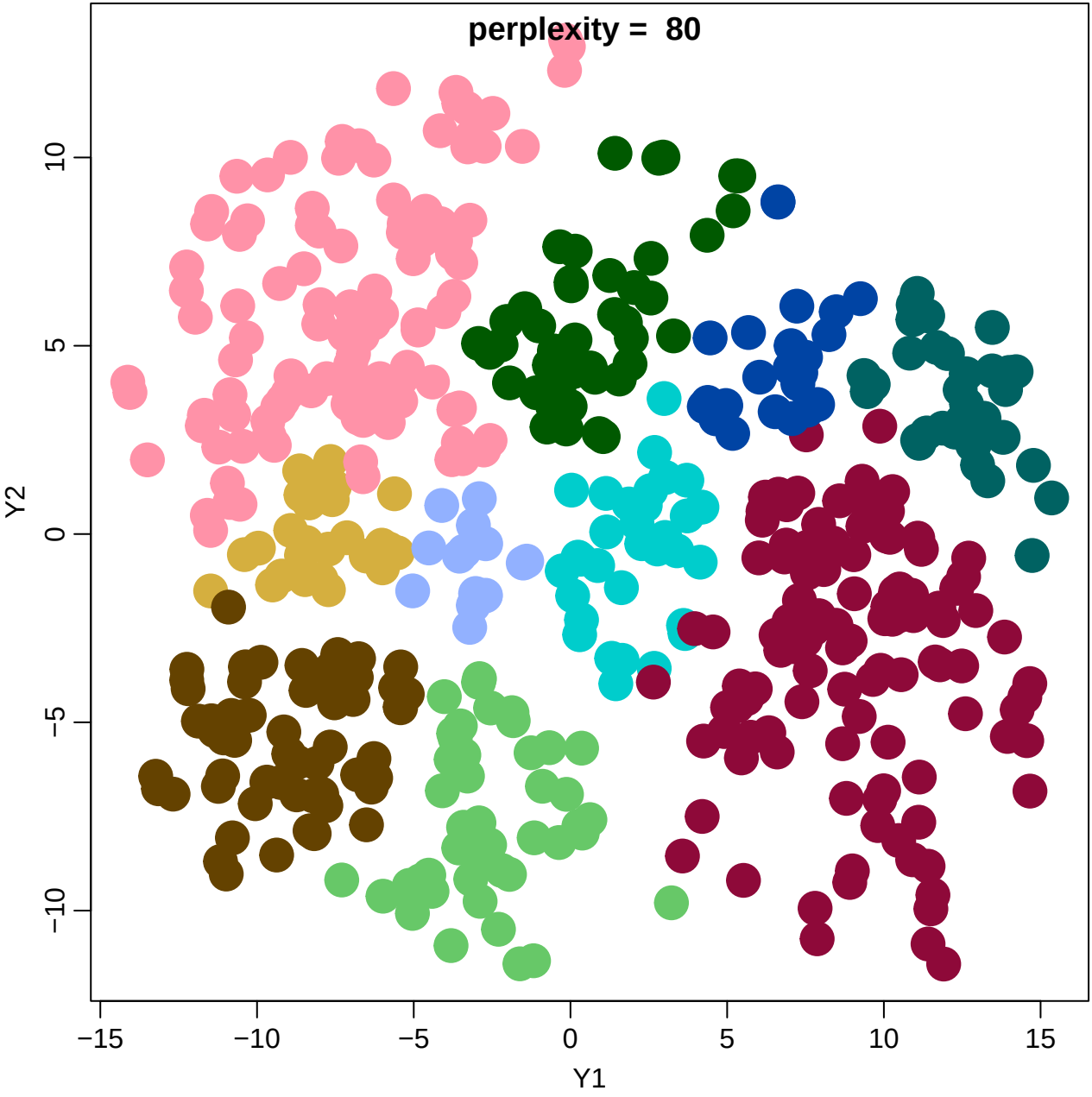

perplexity = 60
5
3
12
11
8
7
2
1
13
14
9
6
10
4

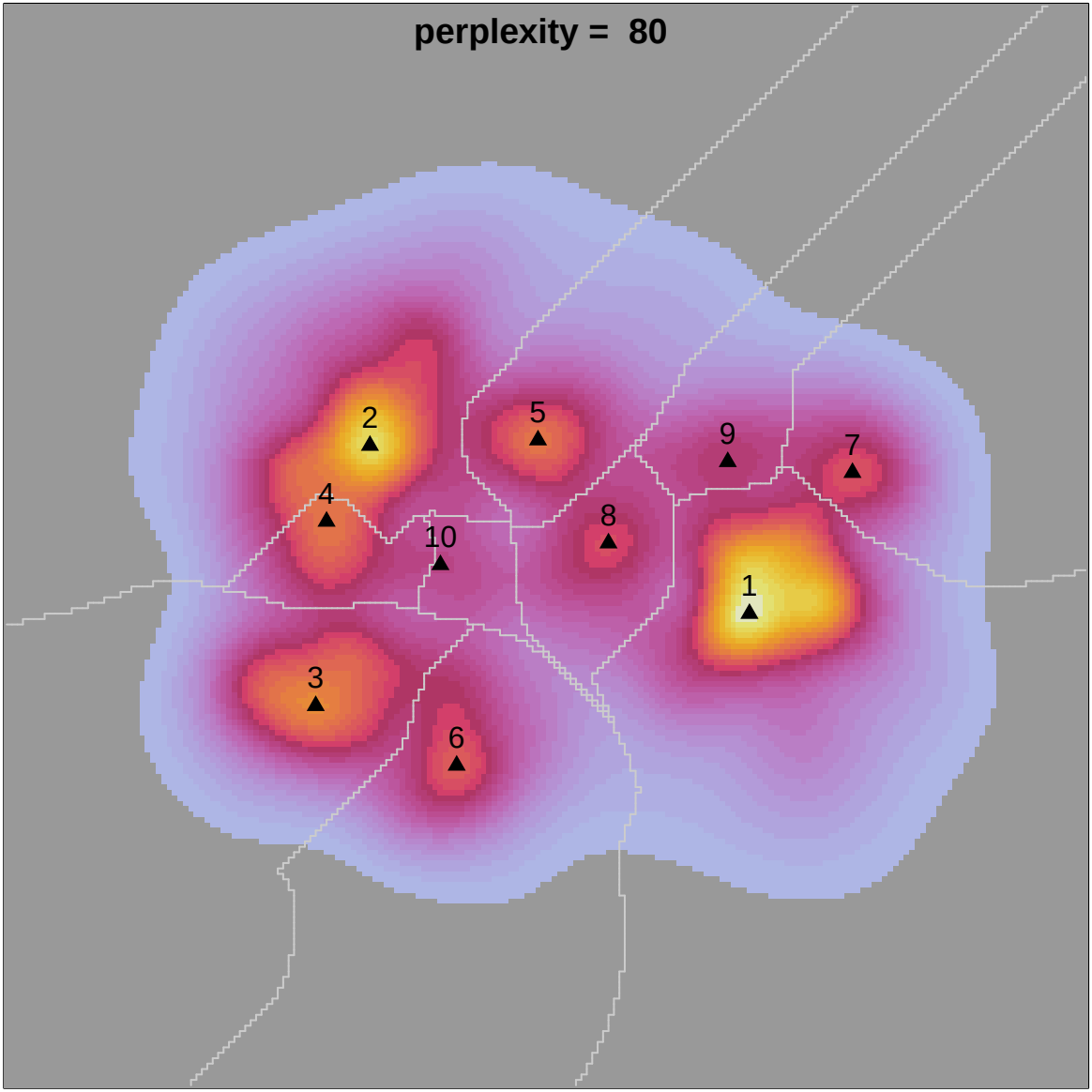

perplexity = 70
Cost
0.9
0.8
0.7
0.6
0.5
Size
50
40
30
20
10
0
20
40
60
80
100
epochs

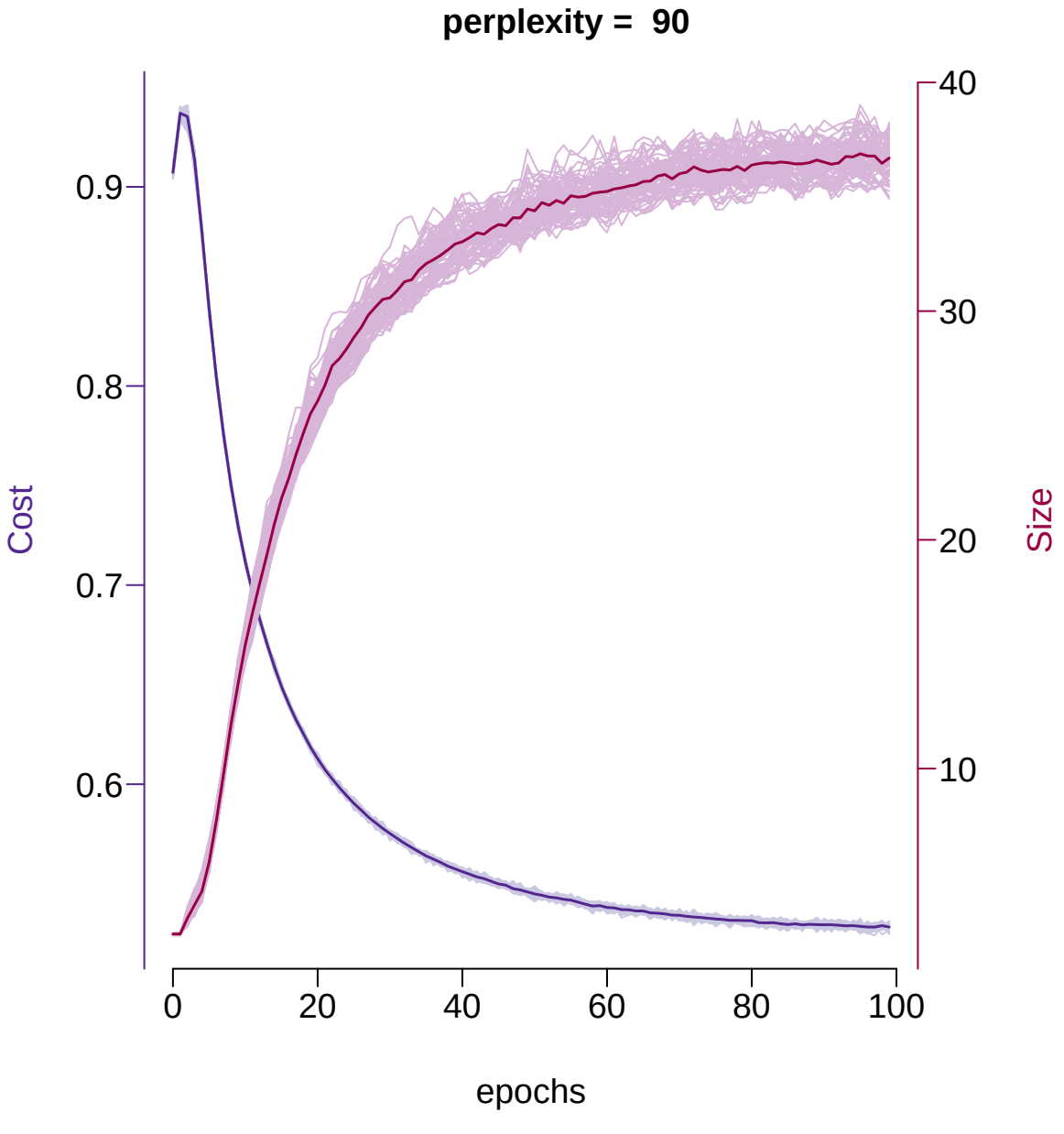

perplexity = 70
Y2
Y1
20
15
10
5
0
−5
−10
−15
−15
−10
−5
0
5
10
15

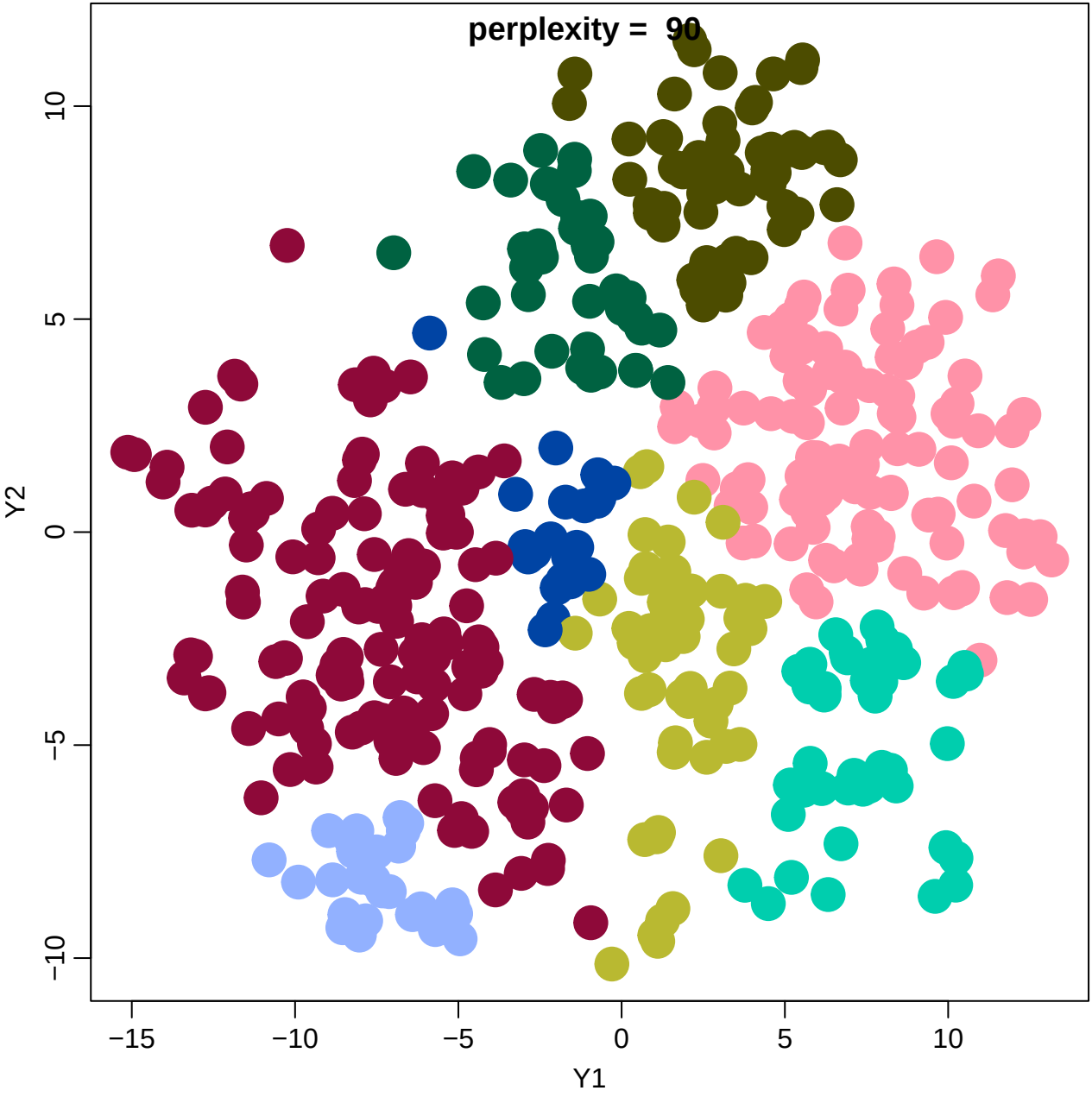

perplexity = 70
8
2
1
10
9
5
7
11
3
4
6

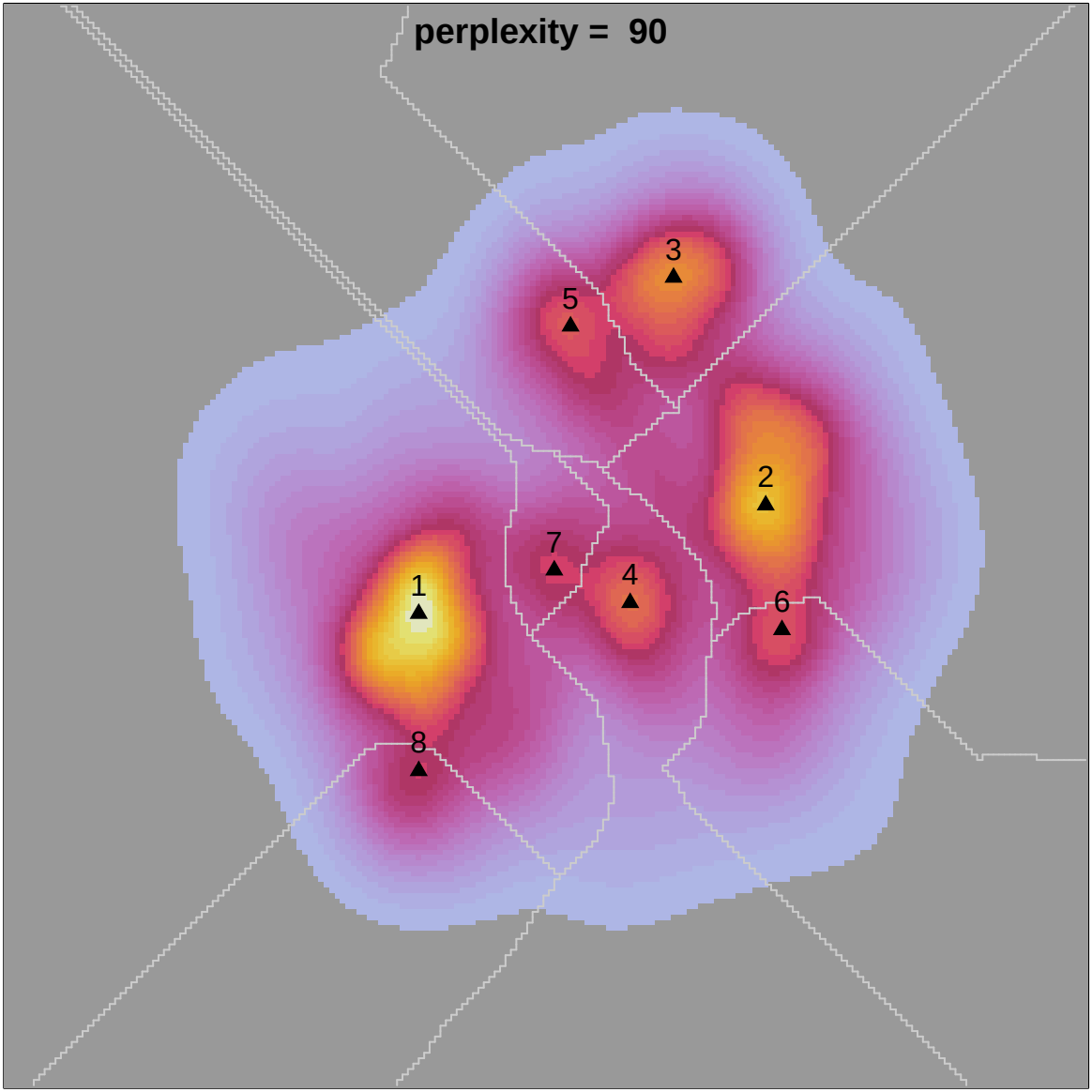

perplexity = 80
Cost
Size
0.9
0.8
0.7
0.6
0.5
40
30
20
10
0
20
40
60
80
100
epochs

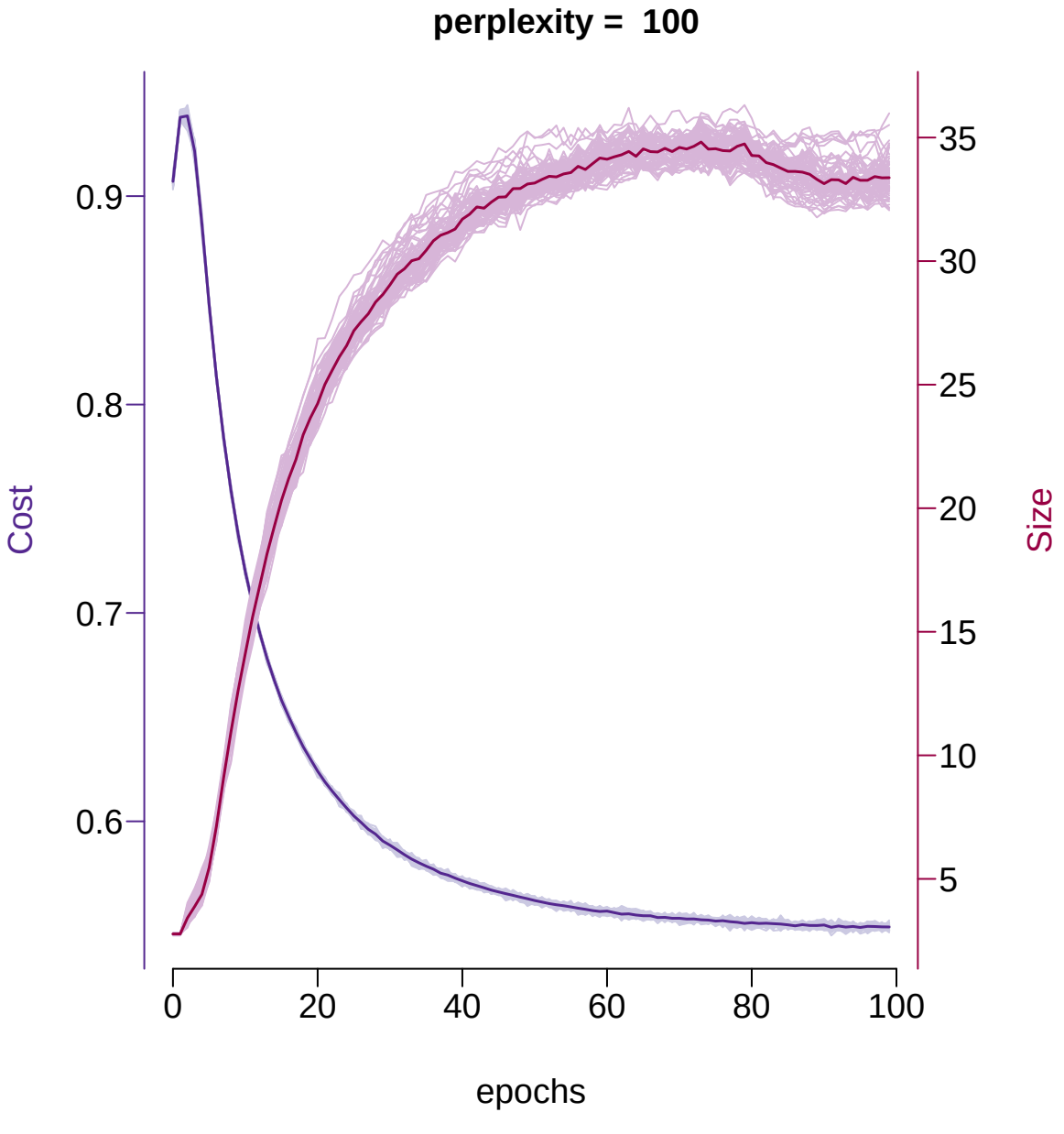

perplexity = 80
Y2
Y1
10
5
0
−5
−10
−15
−10
−5
0
5
10
15

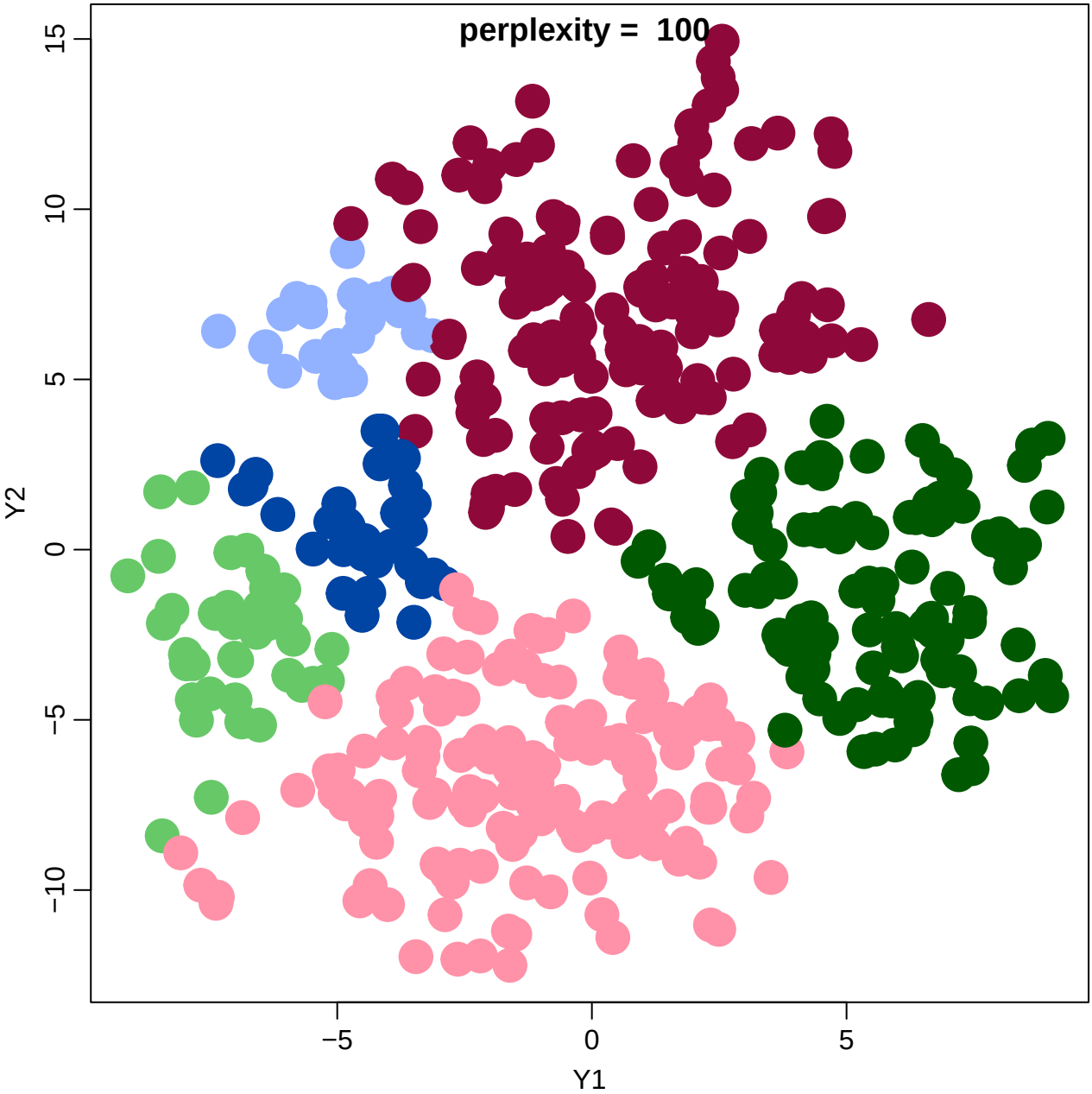

perplexity = 80
2
5
9
7
4
8
10
1
3
6

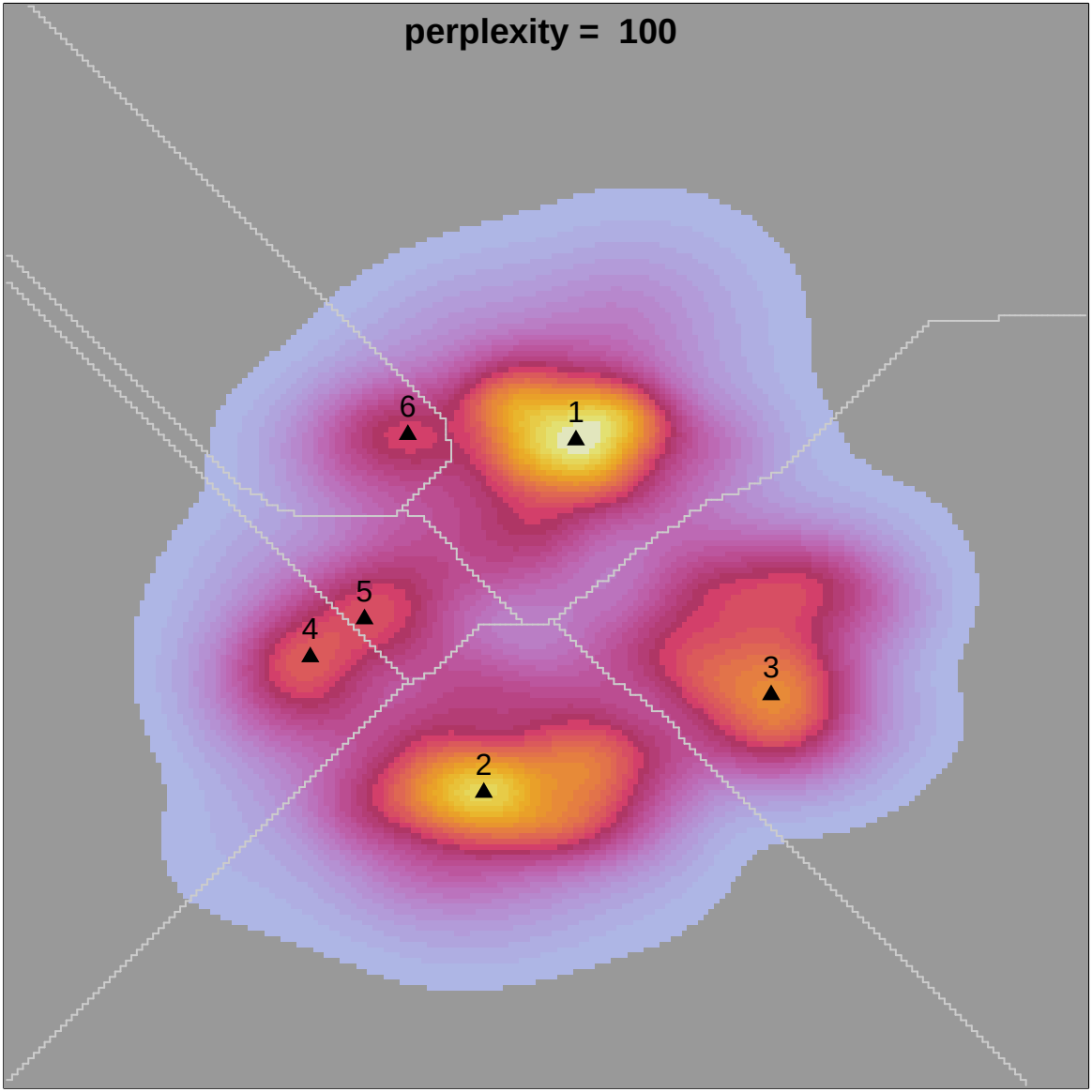

perplexity = 90
Cost
0.9
0.8
0.7
0.6
Size
40
30
20
10
0
20
40
60
80
100
epochs

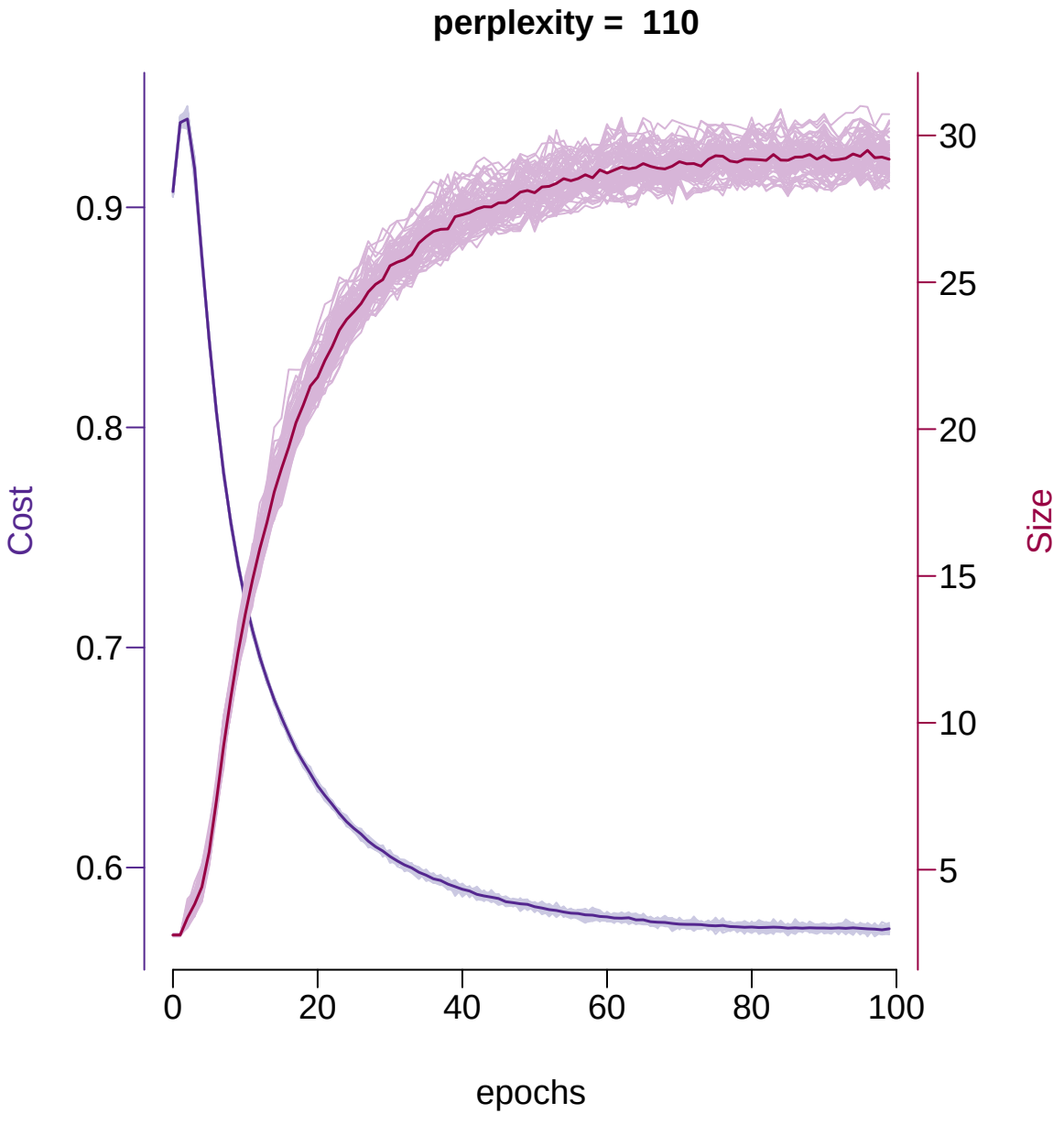

perplexity = 90
Y2
Y1
10
5
0
−5
−10
−15
−10
−5
0
5
10

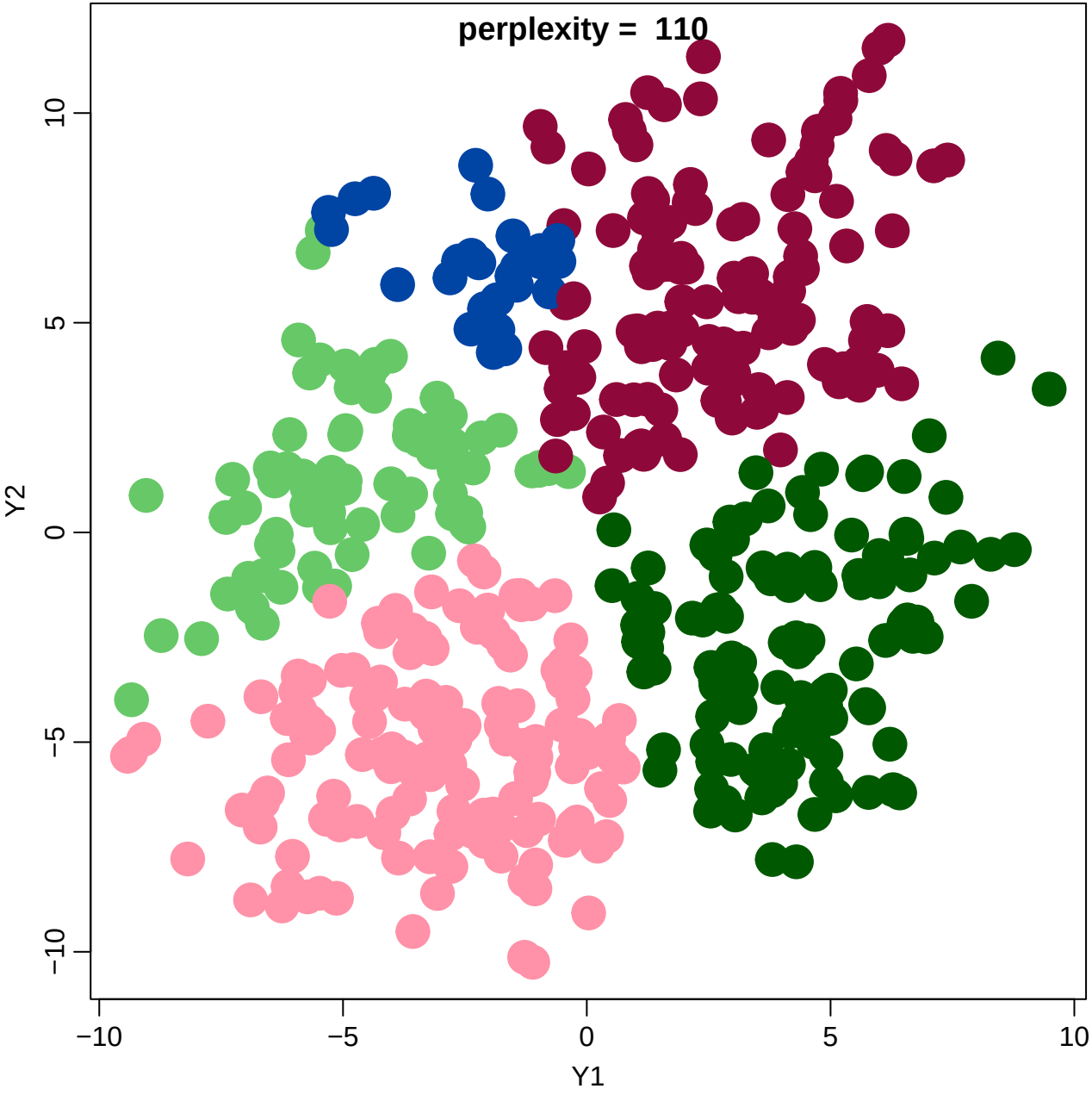

perplexity = 90
3
5
2
7
4
1
6
8

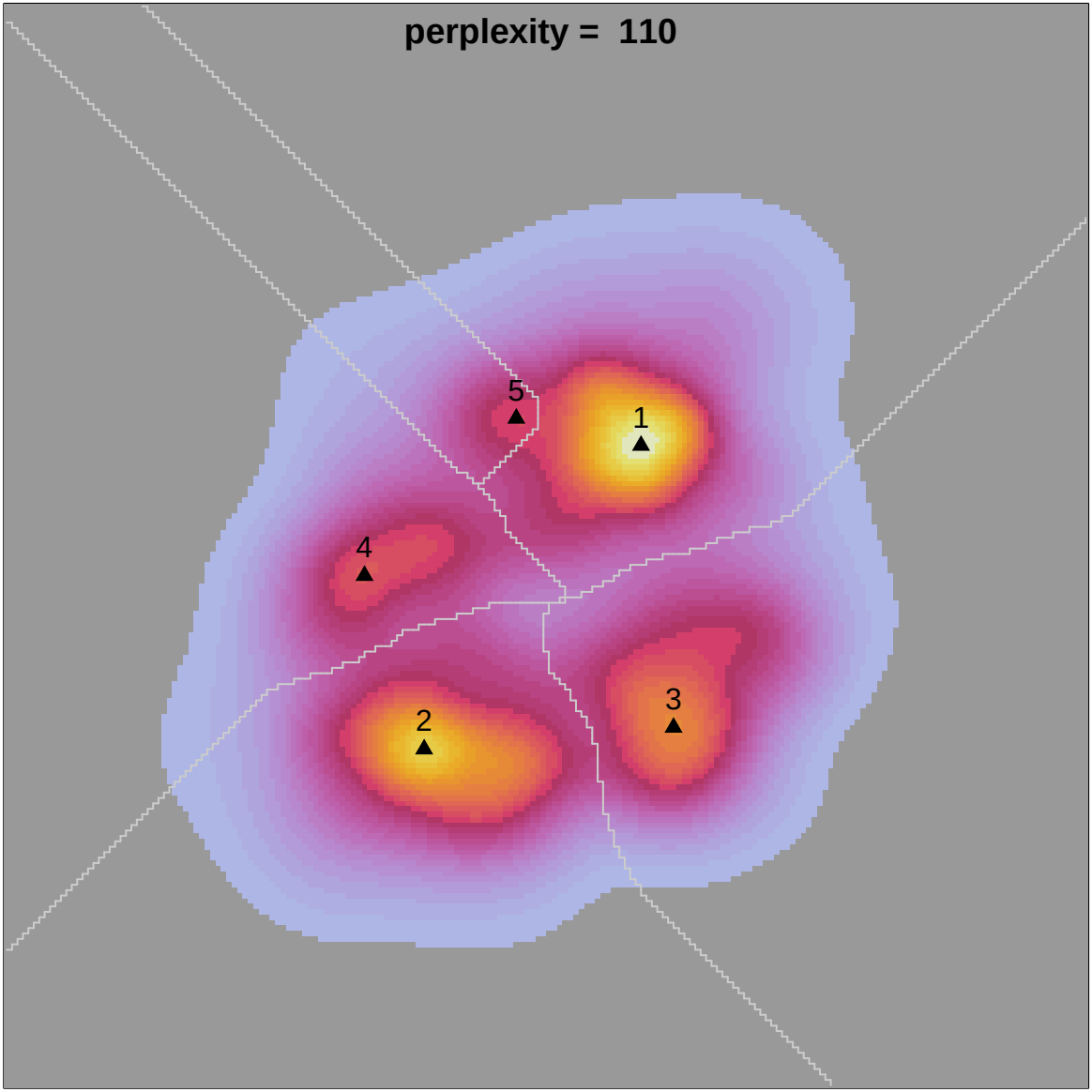

perplexity = 100
Cost
0.9
0.8
0.7
0.6
Size
35
30
25
20
15
10
5
0
20
40
60
80
100
epochs

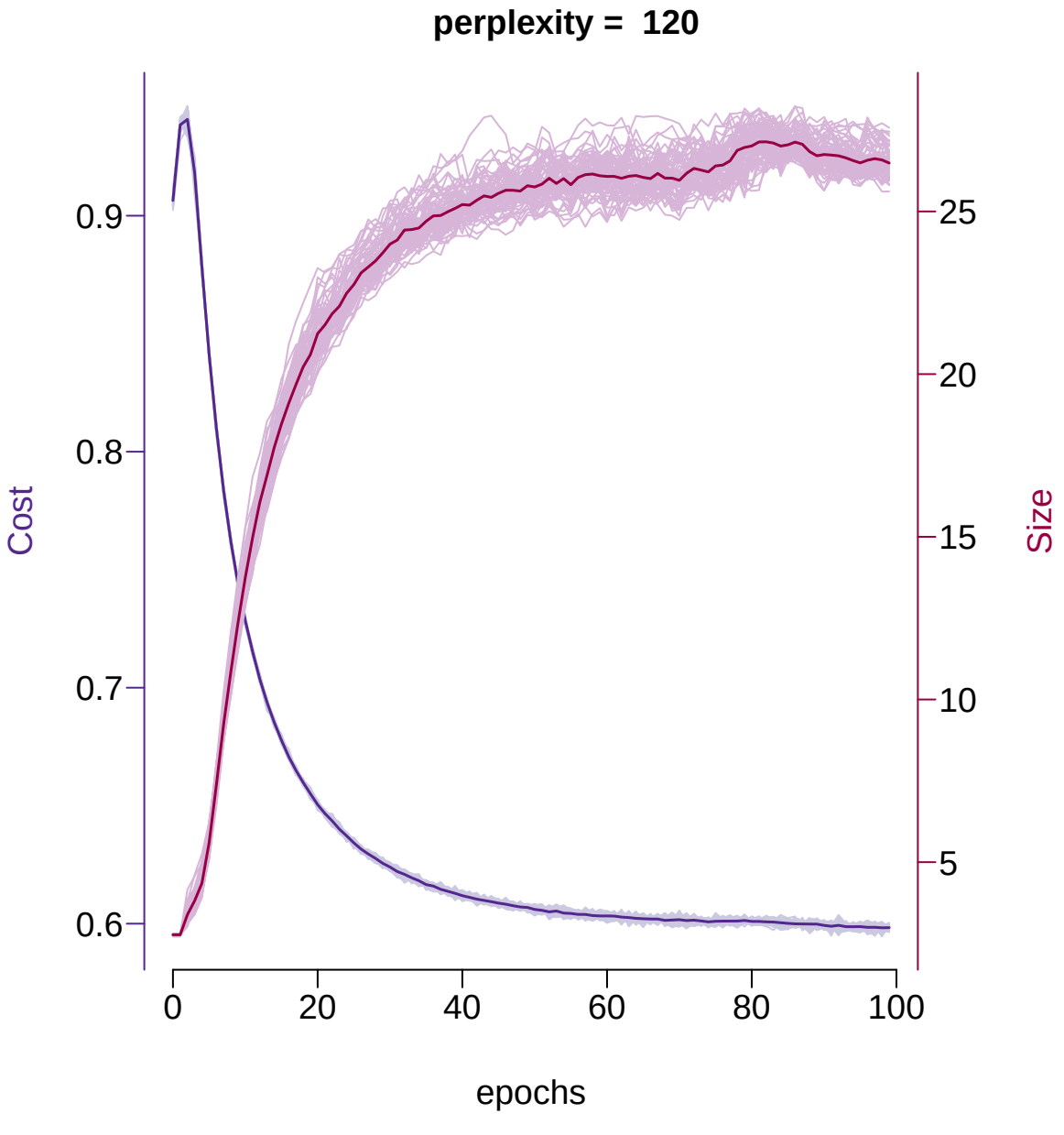

perplexity = 100
Y2
Y1
15
10
5
0
−5
−10
−5
0
5

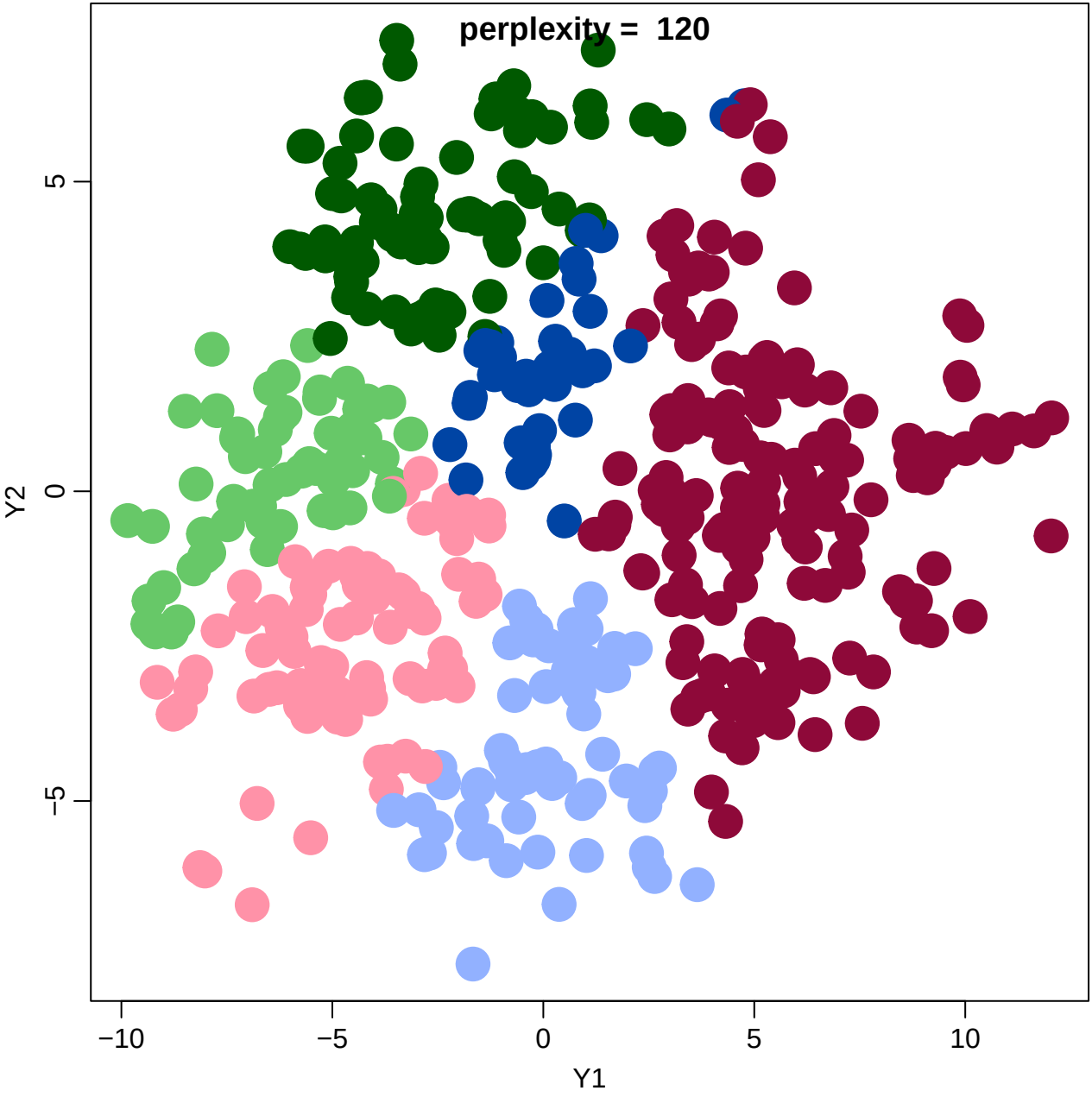

perplexity = 100
1
2
3
4
5
6

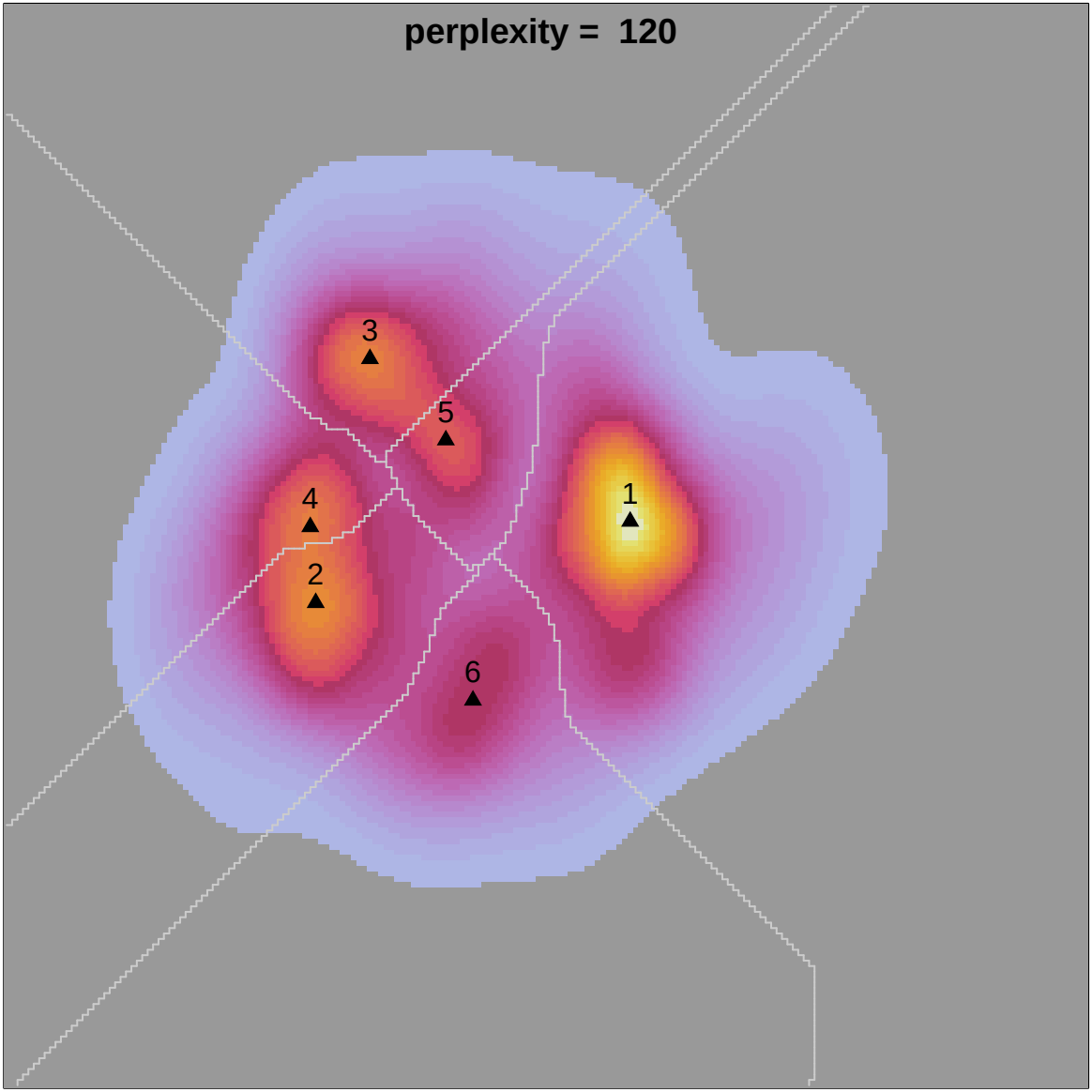

perplexity = 110
Cost
Size
0.9
0.8
0.7
0.6
30
25
20
15
10
5
0
20
40
60
80
100
epochs

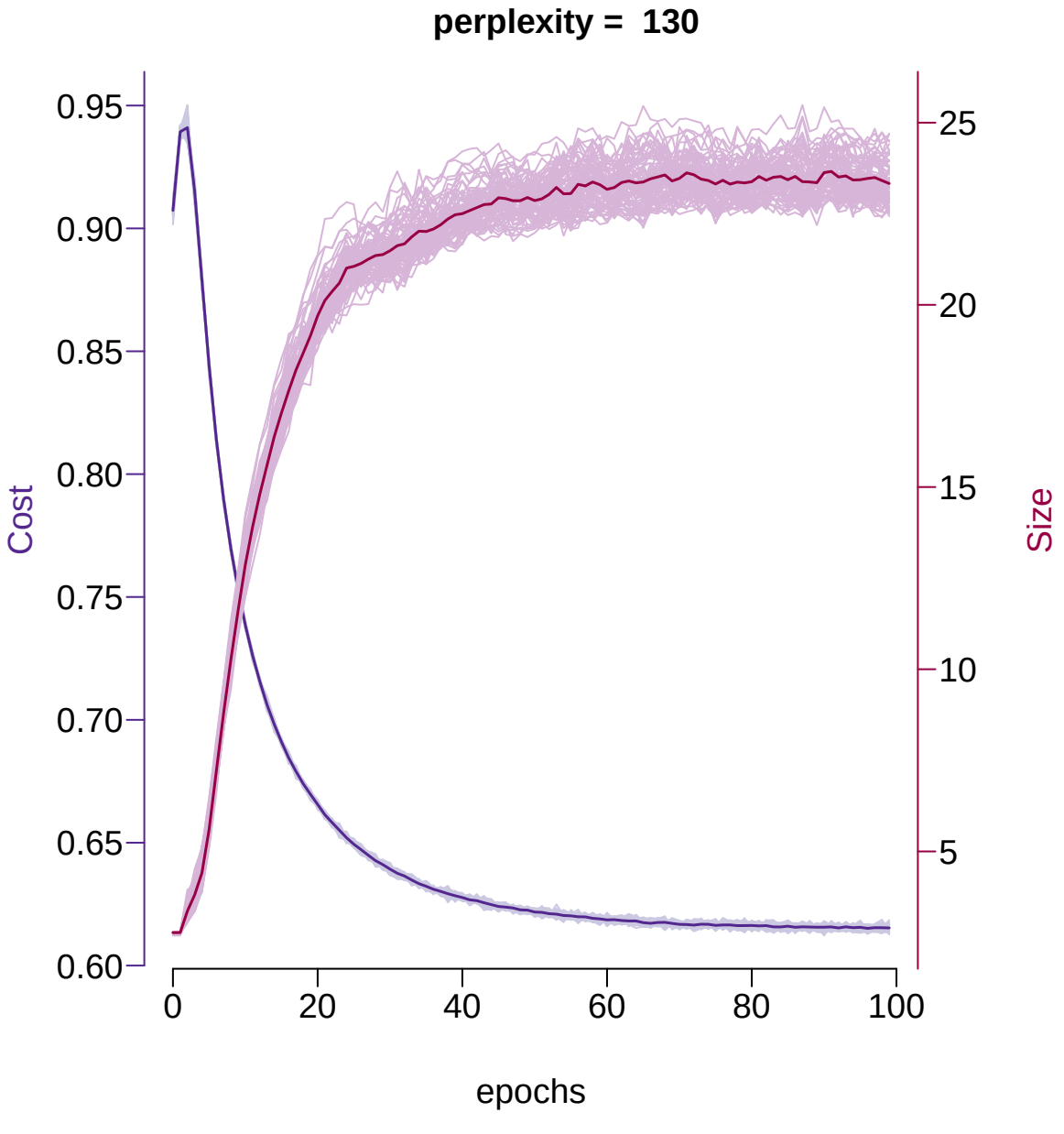

perplexity = 110
Y2
Y1
10
5
0
−5
−10
−10
−5
0
5
10

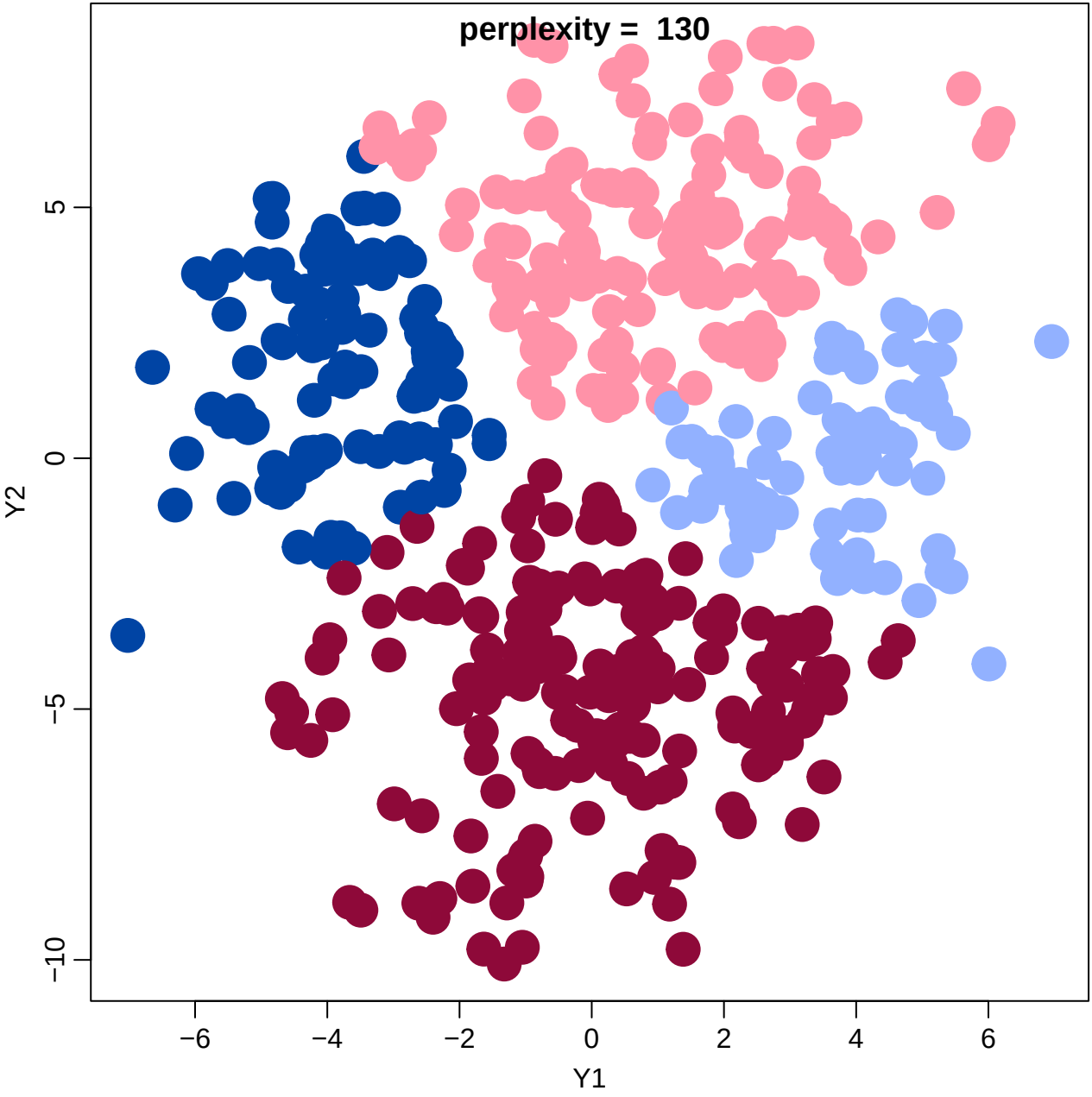

perplexity = 110
5
1
4
3
2

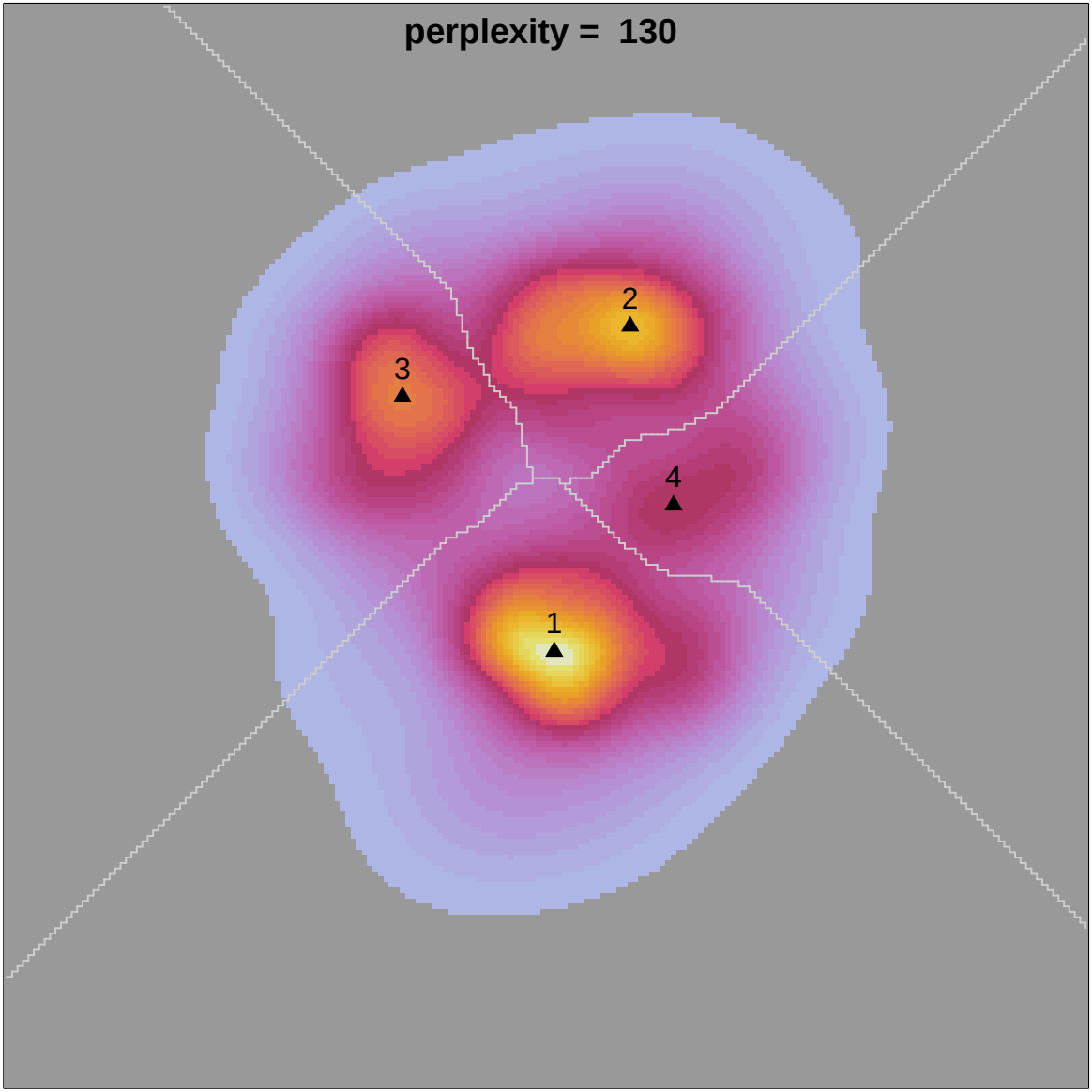

perplexity = 120
Cost
Size
0.9
0.8
0.7
0.6
25
20
15
10
5
0
20
40
60
80
100
epochs

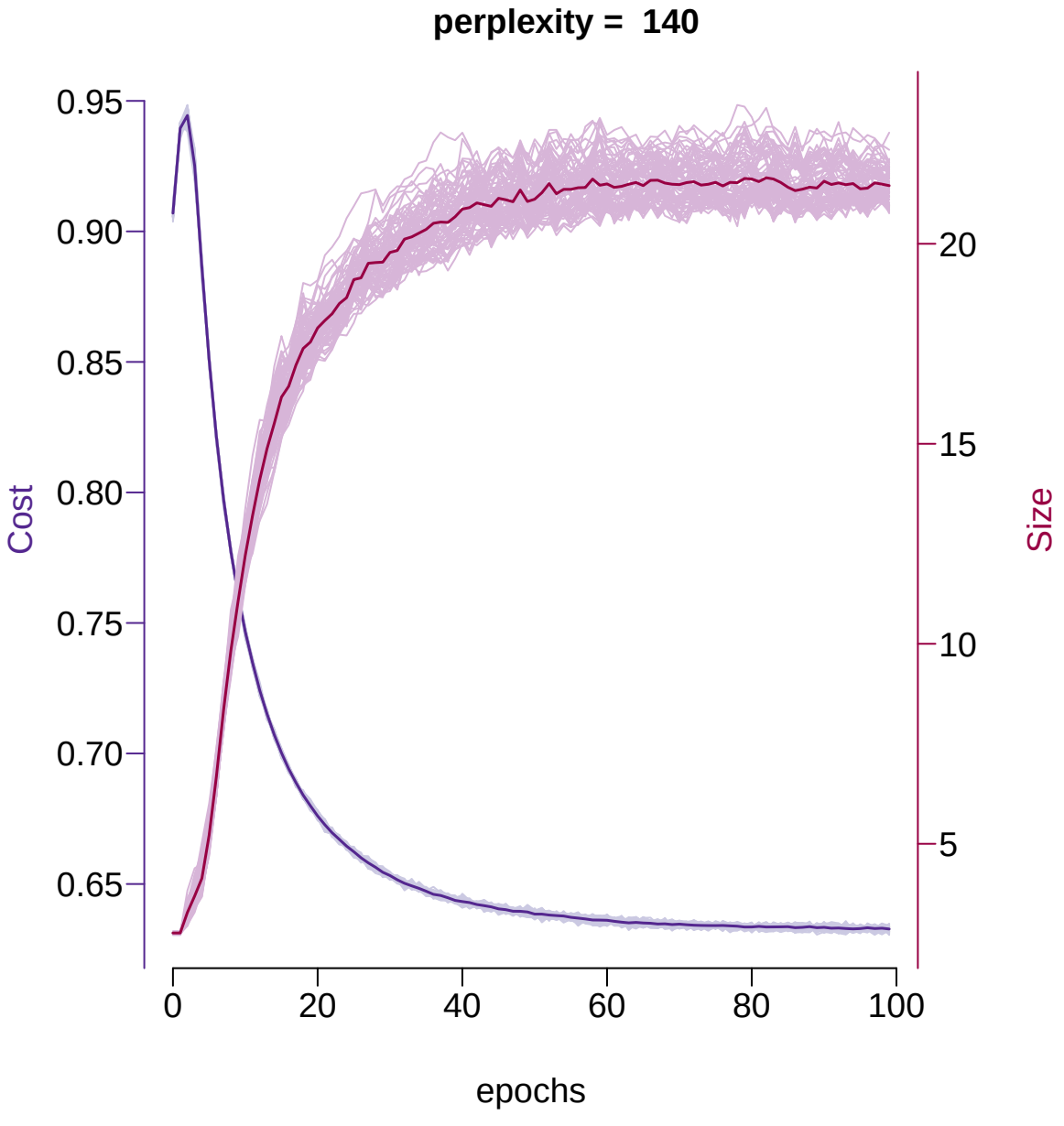

perplexity = 120
Y2
5
0
−5
−10
−5
0
5
10
Y1

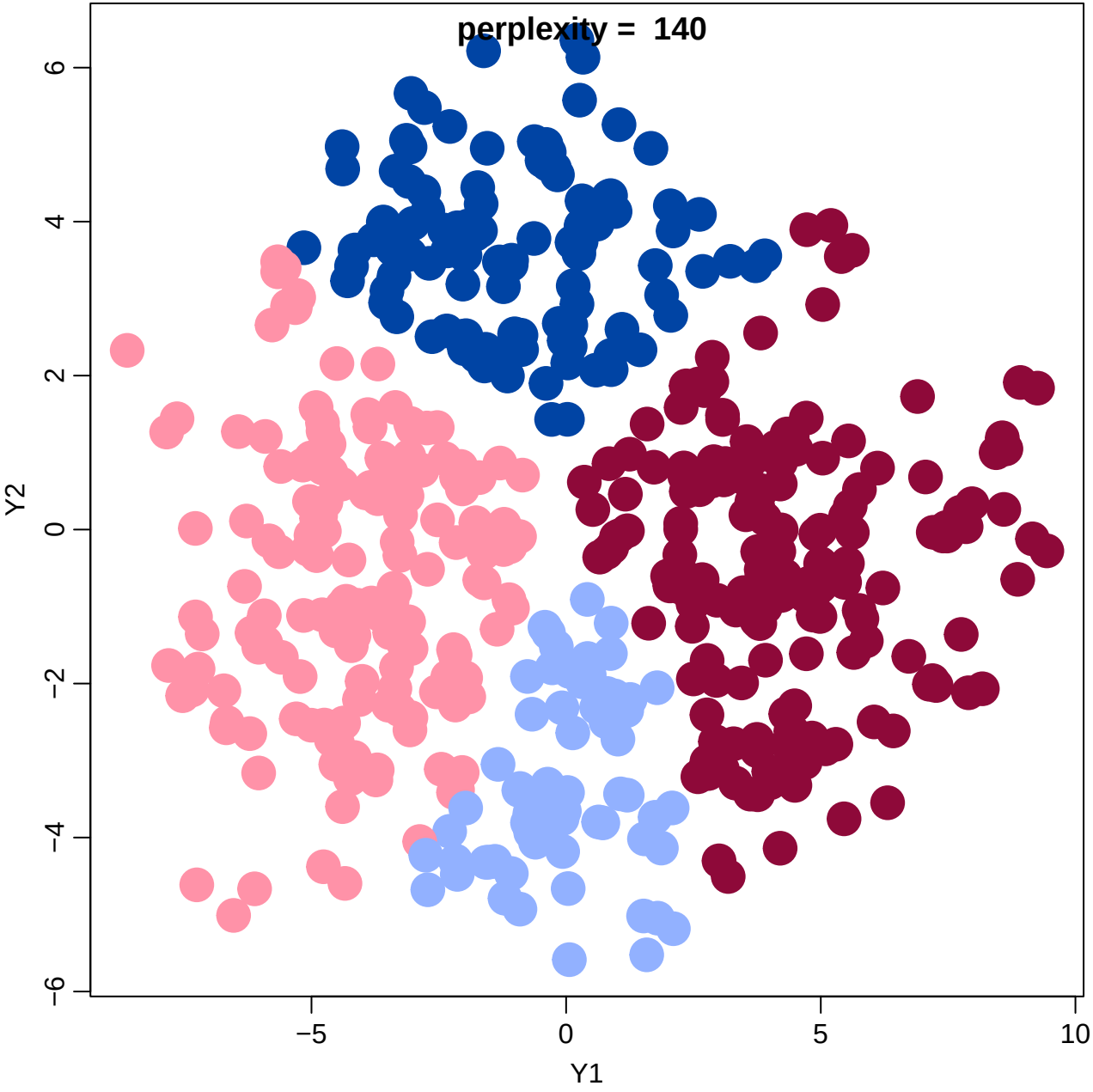

perplexity = 120
3
5
4
1
2
6

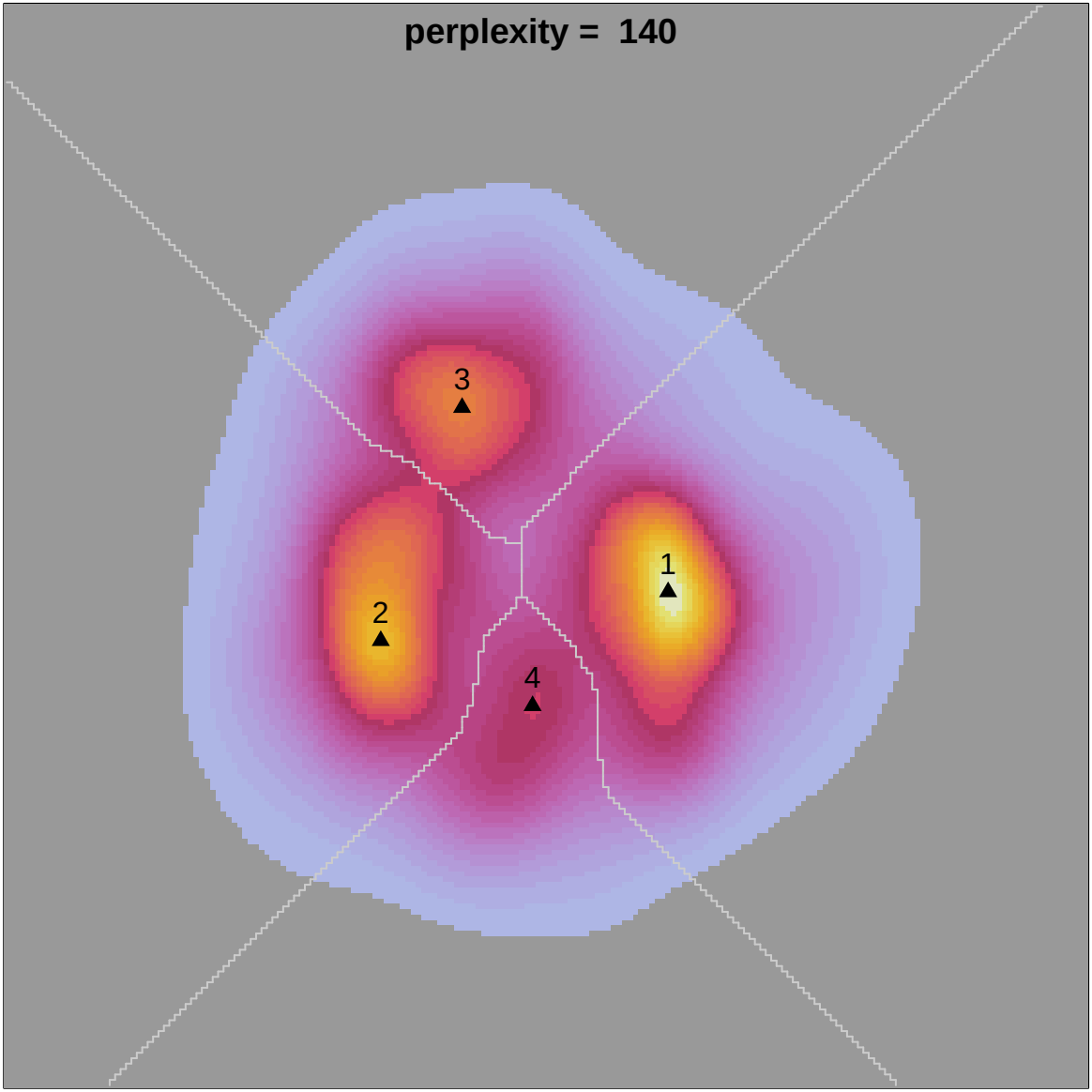

perplexity = 130
Cost
Size
epochs
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
25
20
15
10
5
0
20
40
60
80
100

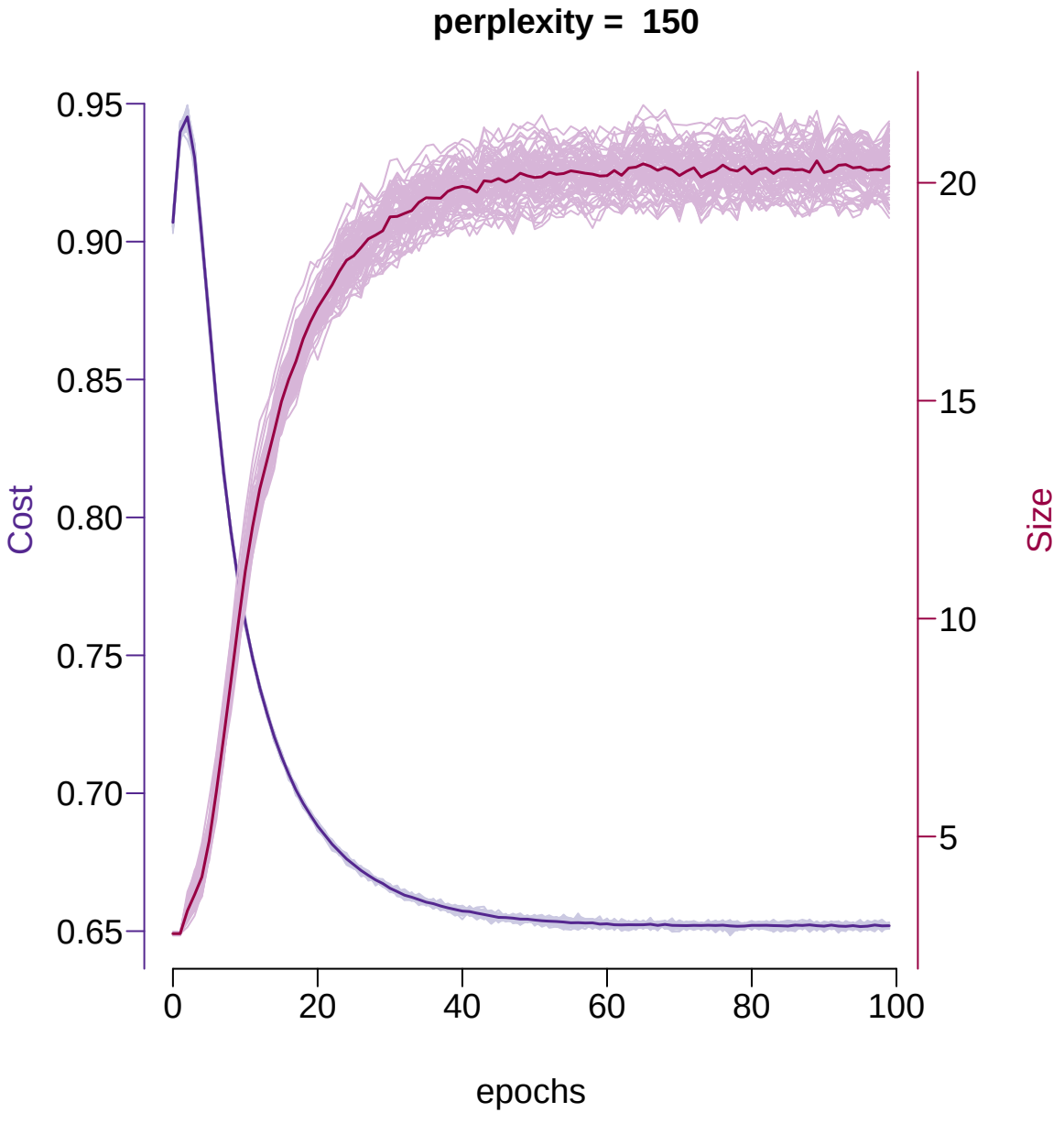

perplexity = 130
Y2
Y1
5
0
−5
−10
−6
−4
−2
0
2
4
6

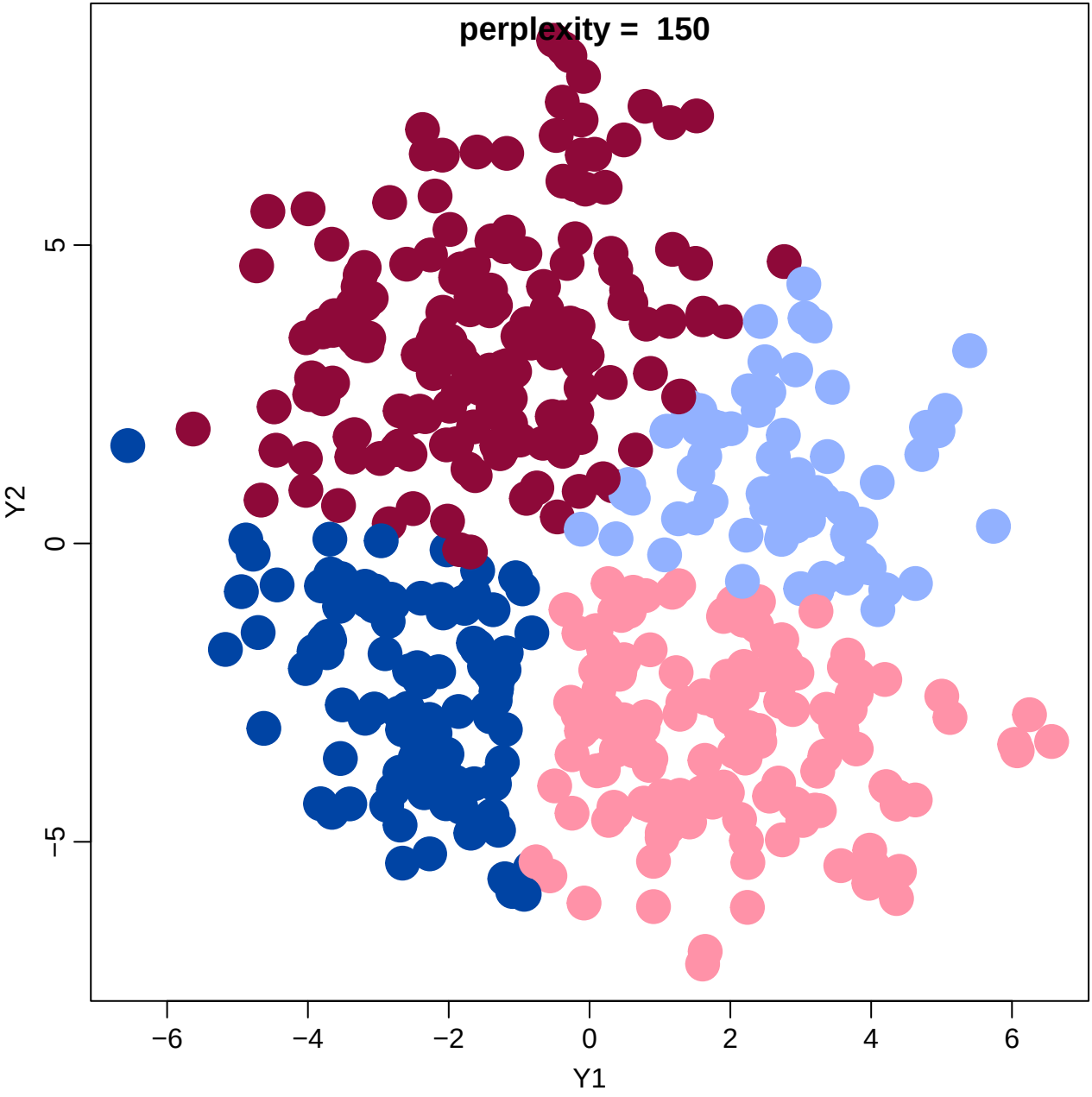

perplexity = 130
2
3
4
1

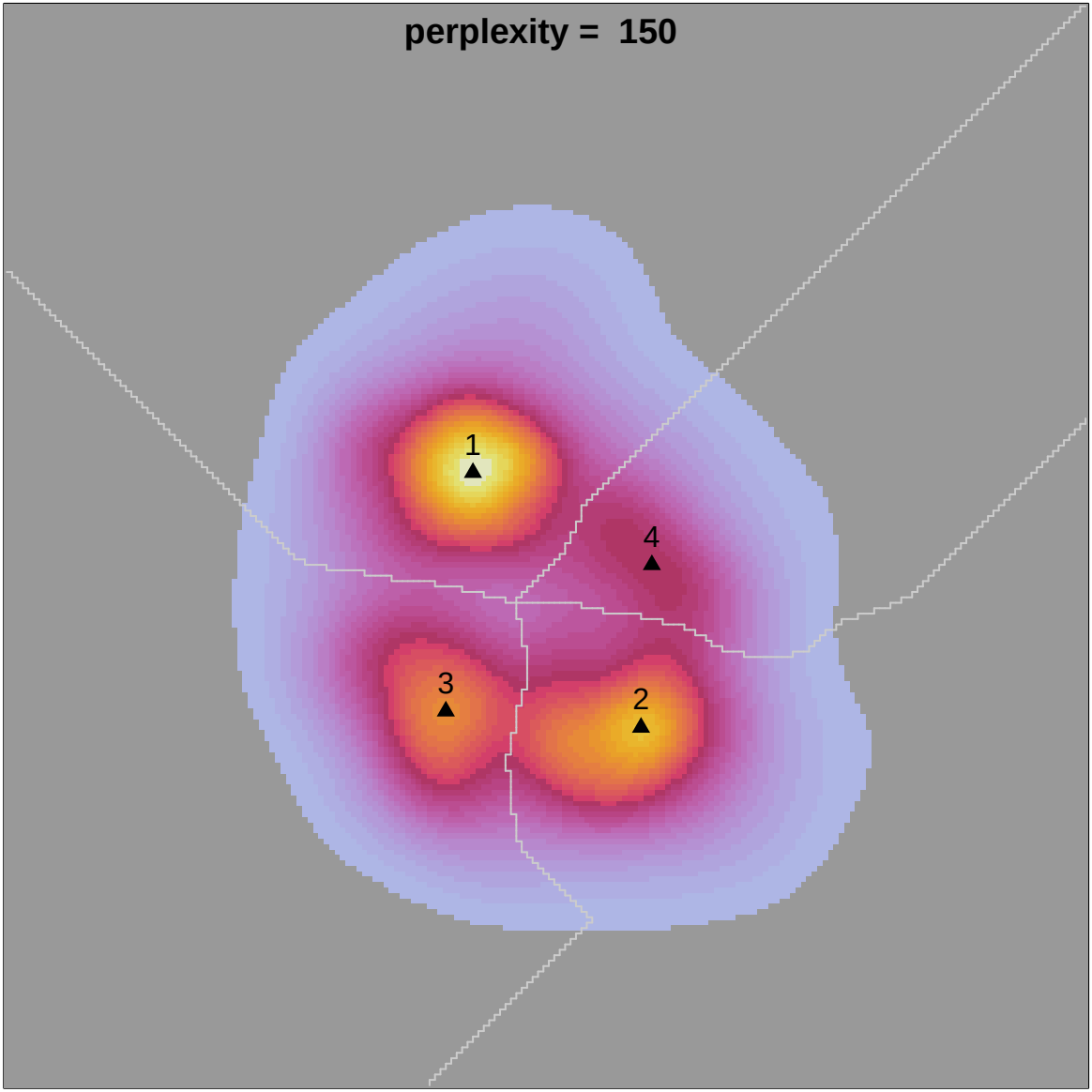

perplexity = 140
Cost
Size
epochs
0.95
0.90
0.85
0.80
0.75
0.70
0.65
20
15
10
5
0
20
40
60
80
100

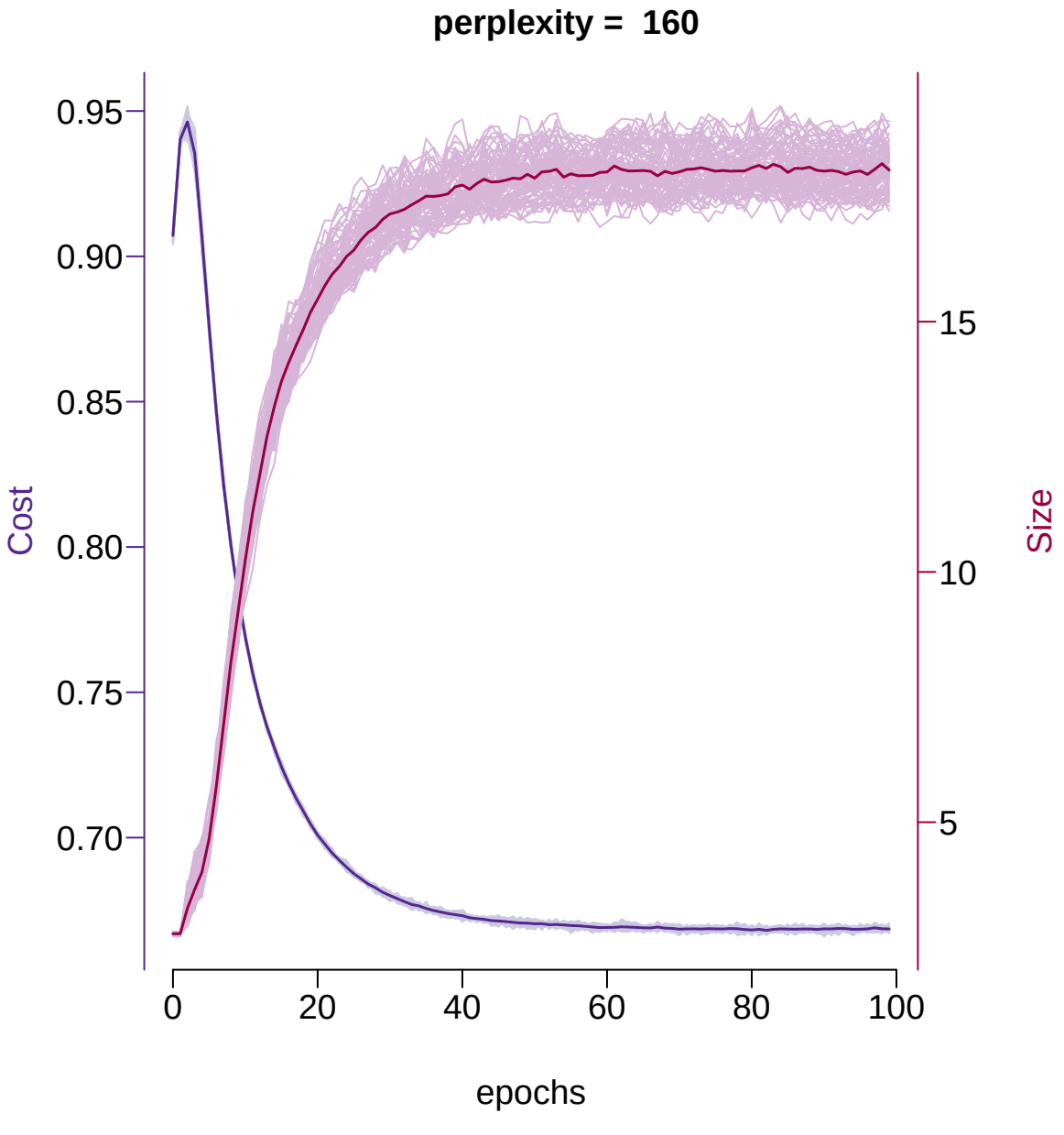

perplexity = 140
Y2
Y1
6
4
2
0
−2
−4
−6
−5
0
5
10

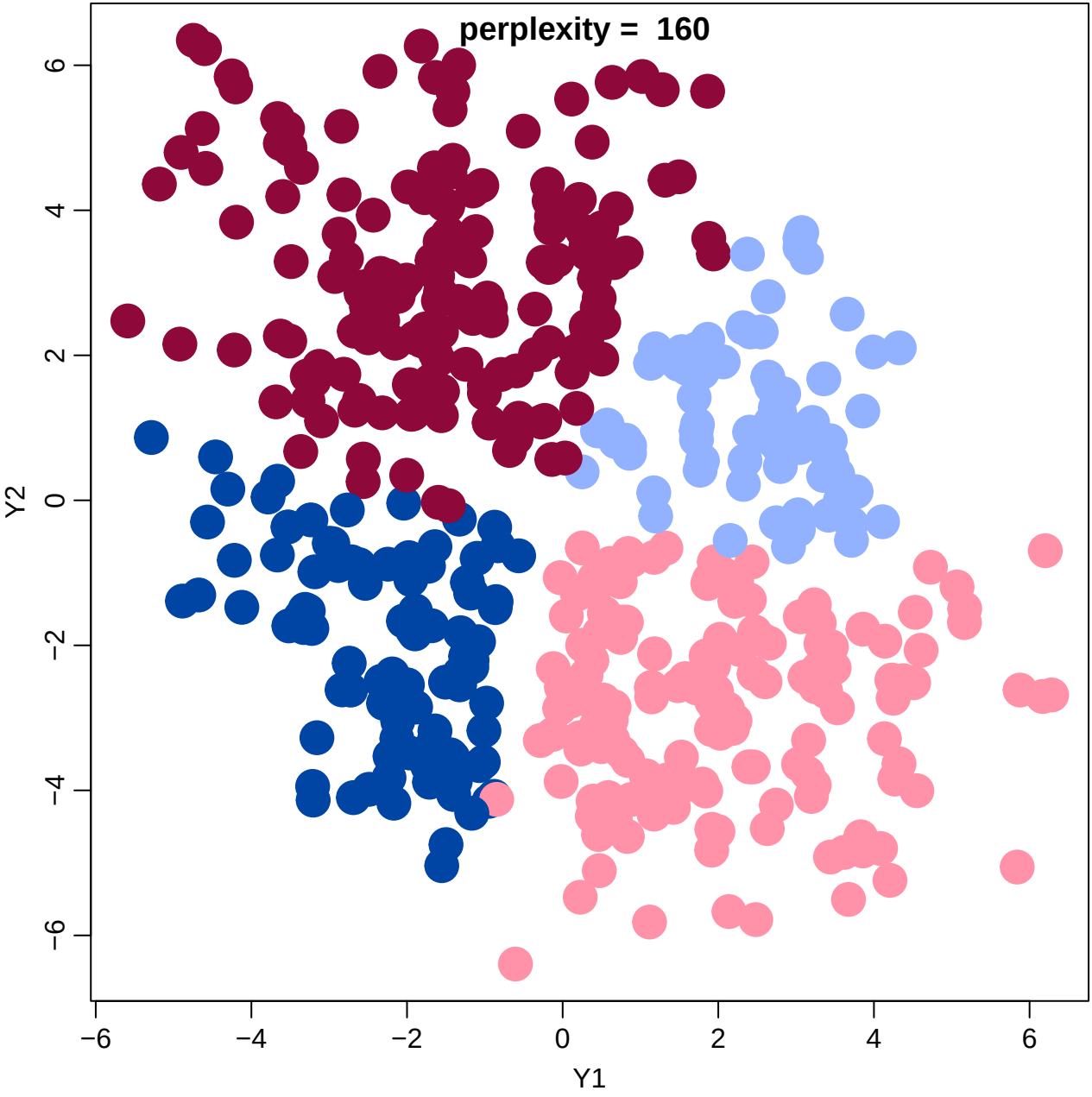

perplexity = 140
3
1
2
4

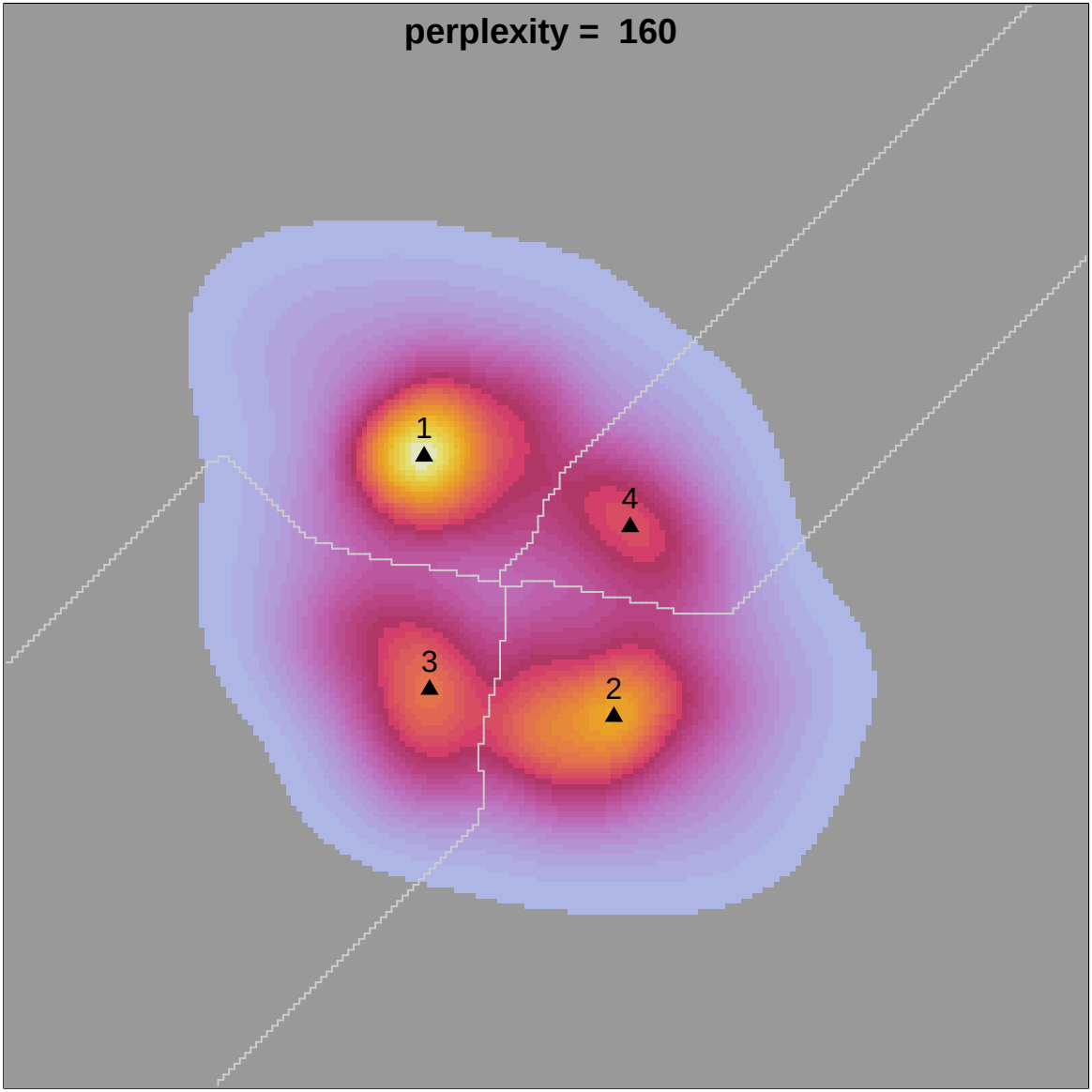

perplexity = 150
Cost
Size
epochs
0.95
0.90
0.85
0.80
0.75
0.70
0.65
20
15
10
5
0
20
40
60
80
100

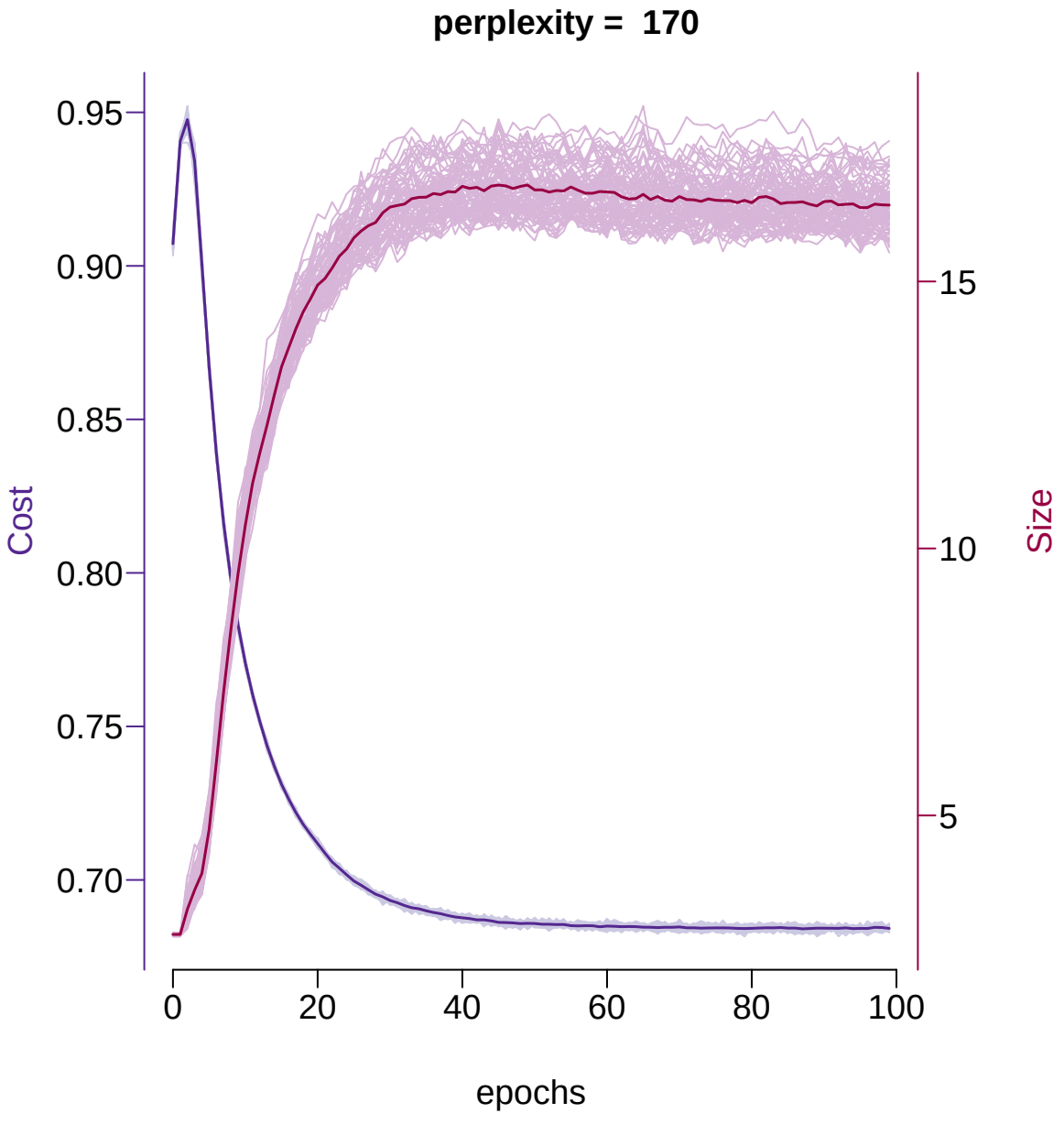

perplexity = 150
Y2
5
0
−5
−6
−4
−2
0
2
4
6
Y1

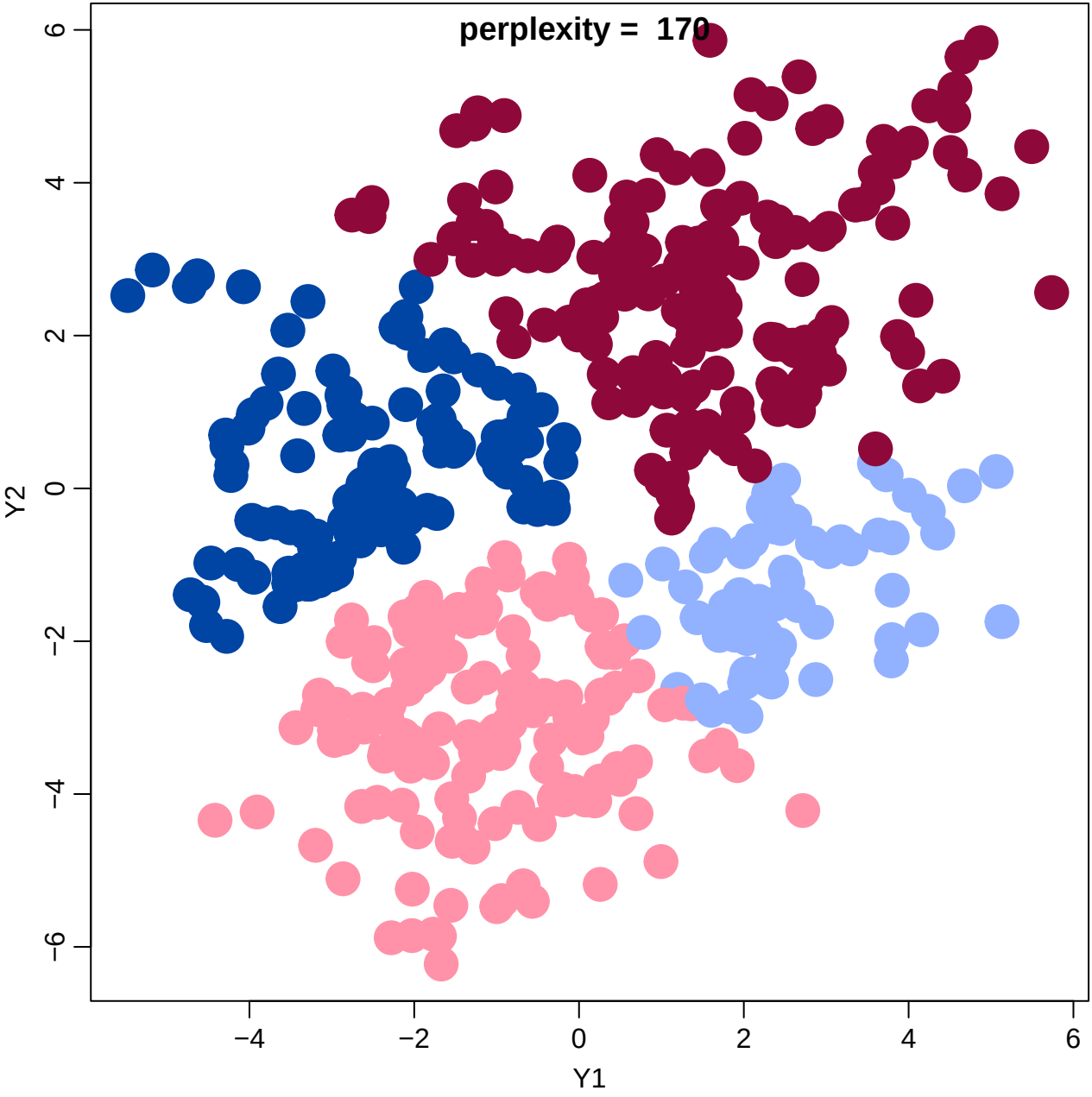

perplexity = 150
1
4
3
2

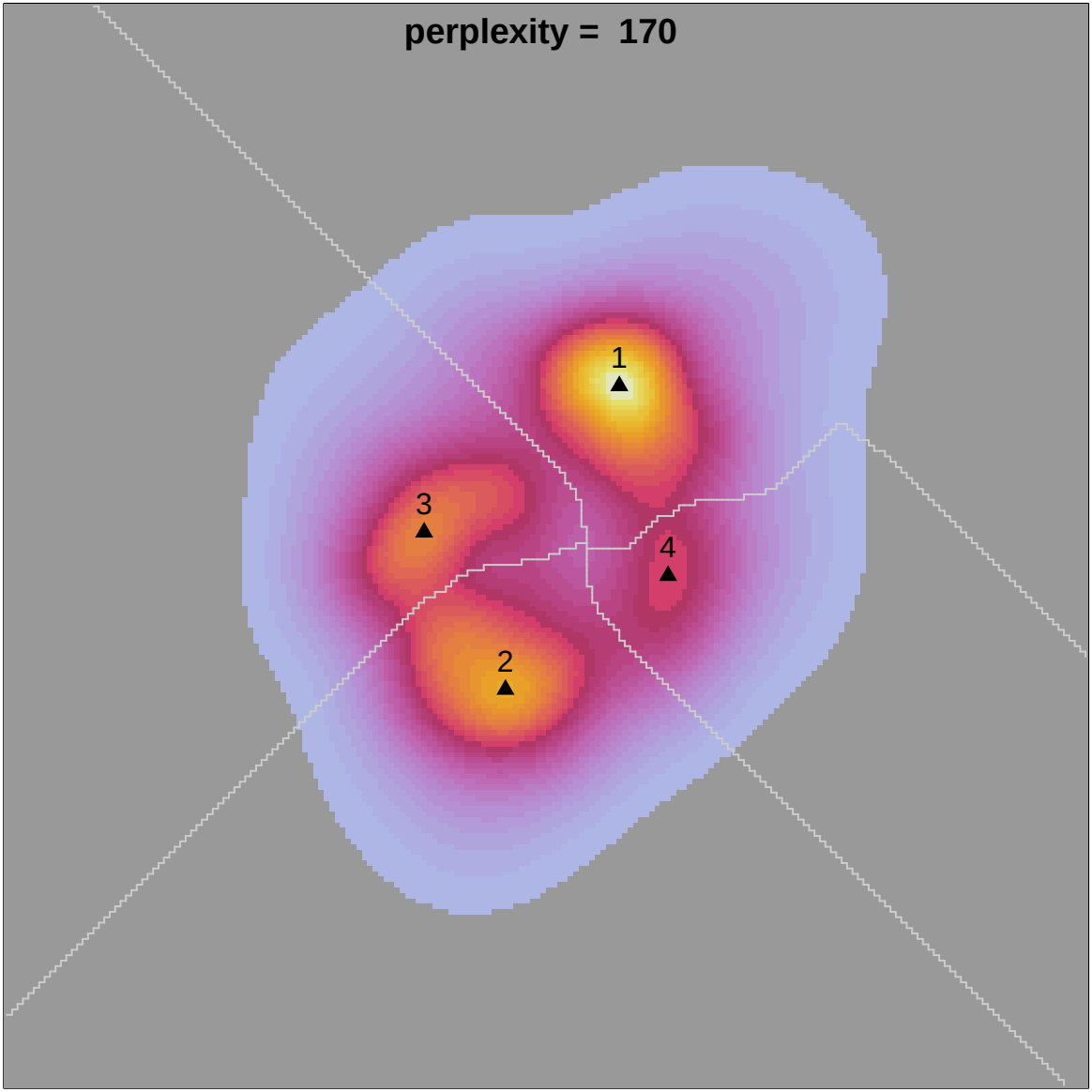

perplexity = 160
Cost
Size
0.95
0.90
0.85
0.80
0.75
0.70
15
10
5
0
20
40
60
80
100
epochs

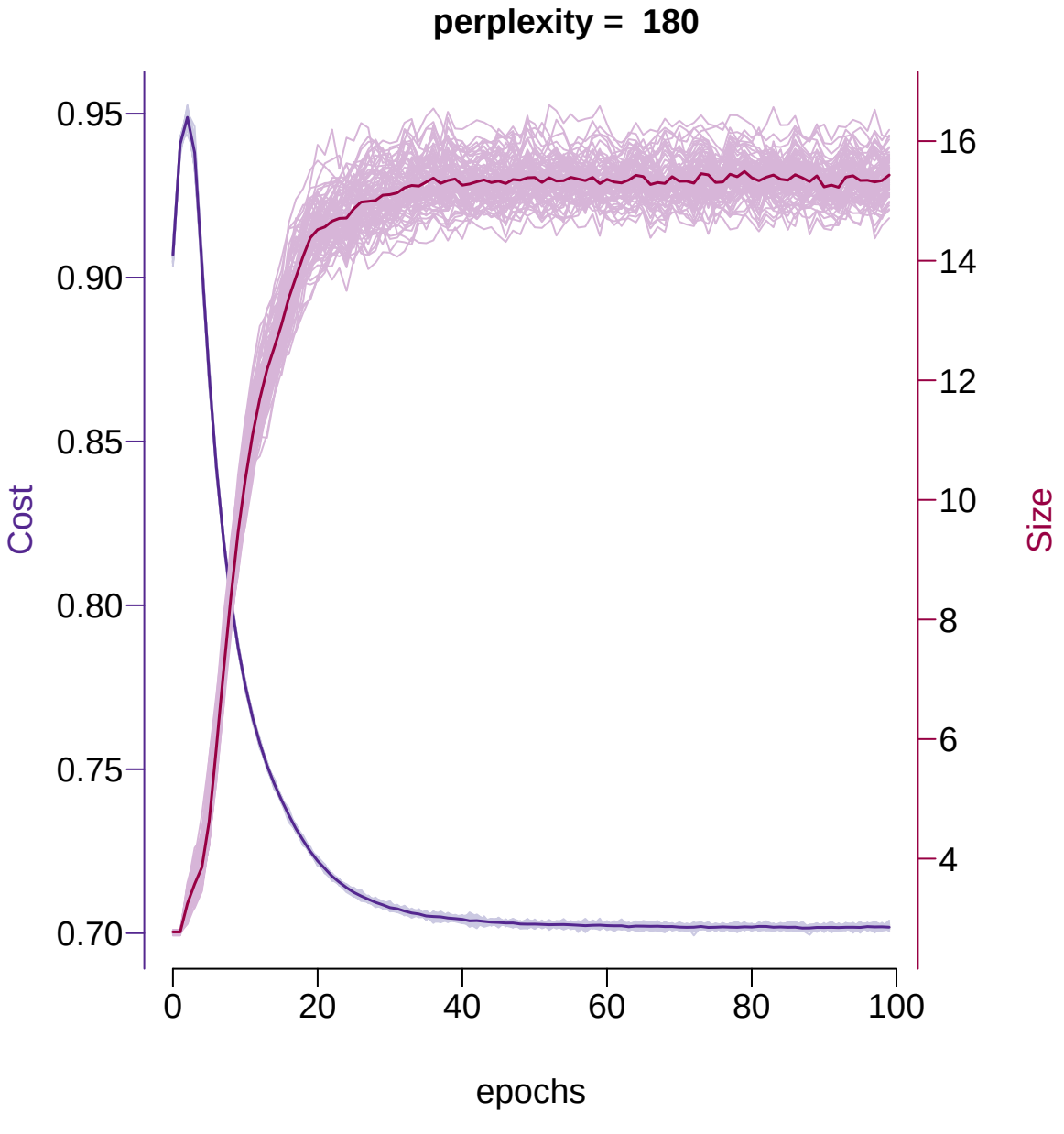

perplexity = 160
Y2
Y1
6
4
2
0
−2
−4
−6
−6
−4
−2
0
2
4
6

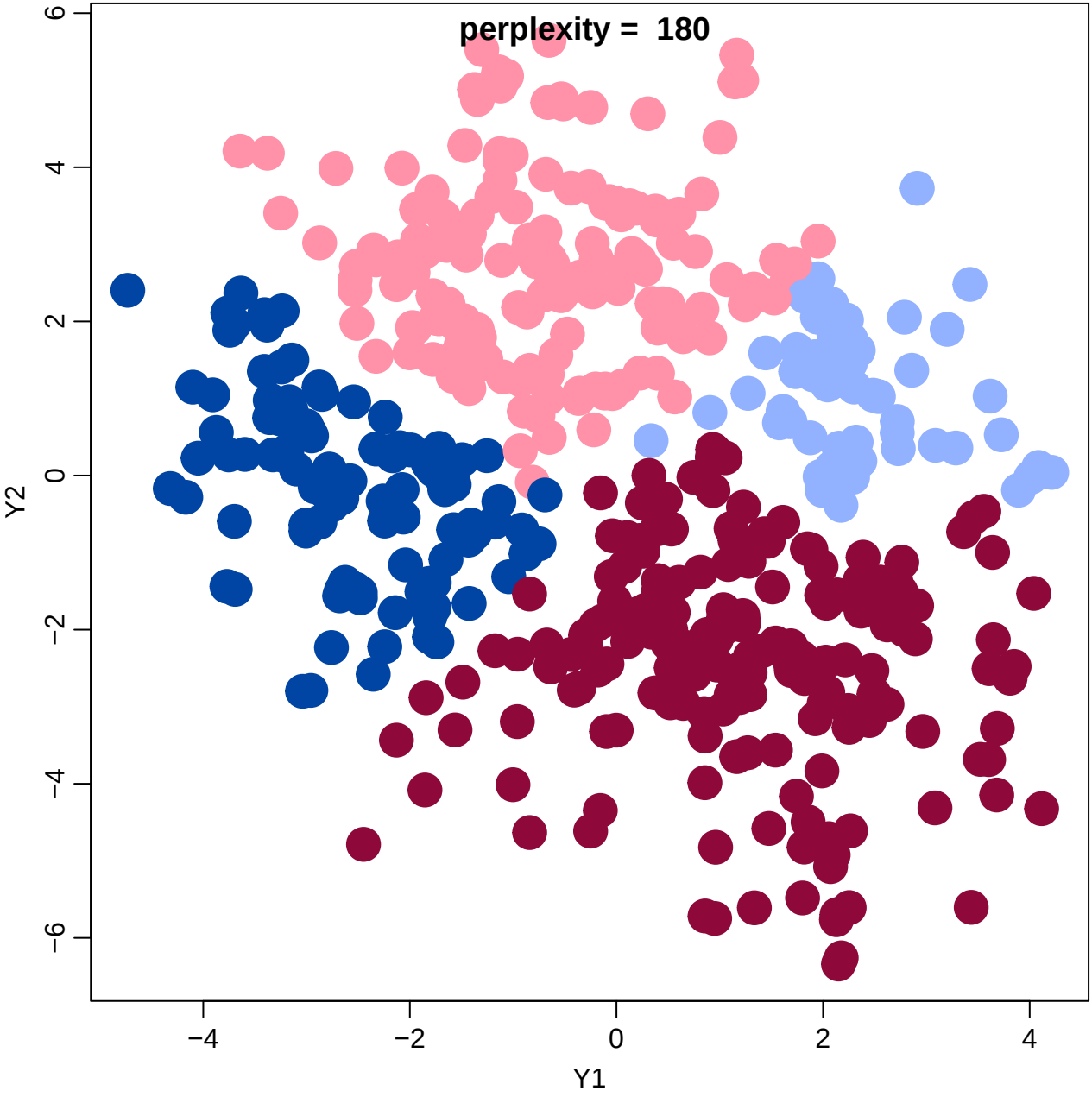

perplexity = 160
1
4
3
2

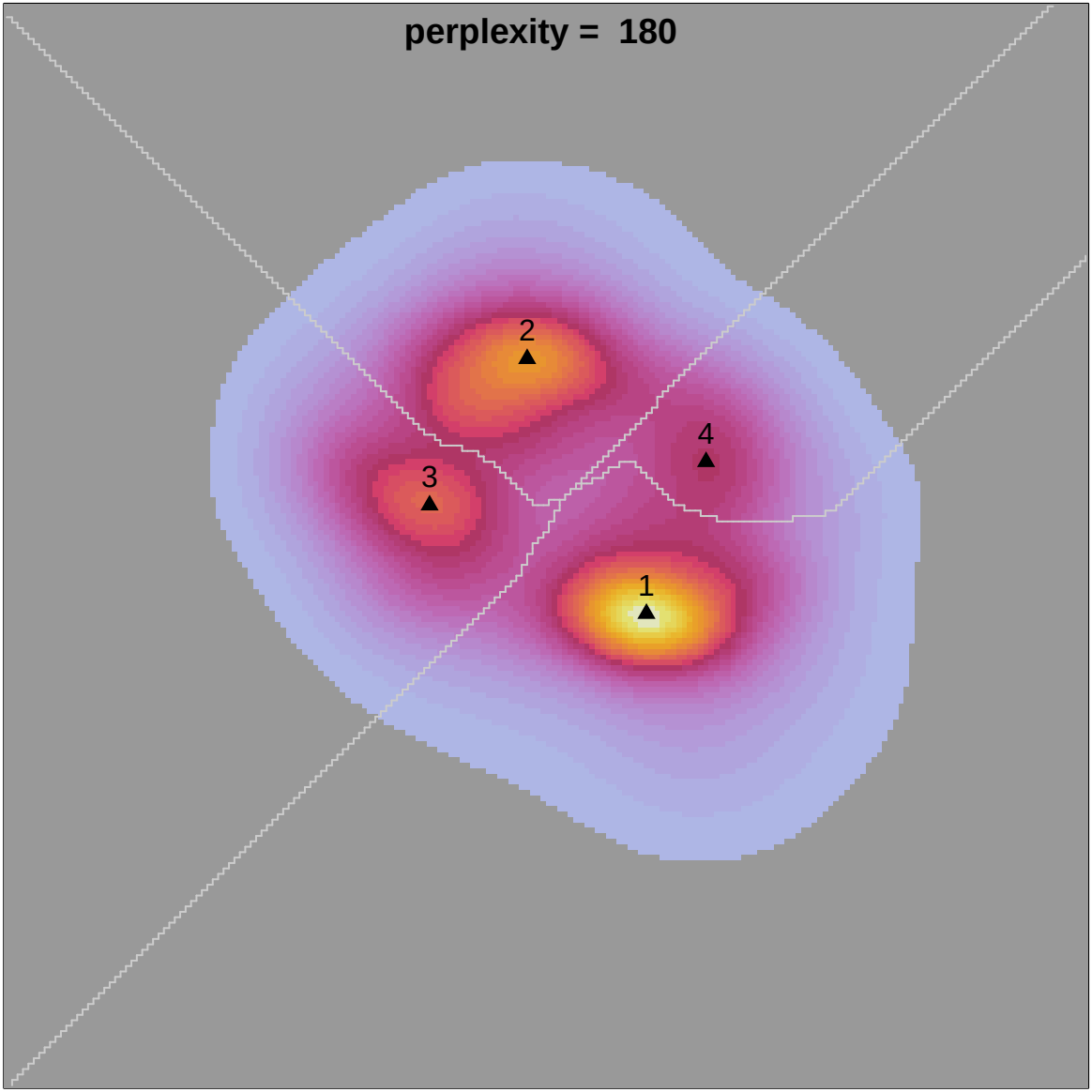

perplexity = 170
0.95
0.90
0.85
0.80
0.75
0.70
Cost
15
10
5
Size
0
20
40
60
80
100
epochs

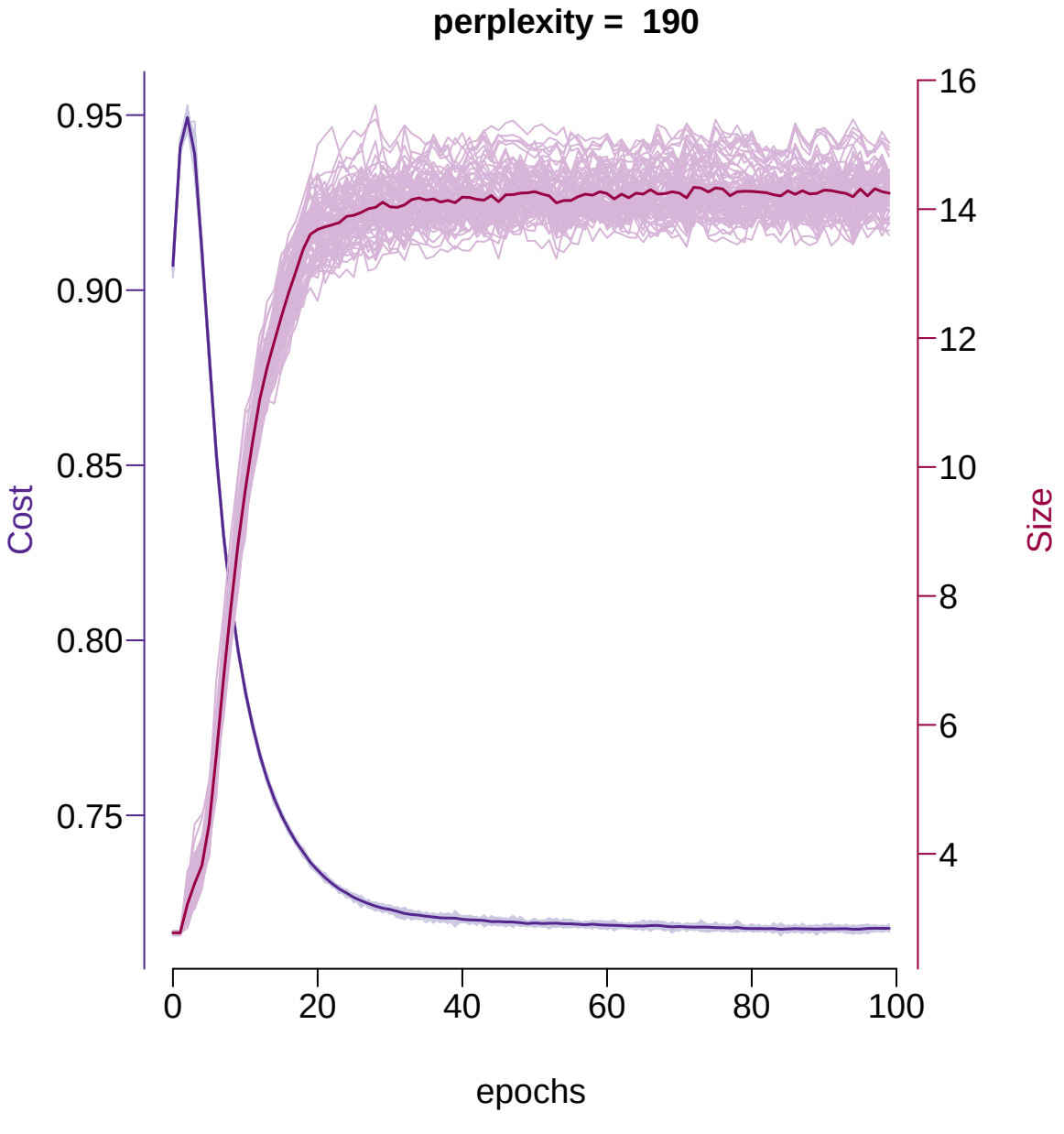

perplexity = 170
Y2
Y1
6
4
2
0
−2
−4
−6
−4
−2
0
2
4
6

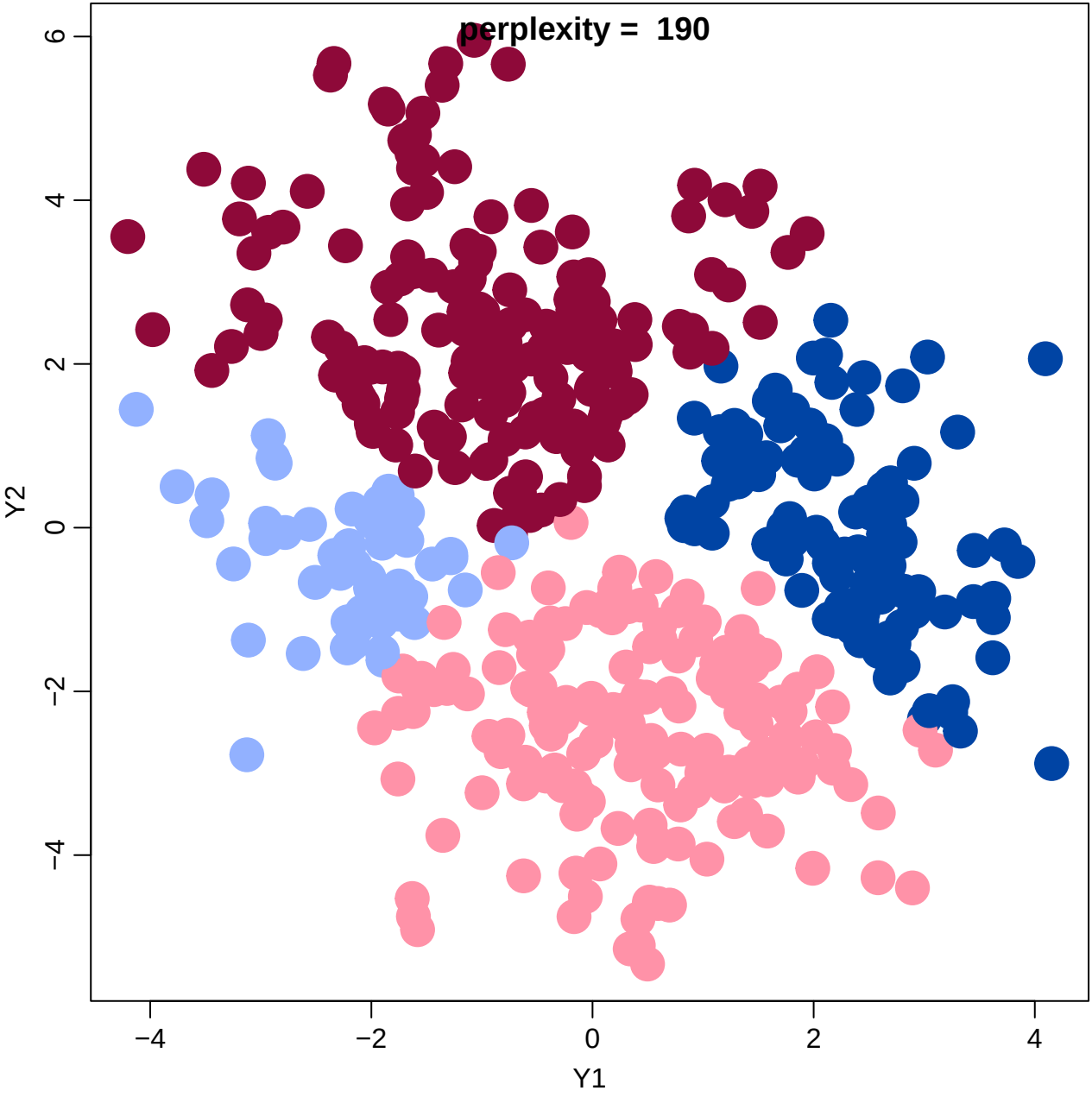

perplexity = 170
1
2
3
4

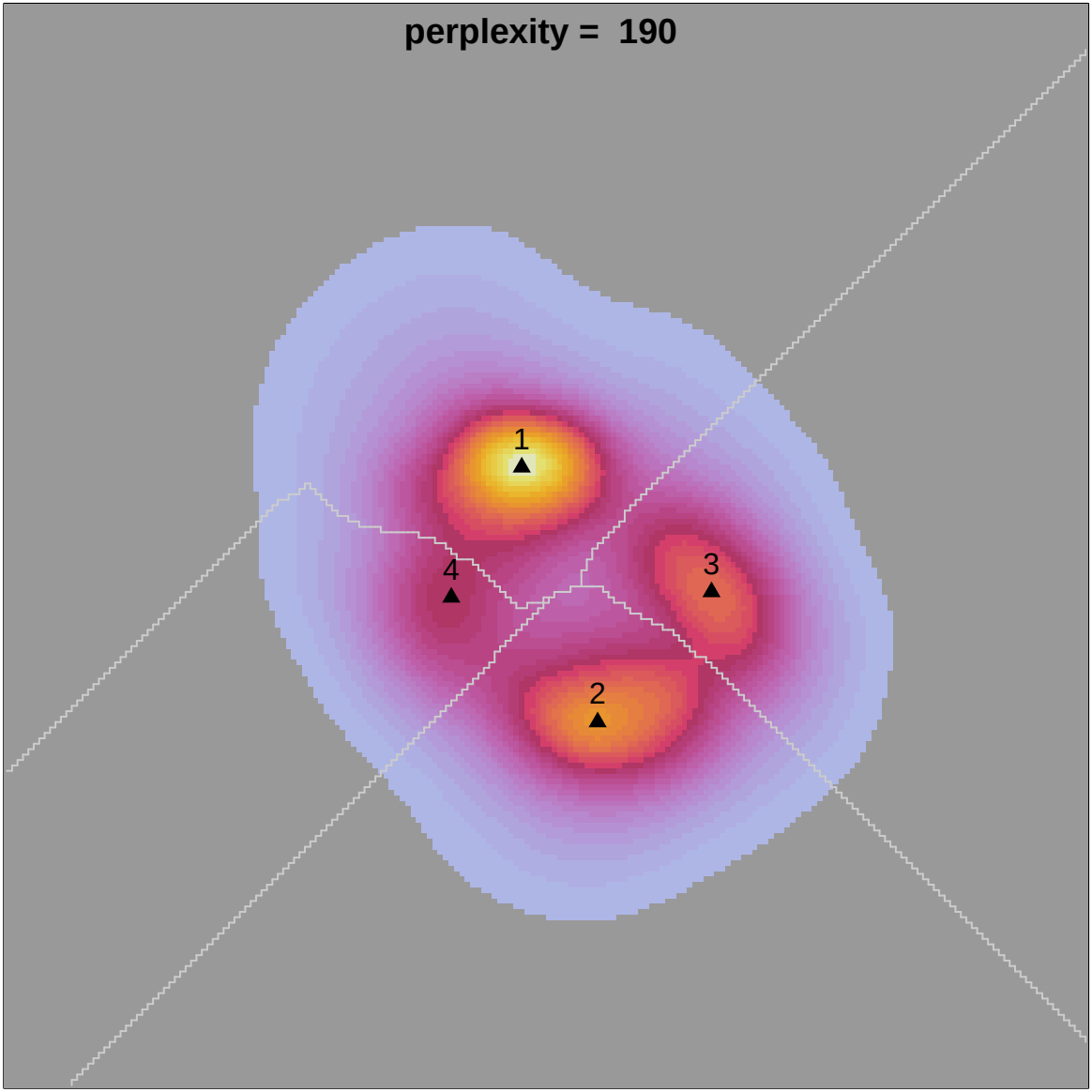

perplexity = 180
Cost
Size
epochs
0.95
0.90
0.85
0.80
0.75
0.70
16
14
12
10
8
6
4
0
20
40
60
80
100

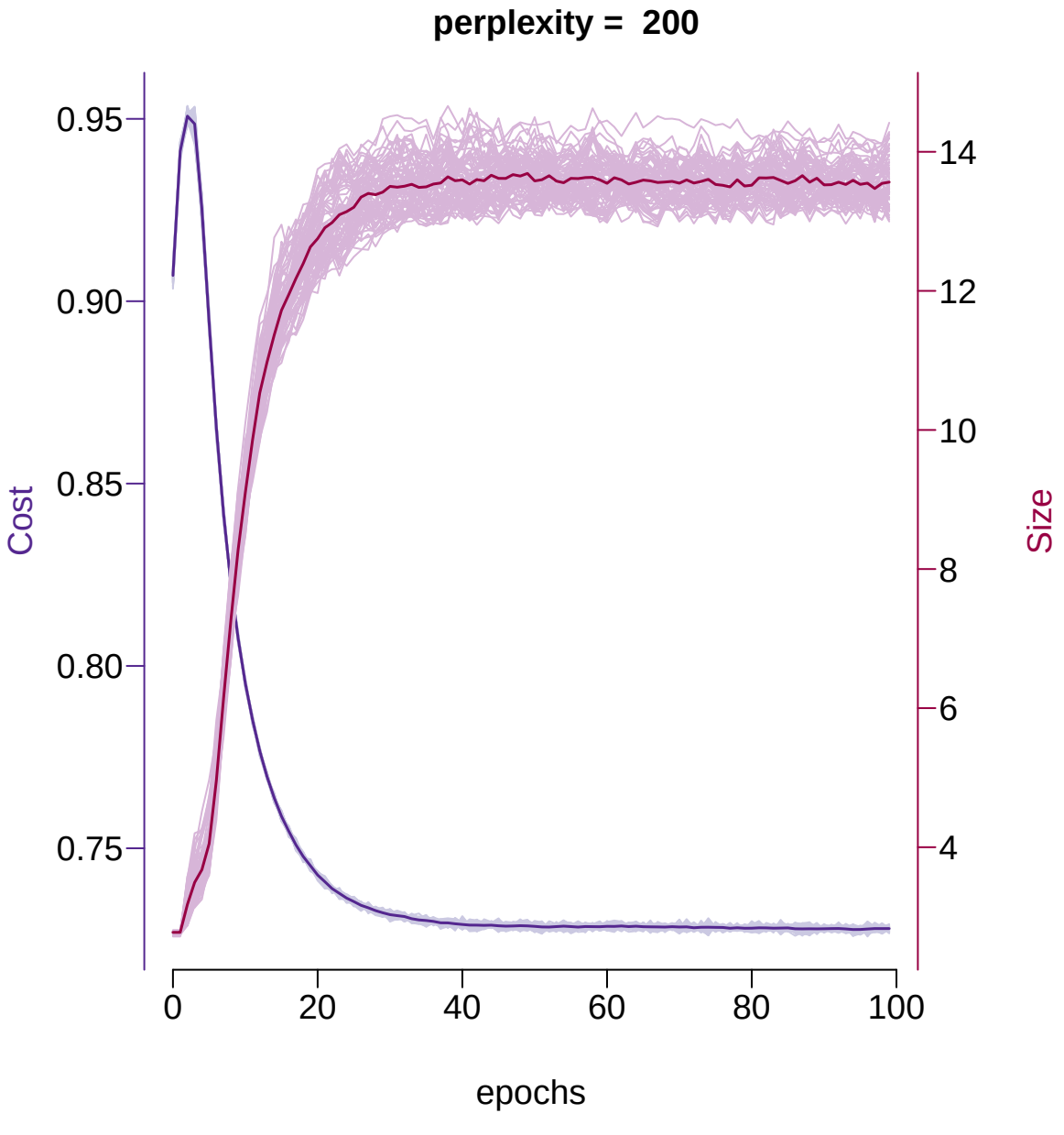

perplexity = 180
Y2
Y1
6
4
2
0
−2
−4
−6
−4
−2
0
2
4

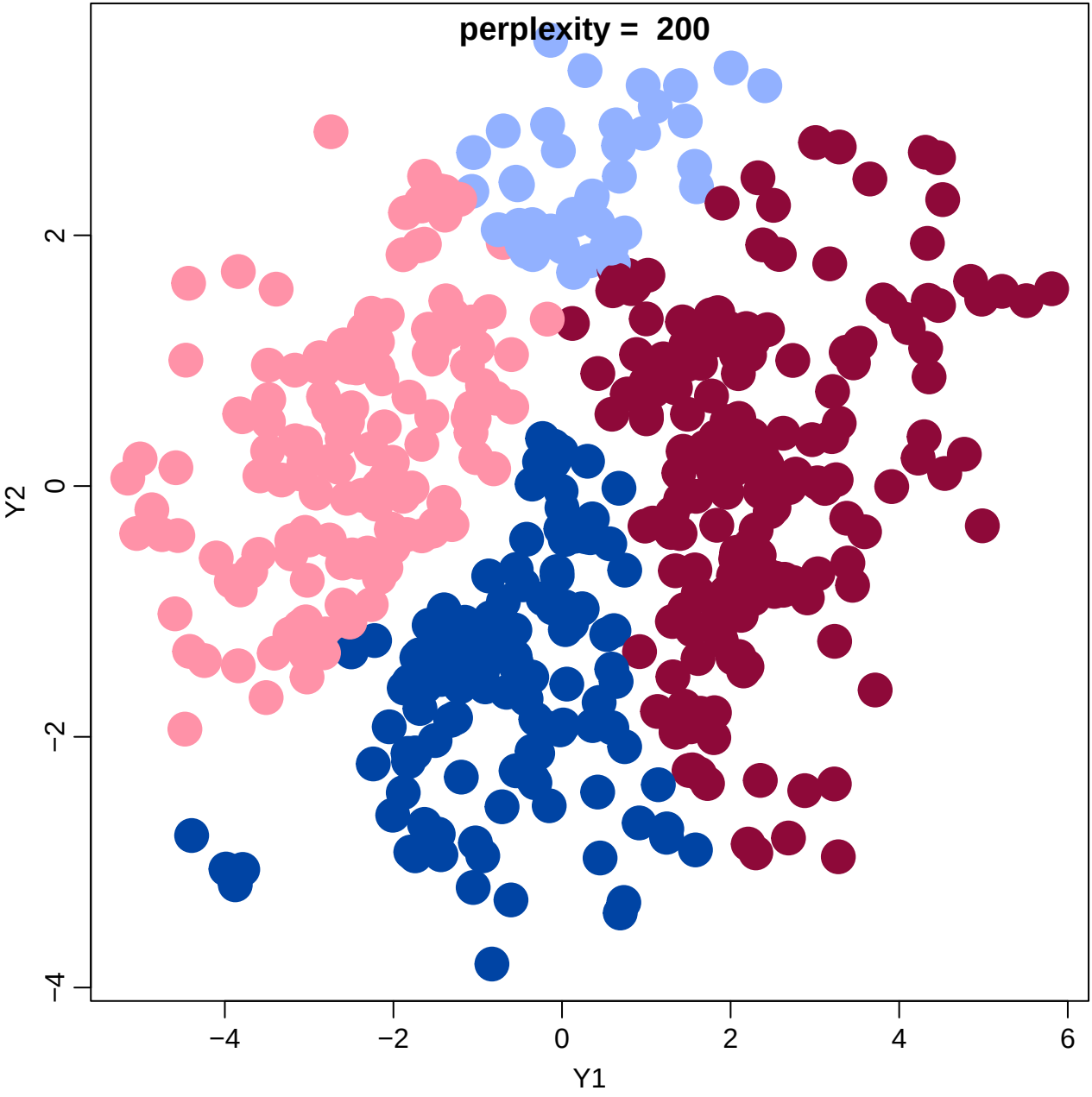

perplexity = 180
2
4
3
1

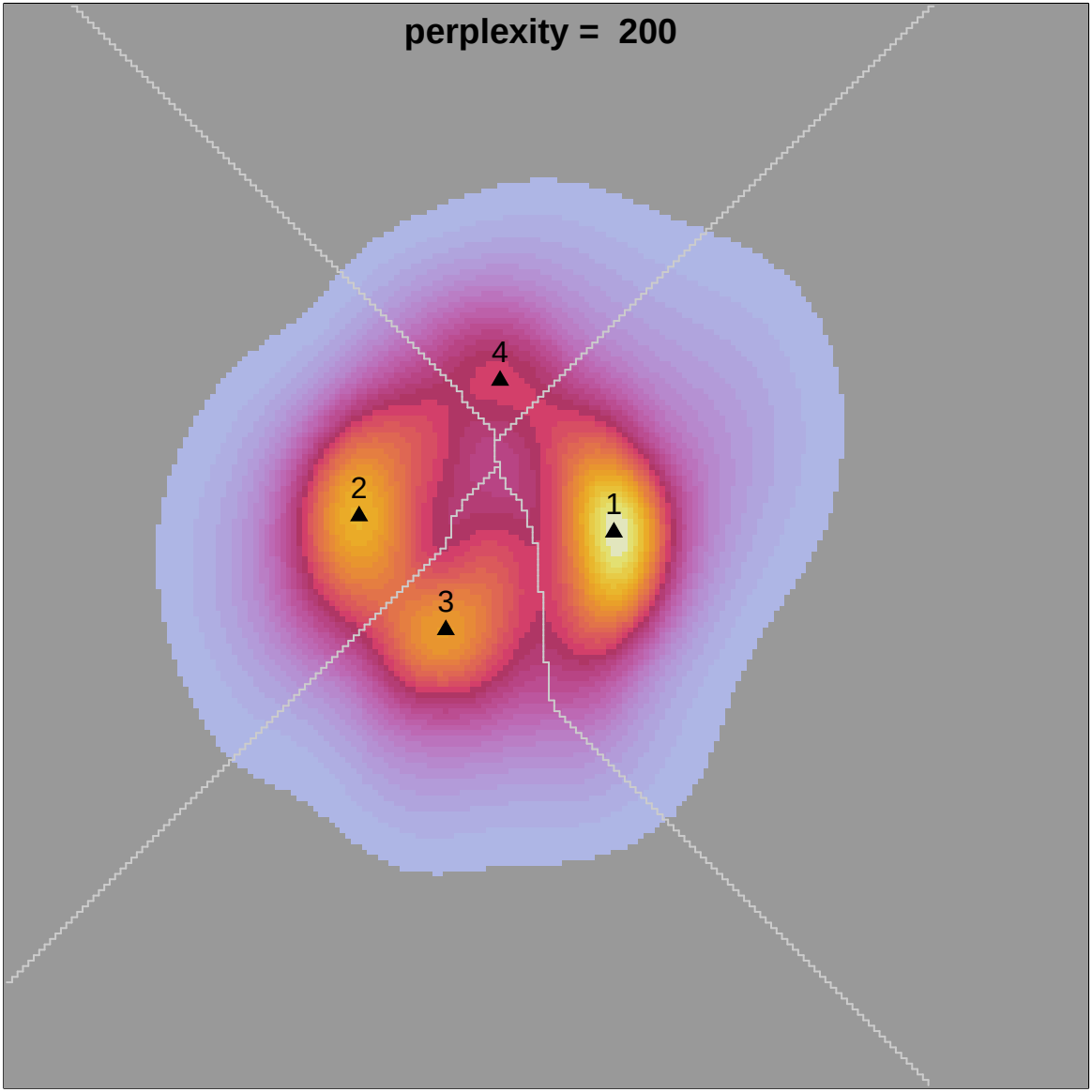

perplexity = 190
Cost
Size
epochs
0.95
0.90
0.85
0.80
0.75
16
14
12
10
8
6
4
0
20
40
60
80
100

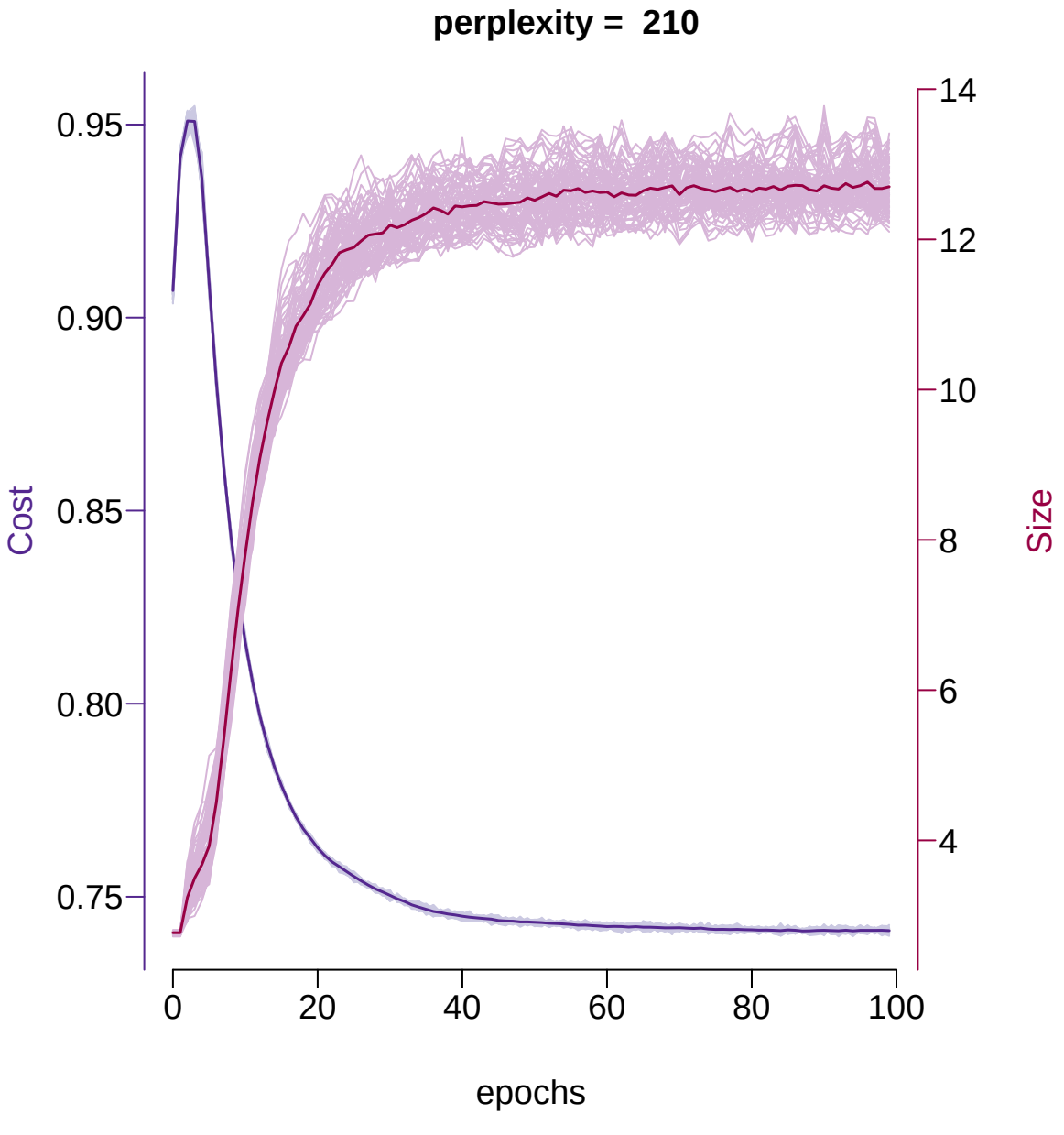

perplexity = 190
Y2
Y1
6
4
2
0
−2
−4
−4
−2
0
2
4

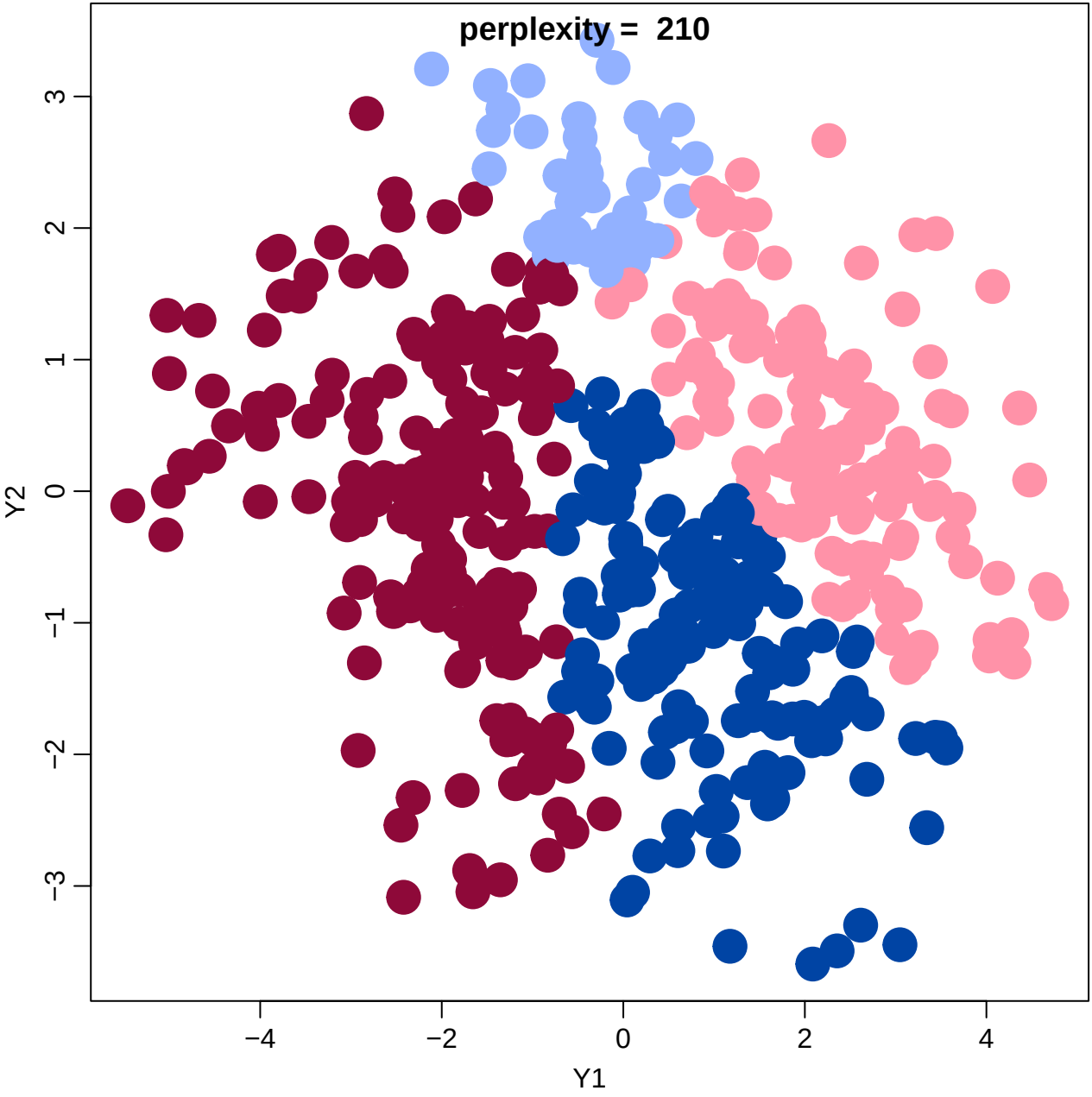

perplexity = 190
1
2
3
4

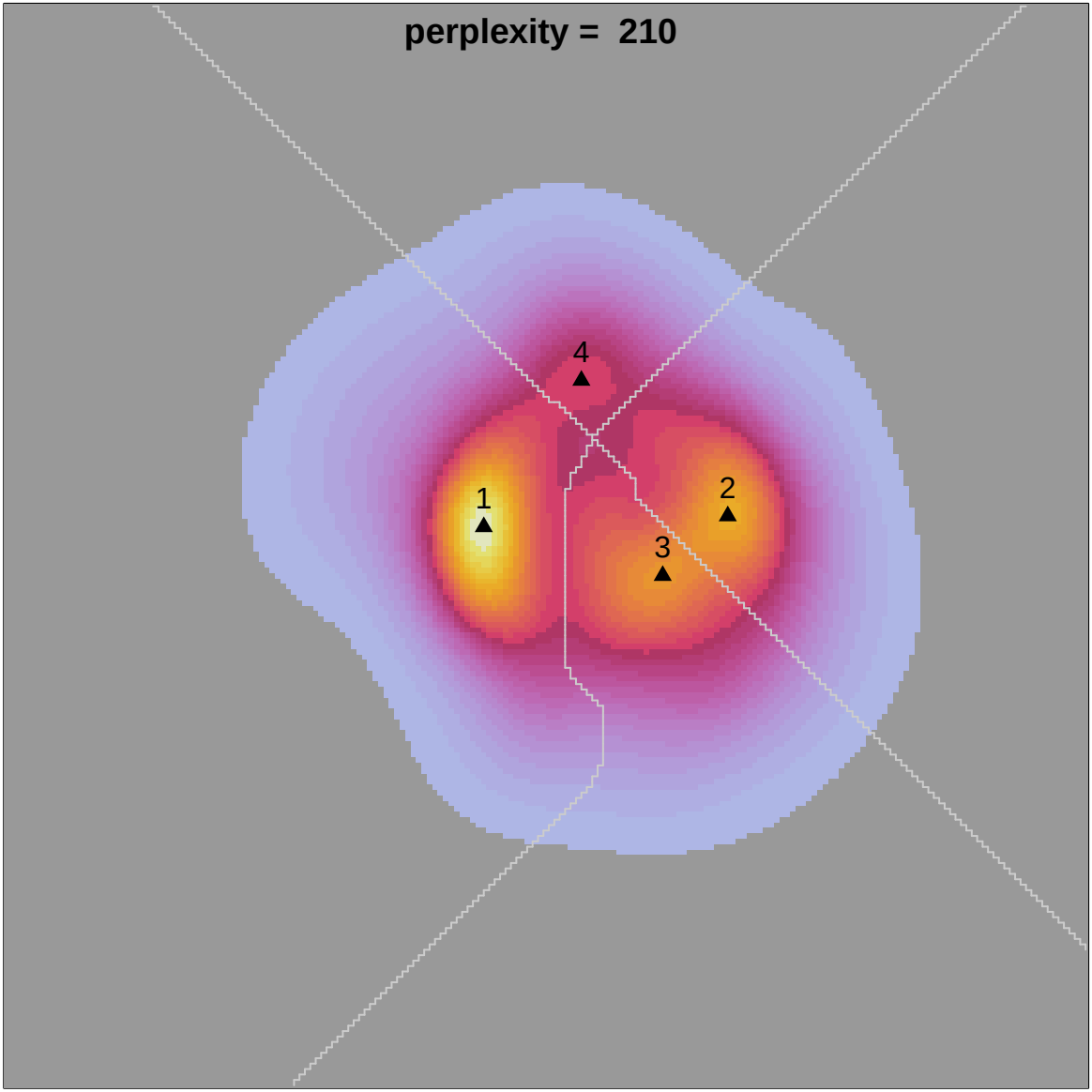

perplexity = 200
Cost
Size
0.95
0.90
0.85
0.80
0.75
14
12
10
8
6
4
0
20
40
60
80
100
epochs

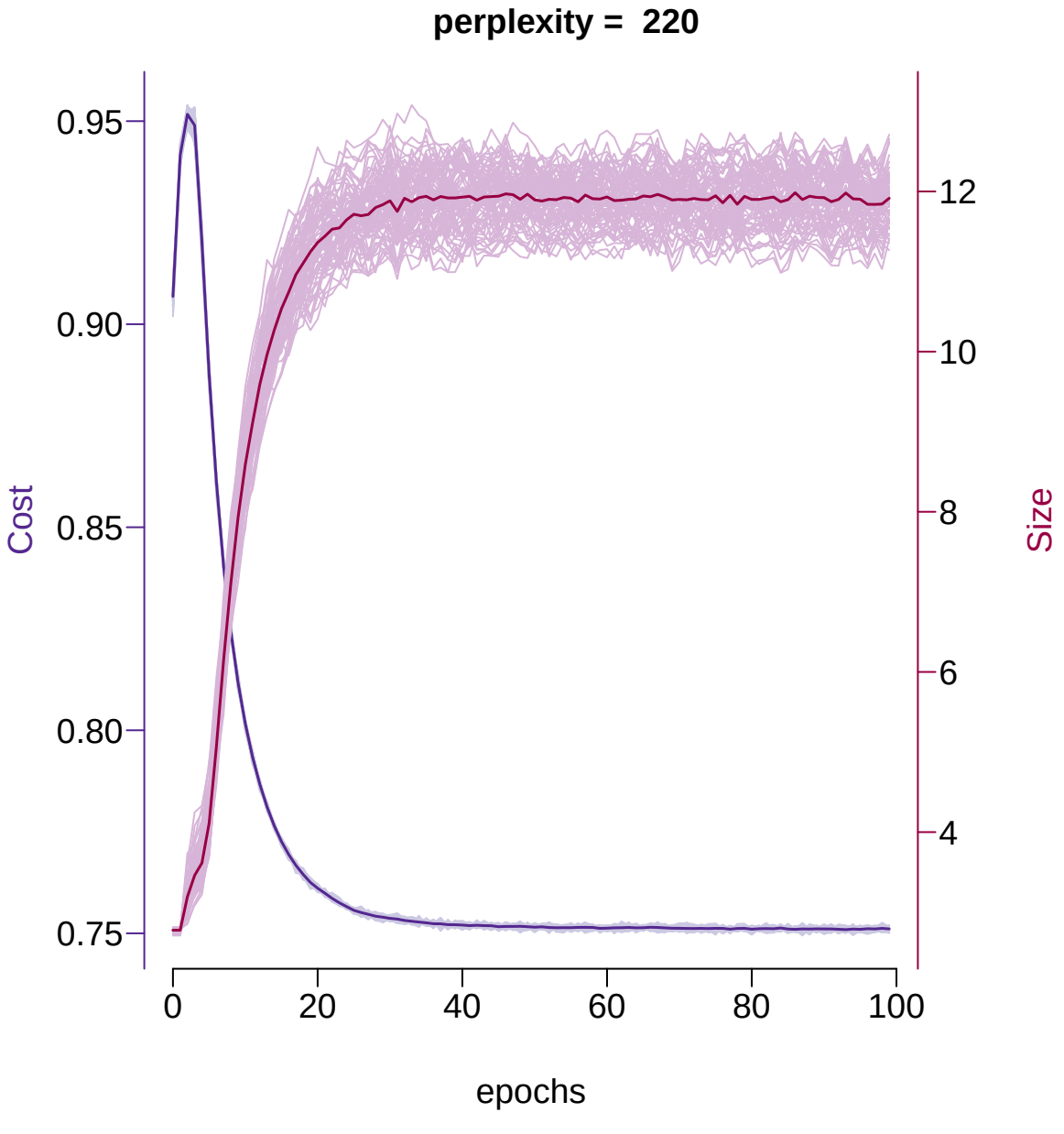

perplexity = 200
Y2
Y1
2
0
−2
−4
−4
−2
0
2
4
6

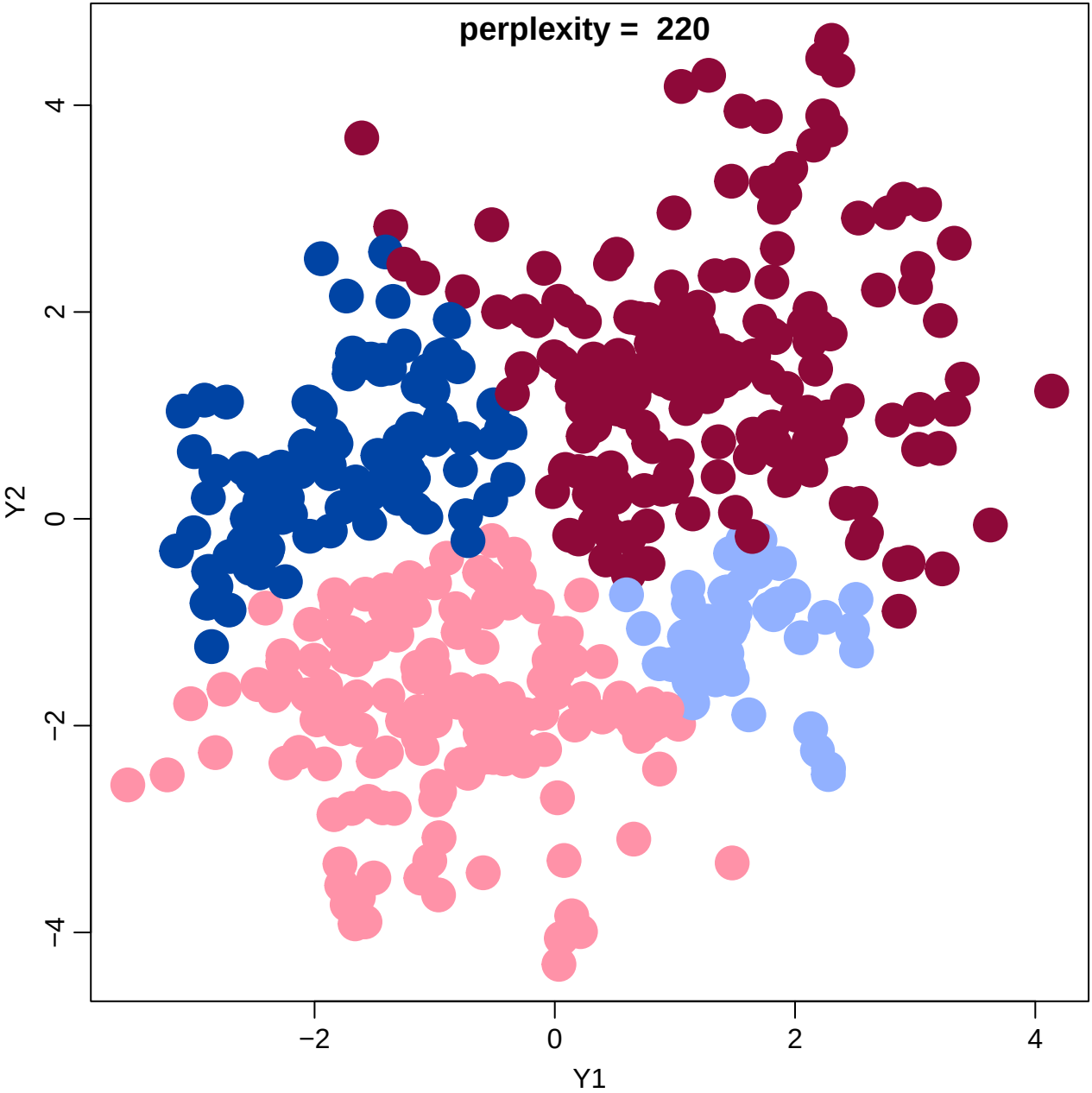

perplexity = 200
4
2
1
3

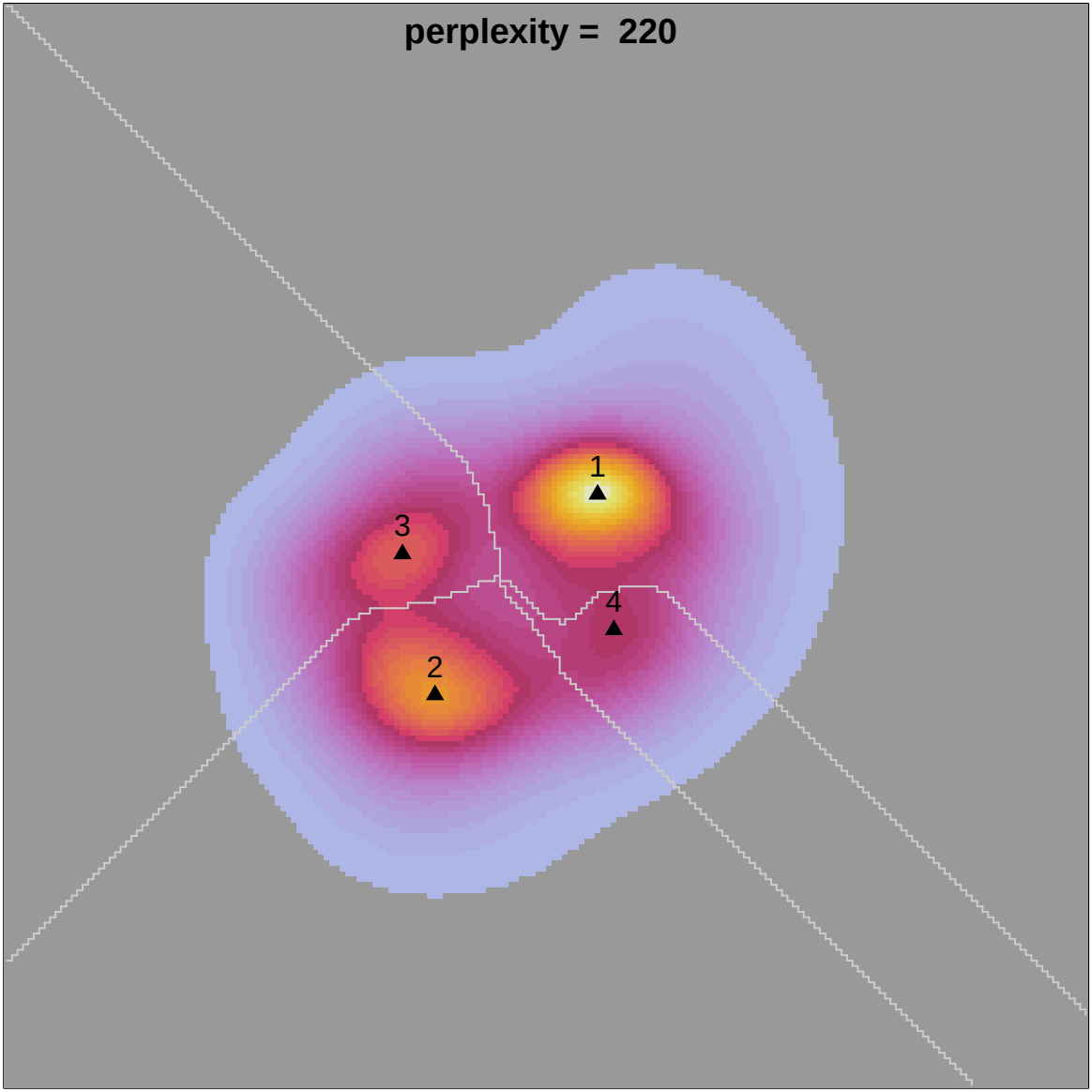

perplexity = 210
Cost
0.95
0.90
0.85
0.80
0.75
Size
14
12
10
8
6
4
0
20
40
60
80
100
epochs

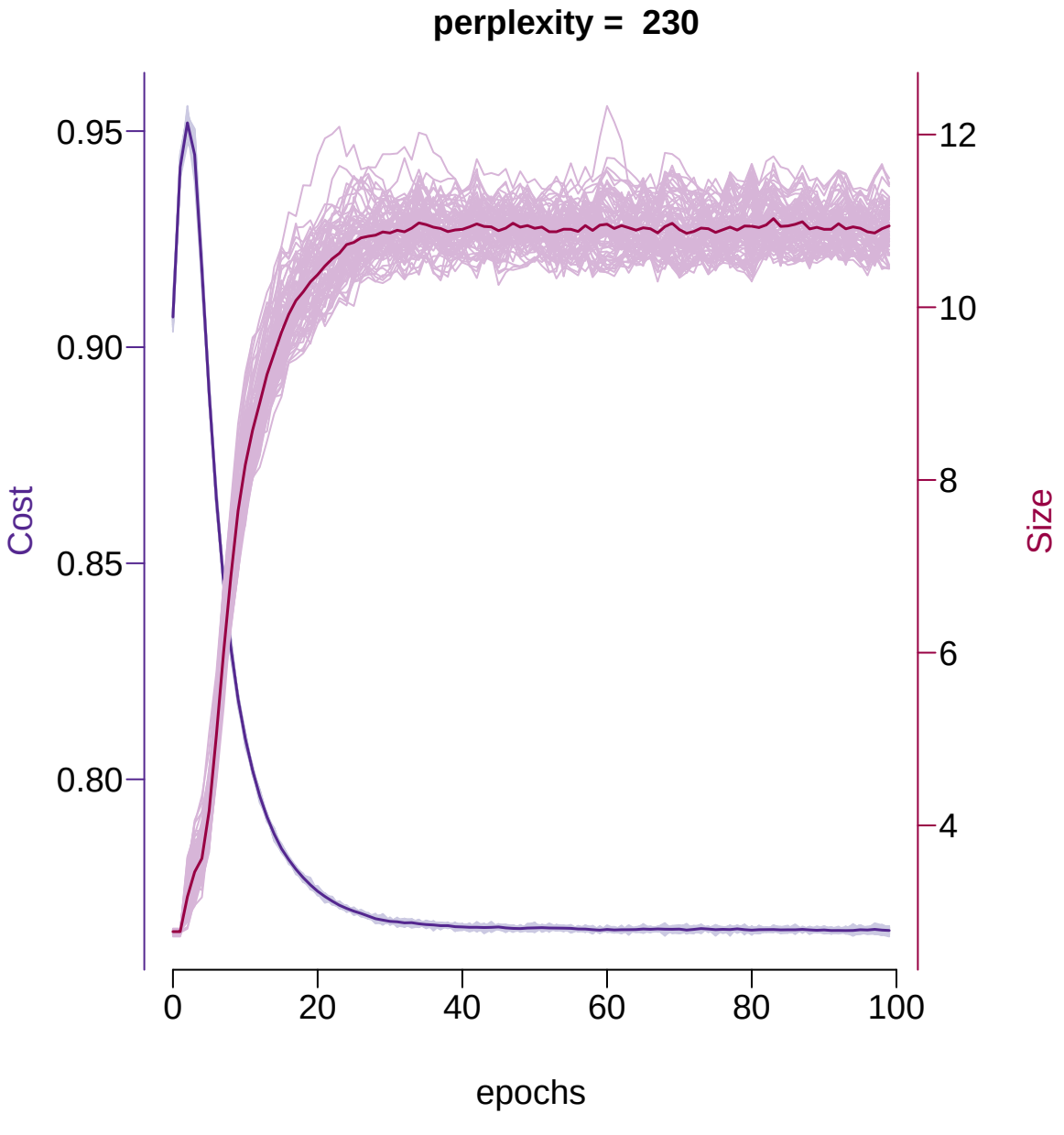

perplexity = 210
Y2
Y1
3
2
1
0
−1
−2
−3
−4
−2
0
2
4

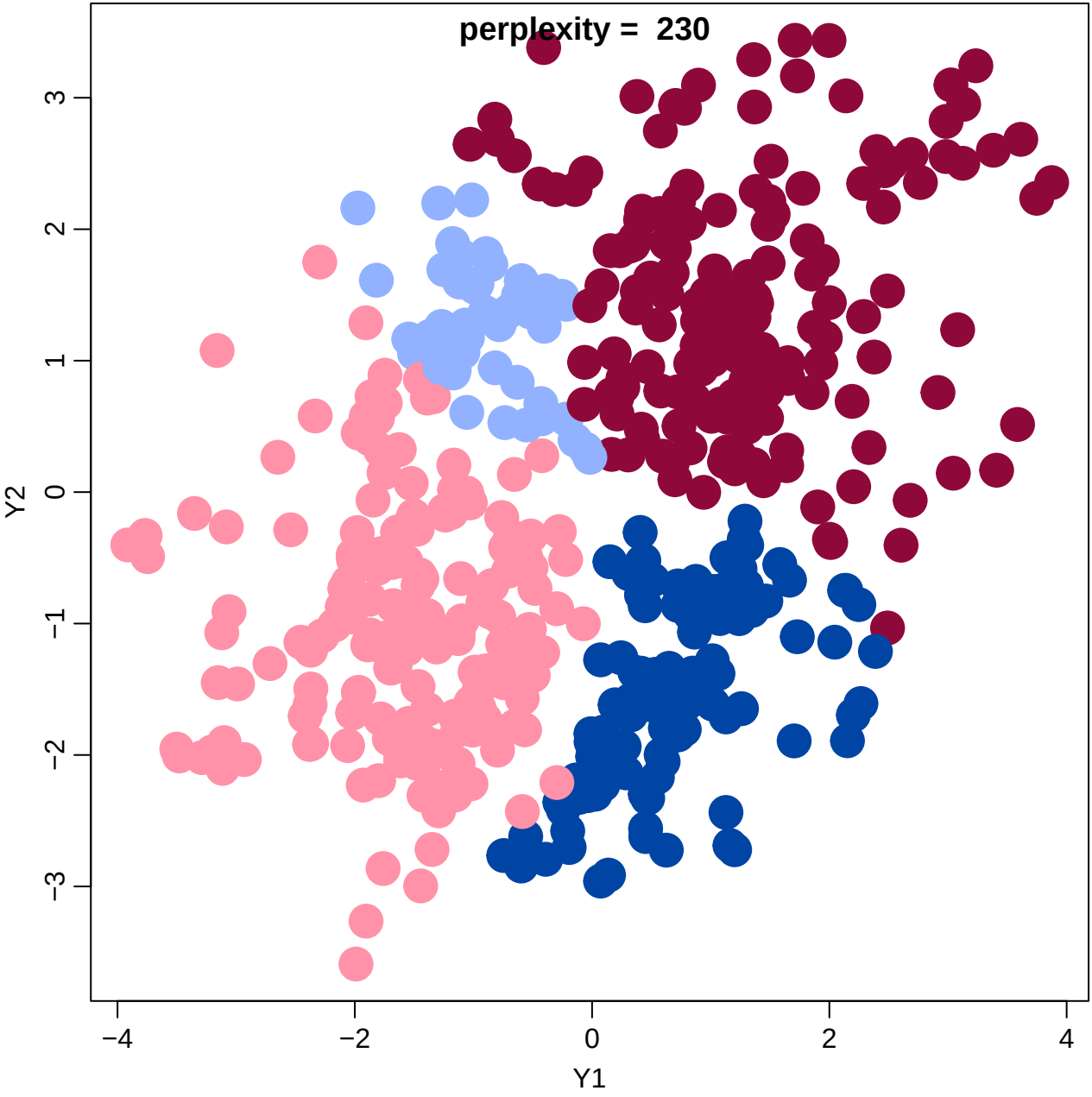

perplexity = 210
4
1
2
3

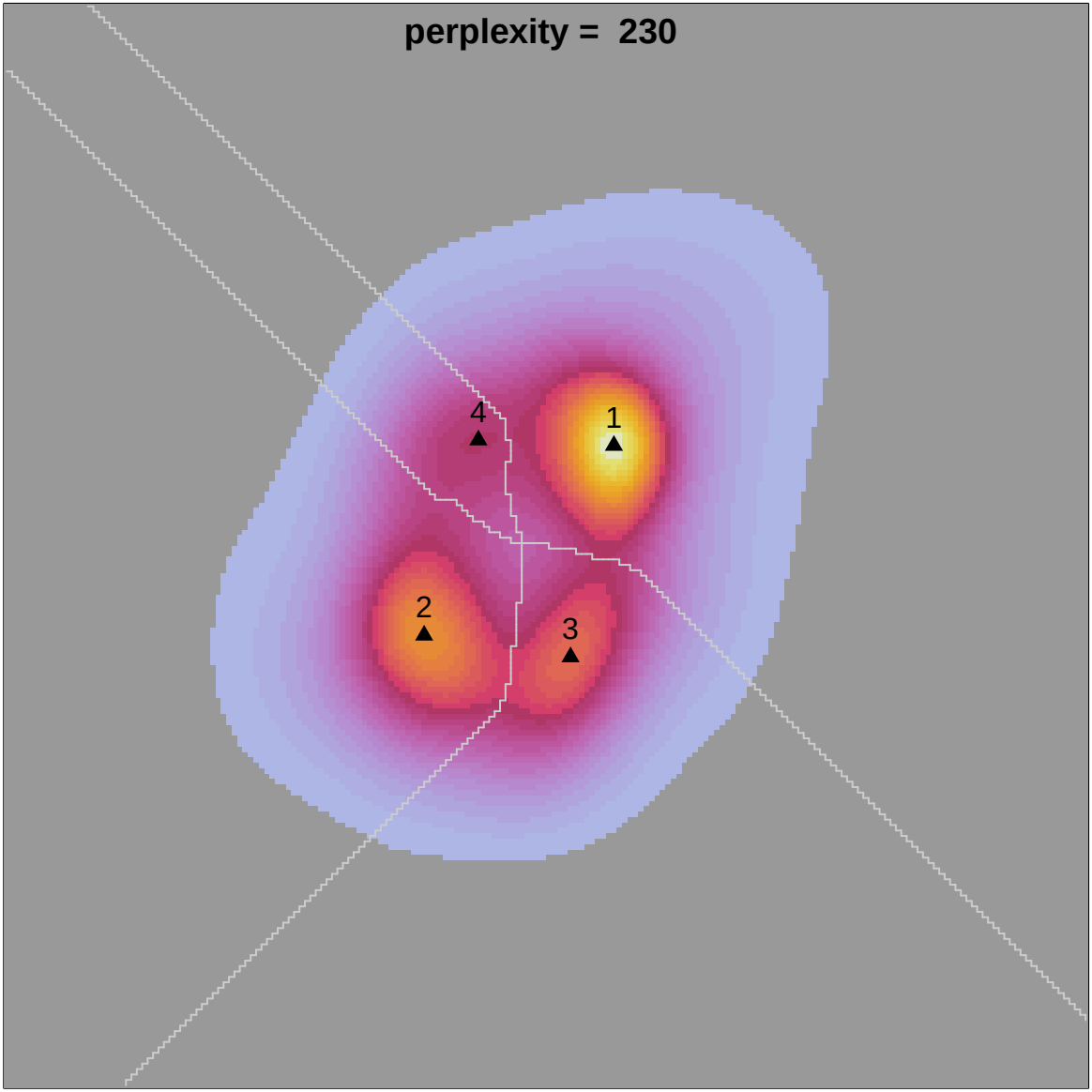

perplexity = 220
Cost
0.95
0.90
0.85
0.80
0.75
Size
12
10
8
6
4
0
20
40
60
80
100
epochs

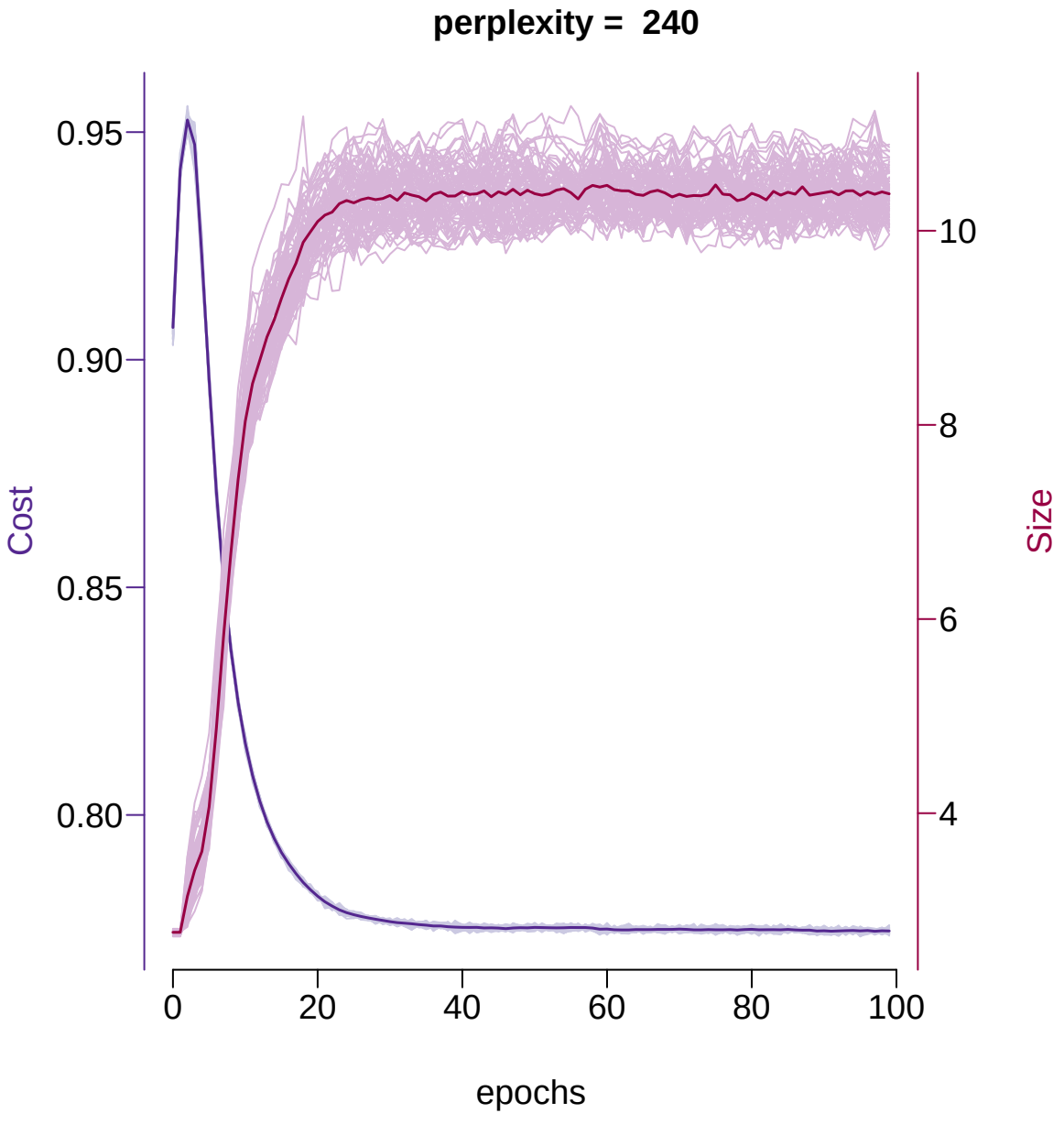

perplexity = 220
Y2
Y1
4
2
0
−2
−4
−2
0
2
4

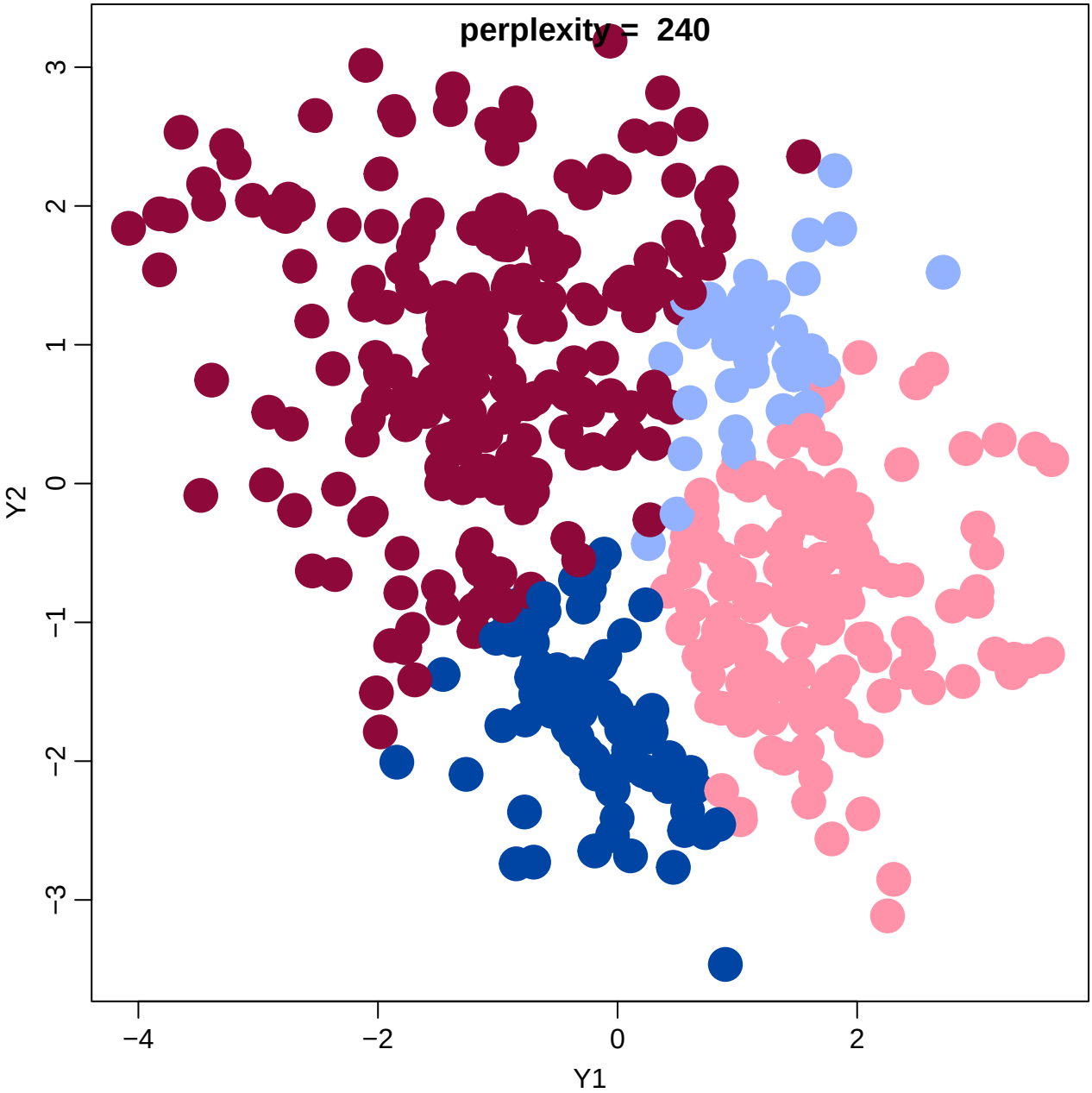

perplexity = 220
1
2
3
4

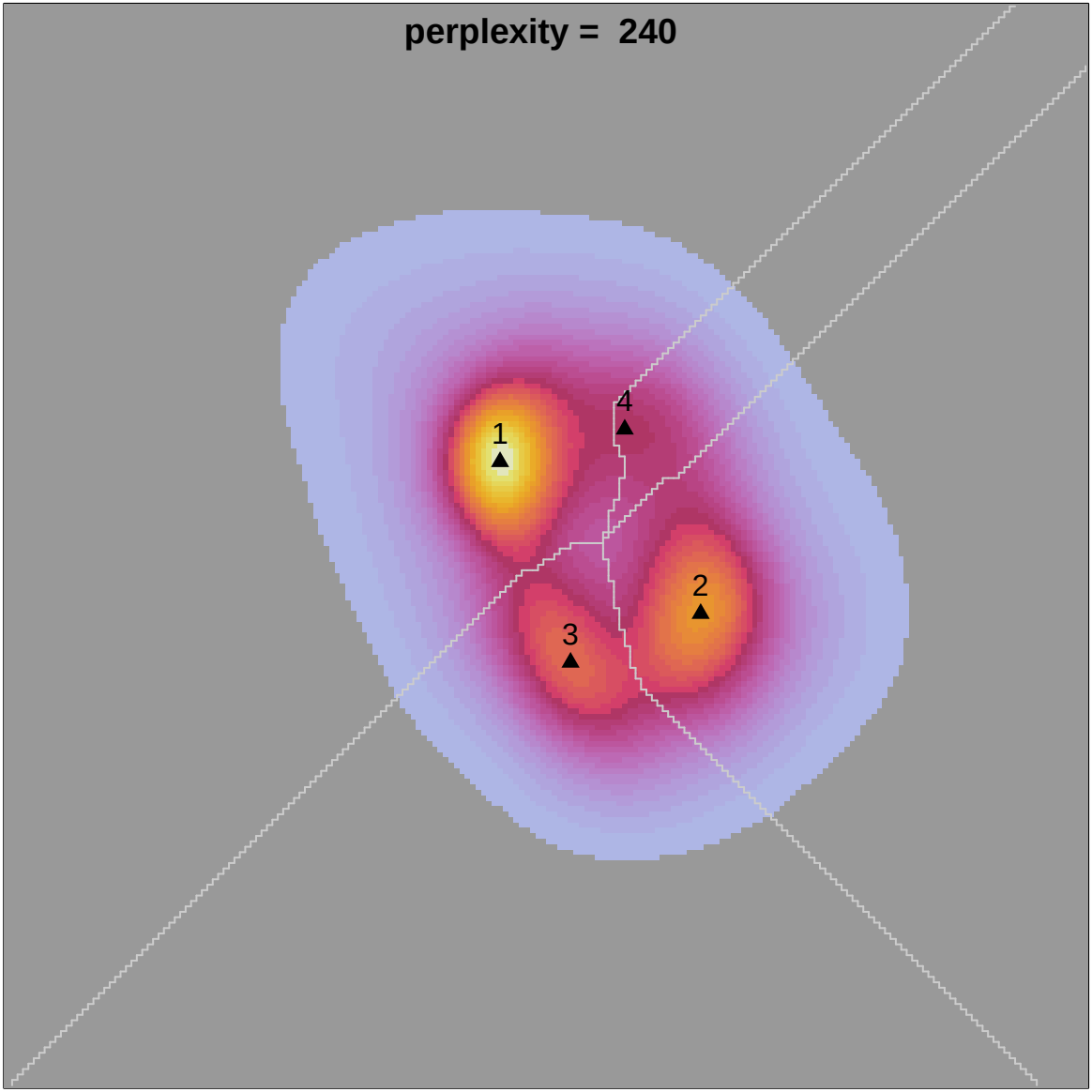

perplexity = 230
Cost
Size
0.95
0.90
0.85
0.80
12
10
8
6
4
0
20
40
60
80
100
epochs

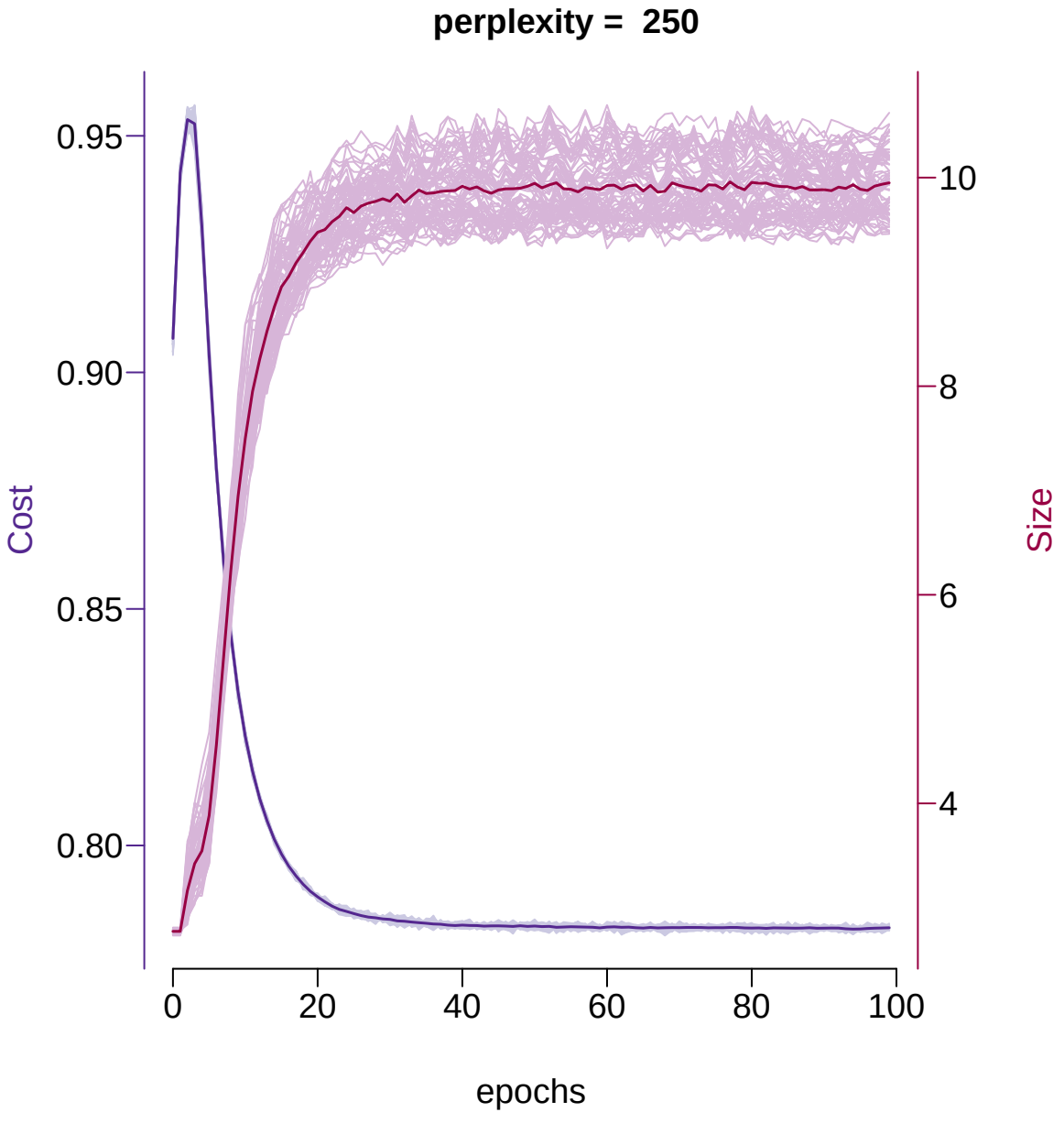

perplexity = 230
Y2
Y1
3
2
1
0
−1
−2
−3
−4
−2
0
2
4

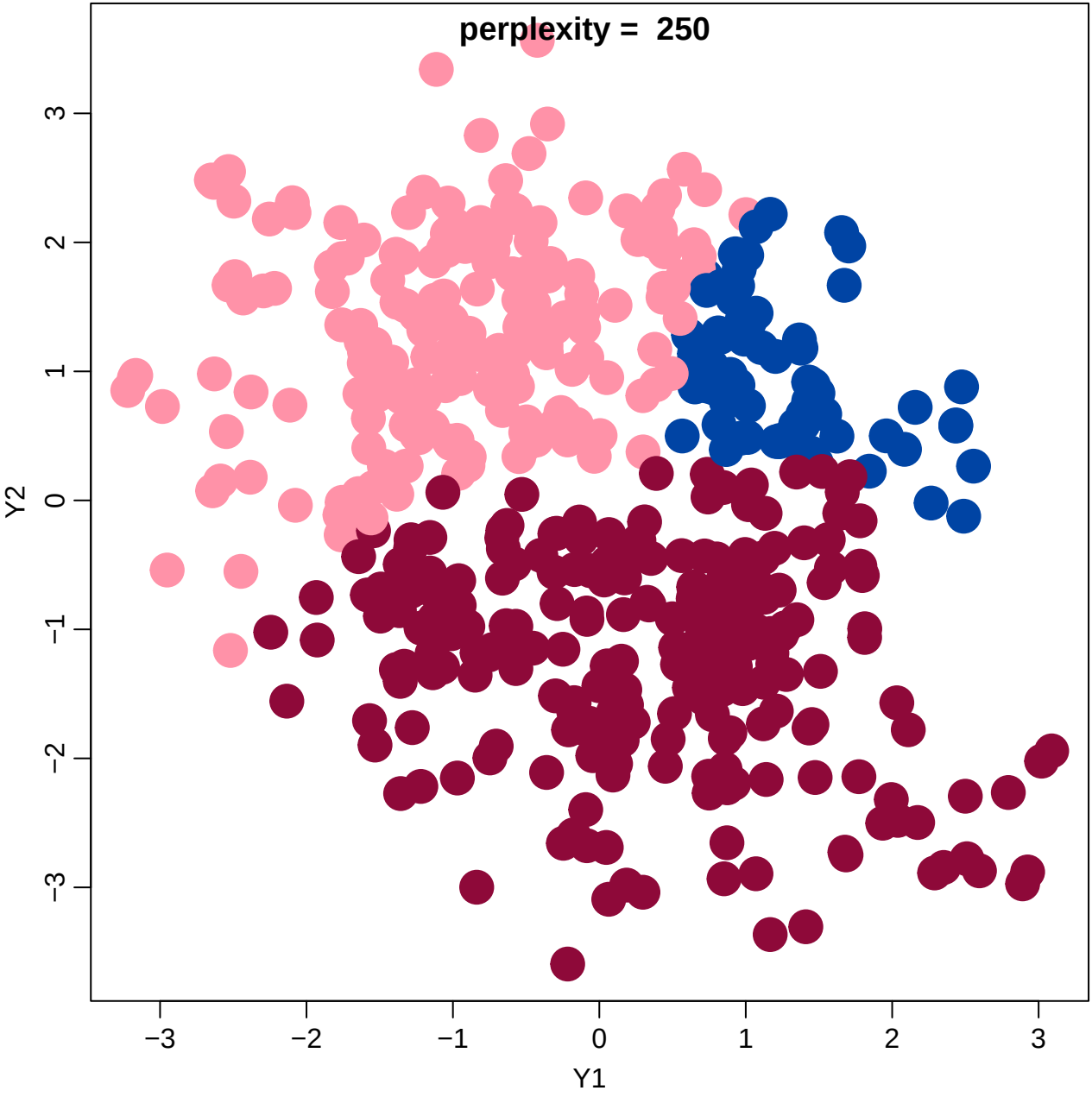

perplexity = 230
4
1
2
3

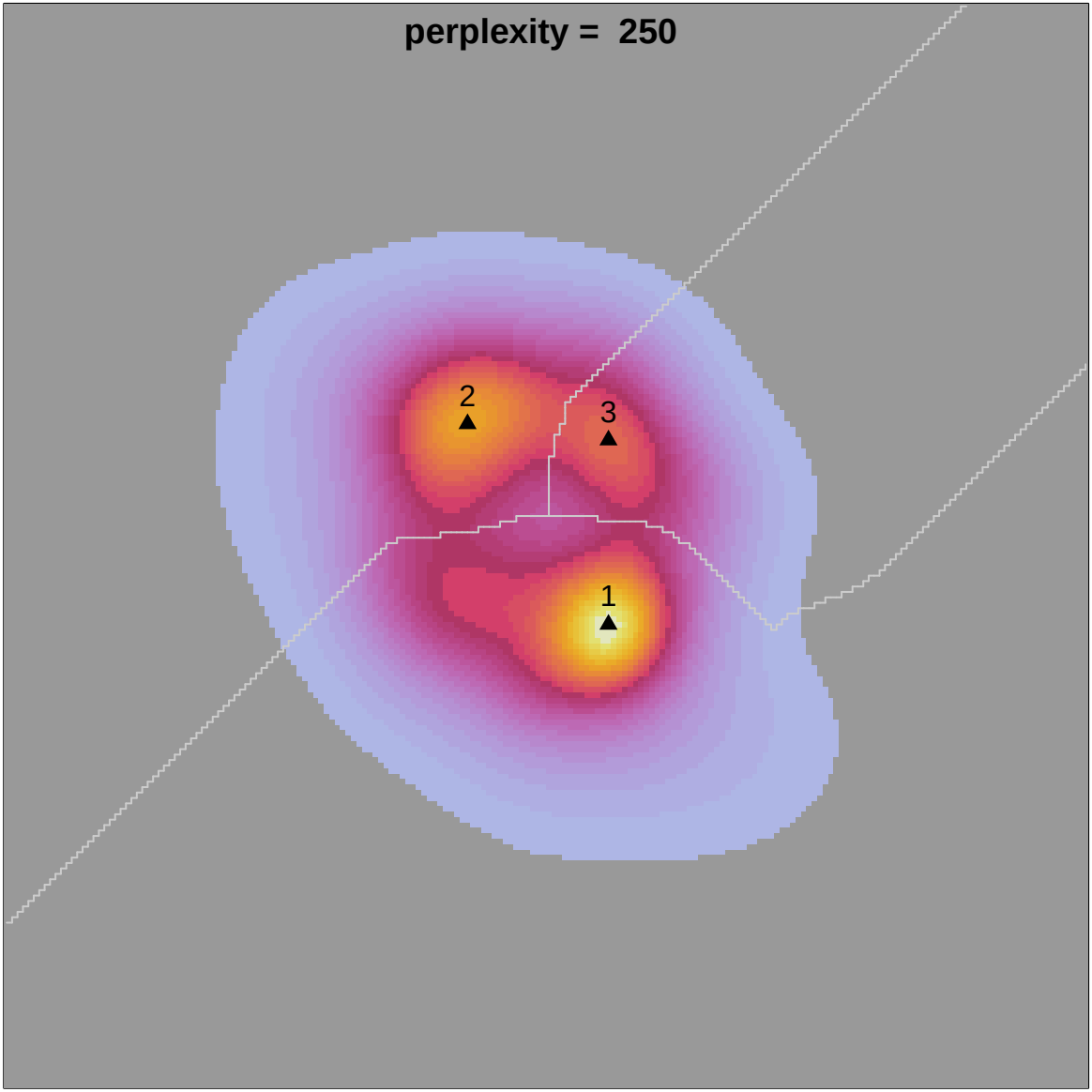

perplexity = 240
Cost
Size
0.95
0.90
0.85
0.80
10
8
6
4
0
20
40
60
80
100
epochs

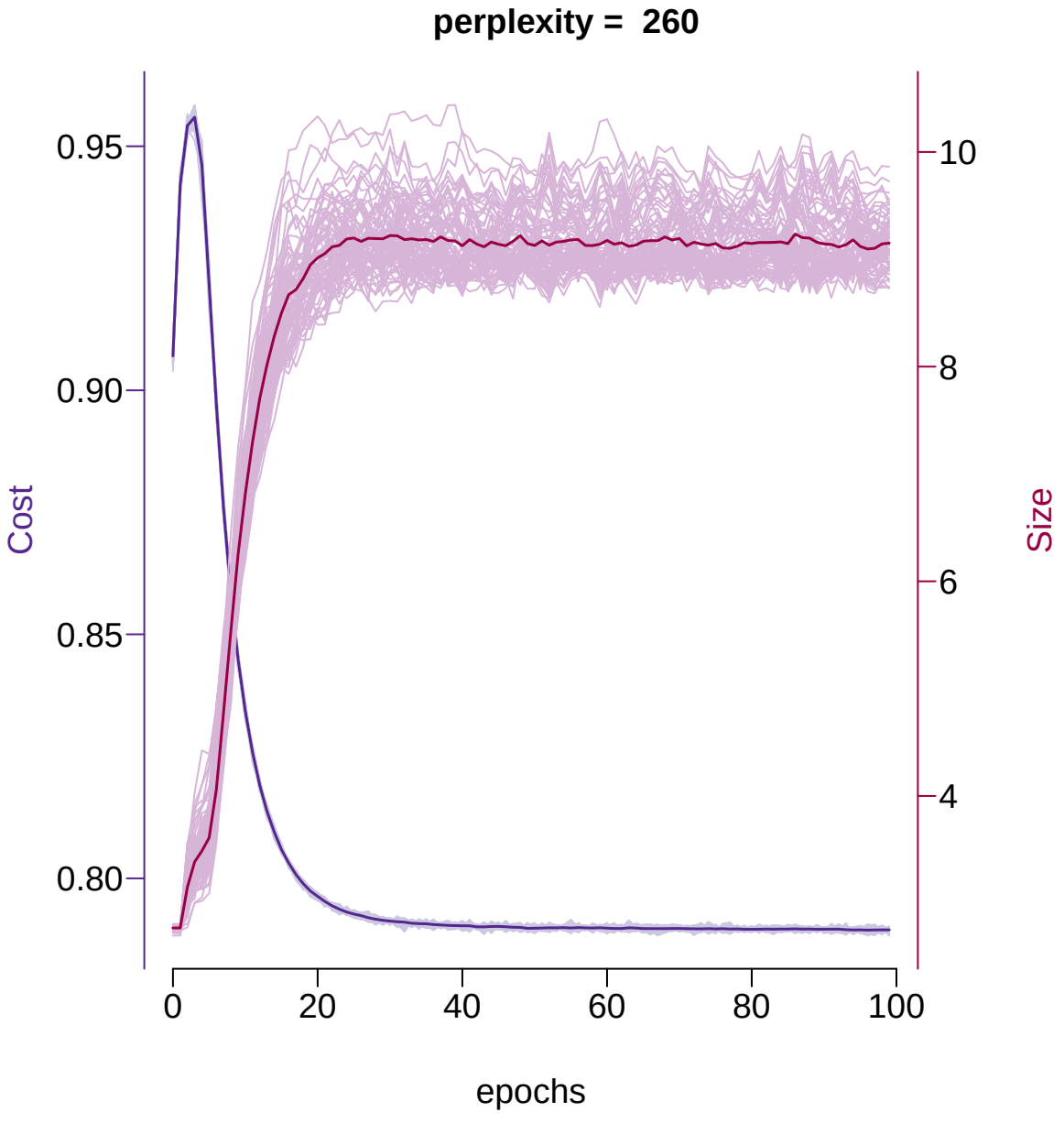

perplexity = 240
Y2
Y1
3
2
1
0
−1
−2
−3
−4
−2
0
2

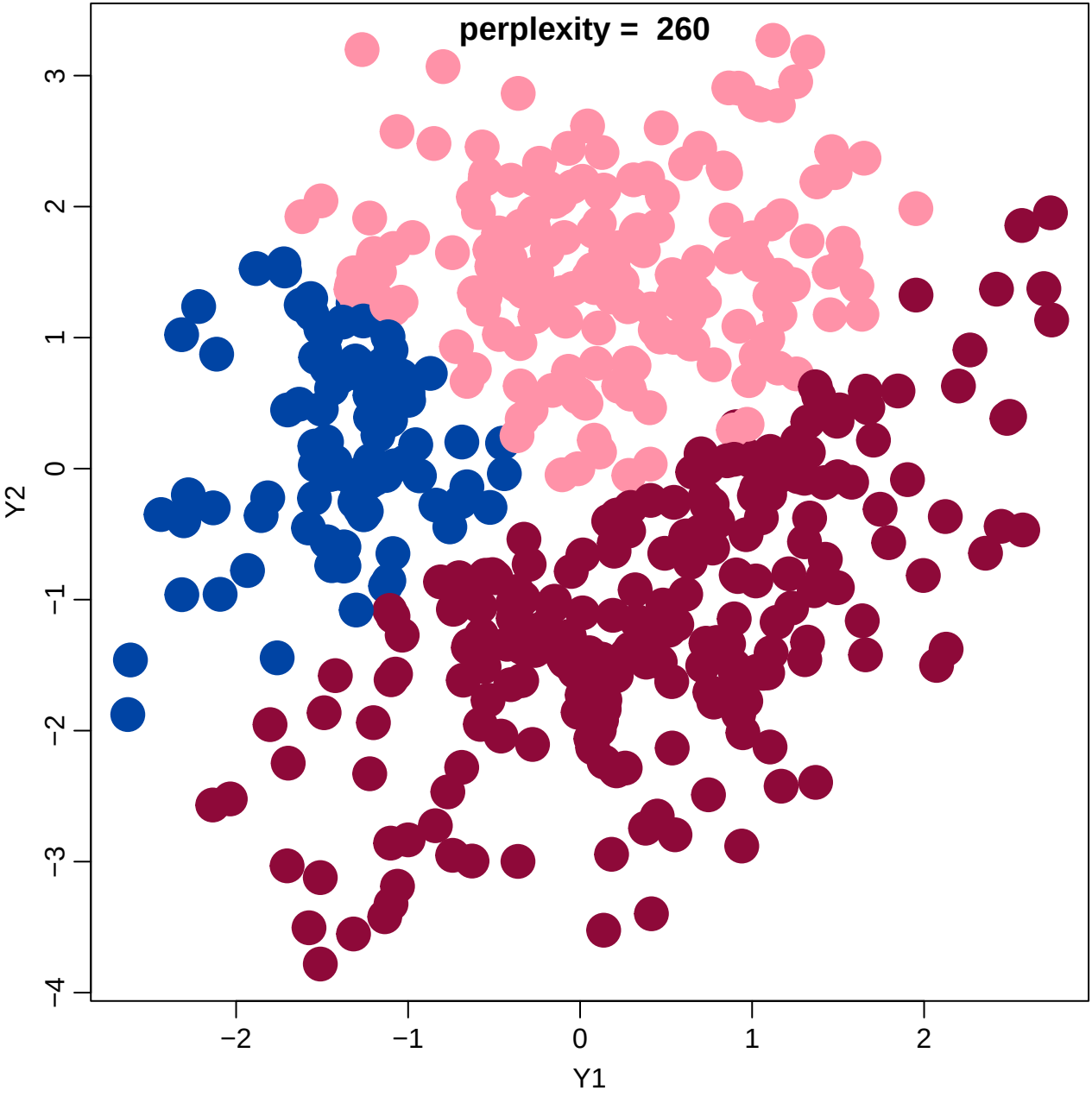

perplexity = 240
1
2
3
4

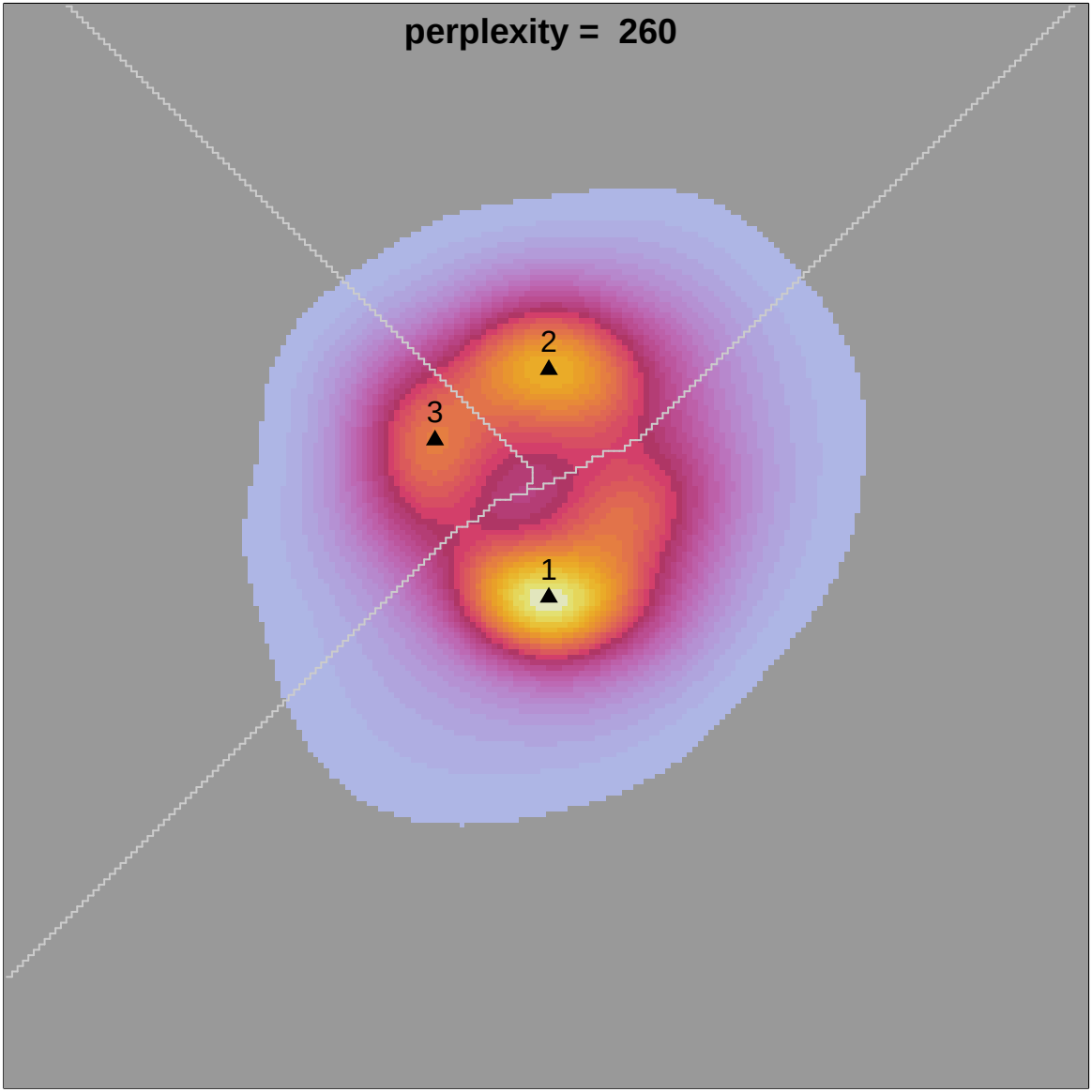

perplexity = 250
Cost
Size
0.95
0.90
0.85
0.80
10
8
6
4
0
20
40
60
80
100
epochs

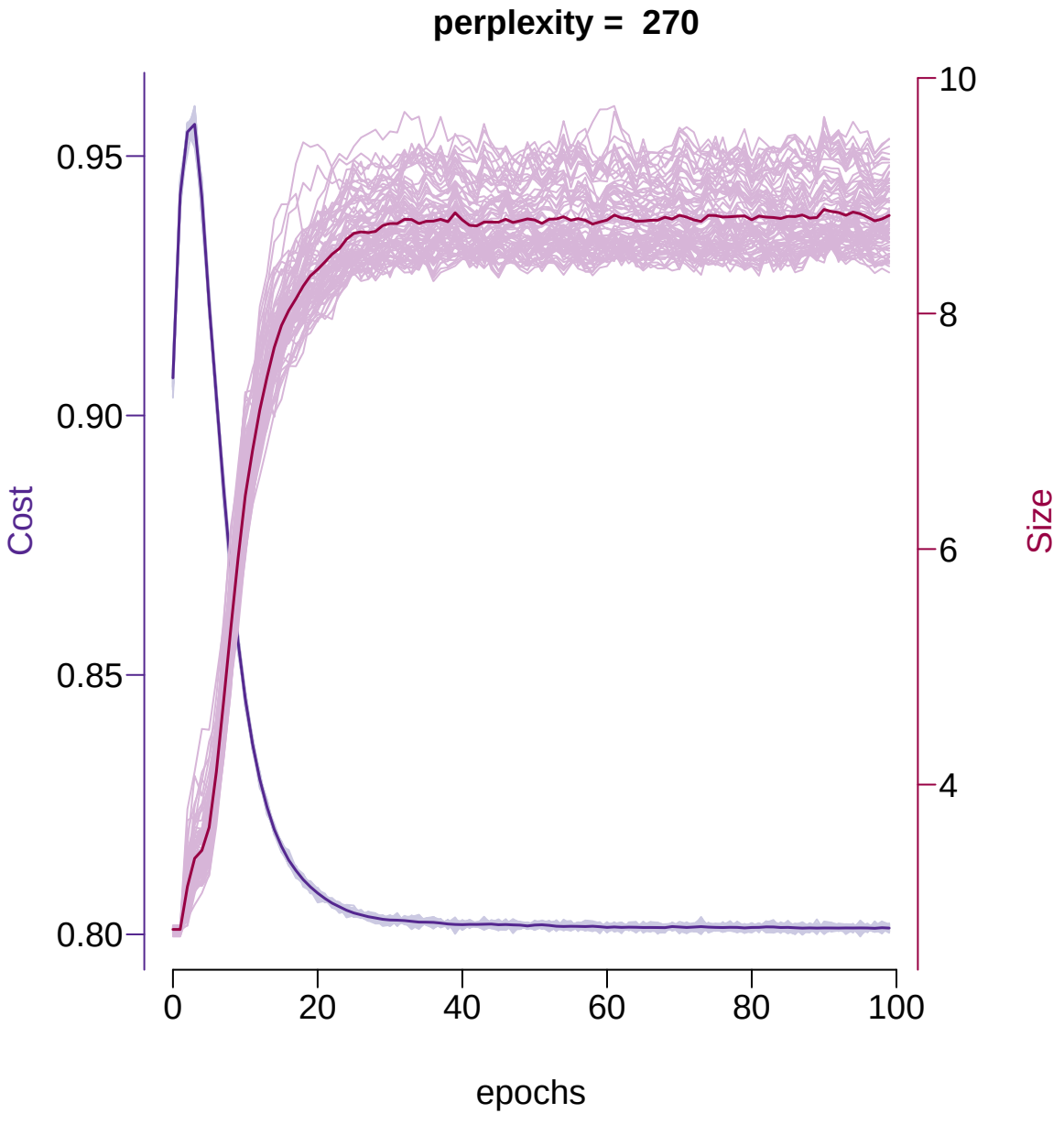

perplexity = 250
Y2
Y1
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

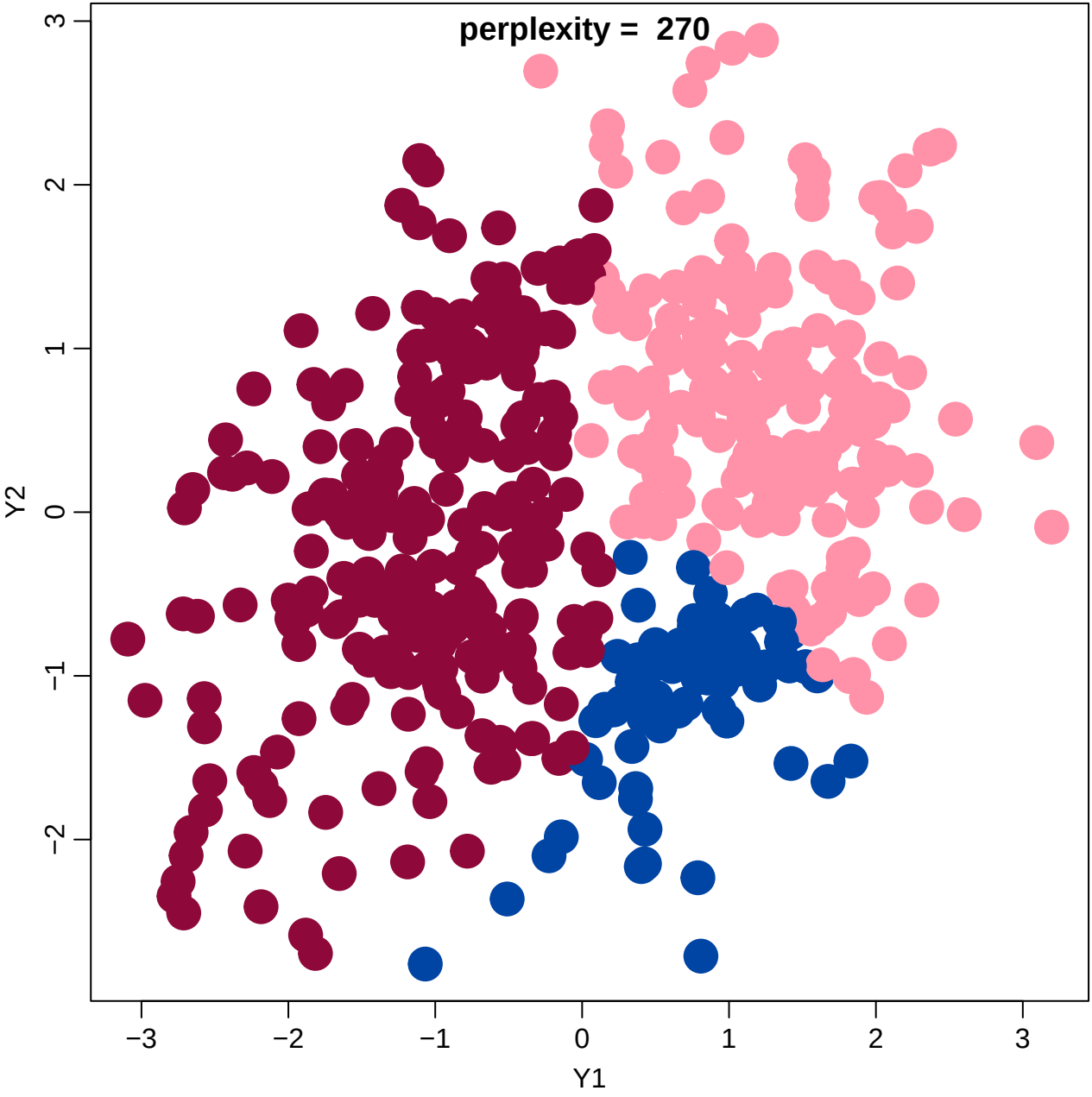

perplexity = 250
2
3
1

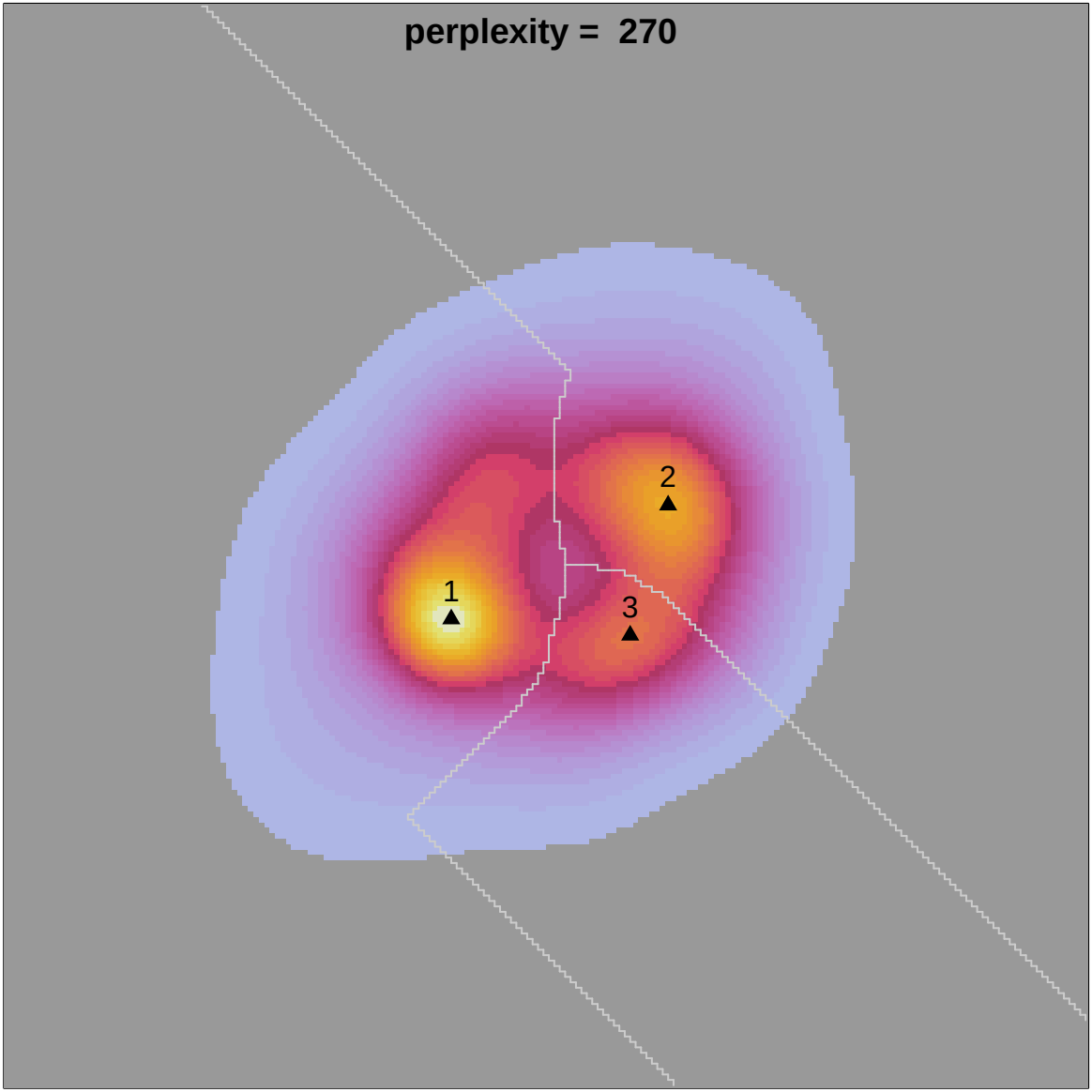

perplexity = 260
Cost
Size
0.95
0.90
0.85
0.80
10
8
6
4
0
20
40
60
80
100
epochs

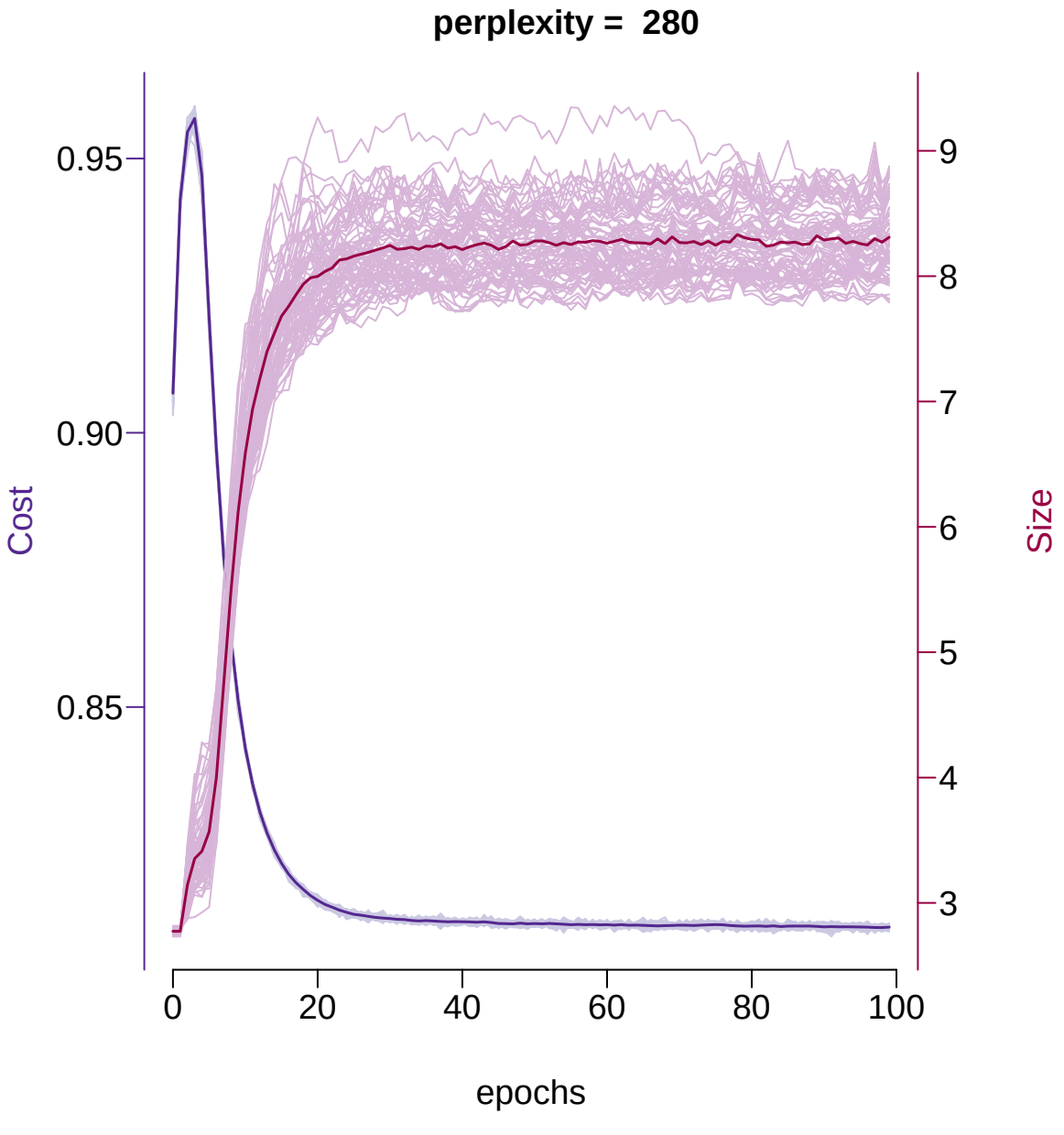

perplexity = 260
Y2
Y1
3
2
1
0
−1
−2
−3
−4
−2
−1
0
1
2

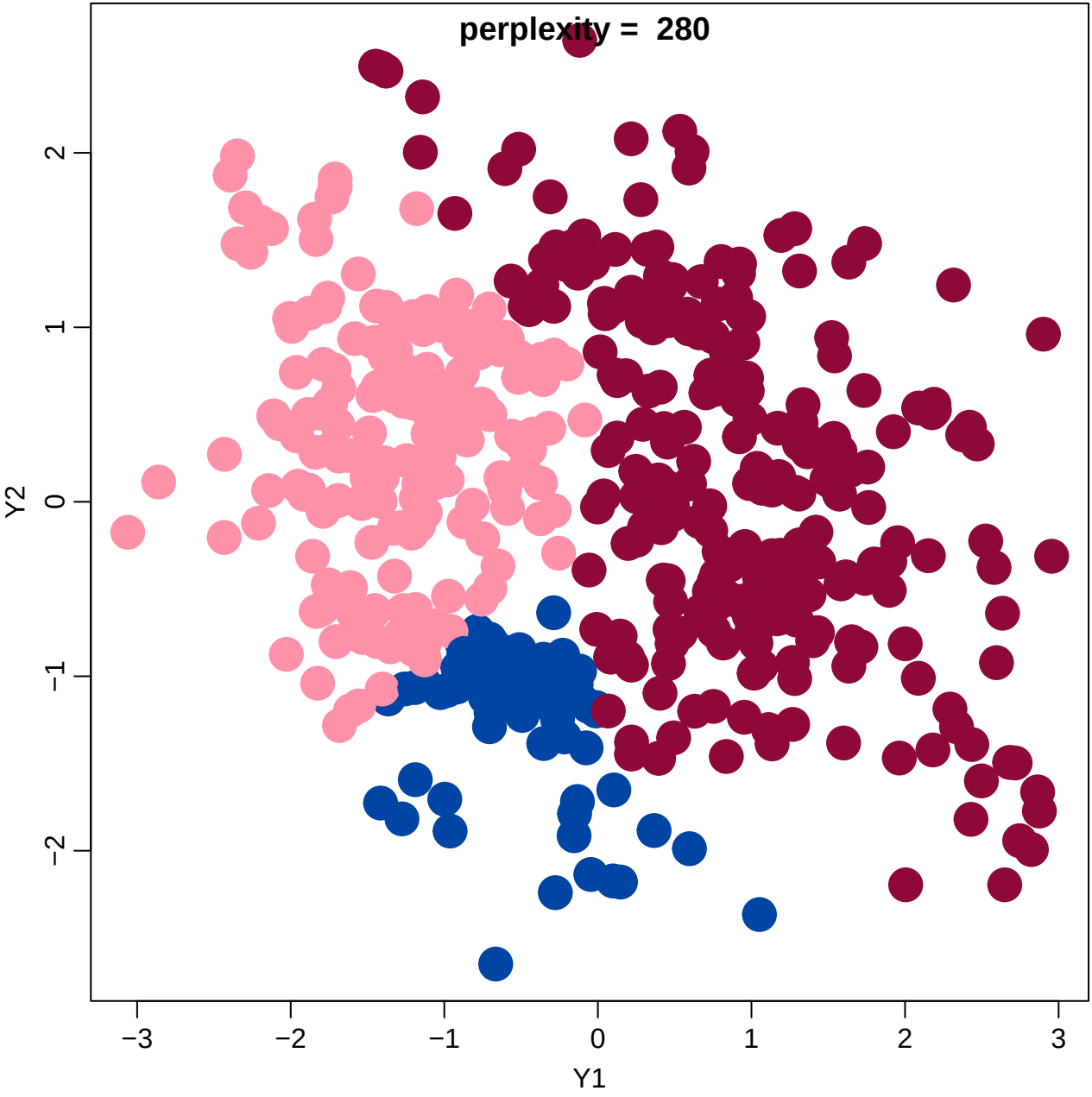

perplexity = 260
2
3
1

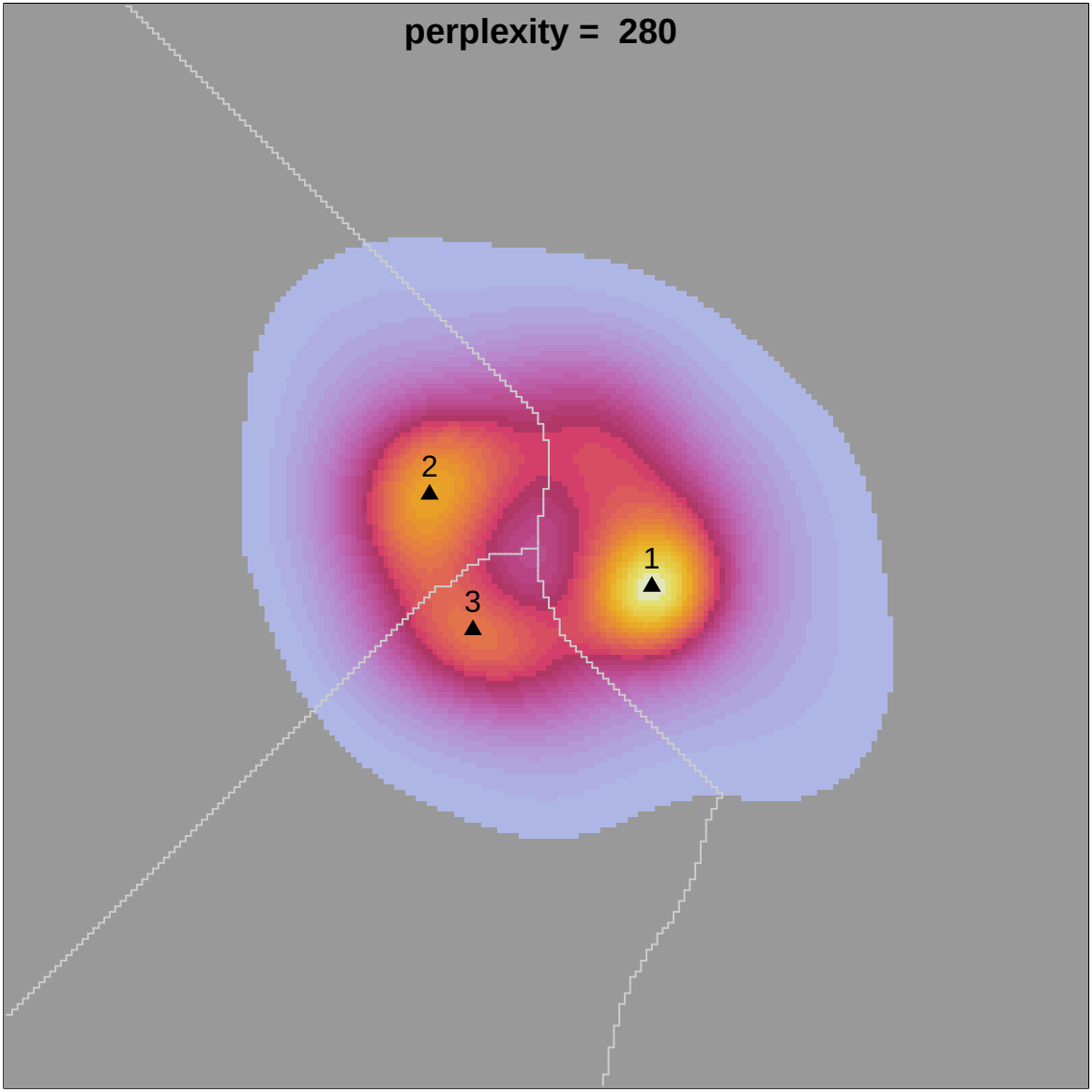

perplexity = 270
Cost
Size
0.95
0.90
0.85
0.80
10
8
6
4
0
20
40
60
80
100
epochs

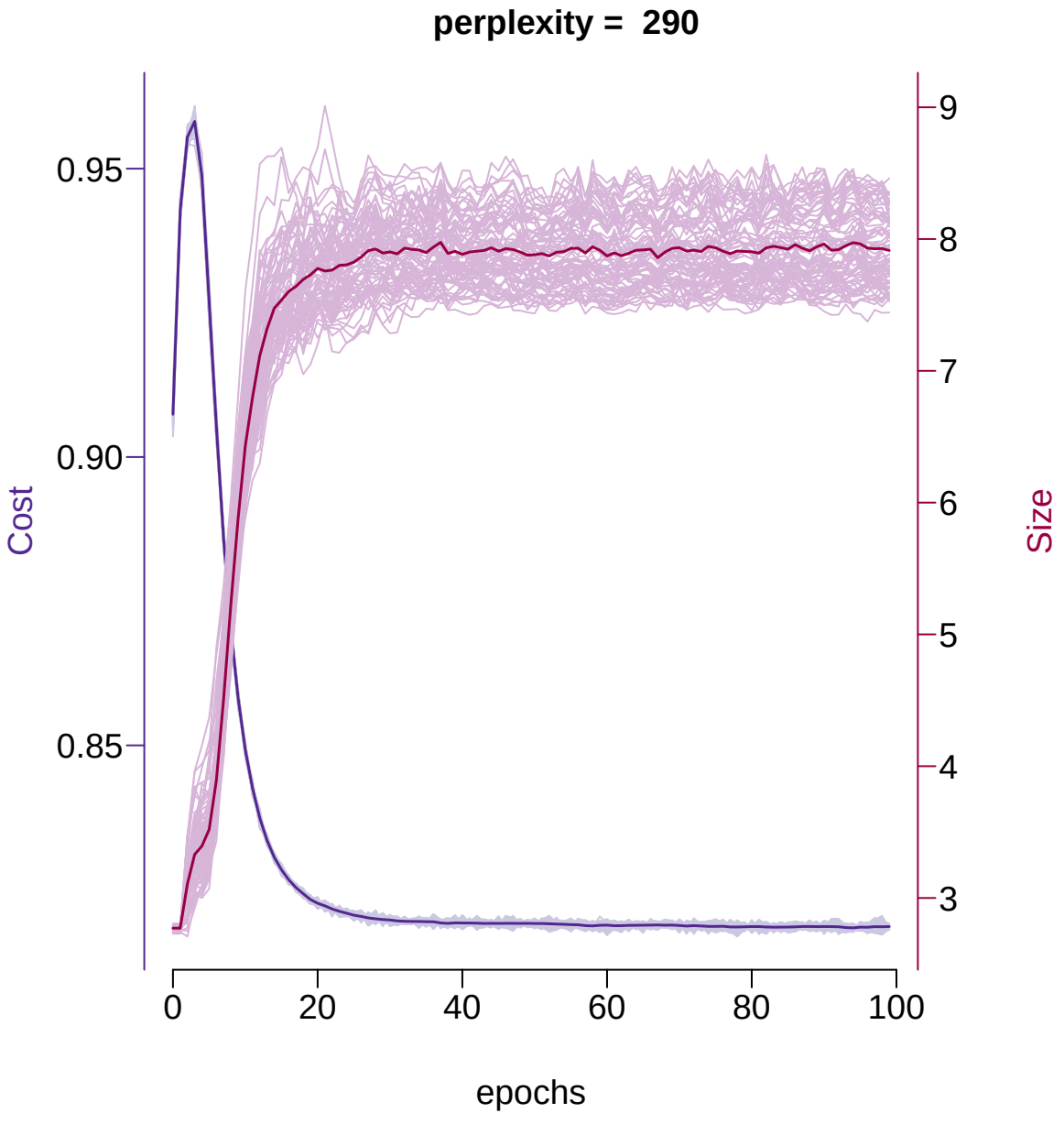

perplexity = 270
Y2
Y1
3
2
1
0
−1
−2
−3
−2
−1
0
1
2
3

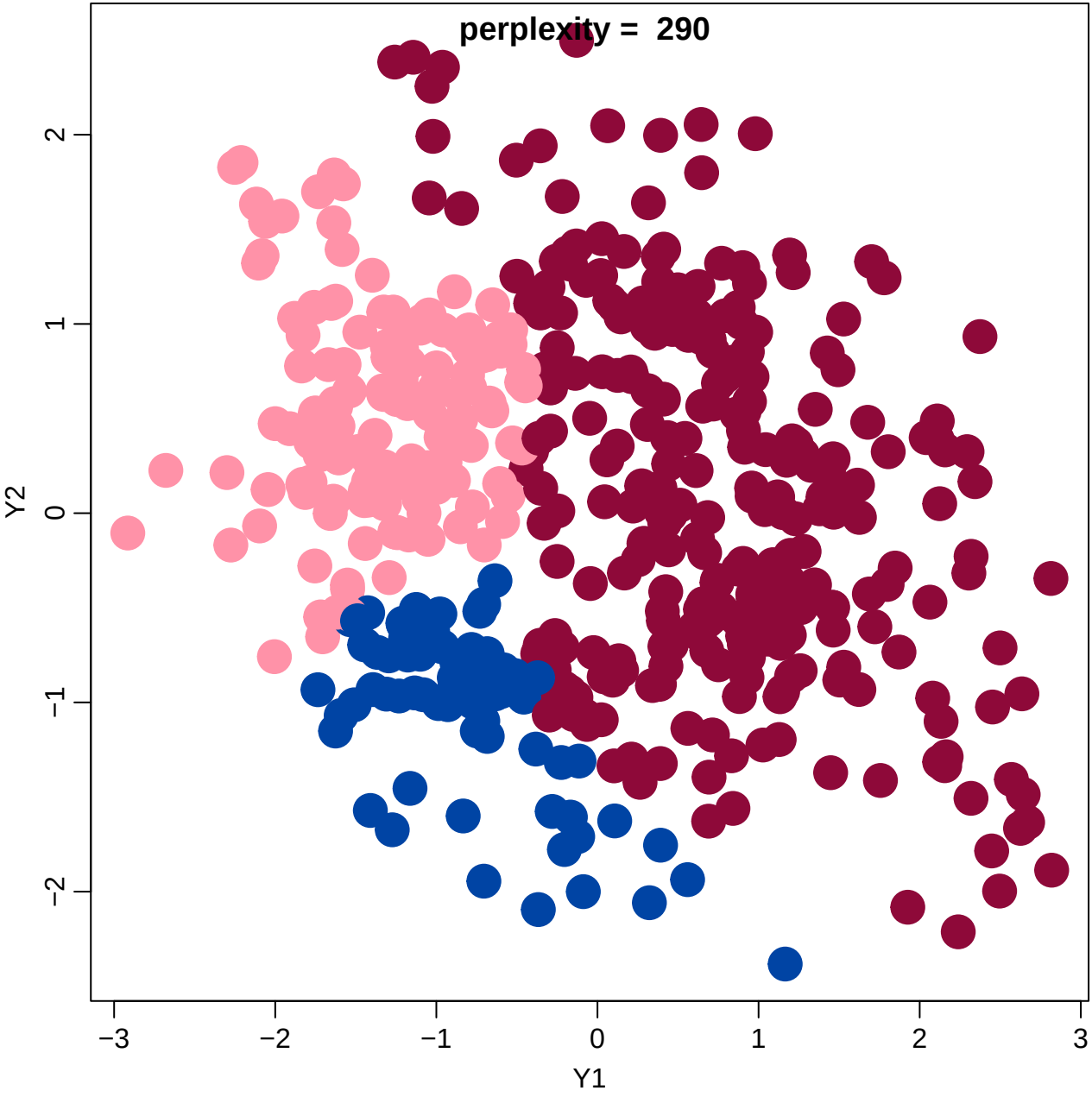

perplexity = 270
1
2
3

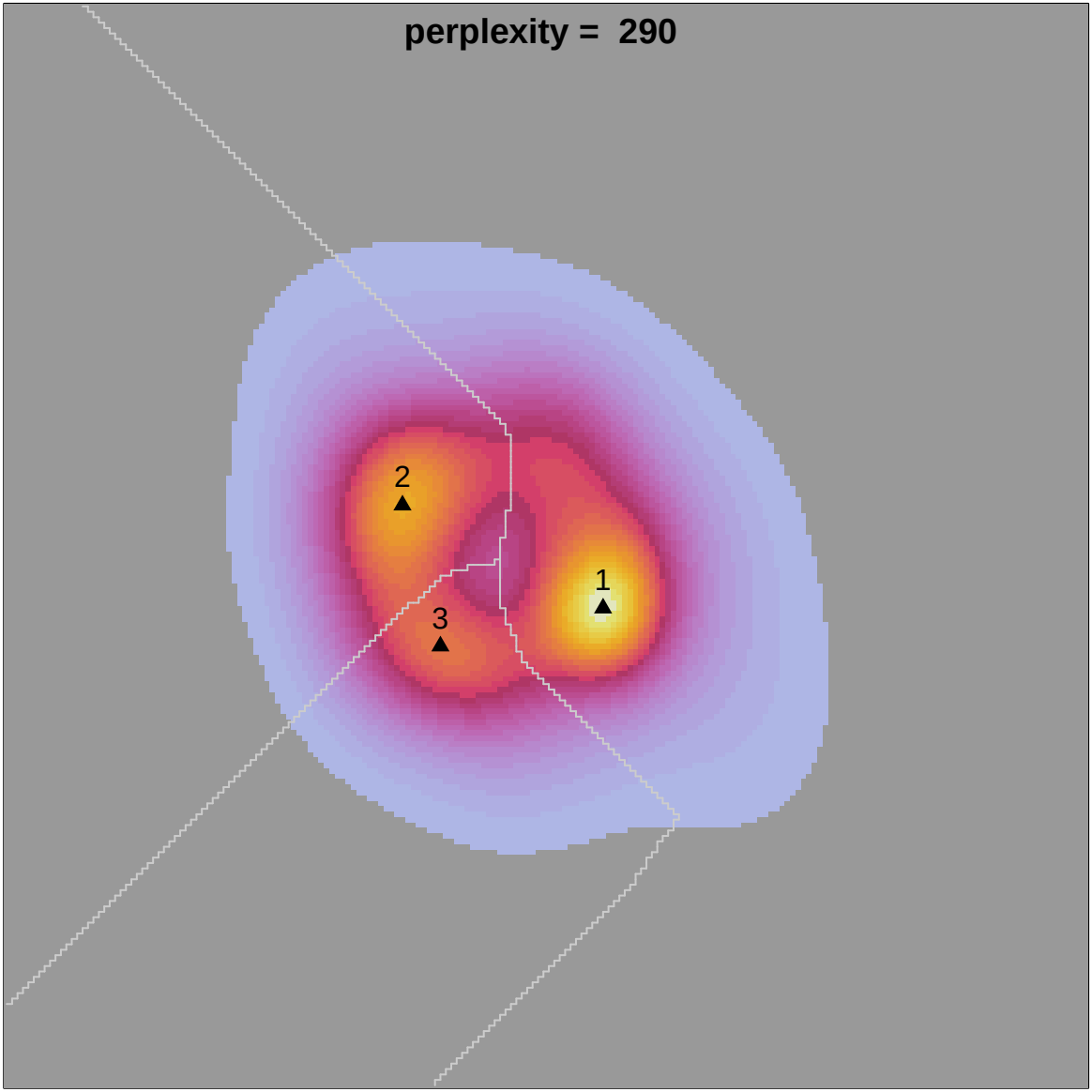

perplexity = 280
Cost
0.95
0.90
0.85
Size
9
8
7
6
5
4
3
0
20
40
60
80
100
epochs

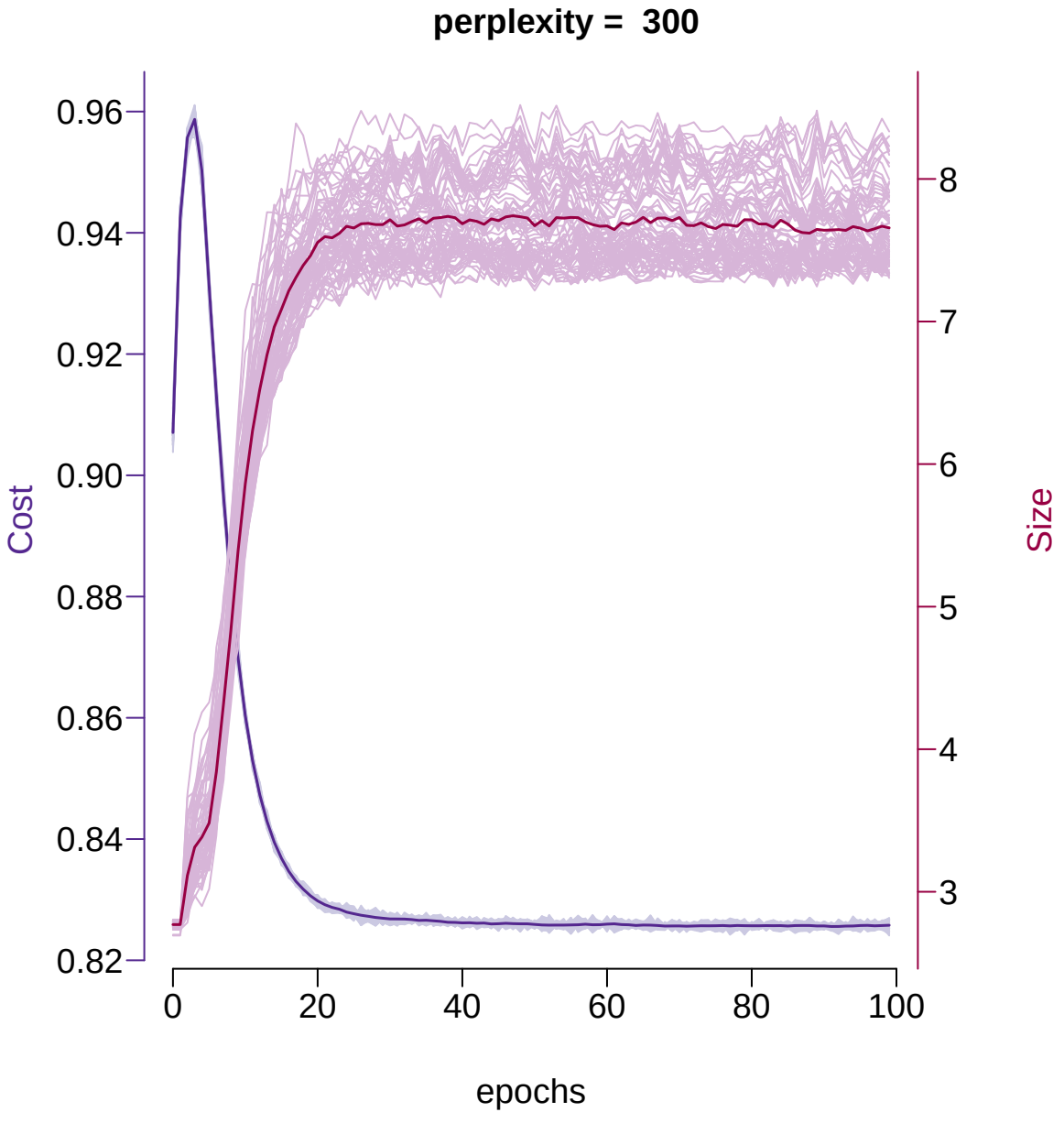

perplexity = 280
Y2
Y1
2
1
0
−1
−2
−3
−2
−1
0
1
2
3

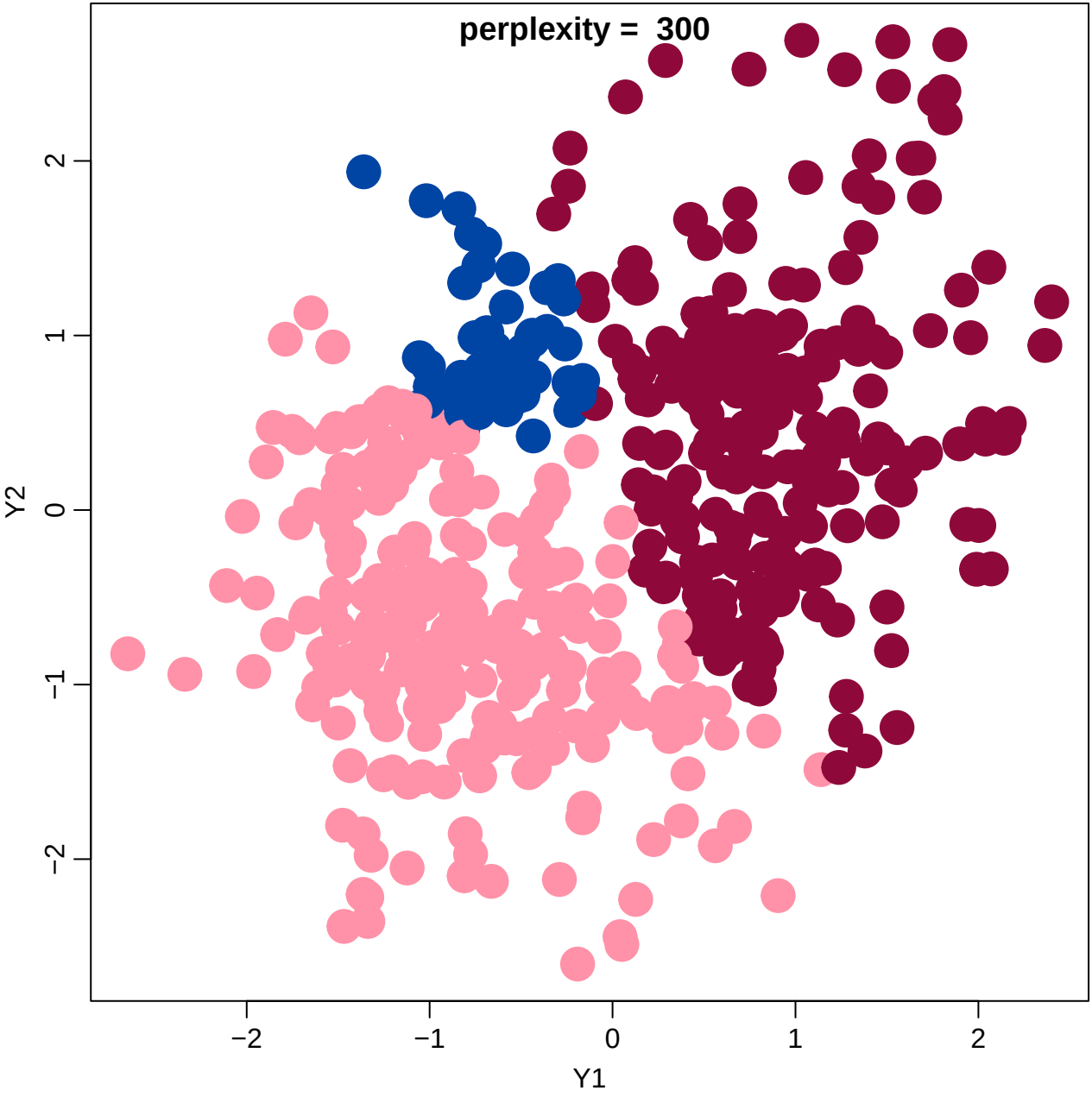

perplexity = 280
2
1
3

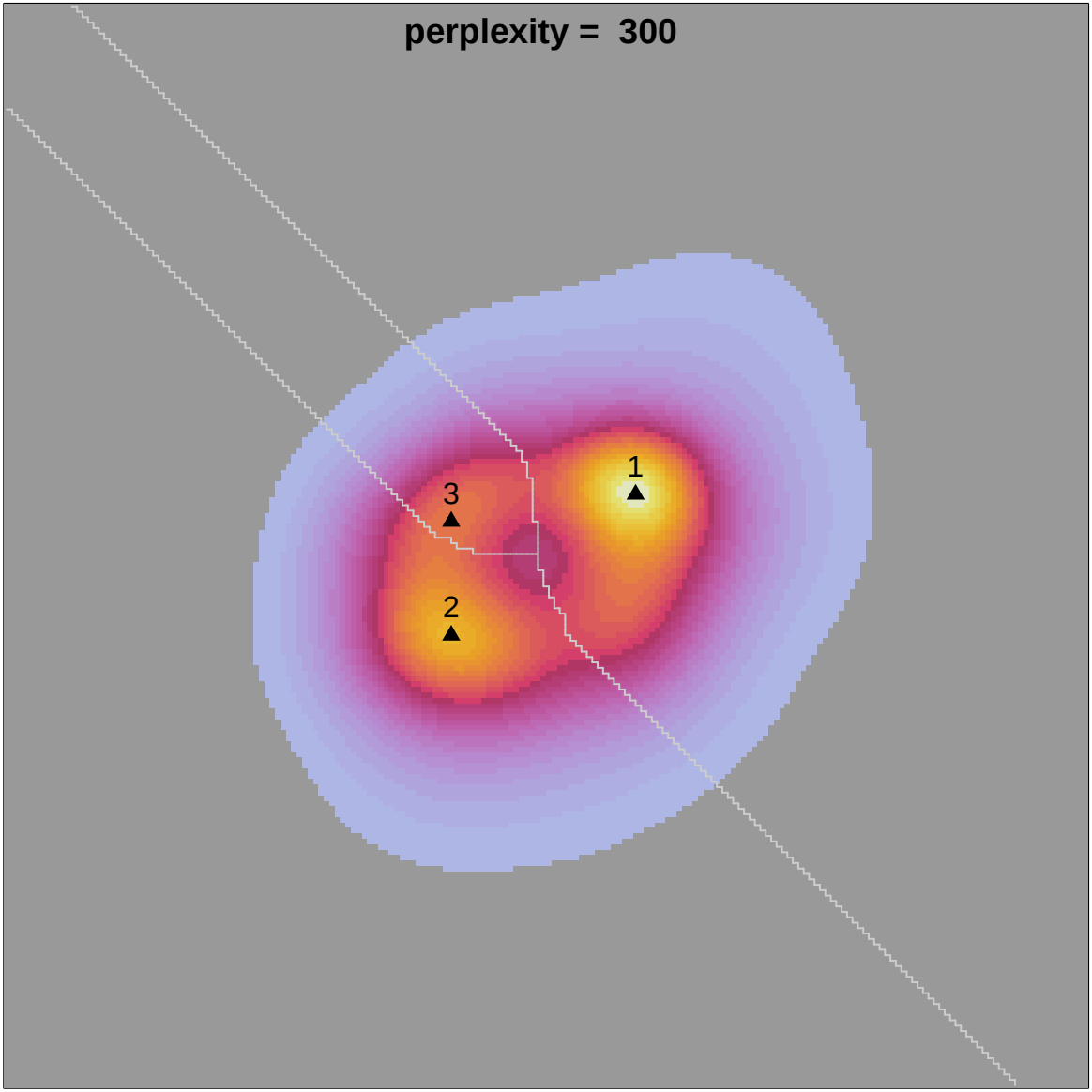

perplexity = 290
Cost
0.95
0.90
0.85
Size
9
8
7
6
5
4
3
0
20
40
60
80
100
epochs

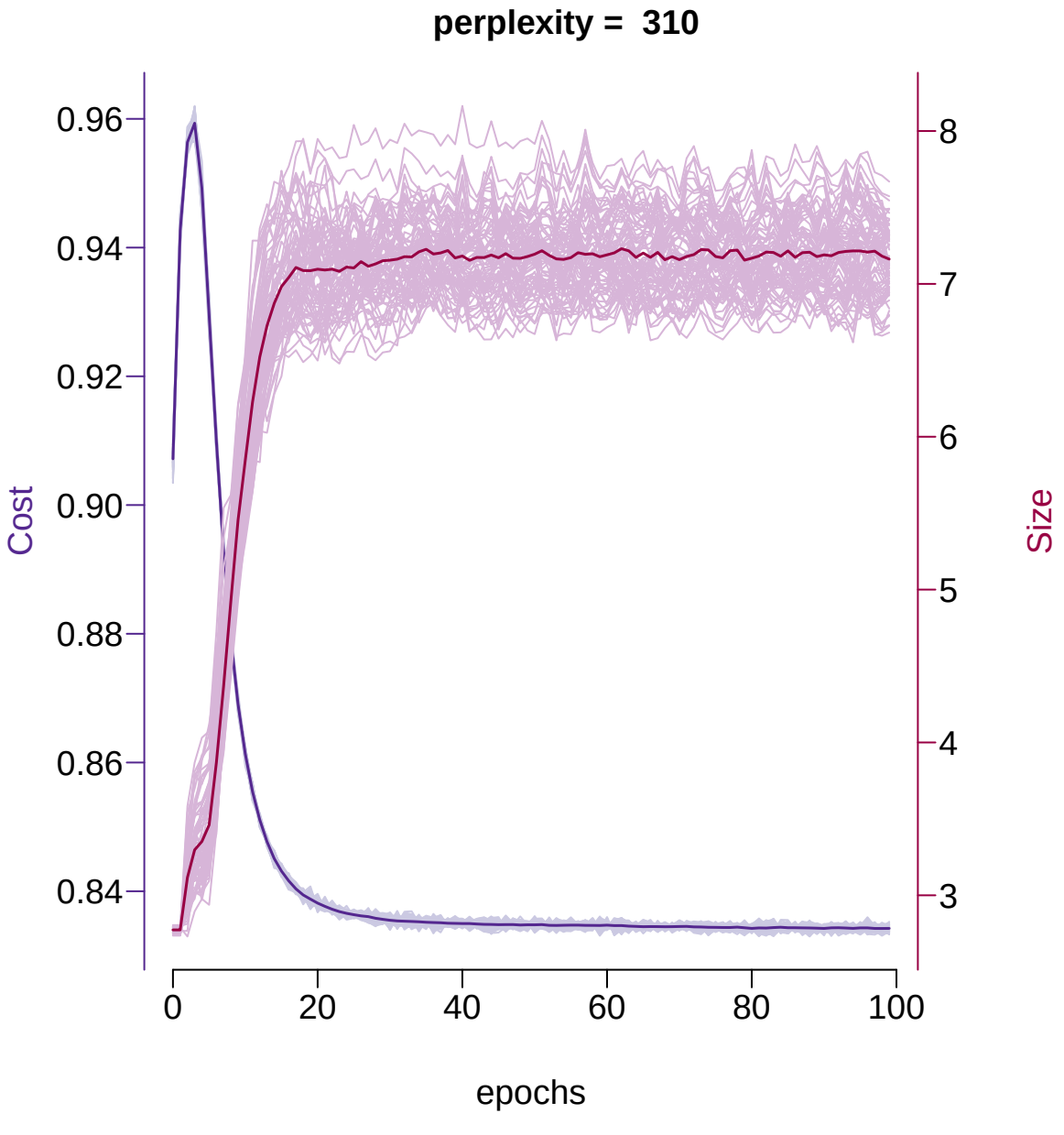

perplexity = 290
Y2
Y1
2
1
0
−1
−2
−3
−2
−1
0
1
2
3

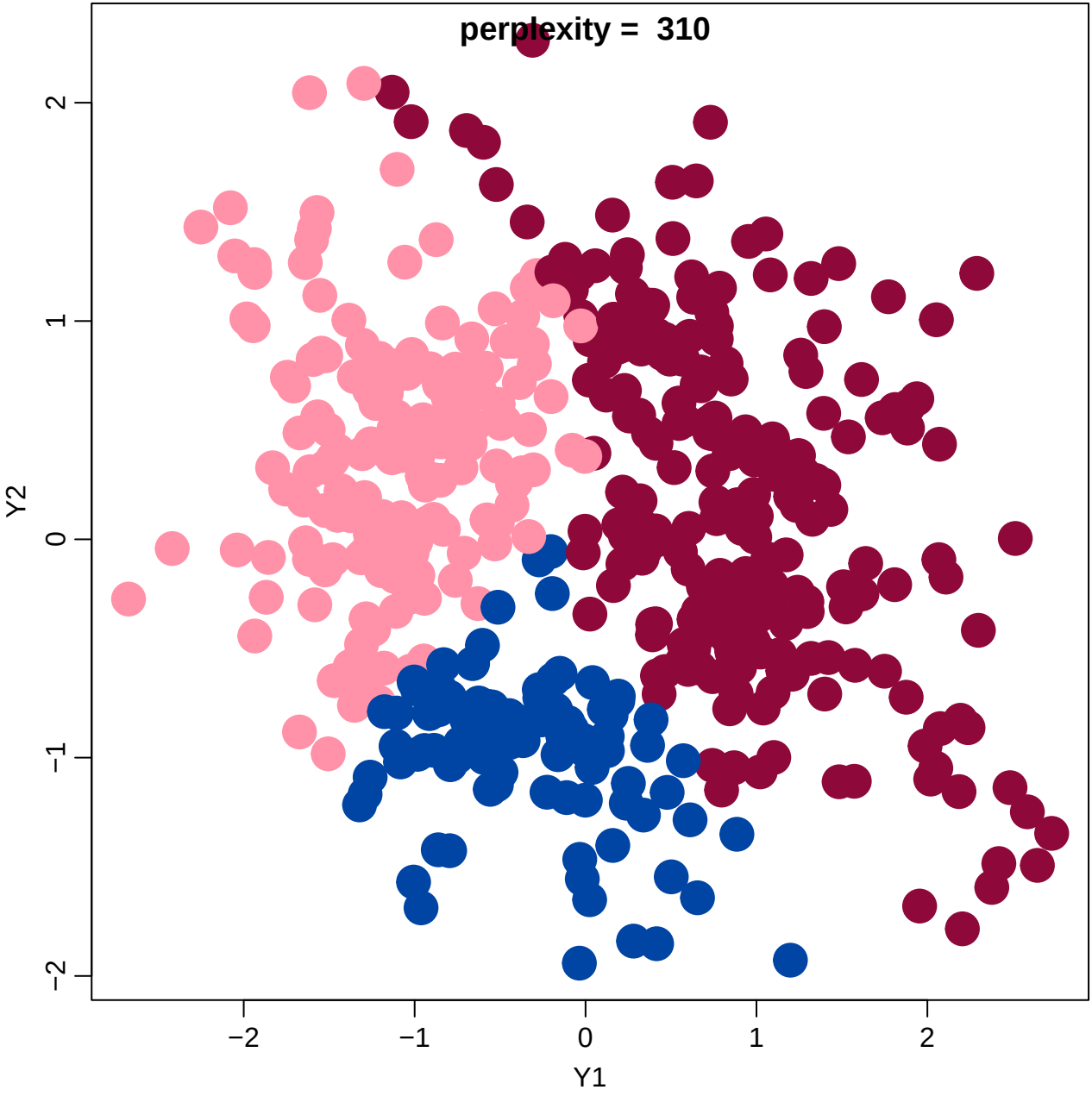

perplexity = 290
2
1
3

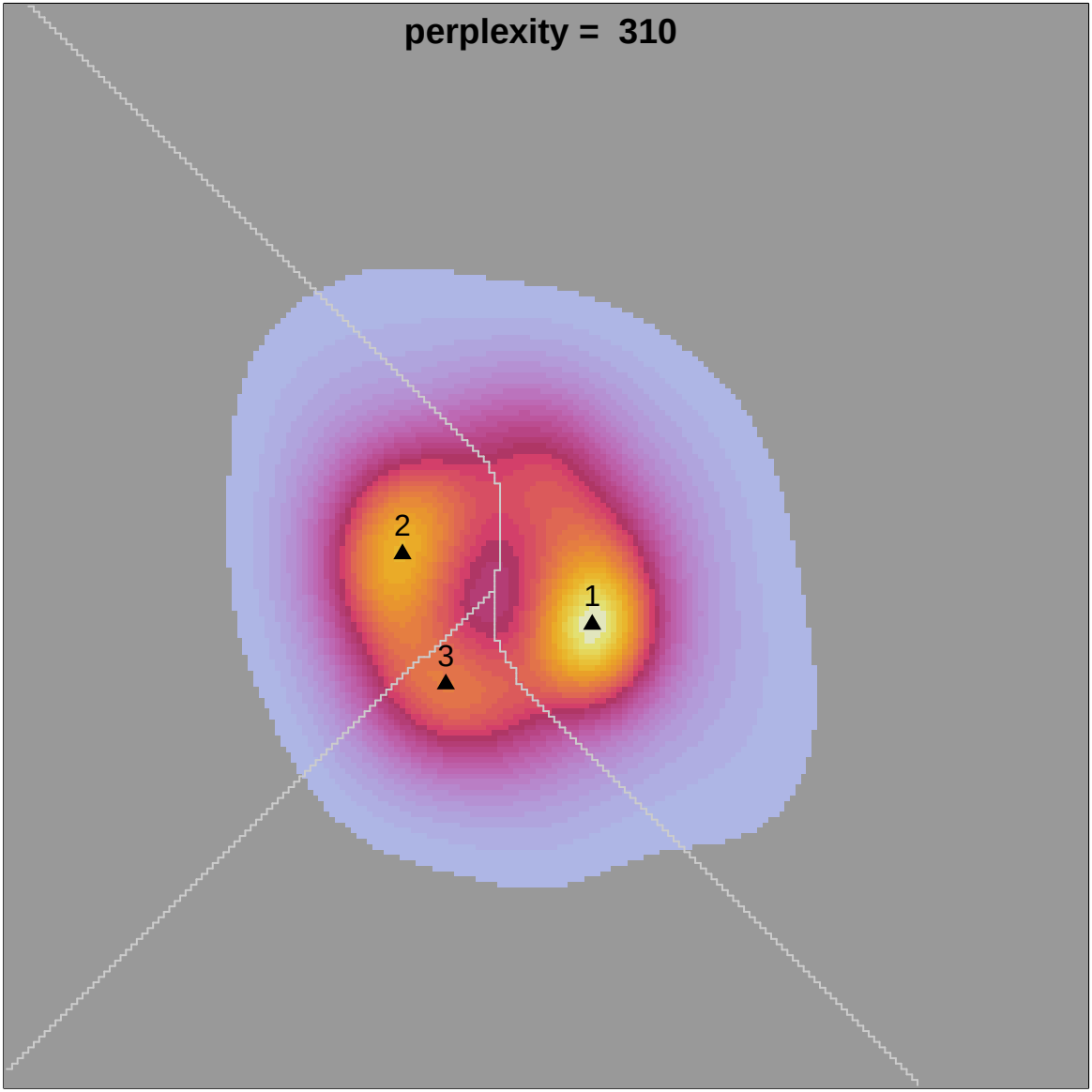

perplexity = 300
Cost
Size
epochs
0.96
0.94
0.92
0.90
0.88
0.86
0.84
0.82
8
7
6
5
4
3
0
20
40
60
80
100

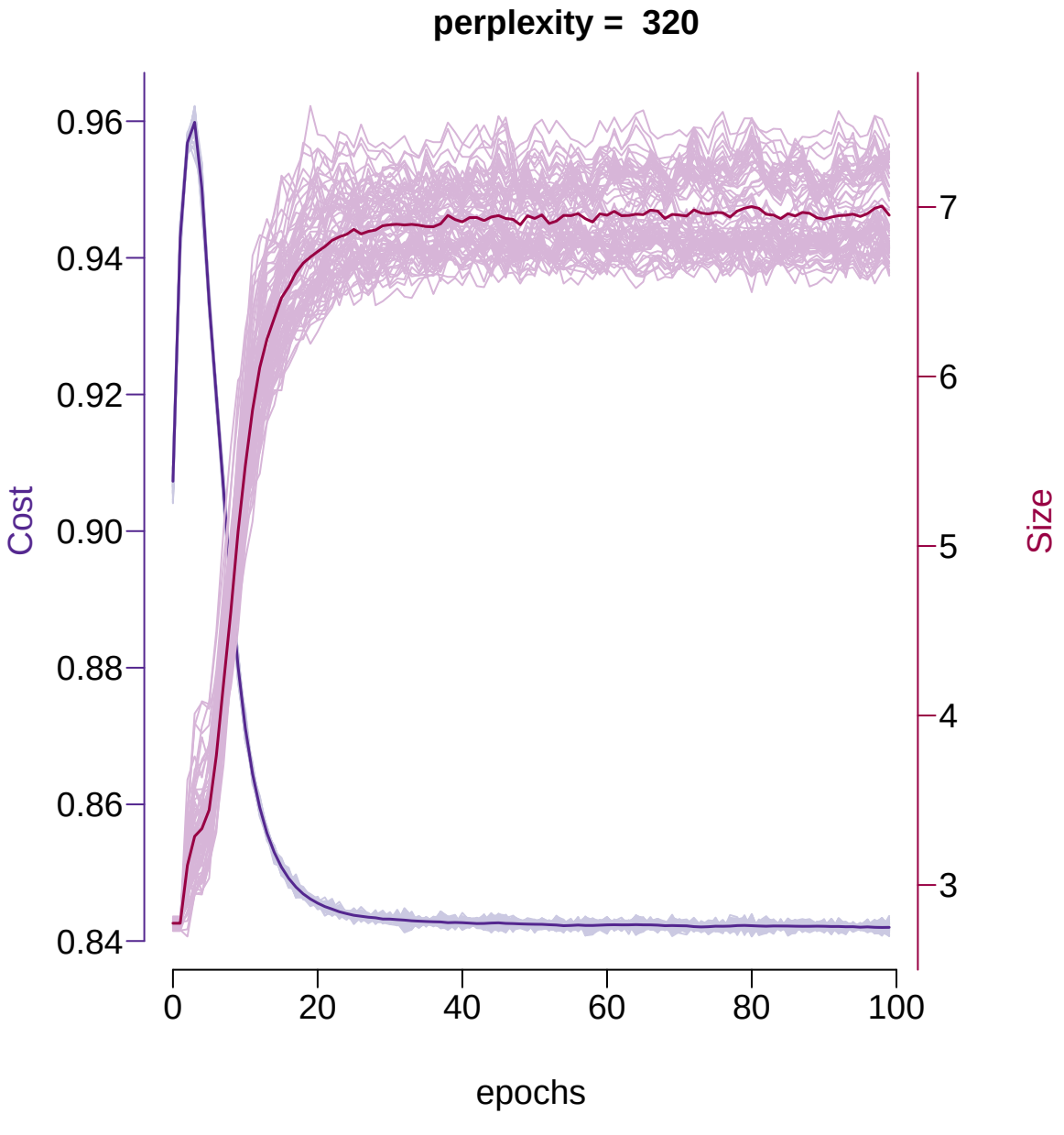

perplexity = 300
Y2
Y1
2
1
0
−1
−2
−2
−1
0
1
2

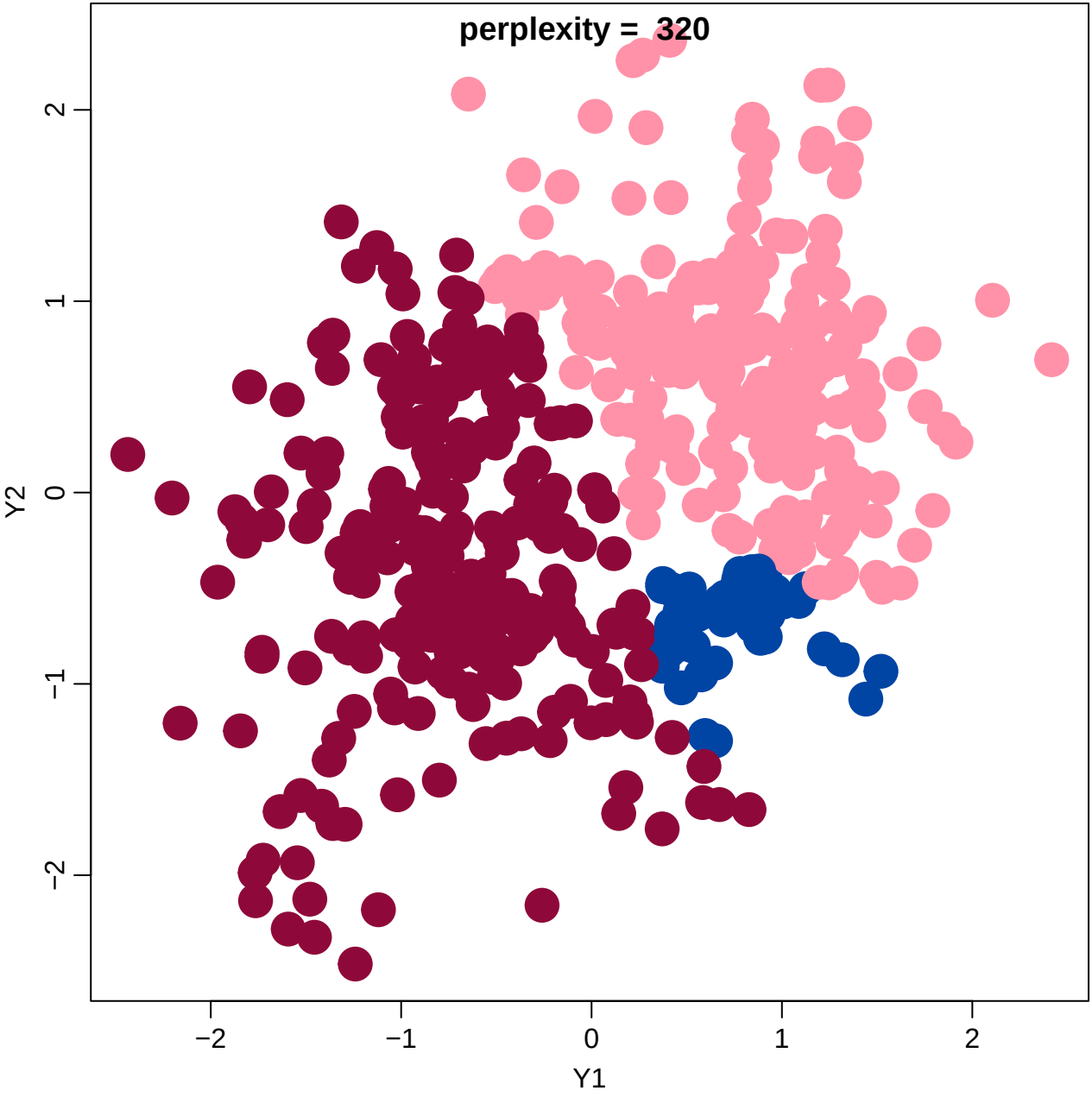

perplexity = 300
1
2
3

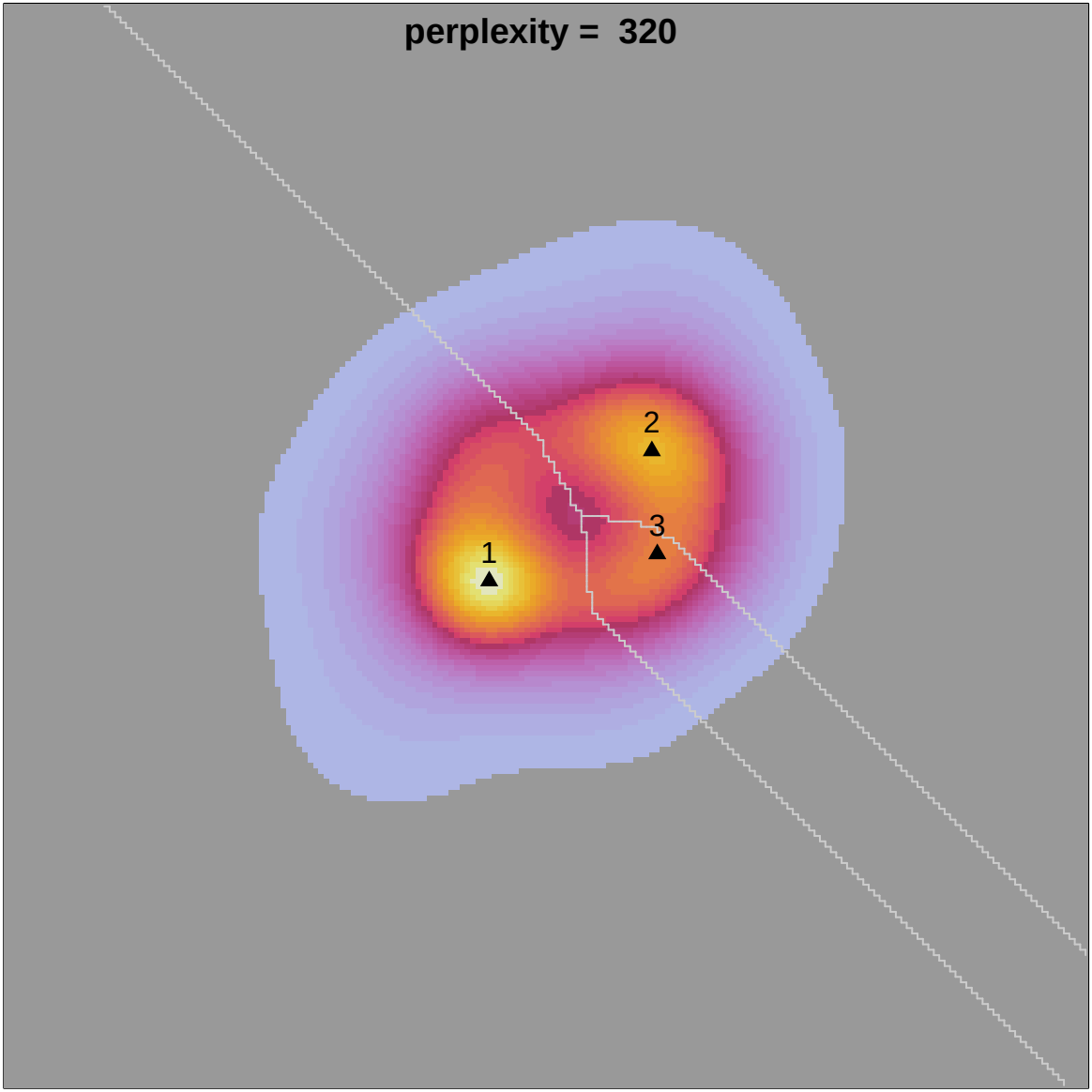

perplexity = 310
Cost
Size
epochs
0.96
0.94
0.92
0.90
0.88
0.86
0.84
8
7
6
5
4
3
0
20
40
60
80
100

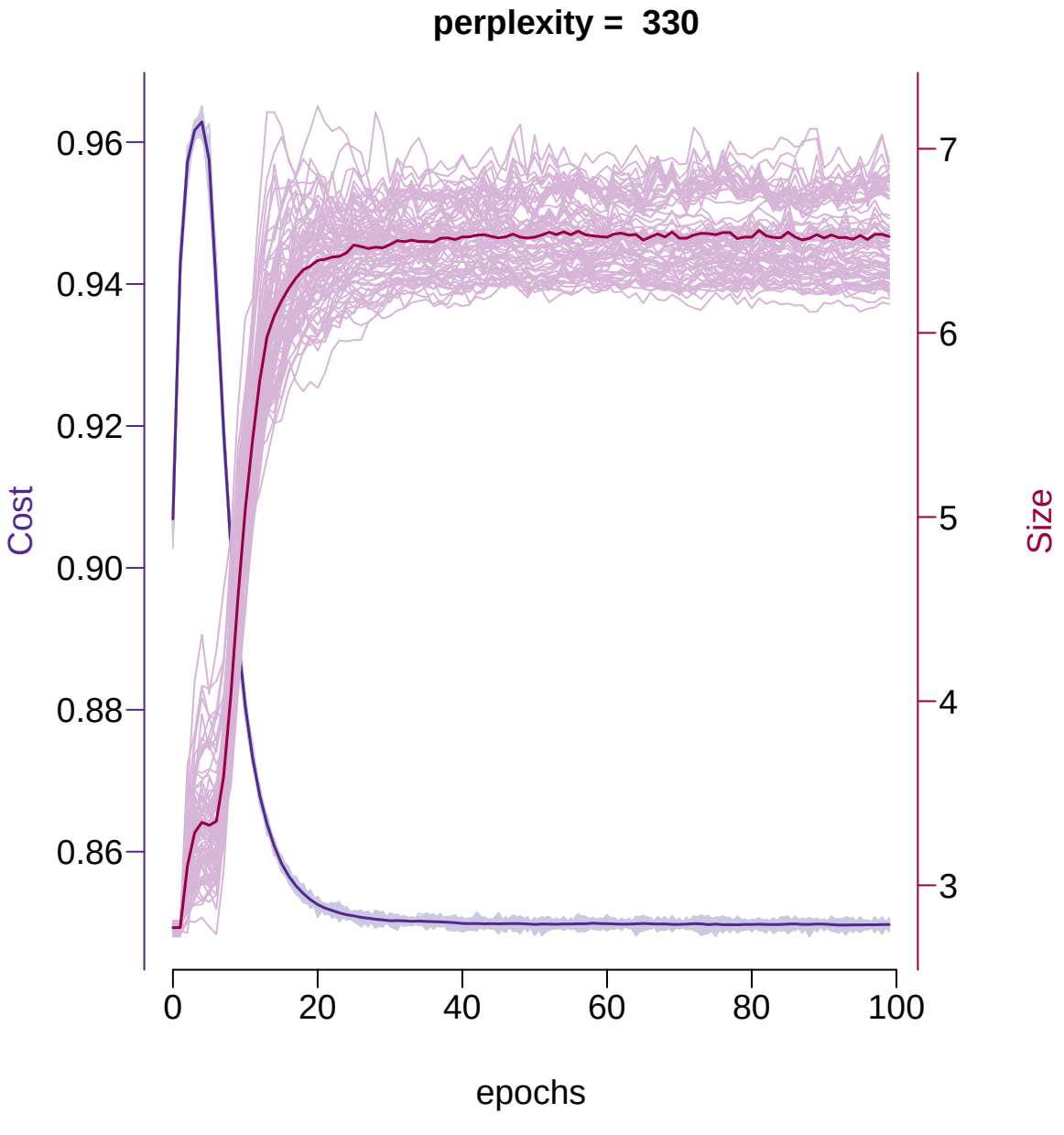

perplexity = 310
Y2
Y1
2
1
0
−1
−2
−2
−1
0
1
2

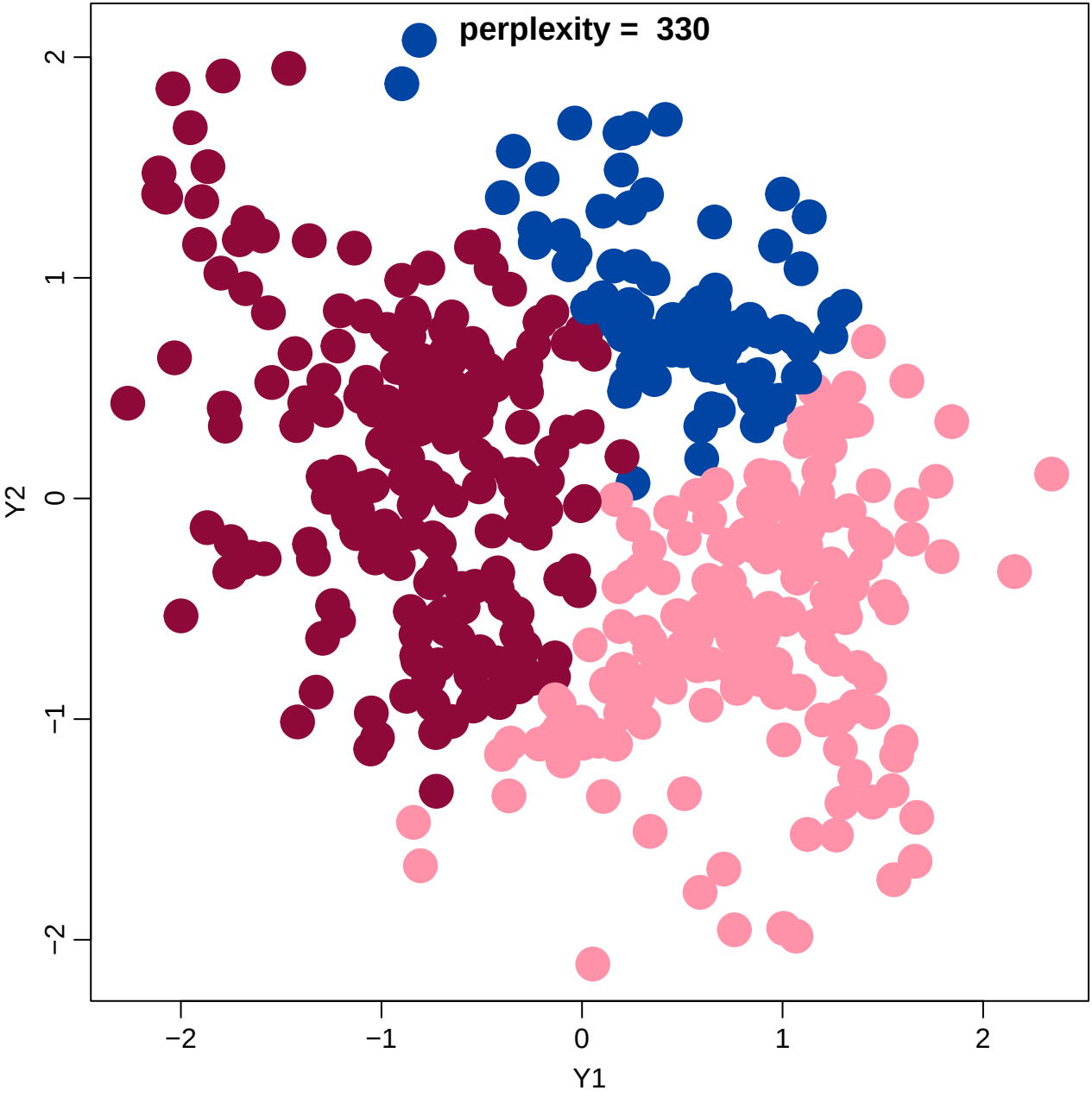

perplexity = 310
2
1
3

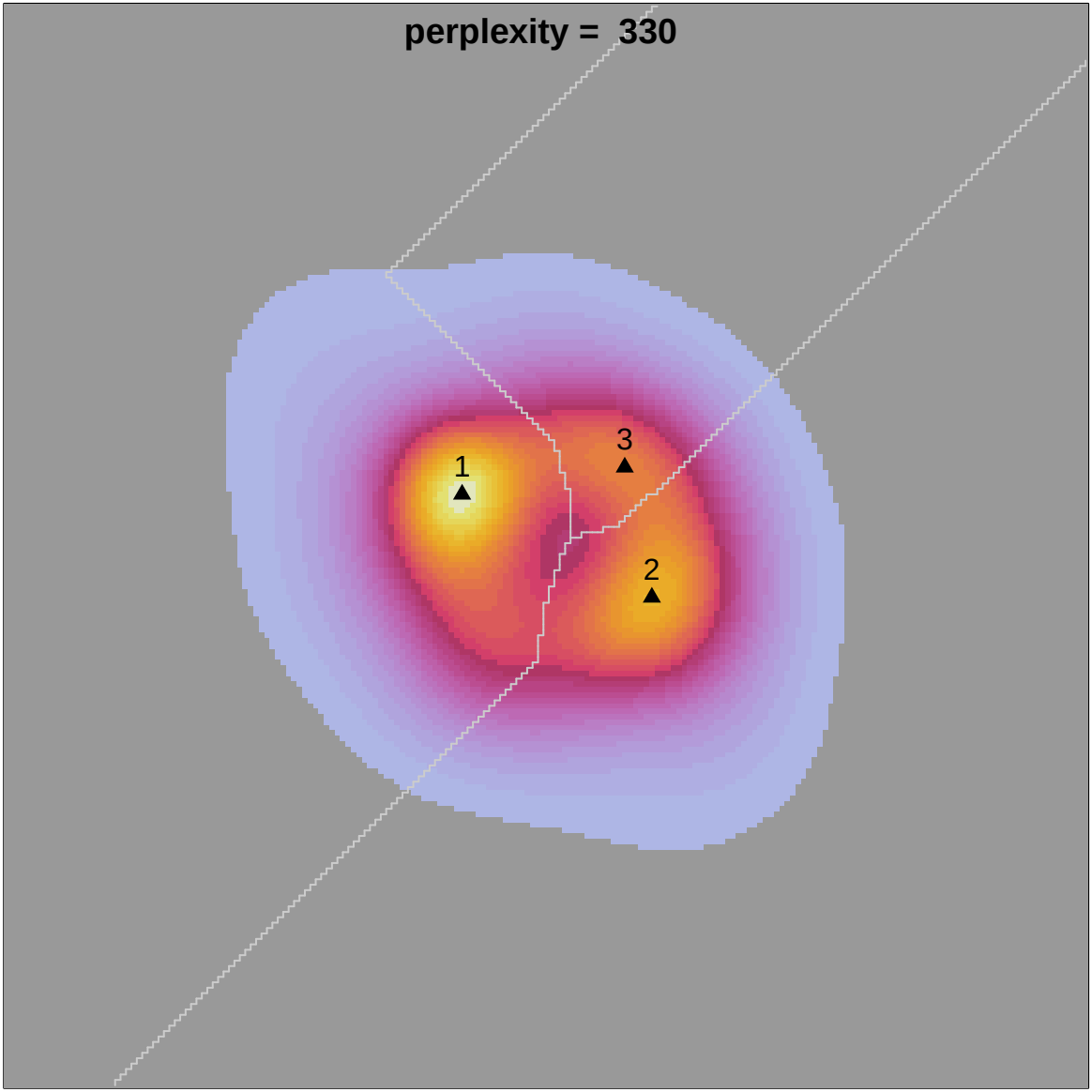

perplexity = 320
Cost
0.96
0.94
0.92
0.90
0.88
0.86
0.84
Size
7
6
5
4
3
0
20
40
60
80
100
epochs

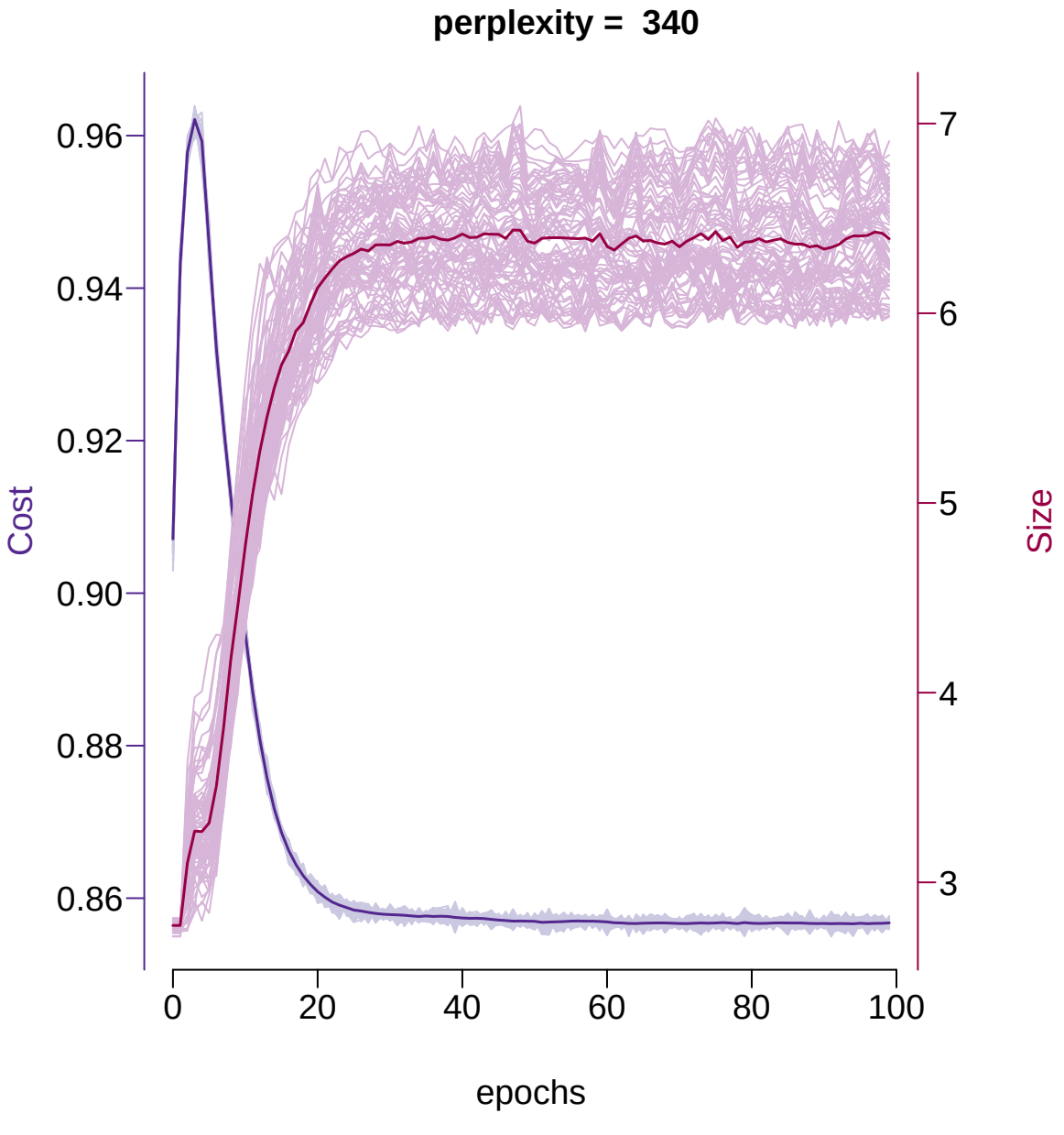

perplexity = 320
Y2
Y1
2
1
0
−1
−2
−2
−1
0
1
2

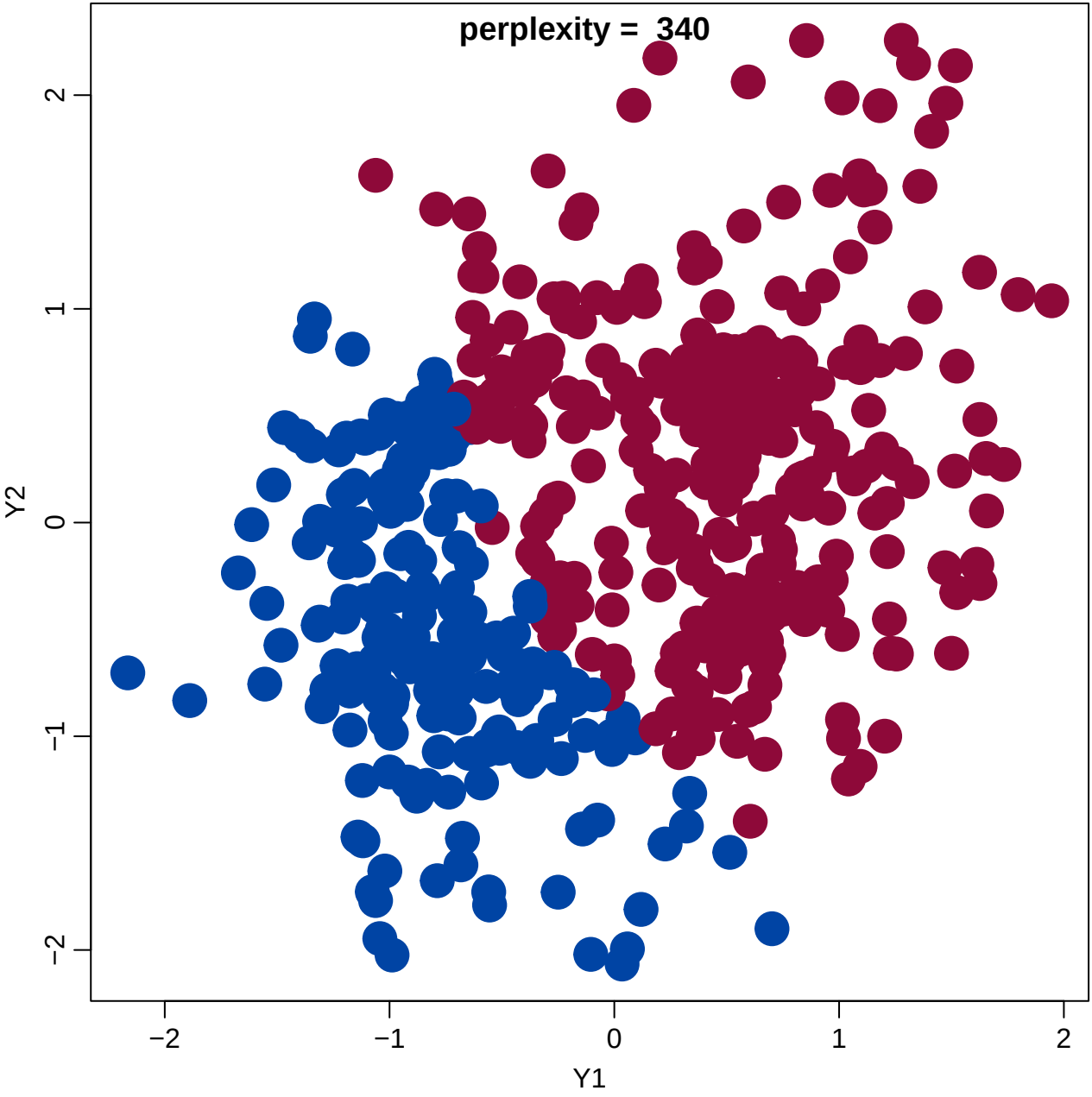

perplexity = 320
1
2
3

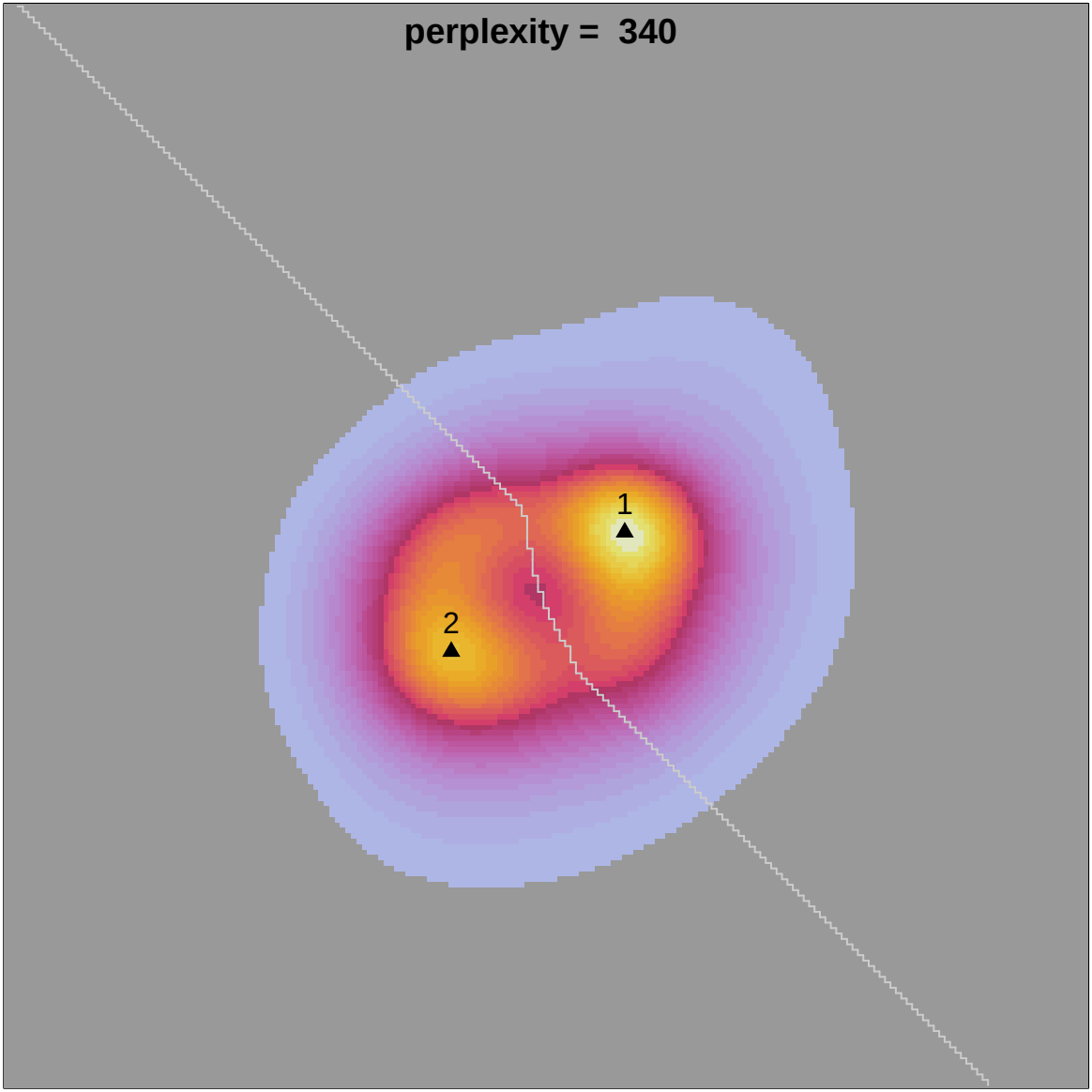

perplexity = 330
Cost
Size
0.96
0.94
0.92
0.90
0.88
0.86
7
6
5
4
3
0
20
40
60
80
100
epochs

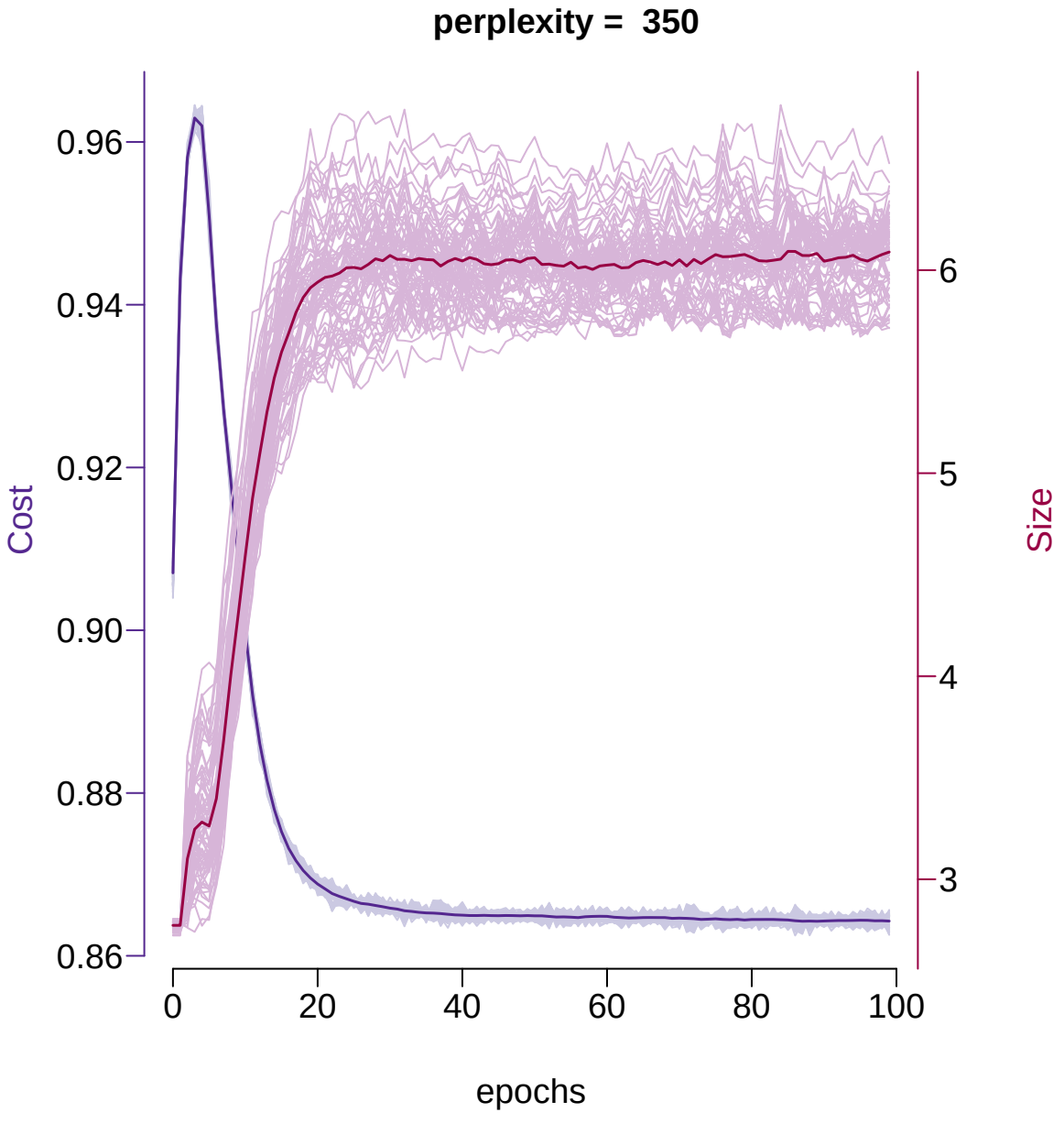

perplexity = 330
Y2
Y1
2
1
0
−1
−2
−2
−1
0
1
2

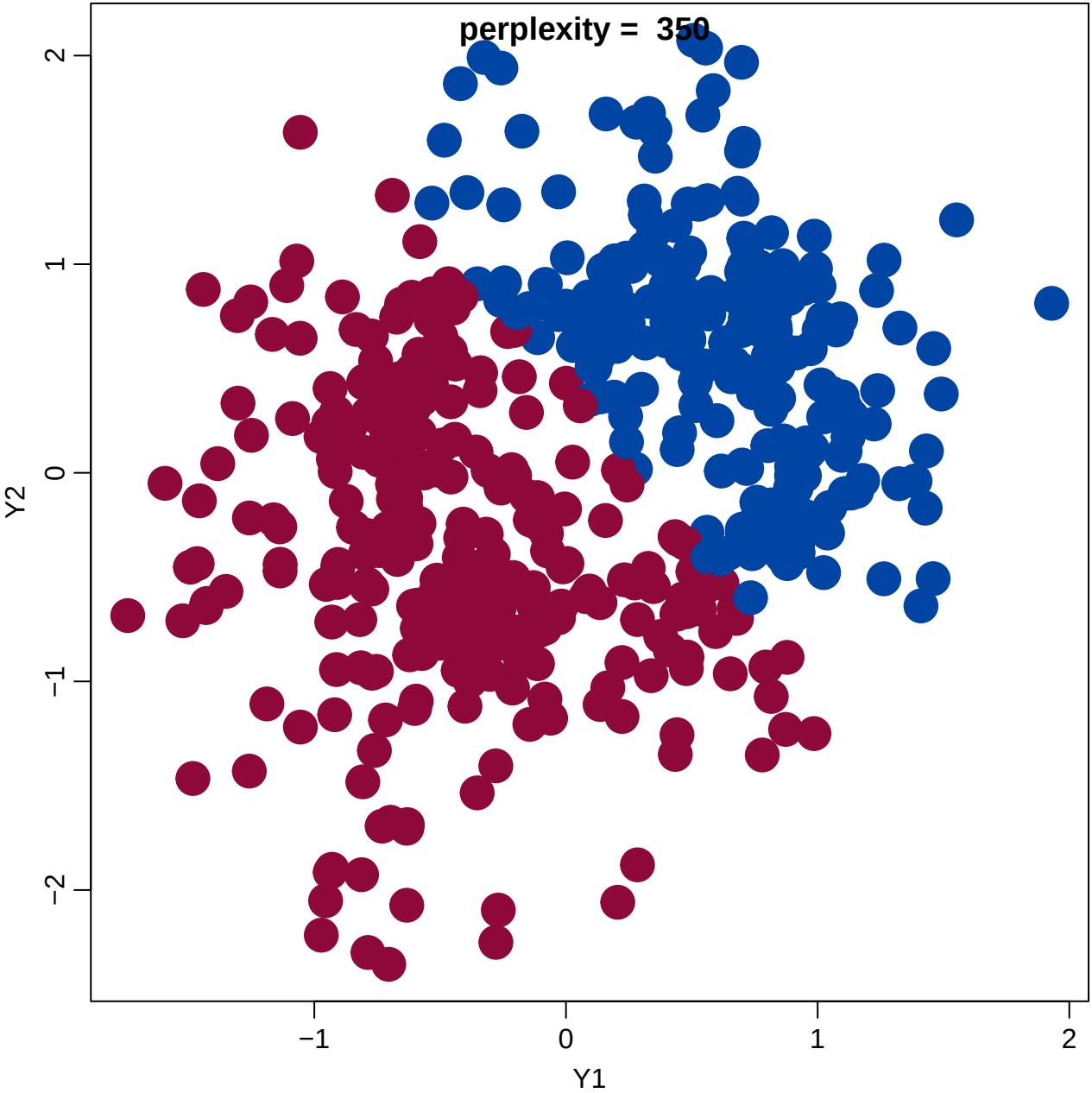

perplexity = 330
1
2
3

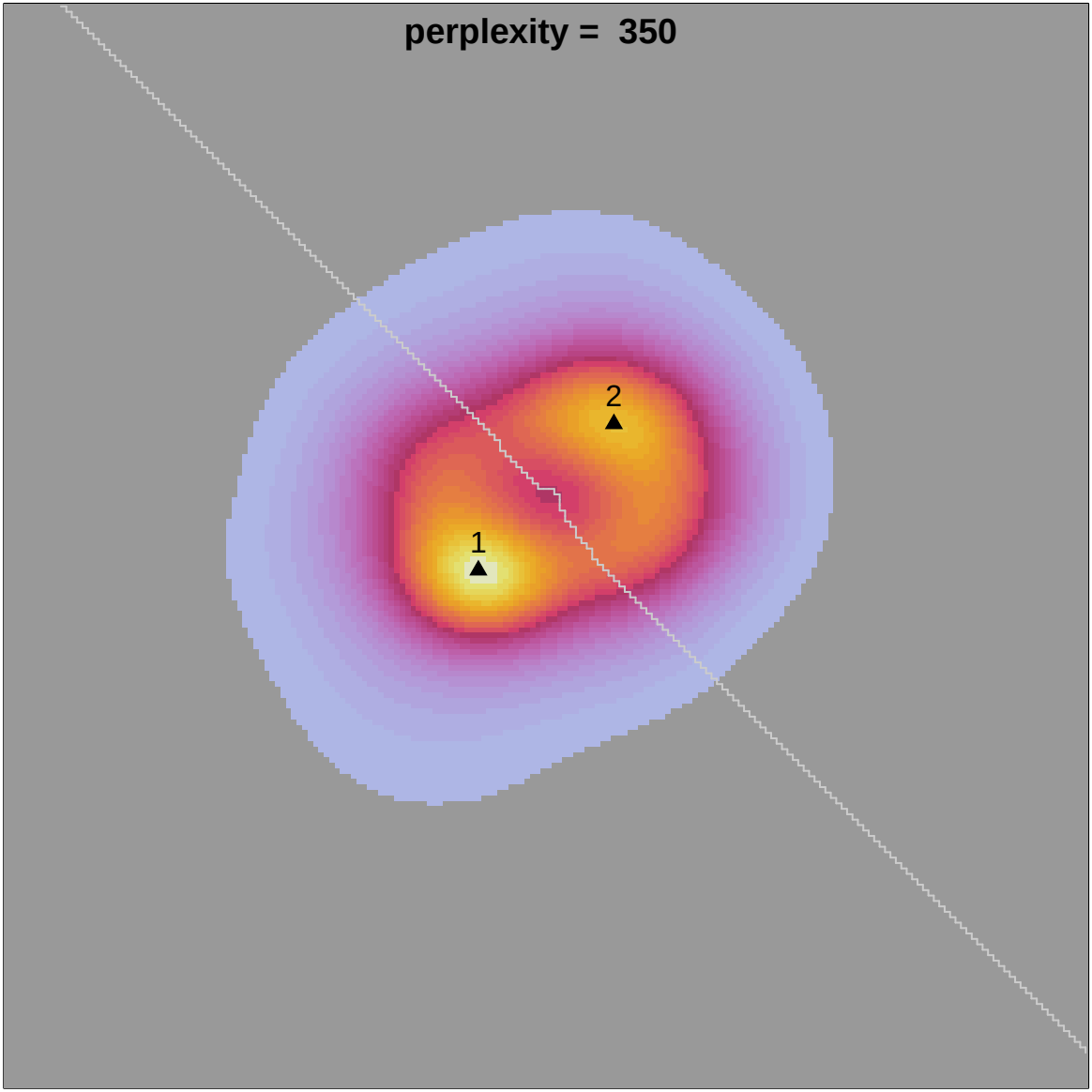

perplexity = 340
Cost
Size
epochs
0.96
0.94
0.92
0.90
0.88
0.86
7
6
5
4
3
0
20
40
60
80
100

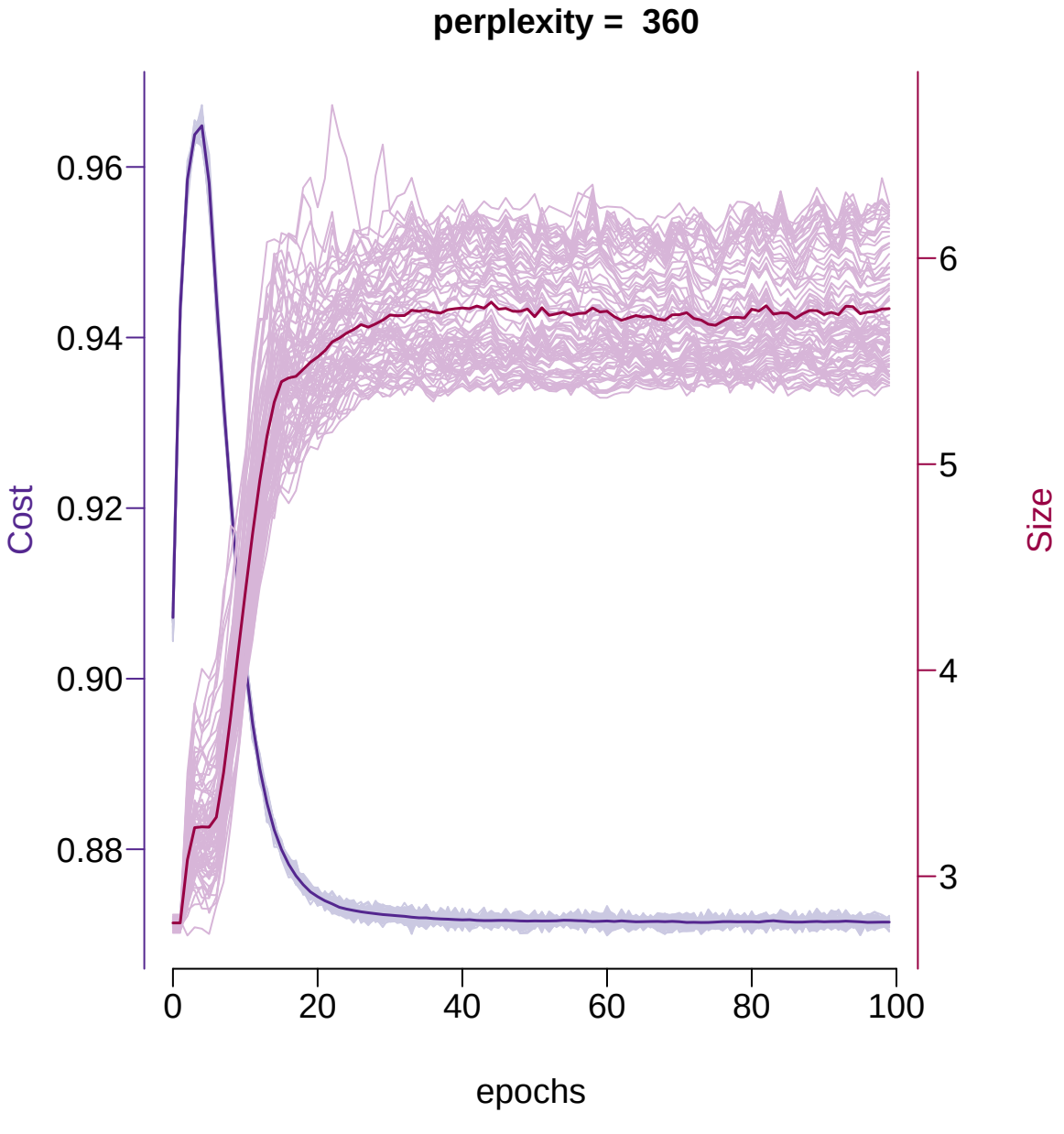

perplexity = 340
Y2
Y1
2
1
0
−1
−2
−2
−1
0
1
2

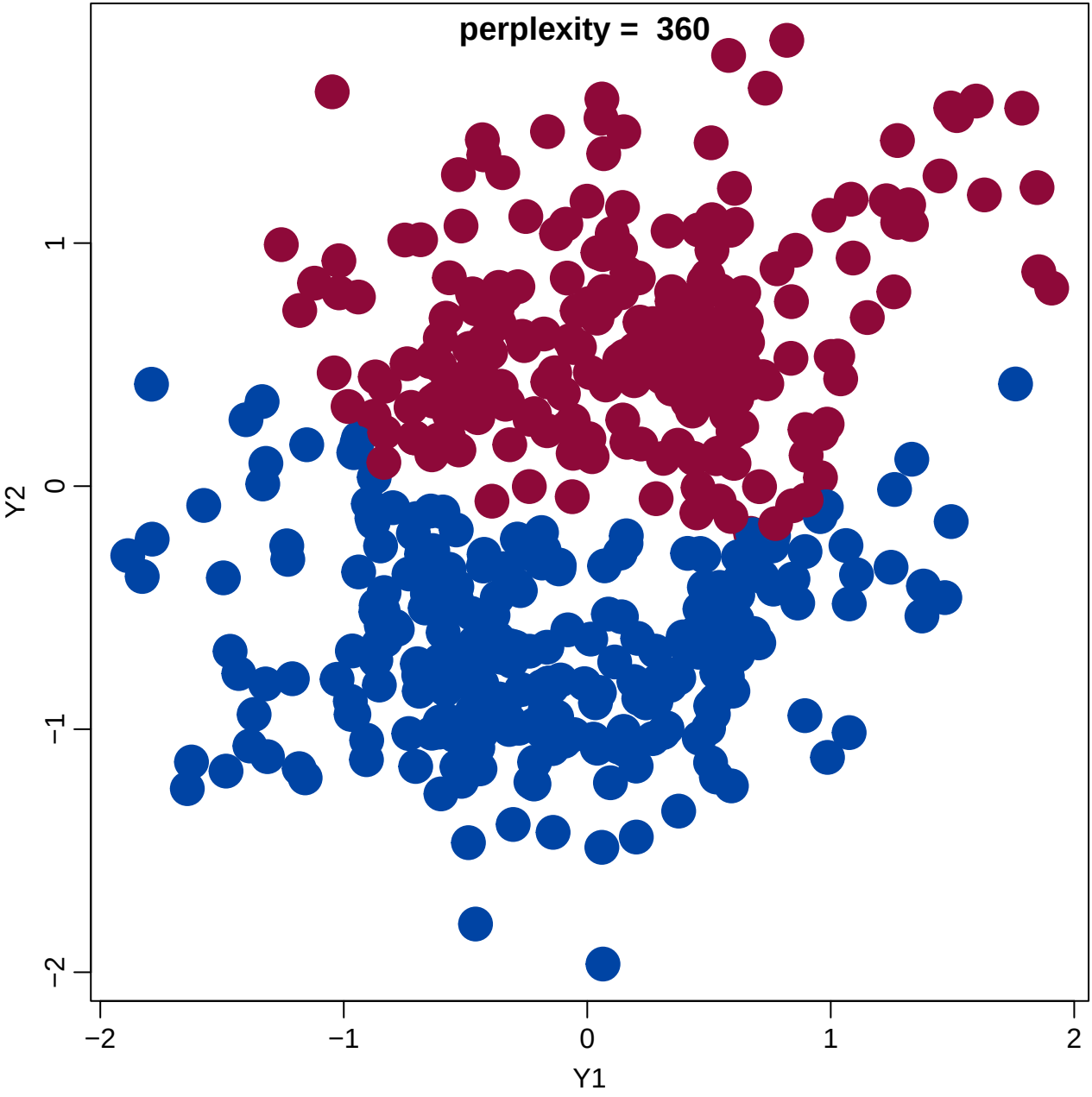

perplexity = 340
1
2

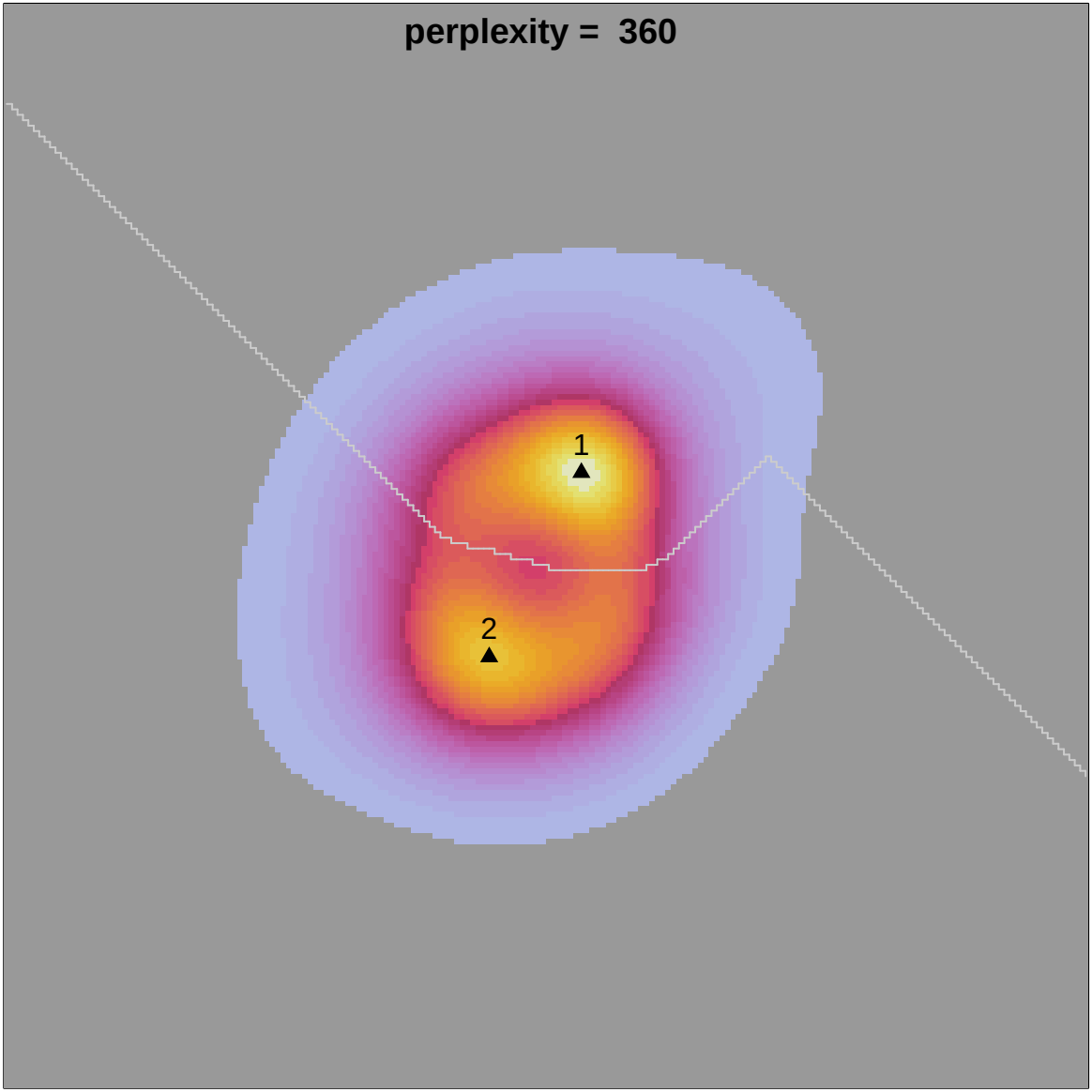

perplexity = 350
Cost
Size
epochs
0.96
0.94
0.92
0.90
0.88
0.86
6
5
4
3
0
20
40
60
80
100

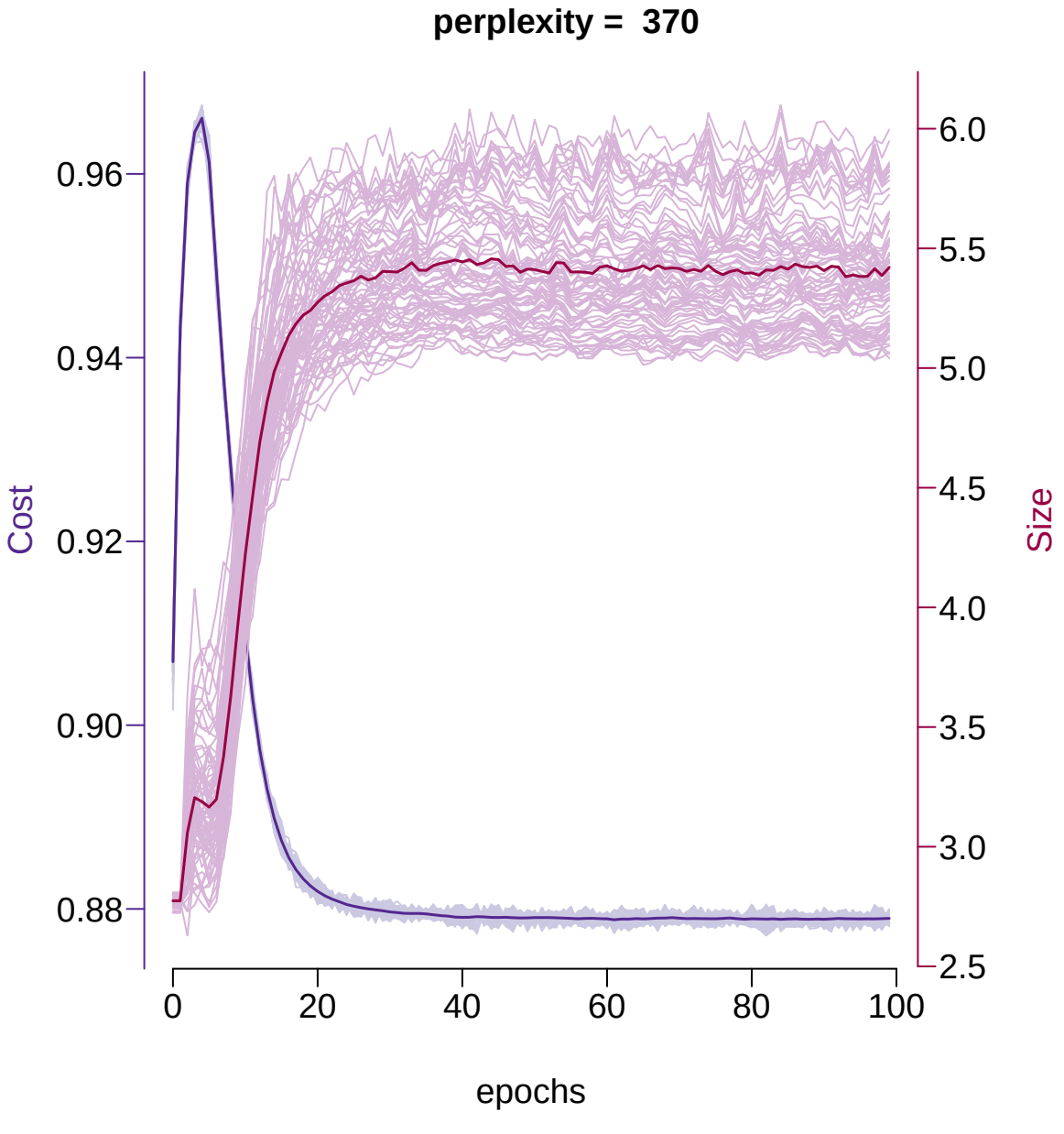

perplexity = 350
Y2
Y1
2
1
0
−1
−2
−1
0
1
2

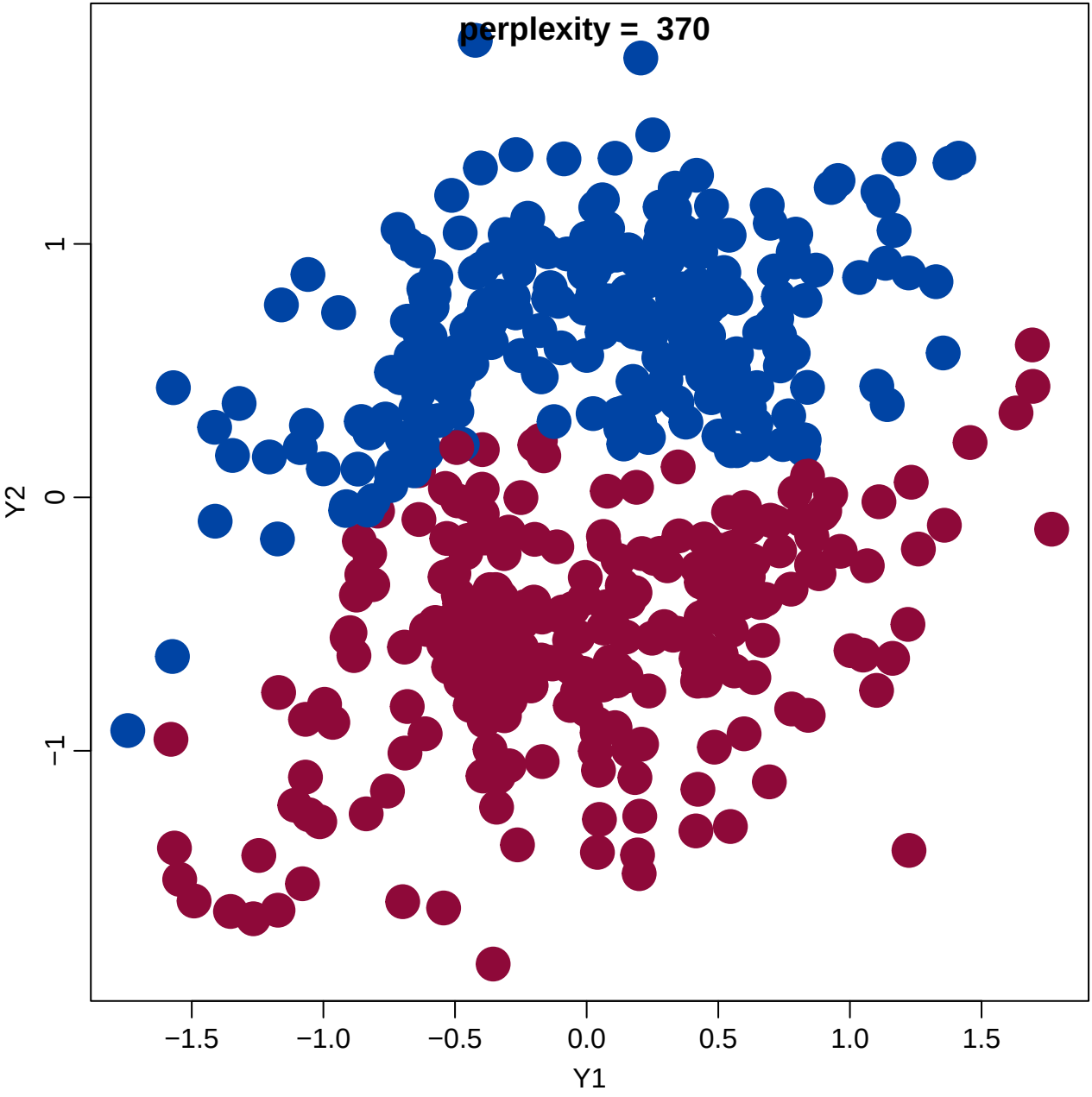

perplexity = 350
2
1

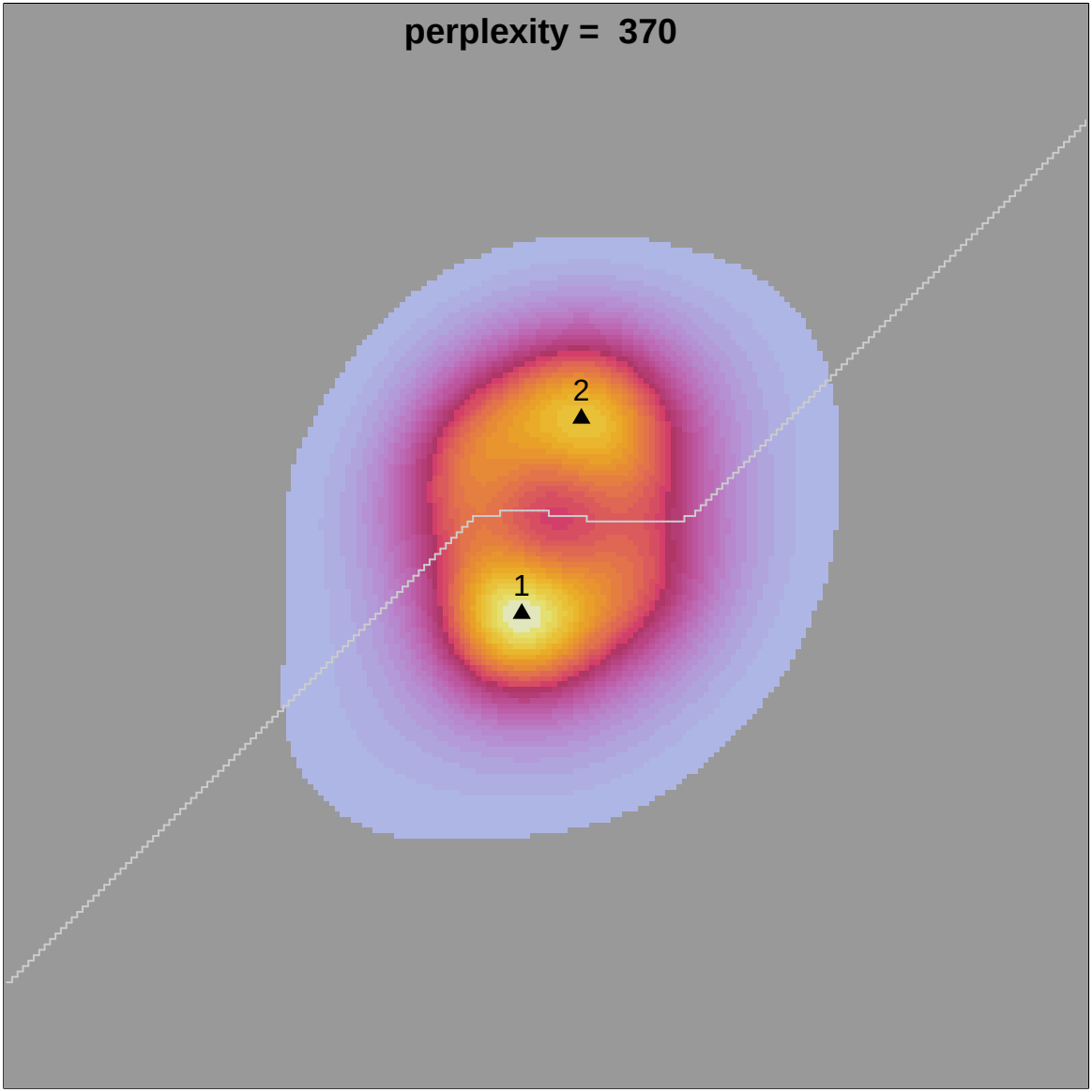

perplexity = 360
Cost
0.96
0.94
0.92
0.90
0.88
Size
6
5
4
3
0
20
40
60
80
100
epochs

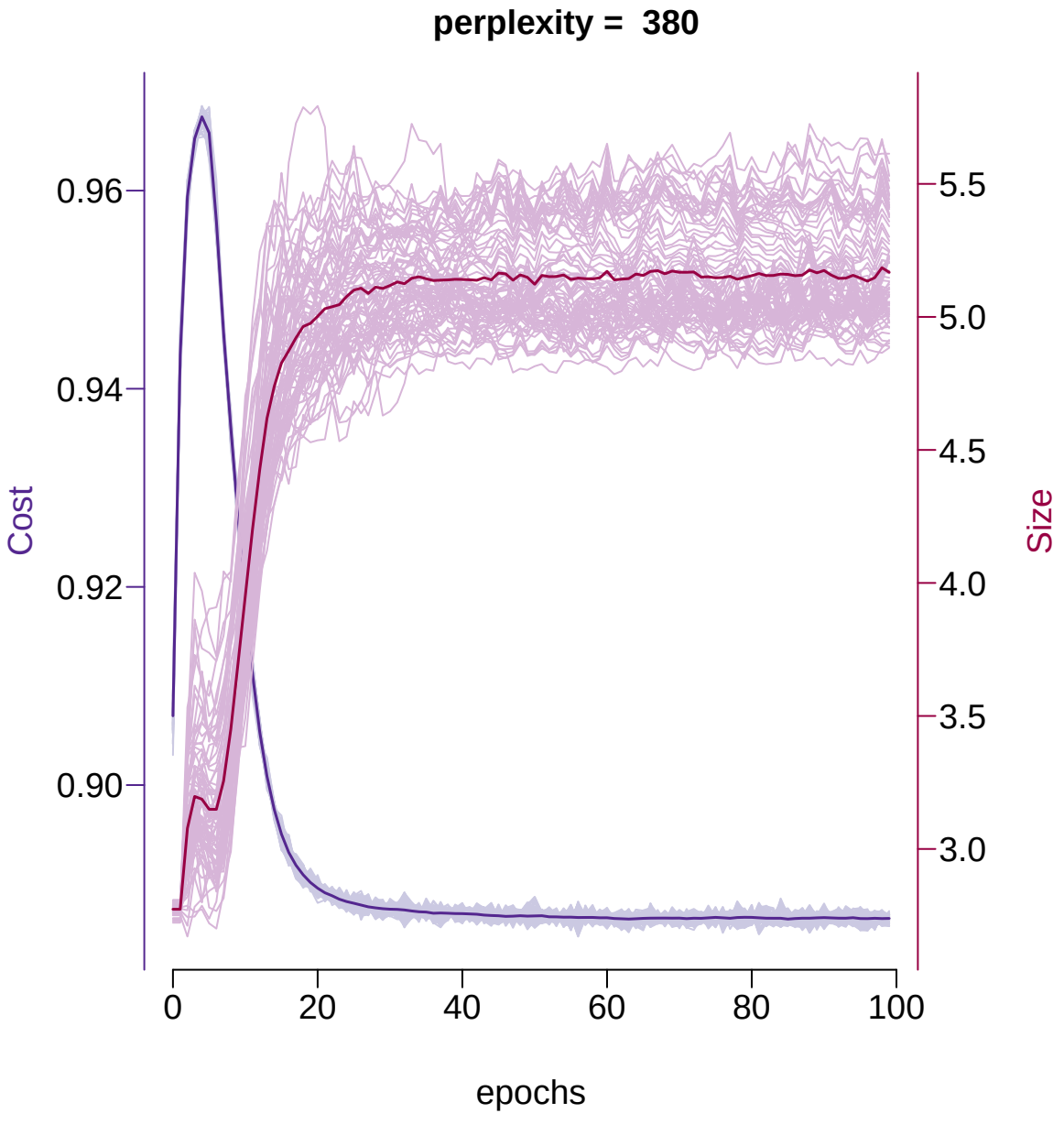

perplexity = 360
Y2
Y1
−2
−1
0
1
2
−2
−1
0
1

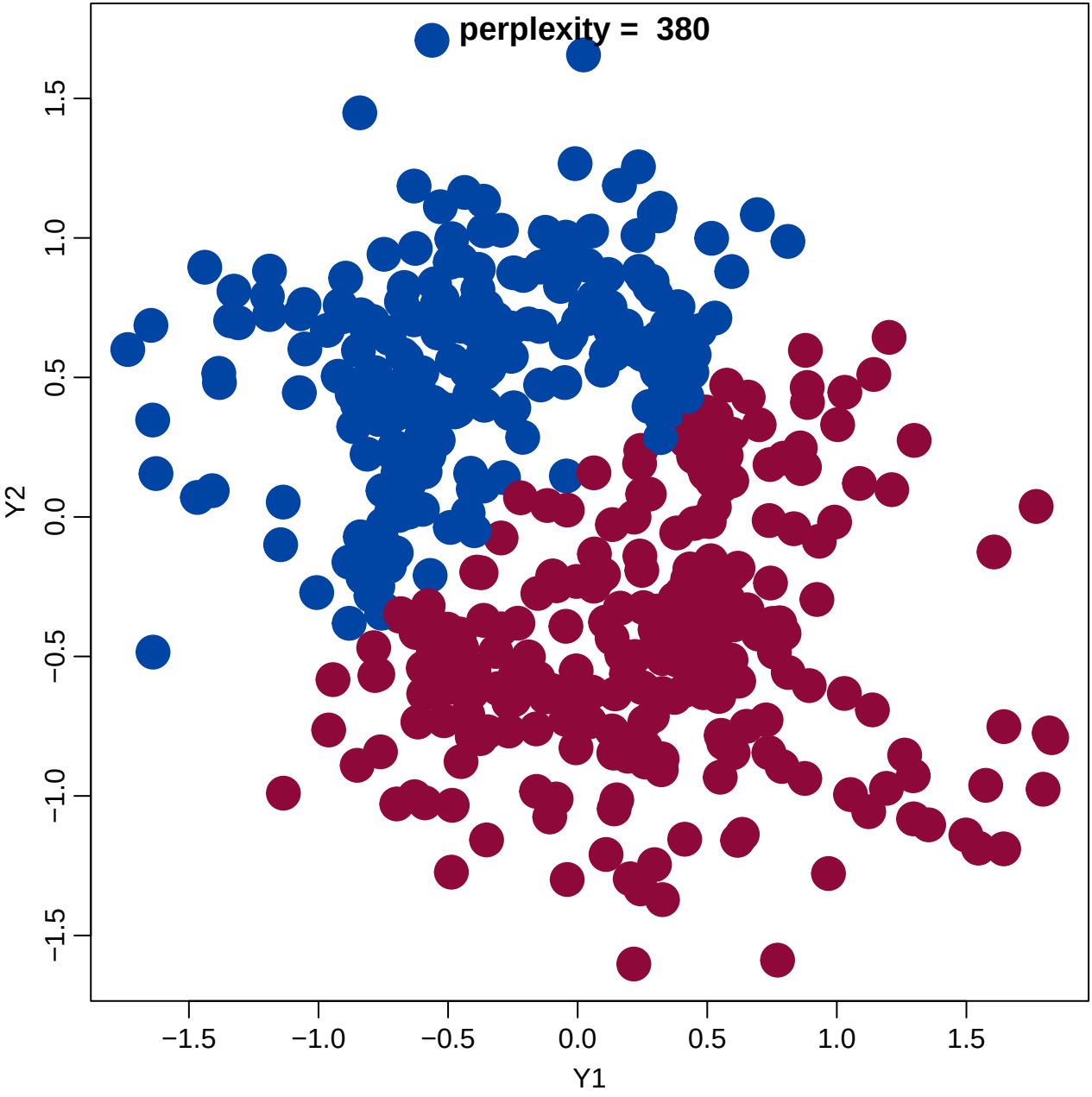

perplexity = 360
1
2

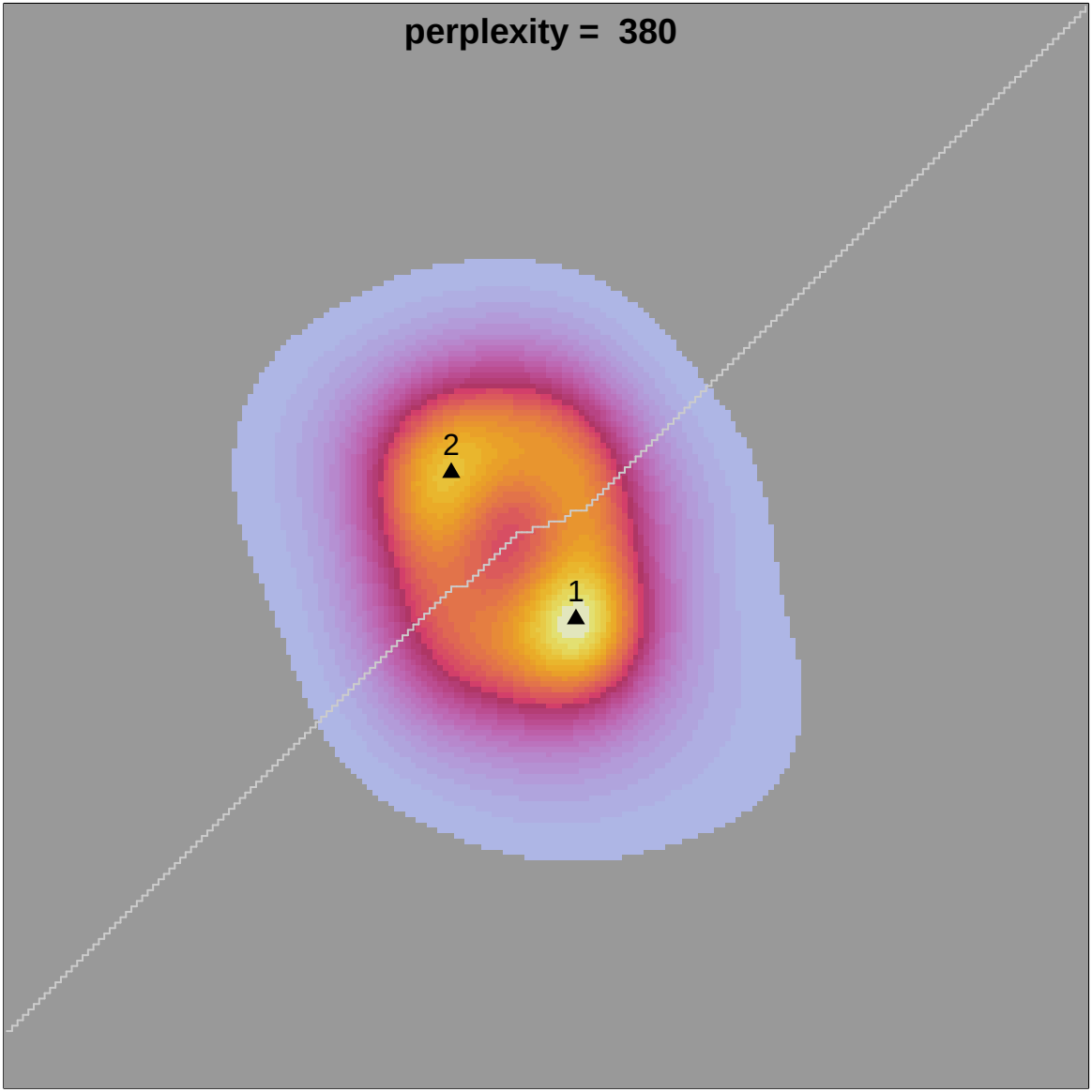

perplexity = 370
Cost
0.96
0.94
0.92
0.90
0.88
Size
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
0
20
40
60
80
100
epochs

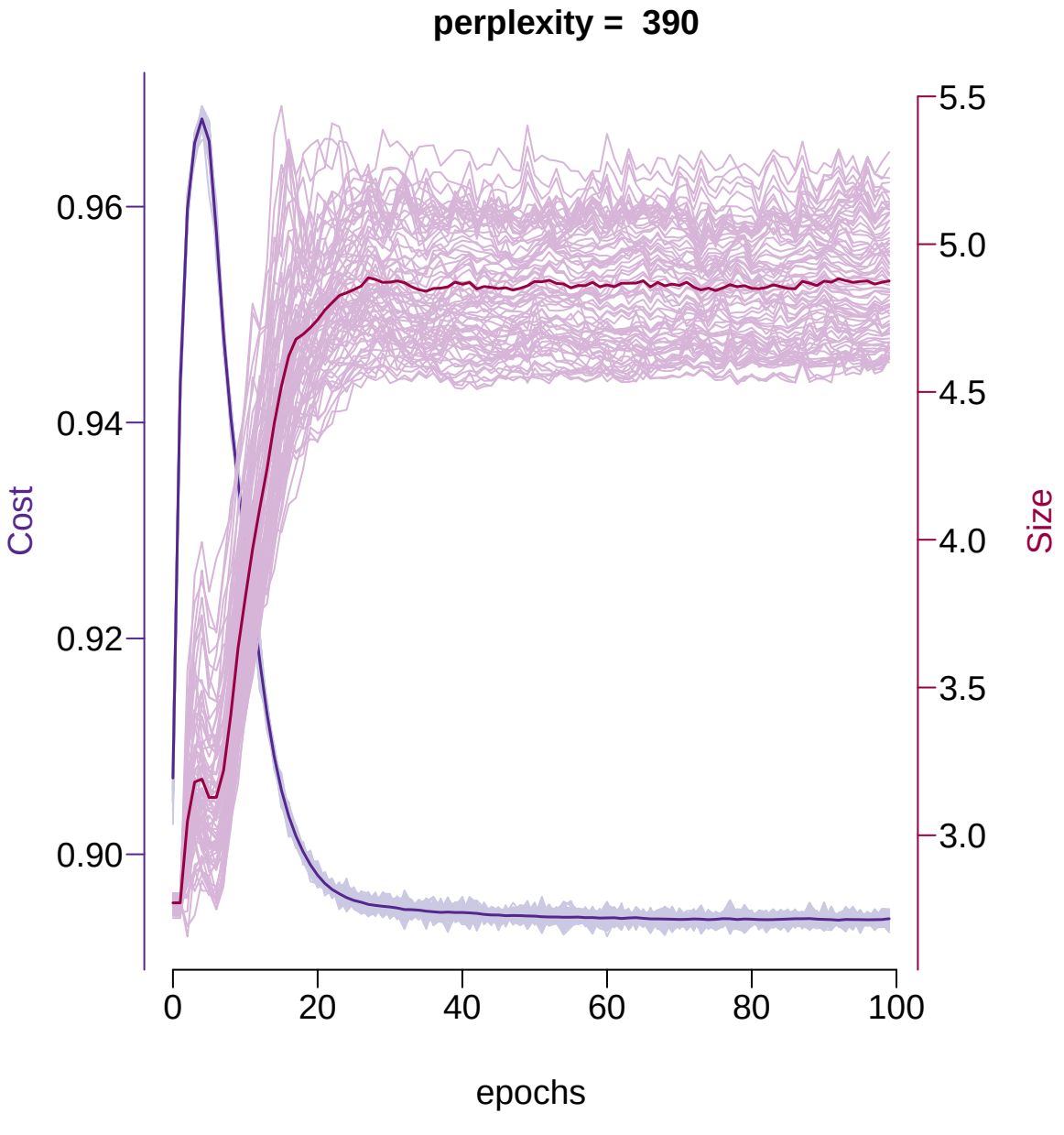

perplexity = 370
Y2
1
0
−1
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Y1

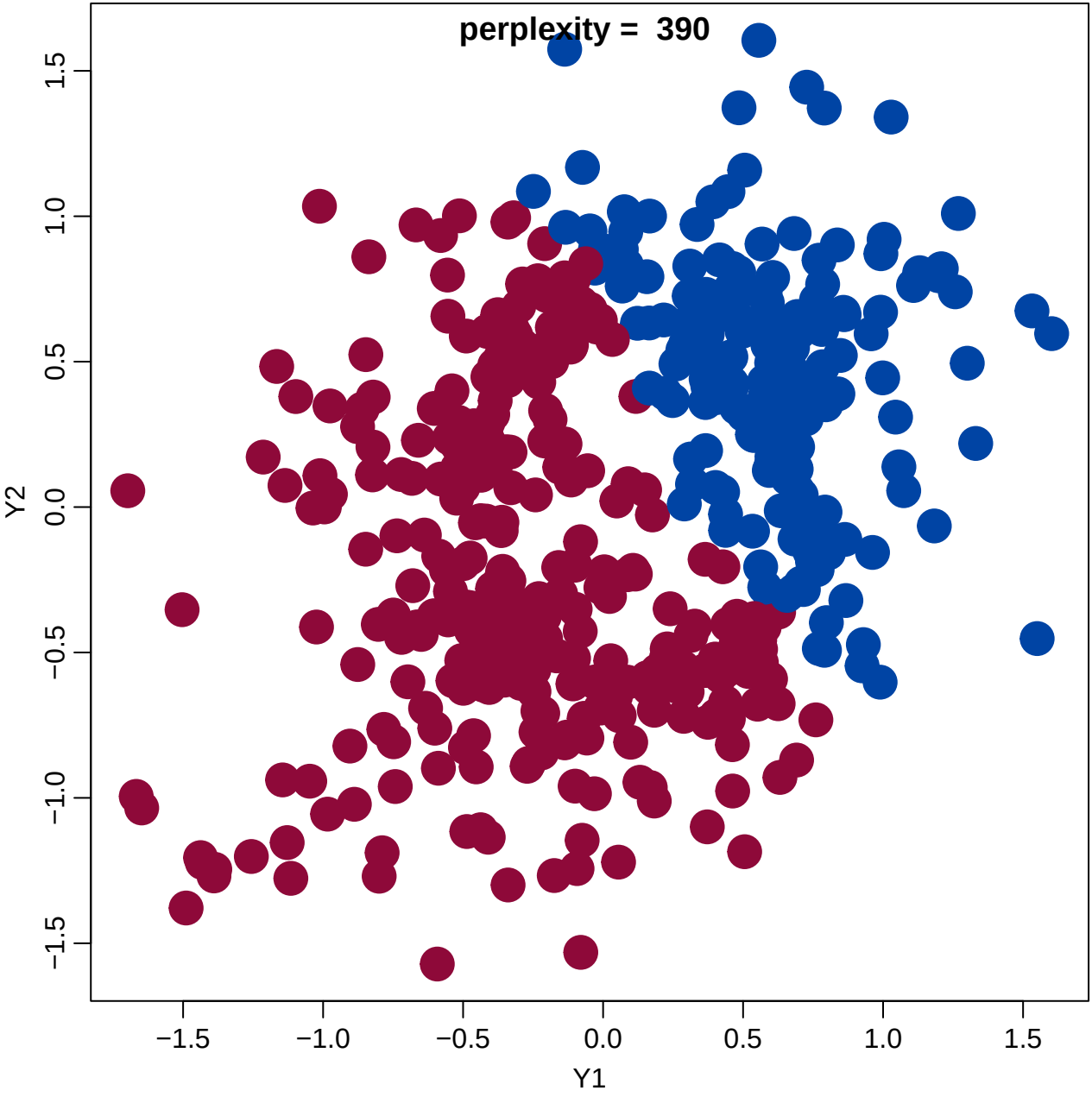

perplexity = 370
2
1

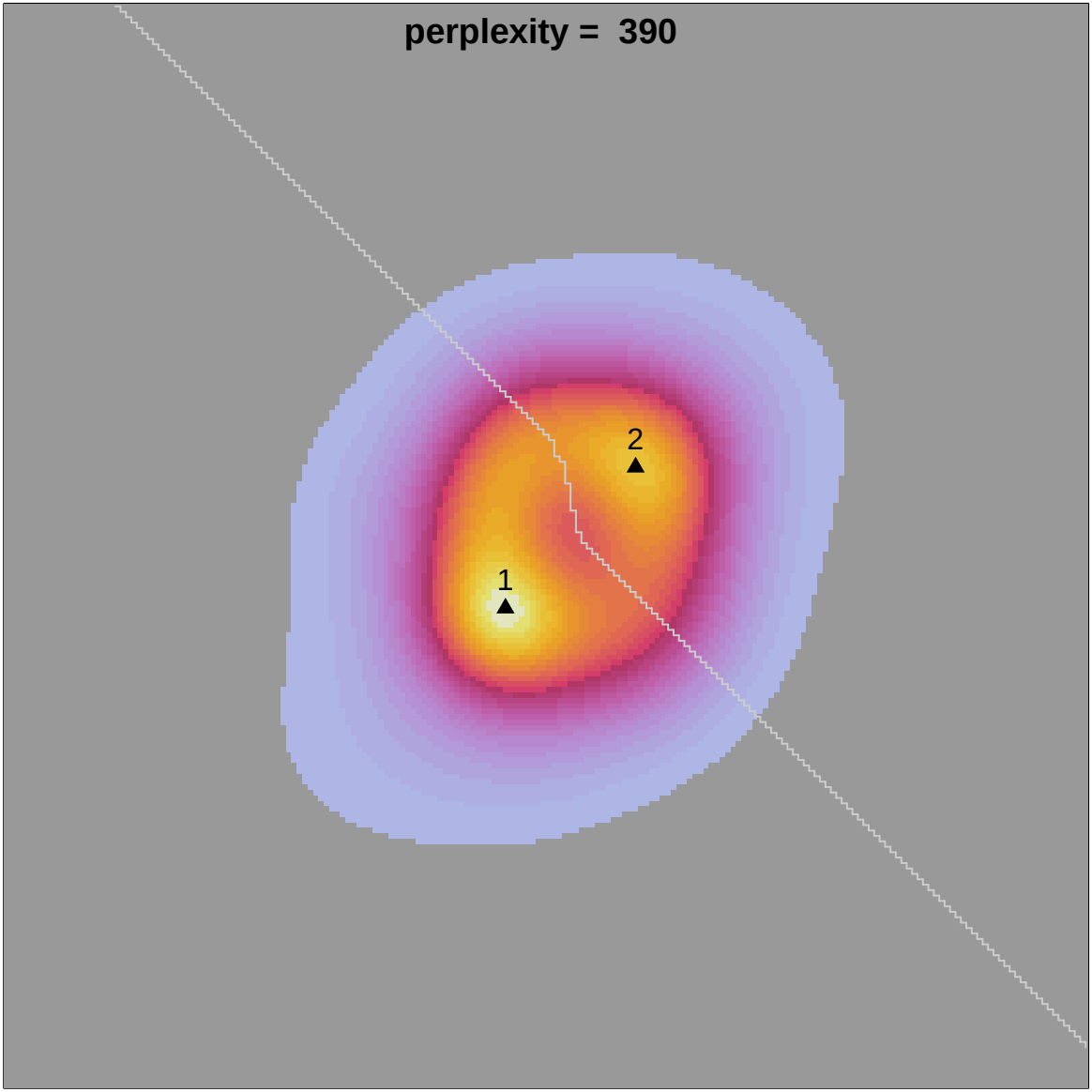

perplexity = 380
Cost
0.96
0.94
0.92
0.90
Size
5.5
5.0
4.5
4.0
3.5
3.0
0
20
40
60
80
100
epochs

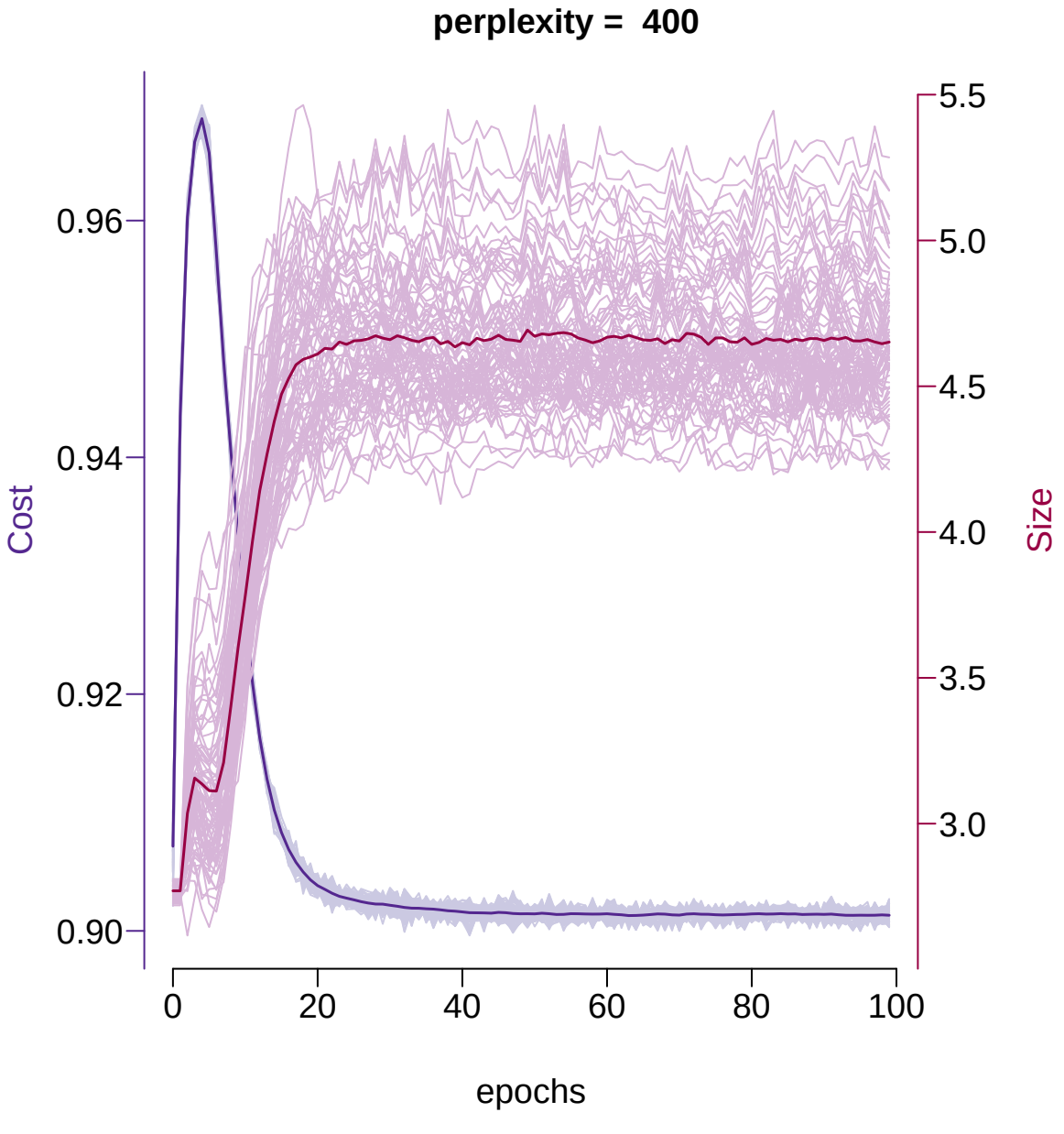

perplexity = 380
Y2
Y1
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5

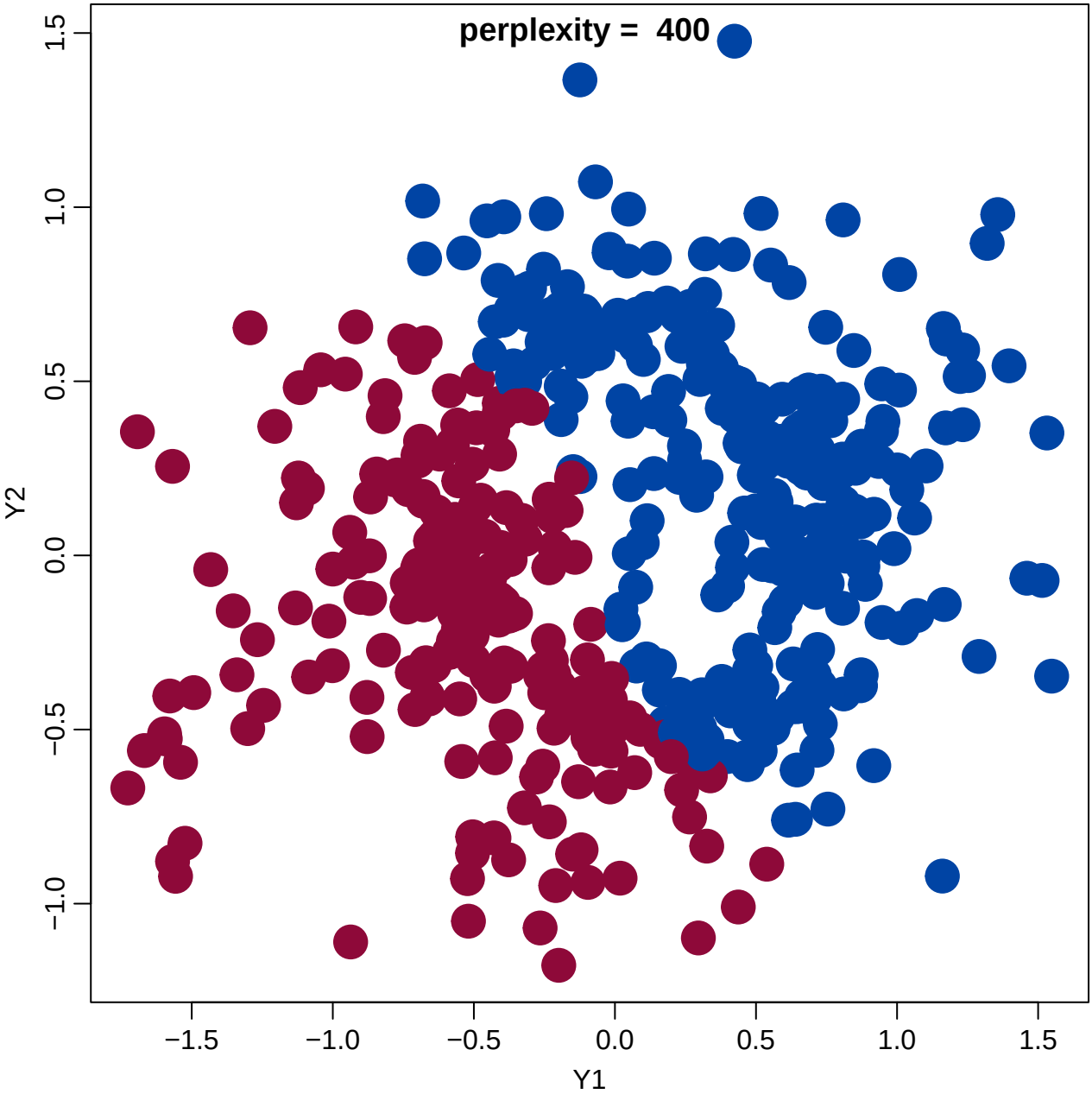

perplexity = 380
2
1

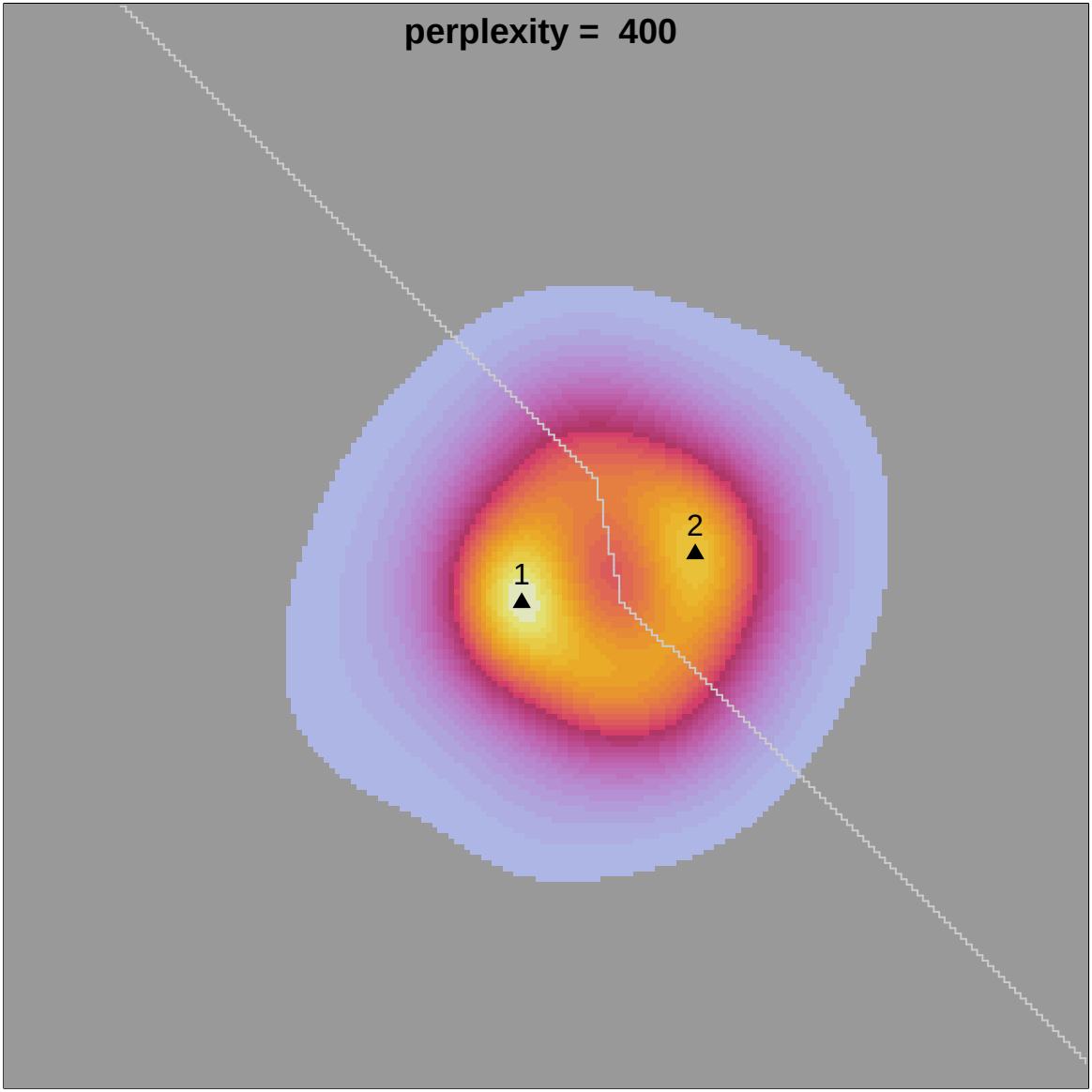

perplexity = 390
Cost
0.96
0.94
0.92
0.90
Size
5.5
5.0
4.5
4.0
3.5
3.0
0
20
40
60
80
100
epochs

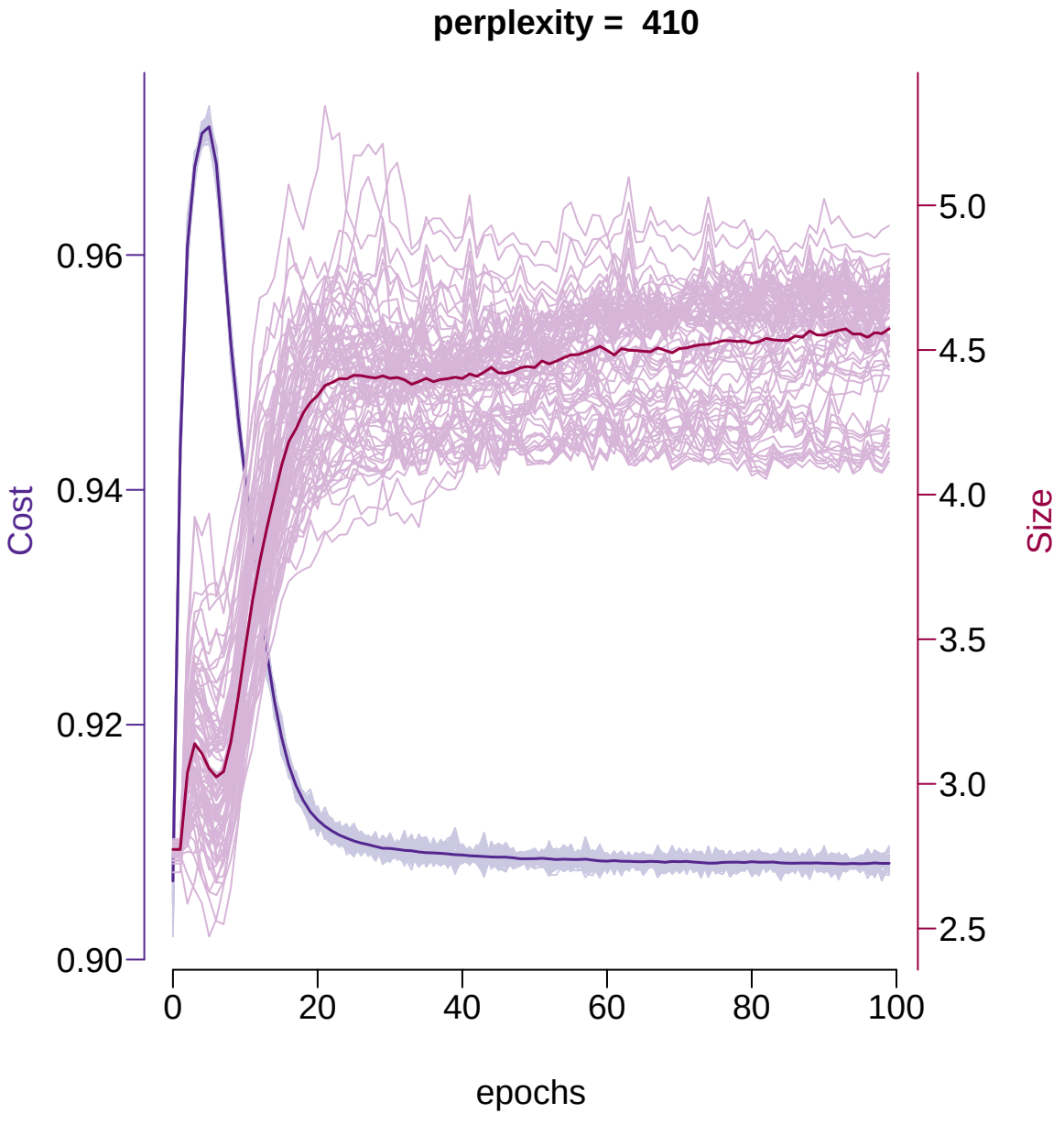

perplexity = 390
Y2
Y1
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5

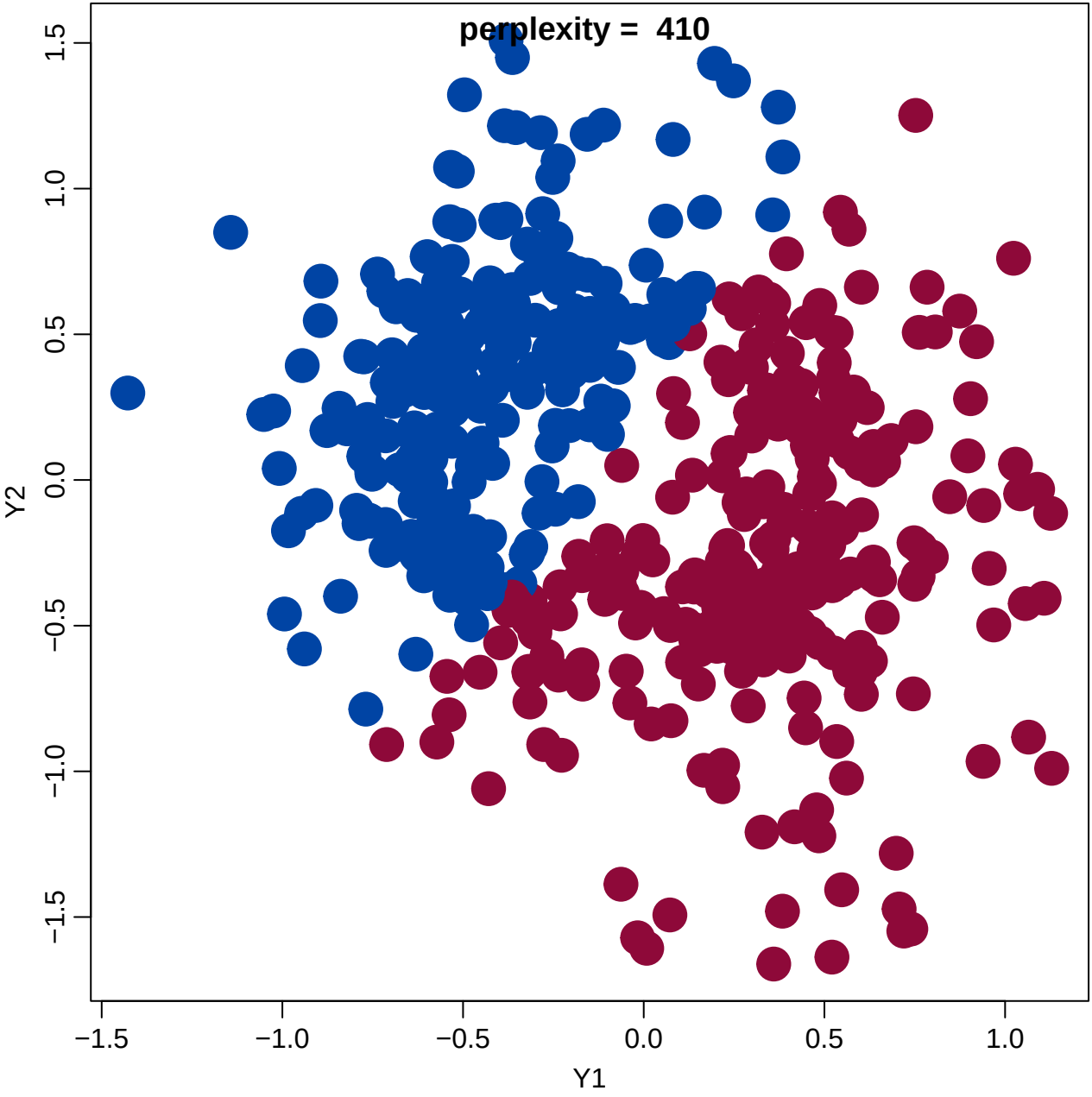

perplexity = 390
2
1

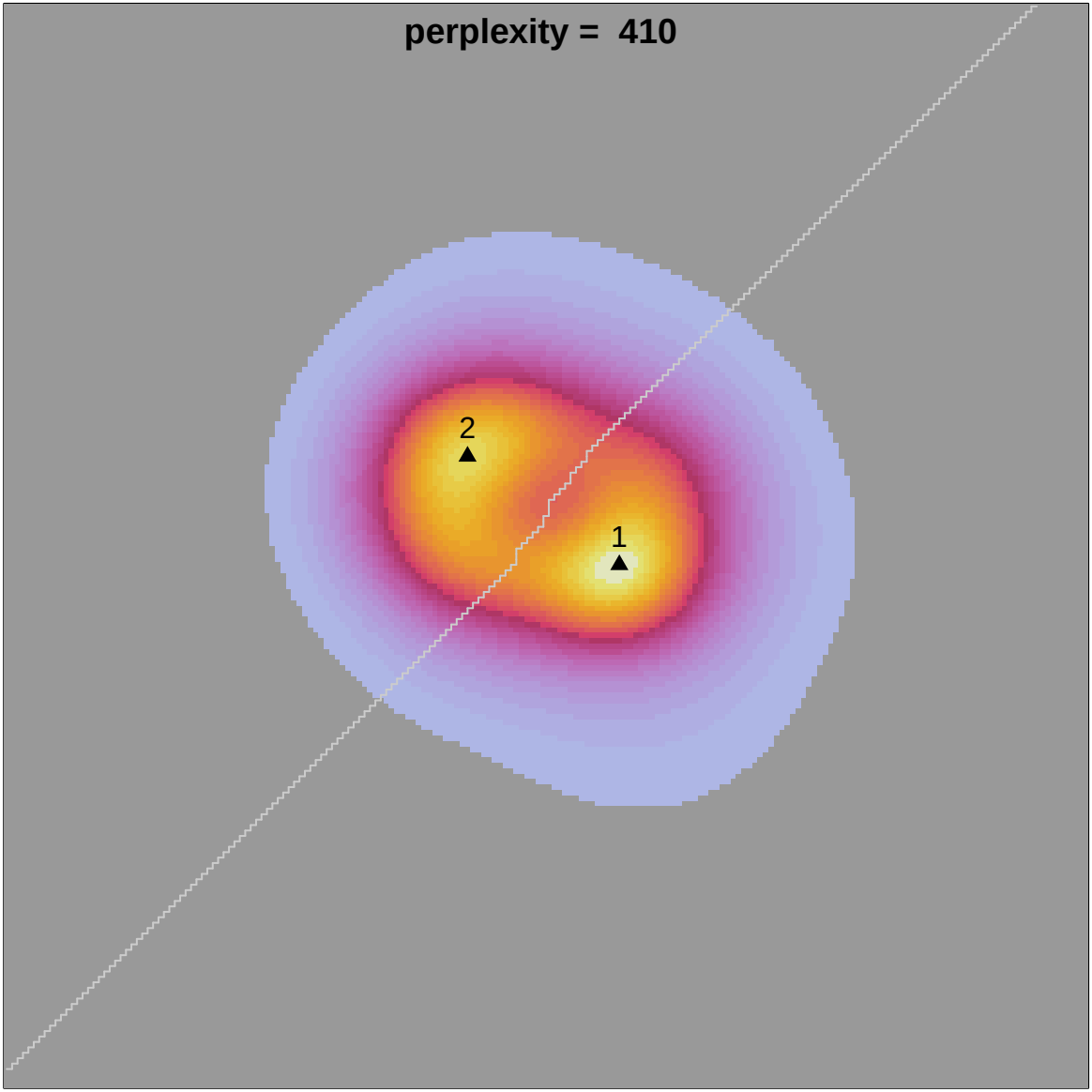

perplexity = 400
Cost
0.96
0.94
0.92
0.90
Size
5.5
5.0
4.5
4.0
3.5
3.0
0
20
40
60
80
100
epochs

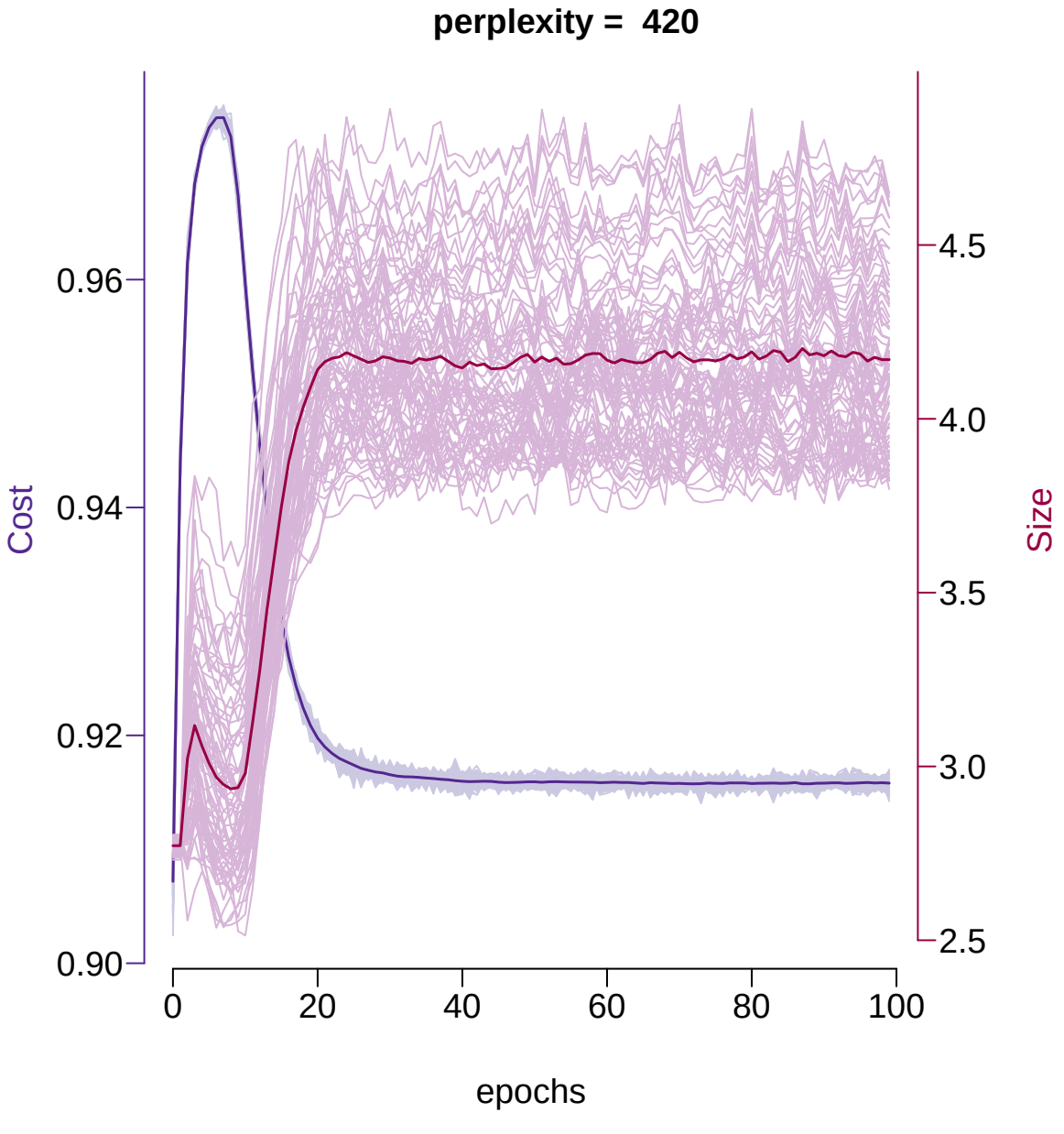

perplexity = 400
Y2
Y1
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5

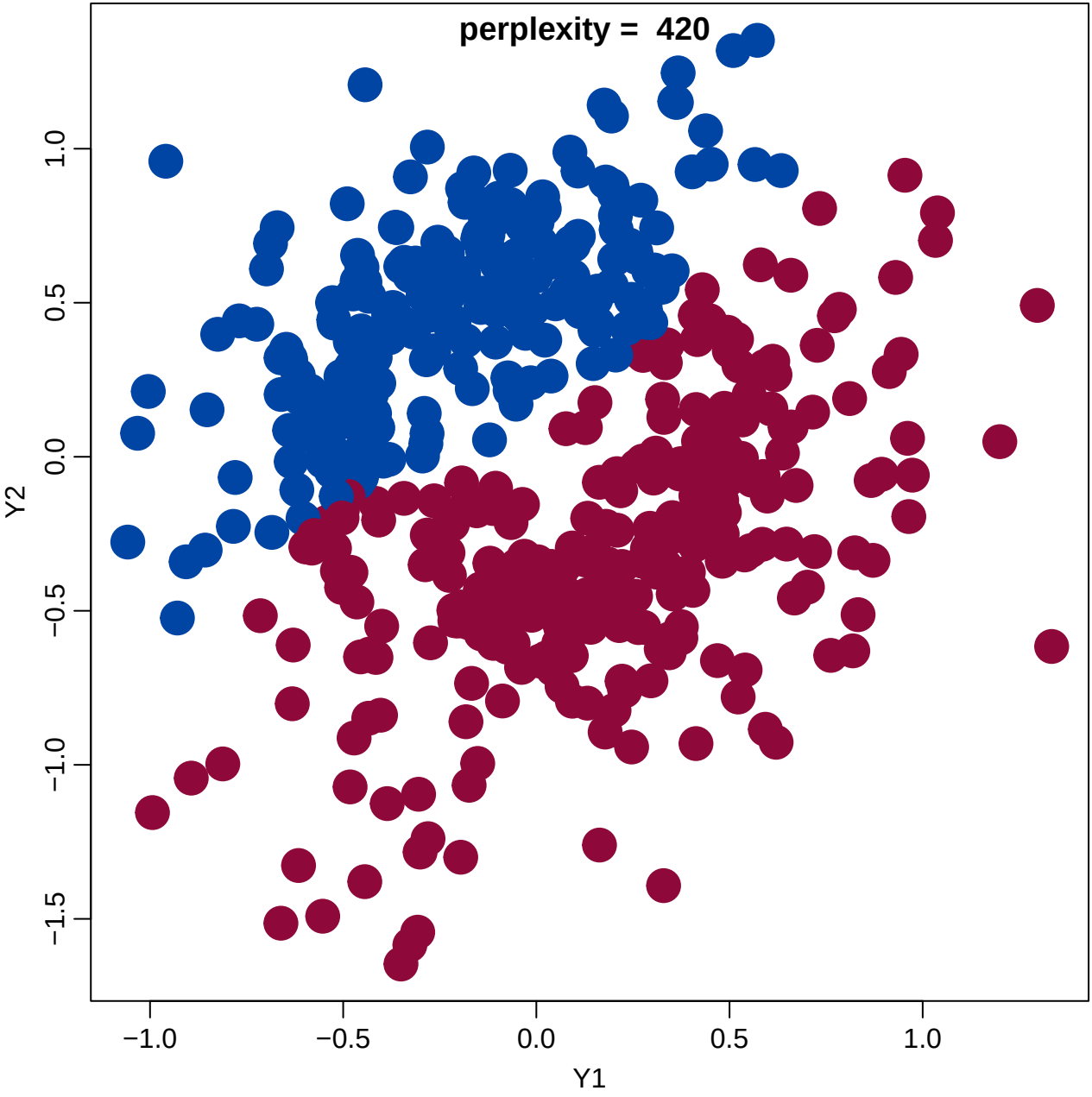

perplexity = 400
1
2

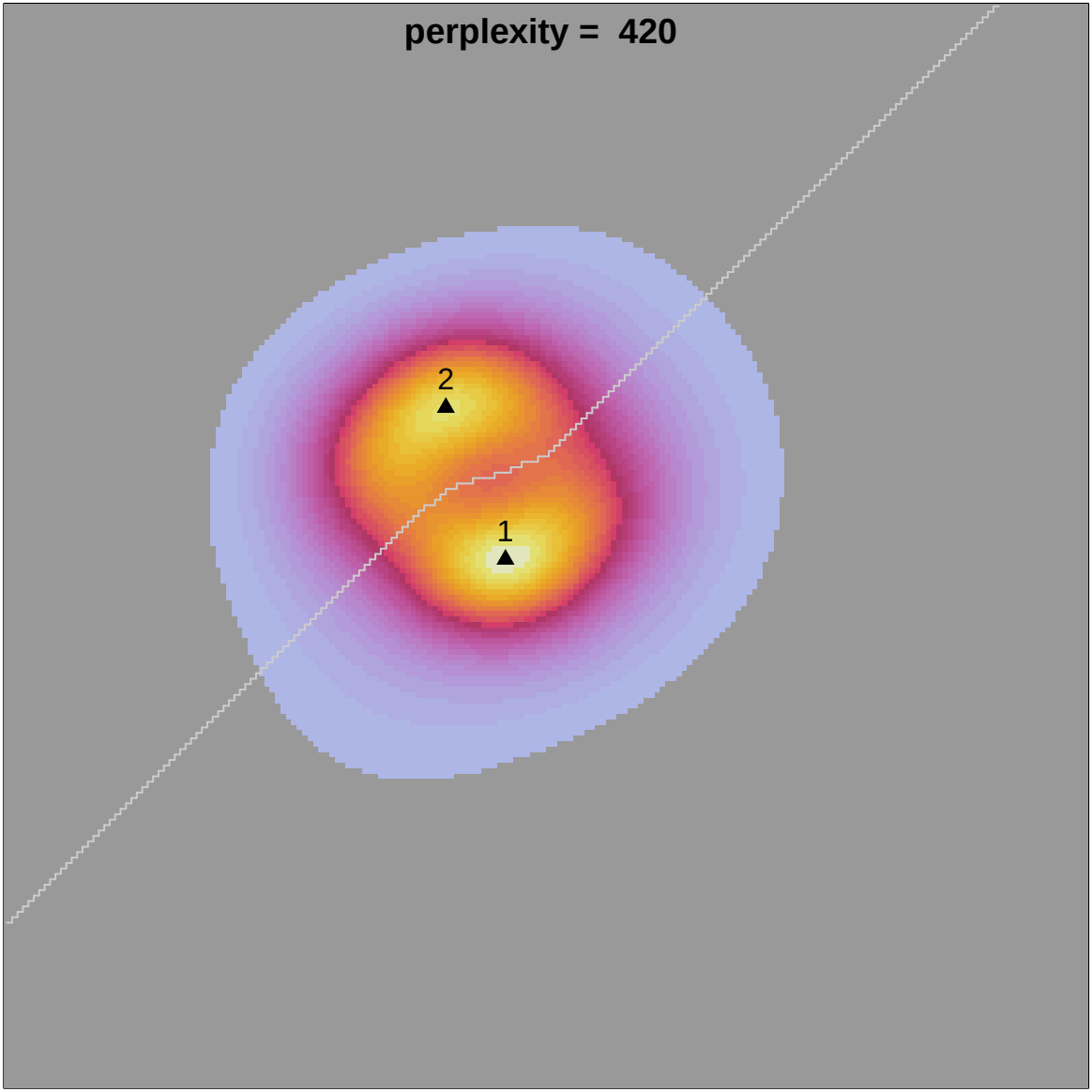

perplexity = 410
Cost
0.96
0.94
0.92
0.90
Size
5.0
4.5
4.0
3.5
3.0
2.5
0
20
40
60
80
100
epochs

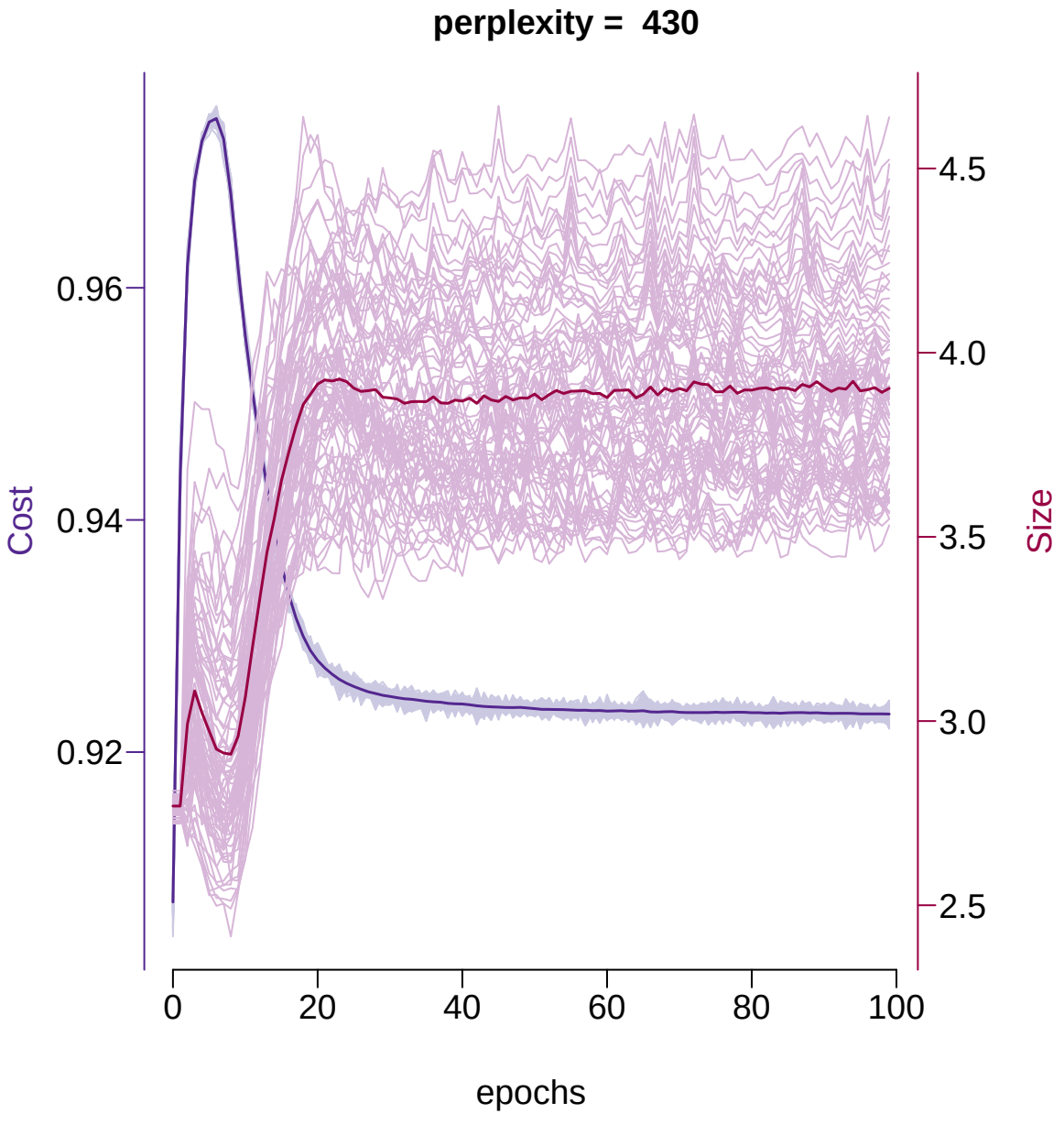

perplexity = 410
Y2
Y1
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5
−1.0
−0.5
0.0
0.5
1.0

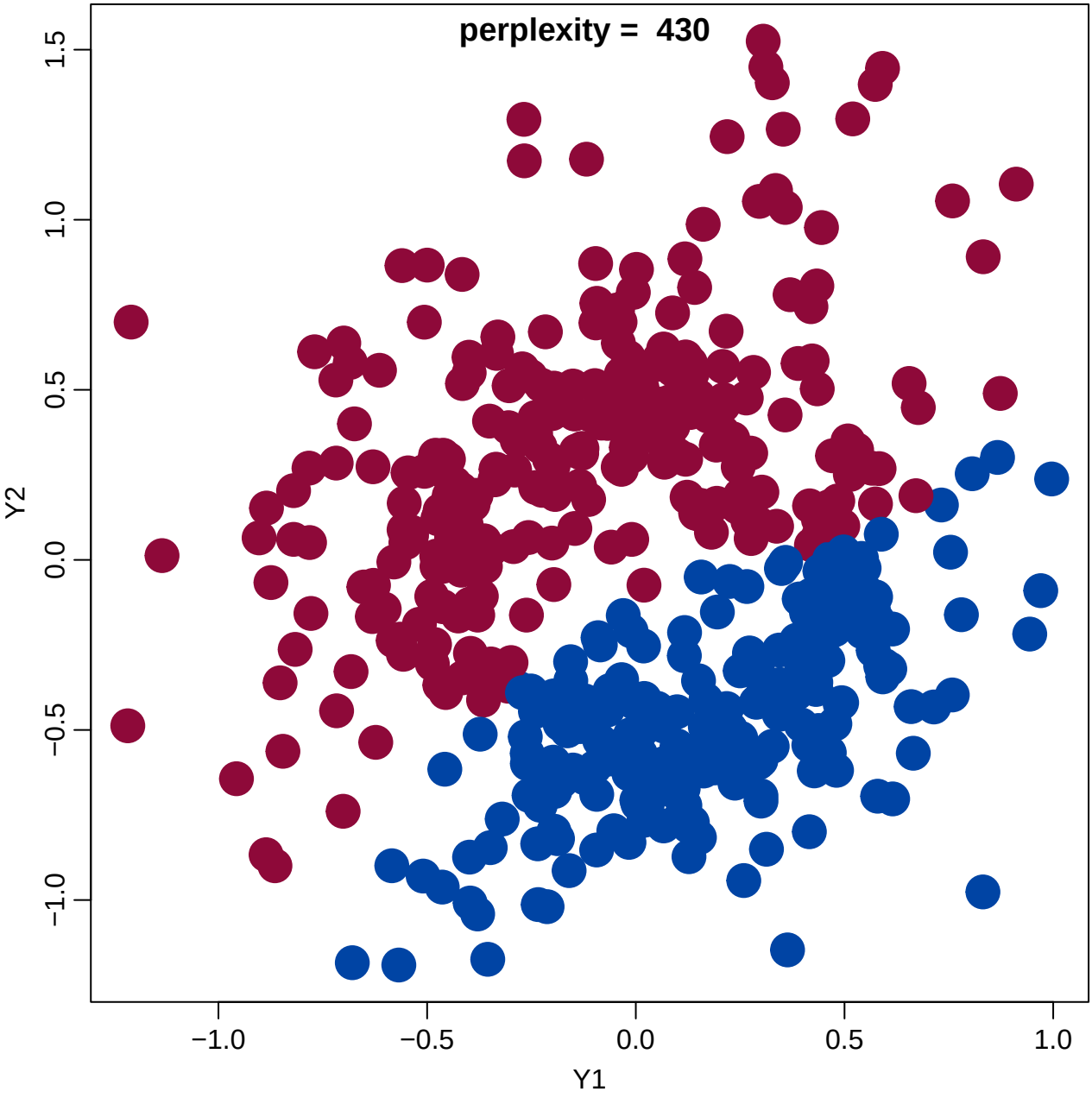

perplexity = 410
2
1

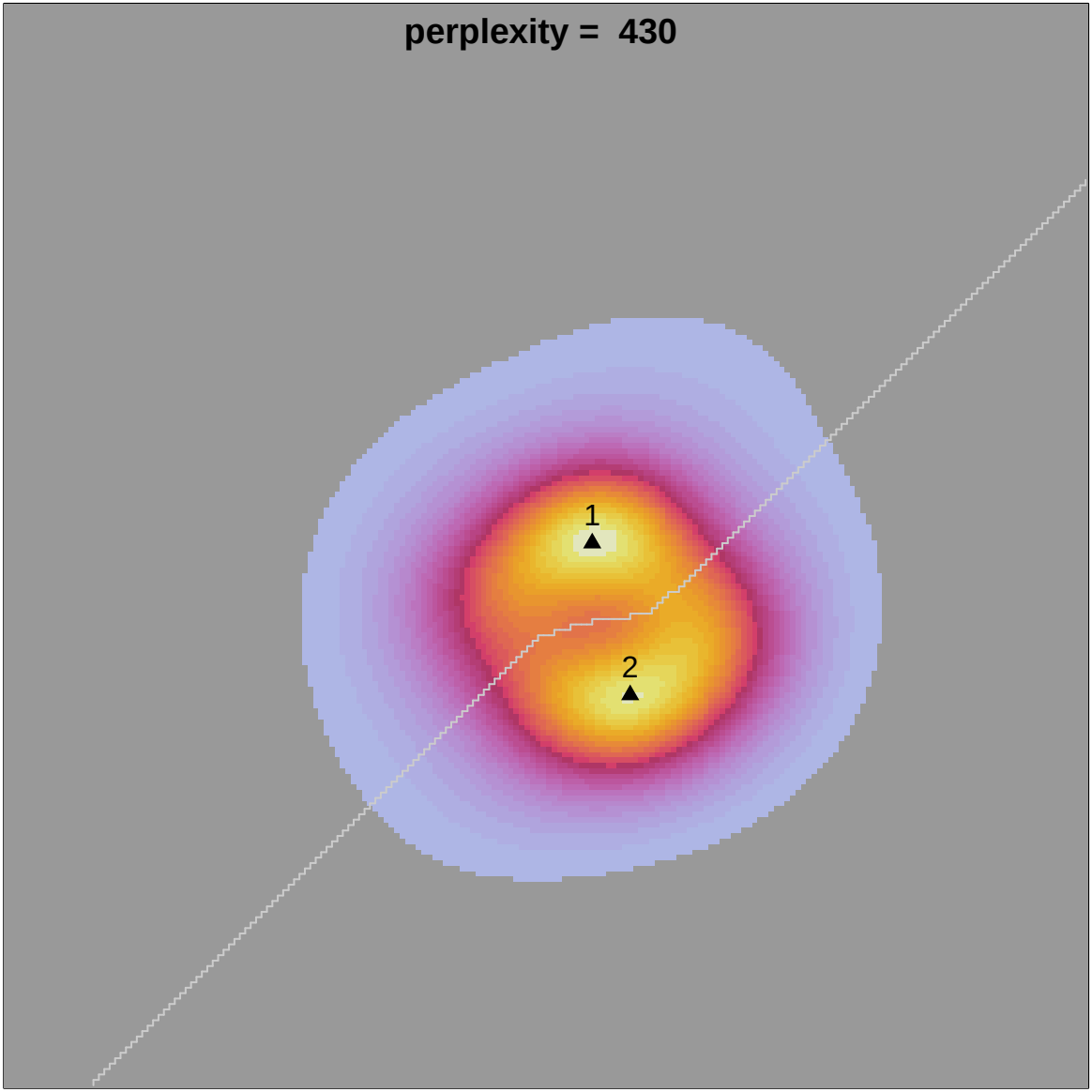

perplexity = 420
Cost
0.96
0.94
0.92
0.90
Size
4.5
4.0
3.5
3.0
2.5
0
20
40
60
80
100
epochs

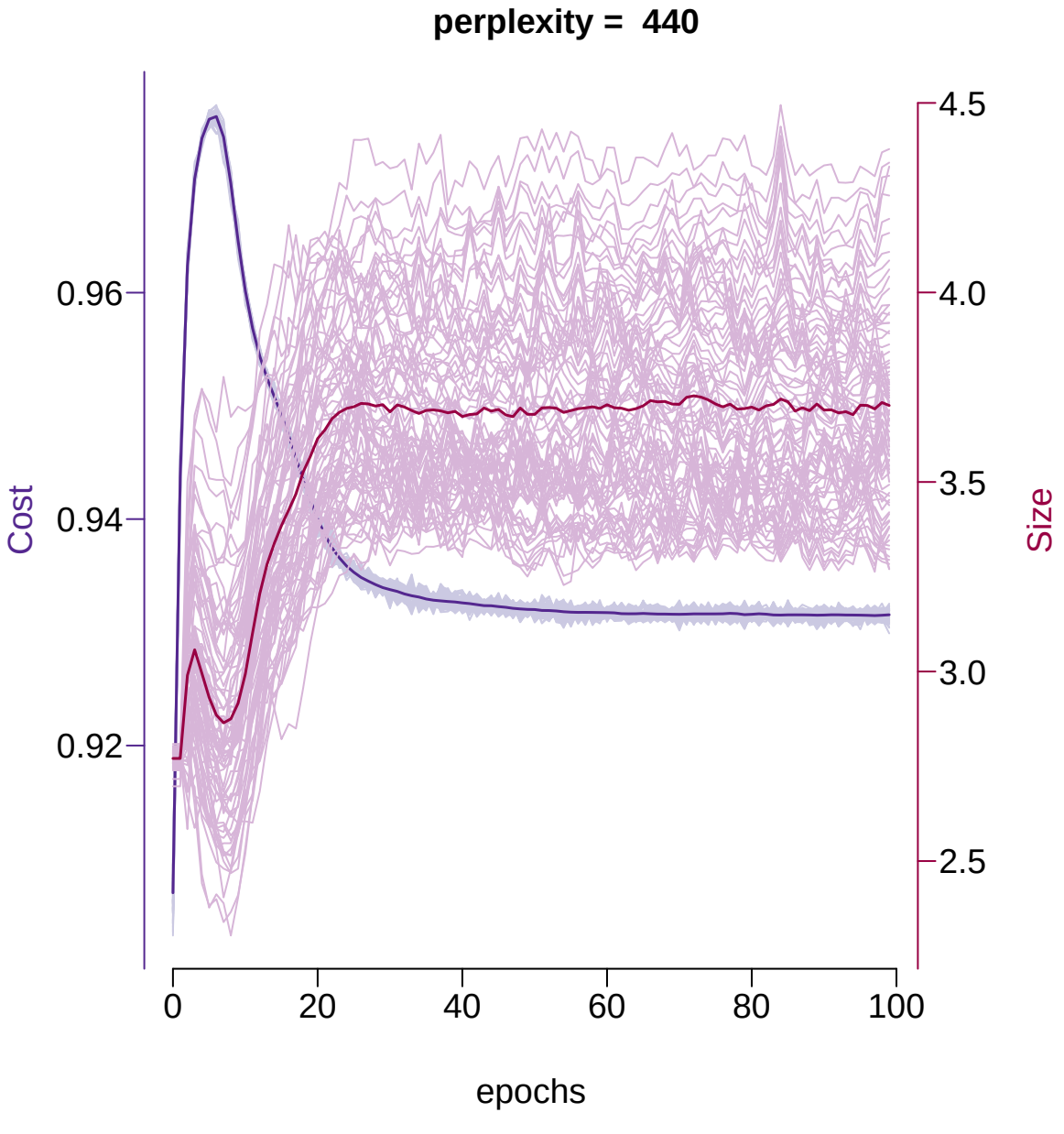

perplexity = 420
Y2
Y1
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.0
−0.5
0.0
0.5
1.0

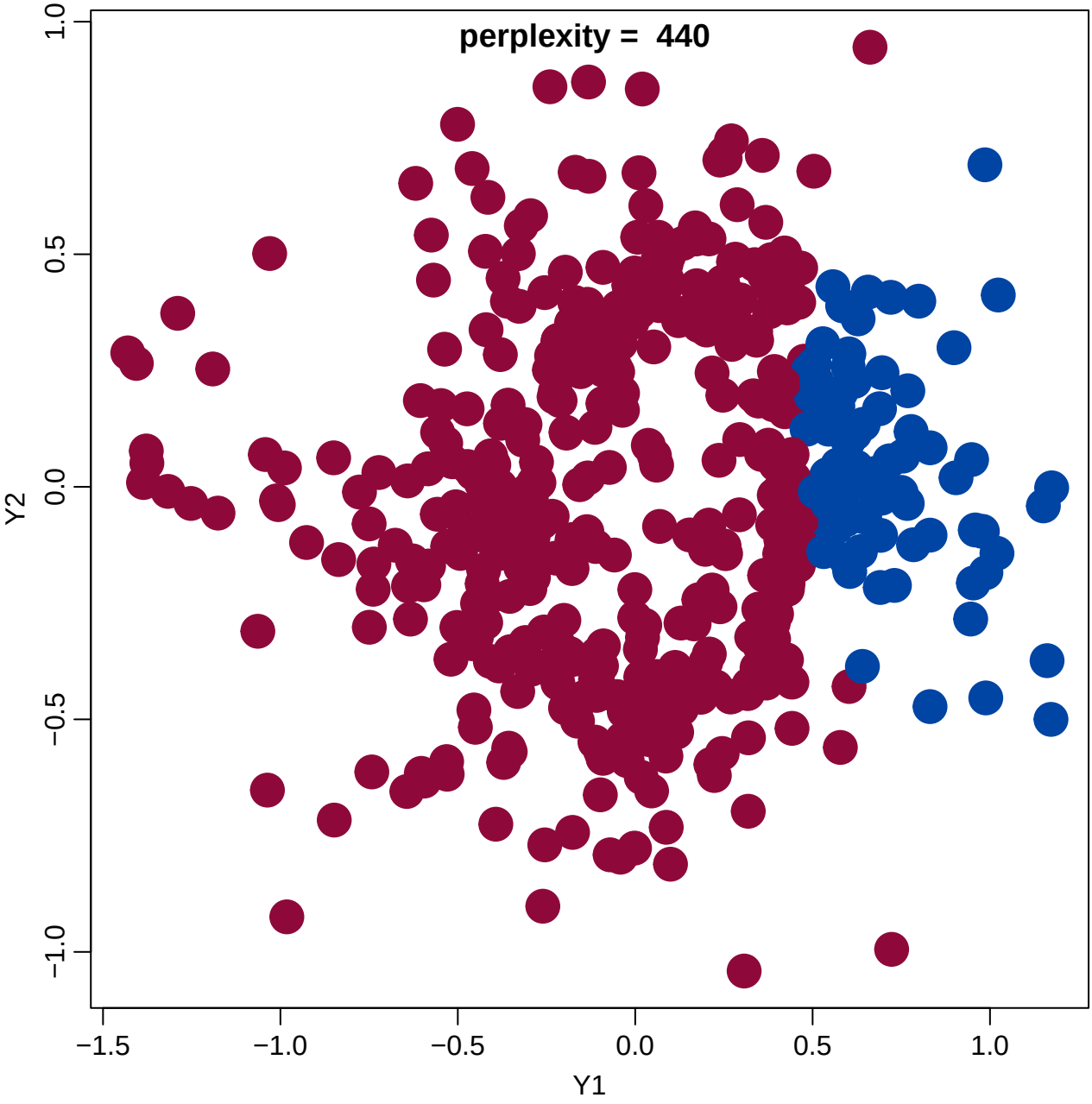

perplexity = 420
2
1

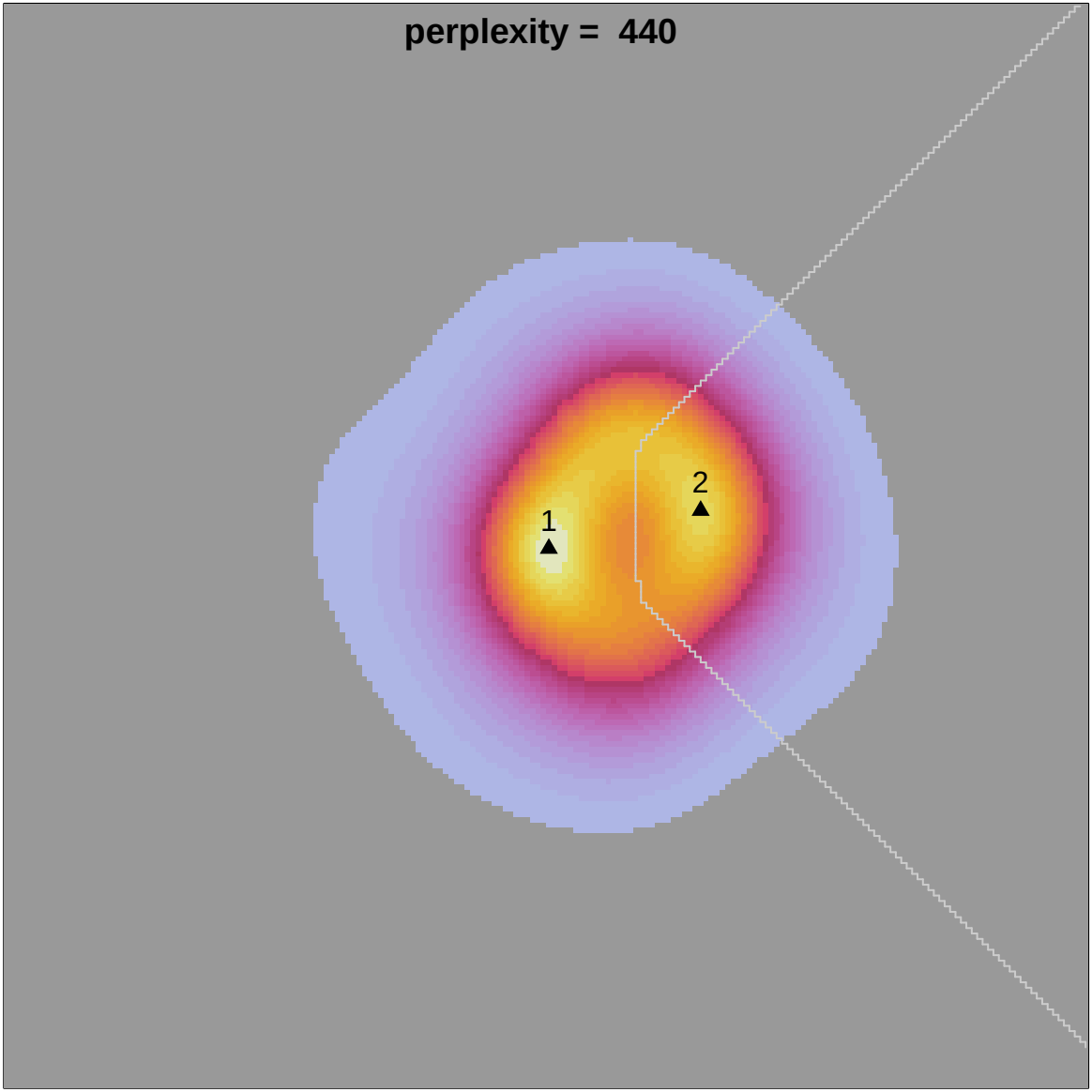

perplexity = 430
Cost
0.96
0.94
0.92
Size
4.5
4.0
3.5
3.0
2.5
0
20
40
60
80
100
epochs

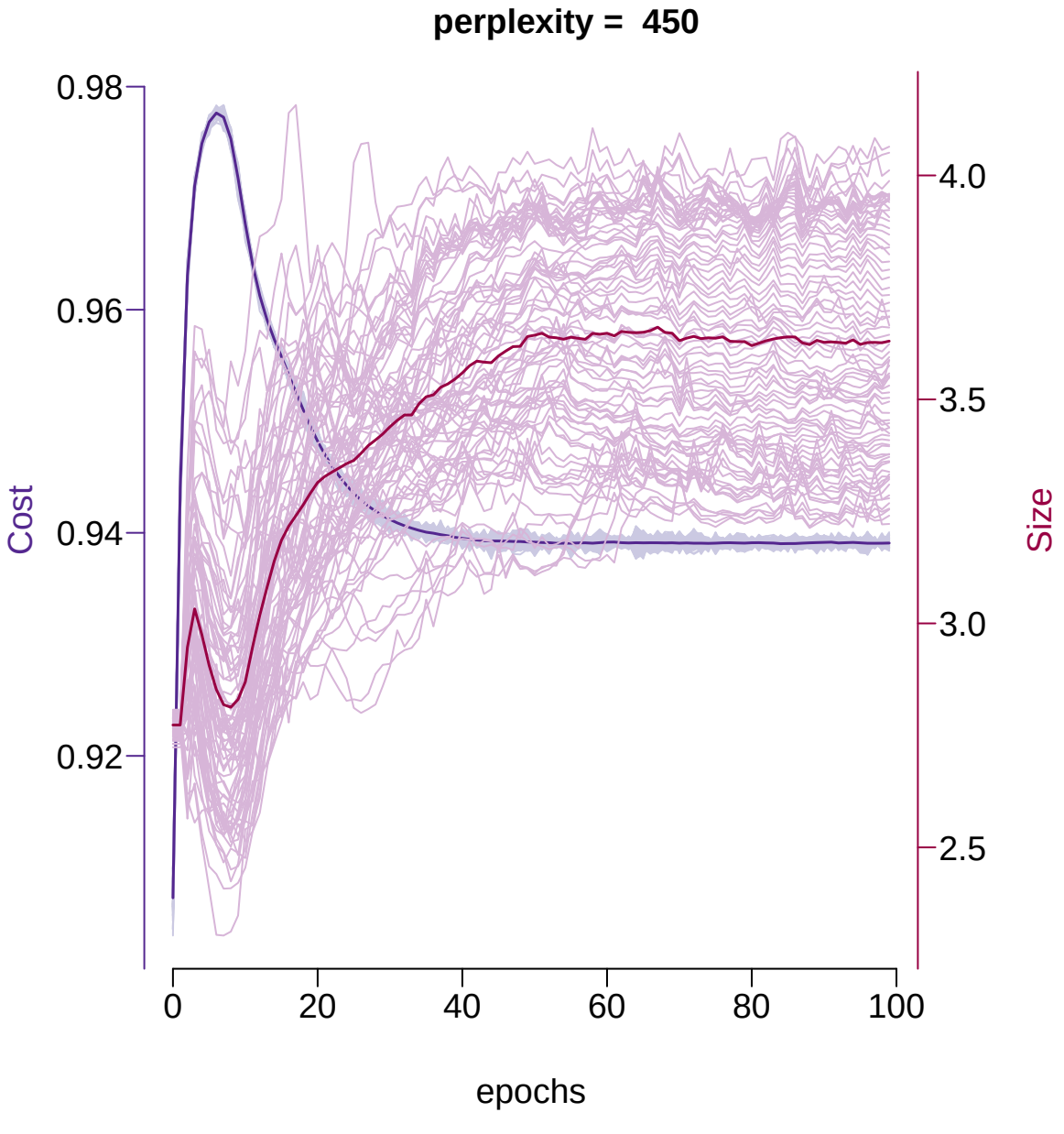

perplexity = 430
Y2
Y1
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0

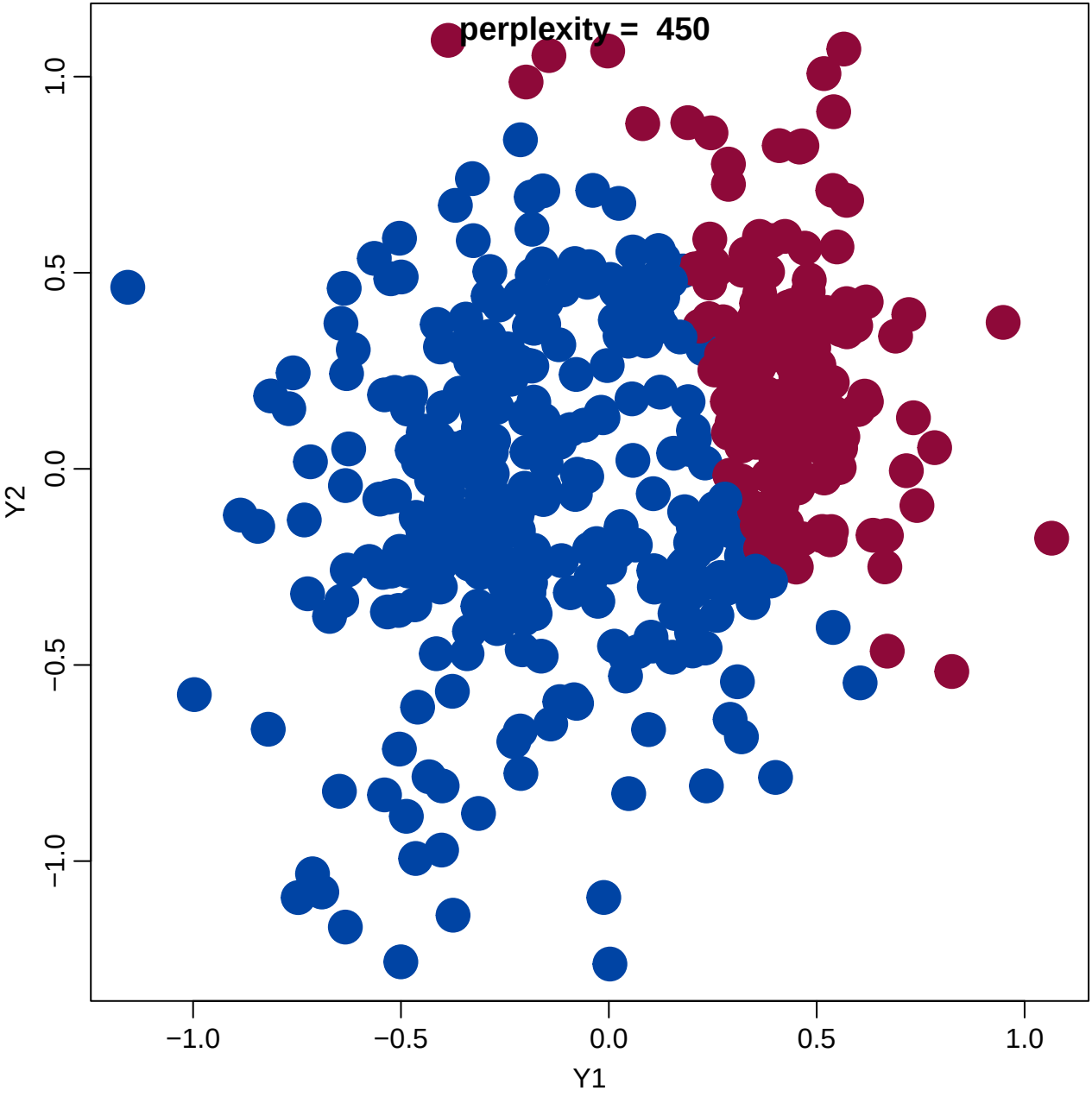

perplexity = 430
1
2

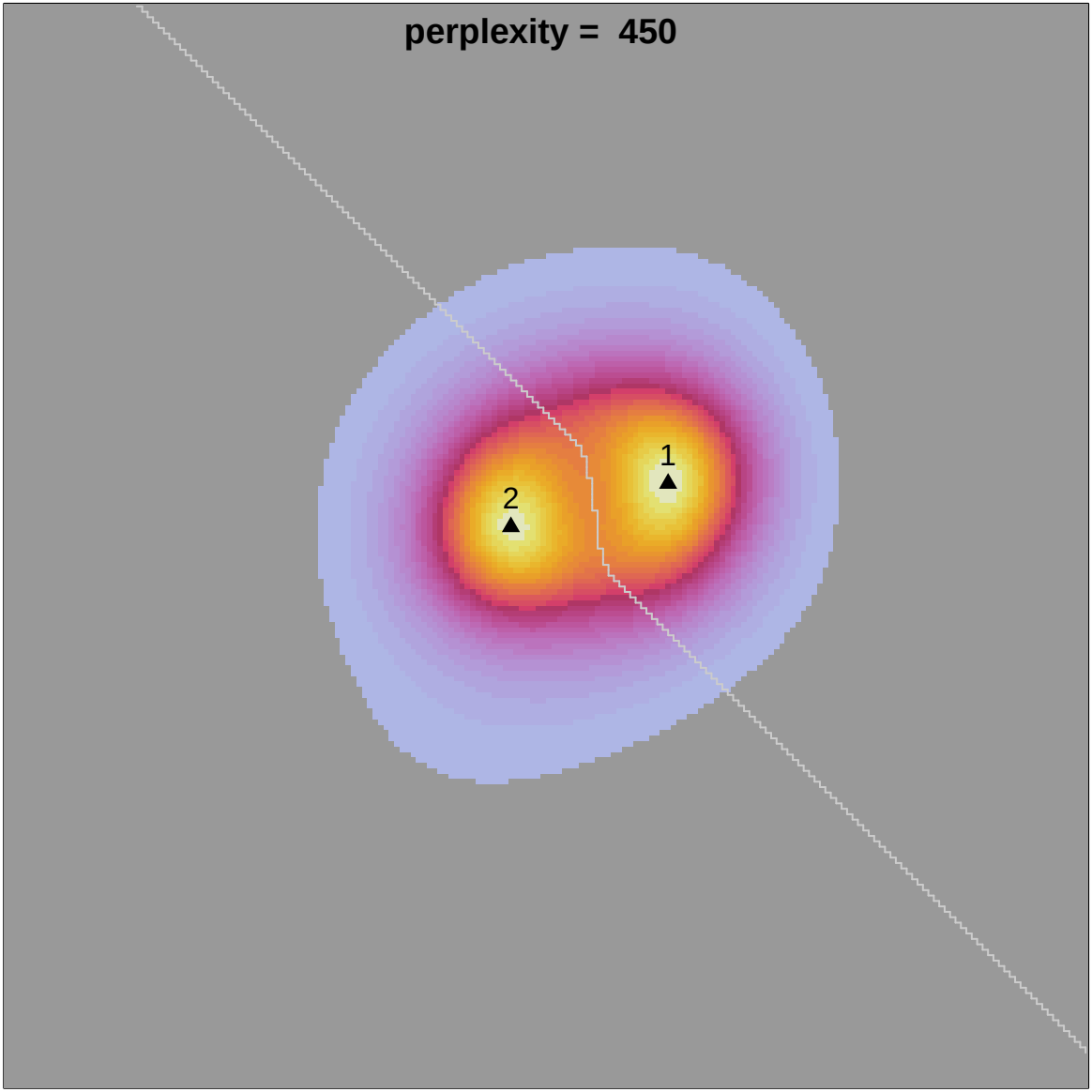

perplexity = 440
Cost
0.96
0.94
0.92
Size
4.5
4.0
3.5
3.0
2.5
0
20
40
60
80
100
epochs

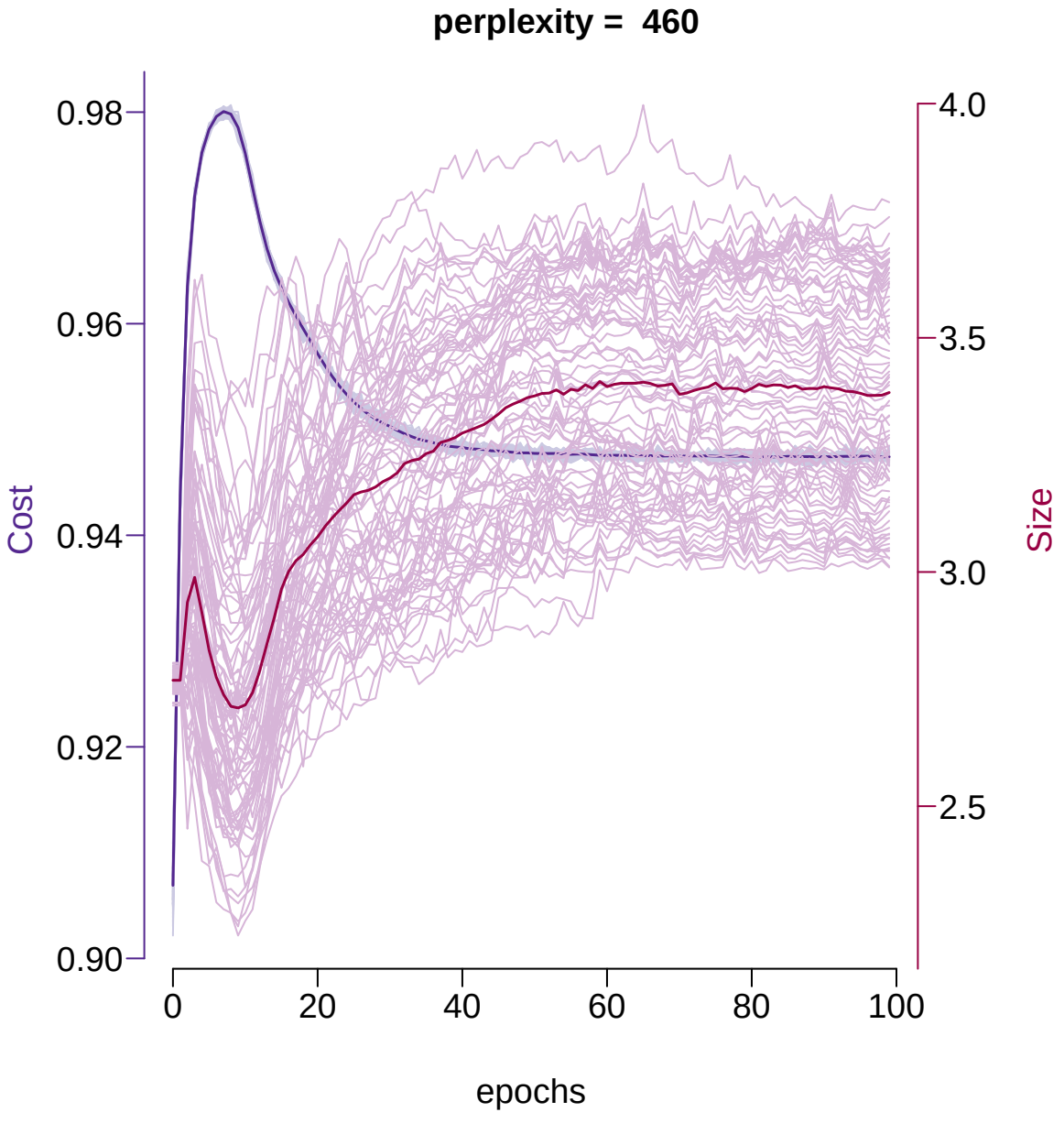

perplexity = 440
Y2
Y1
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.0
−0.5
0.0
0.5
1.0

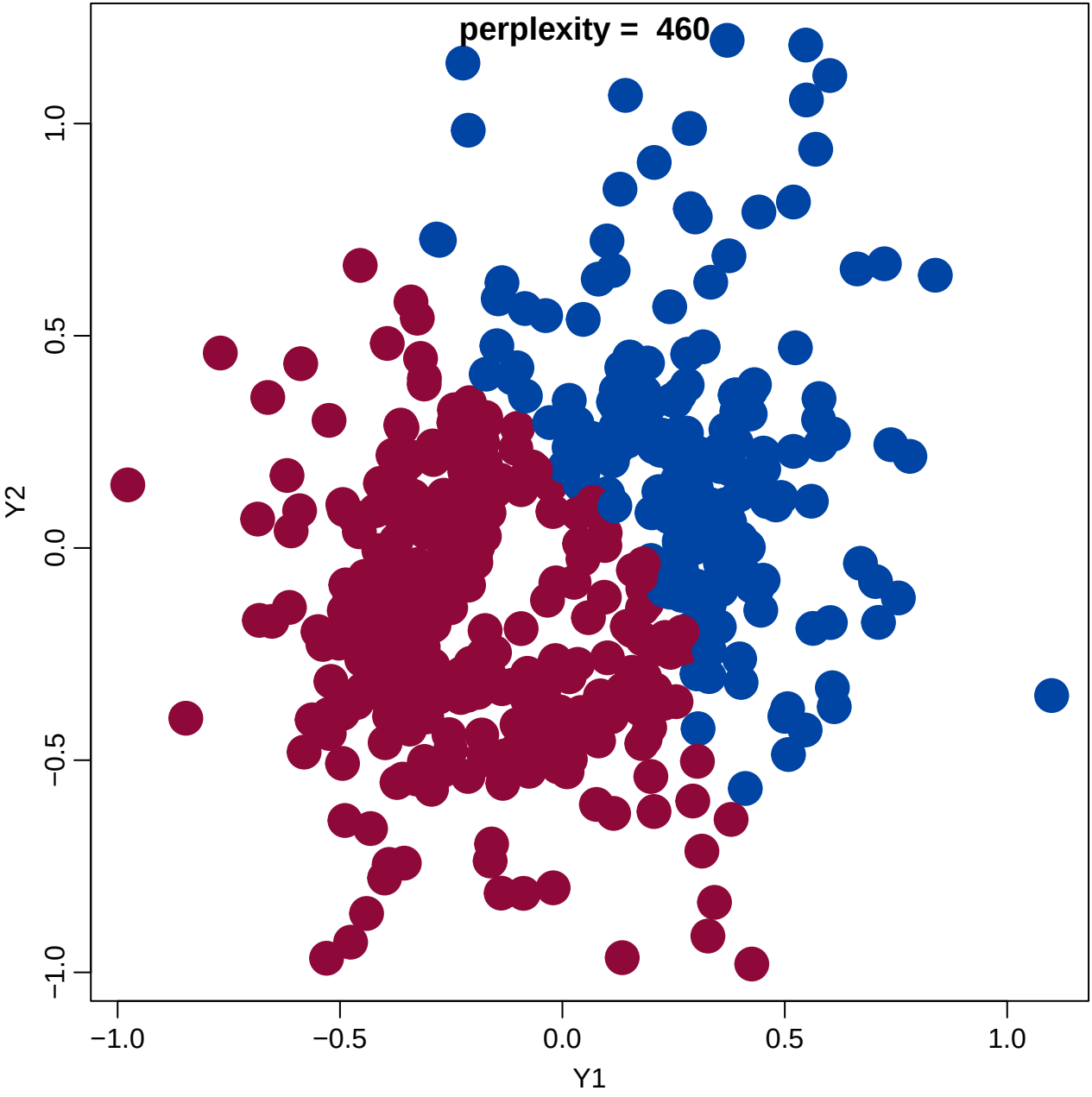

perplexity = 440
1
2

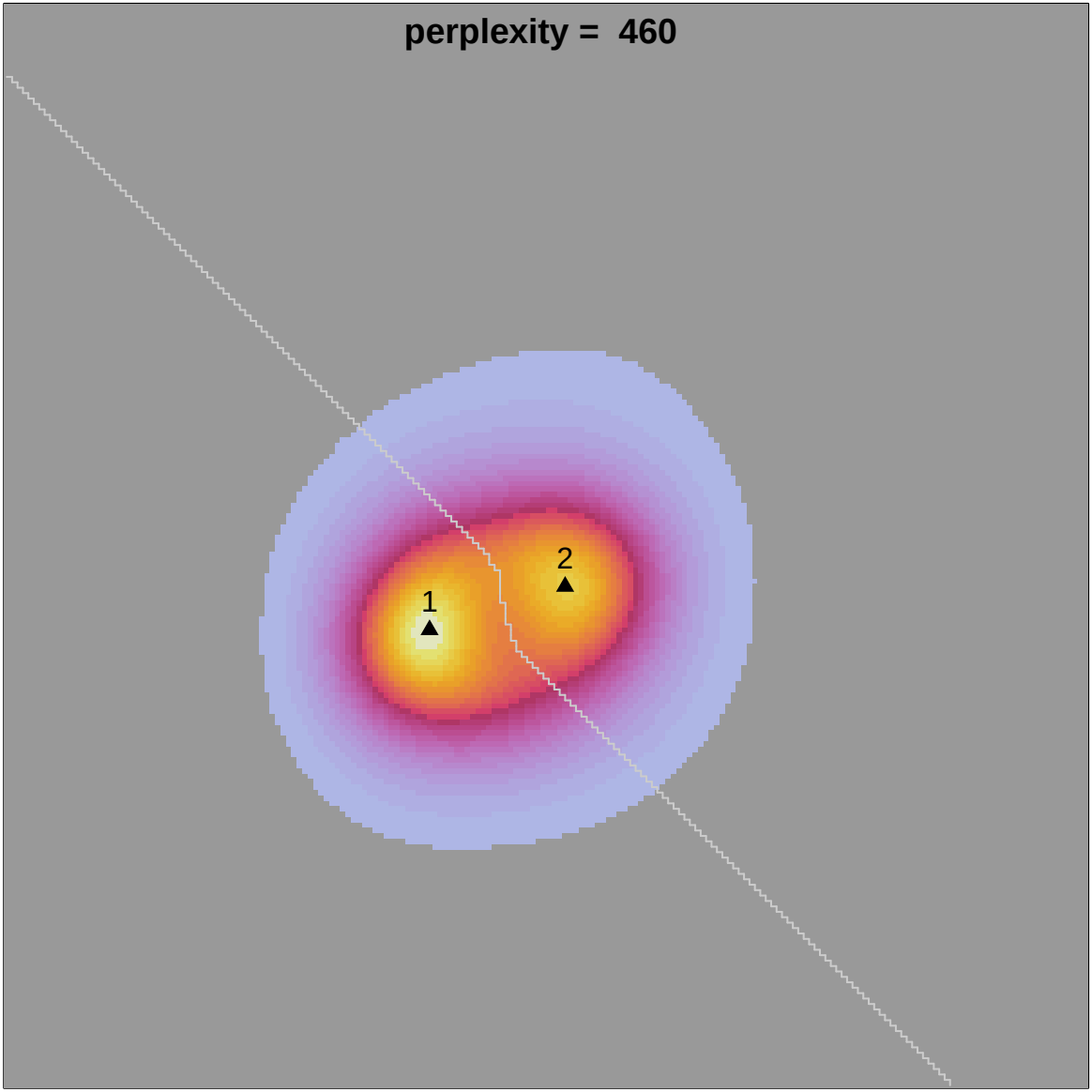

perplexity = 450
0.98
0.96
0.94
0.92
Cost
4.0
3.5
3.0
2.5
Size
0
20
40
60
80
100
epochs

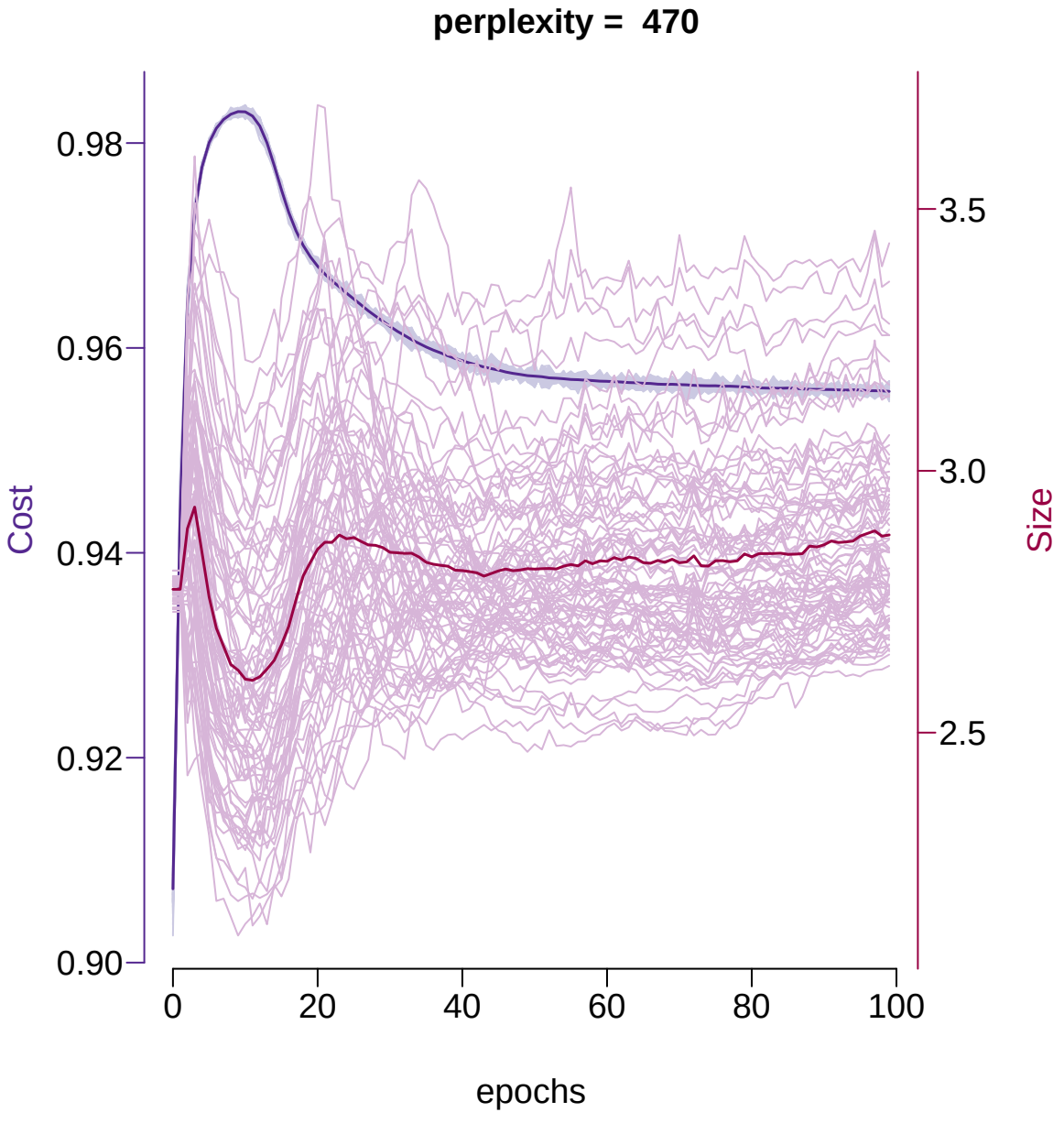

perplexity = 450
Y2
Y1
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0

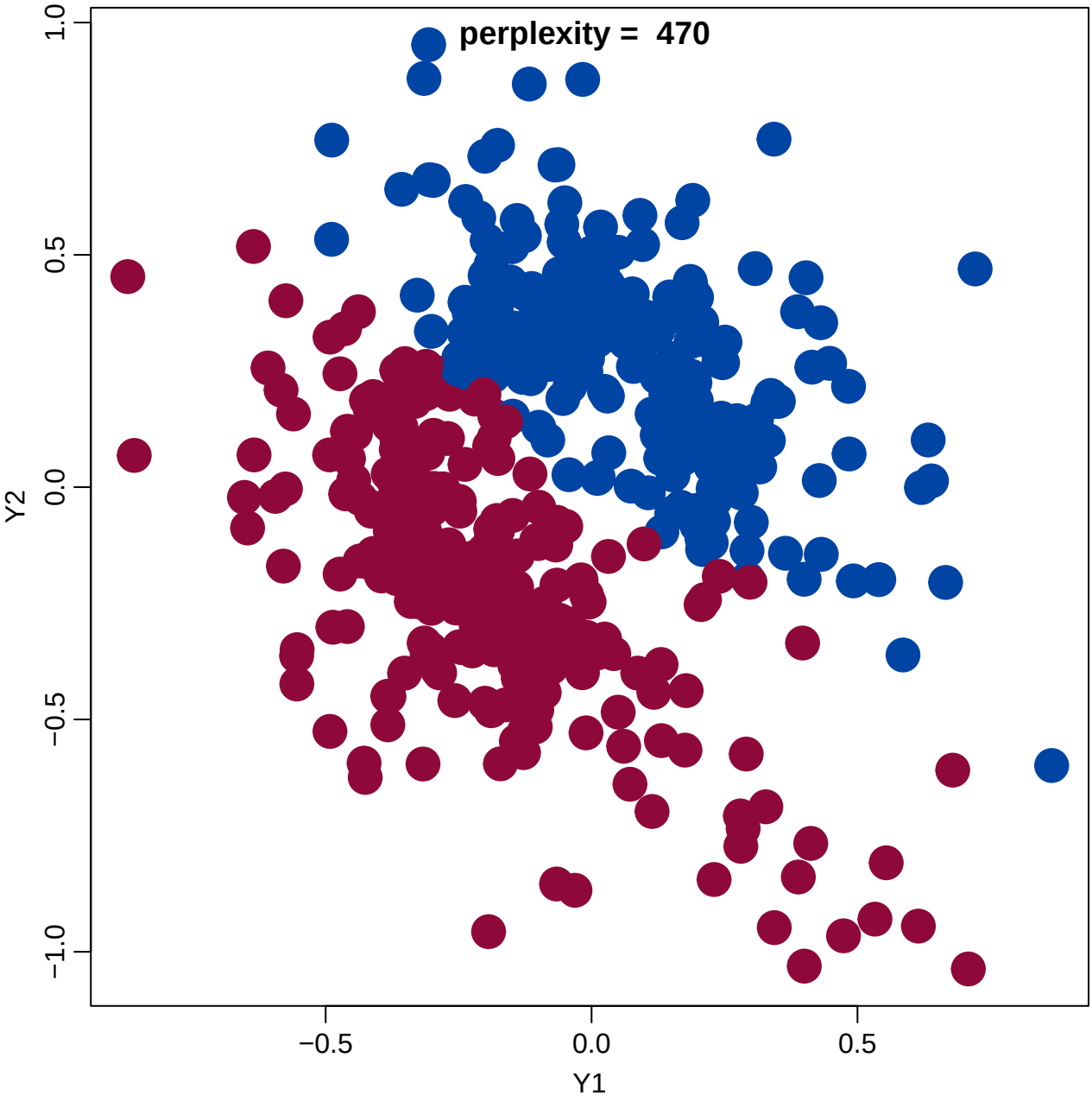

perplexity = 450
1
2

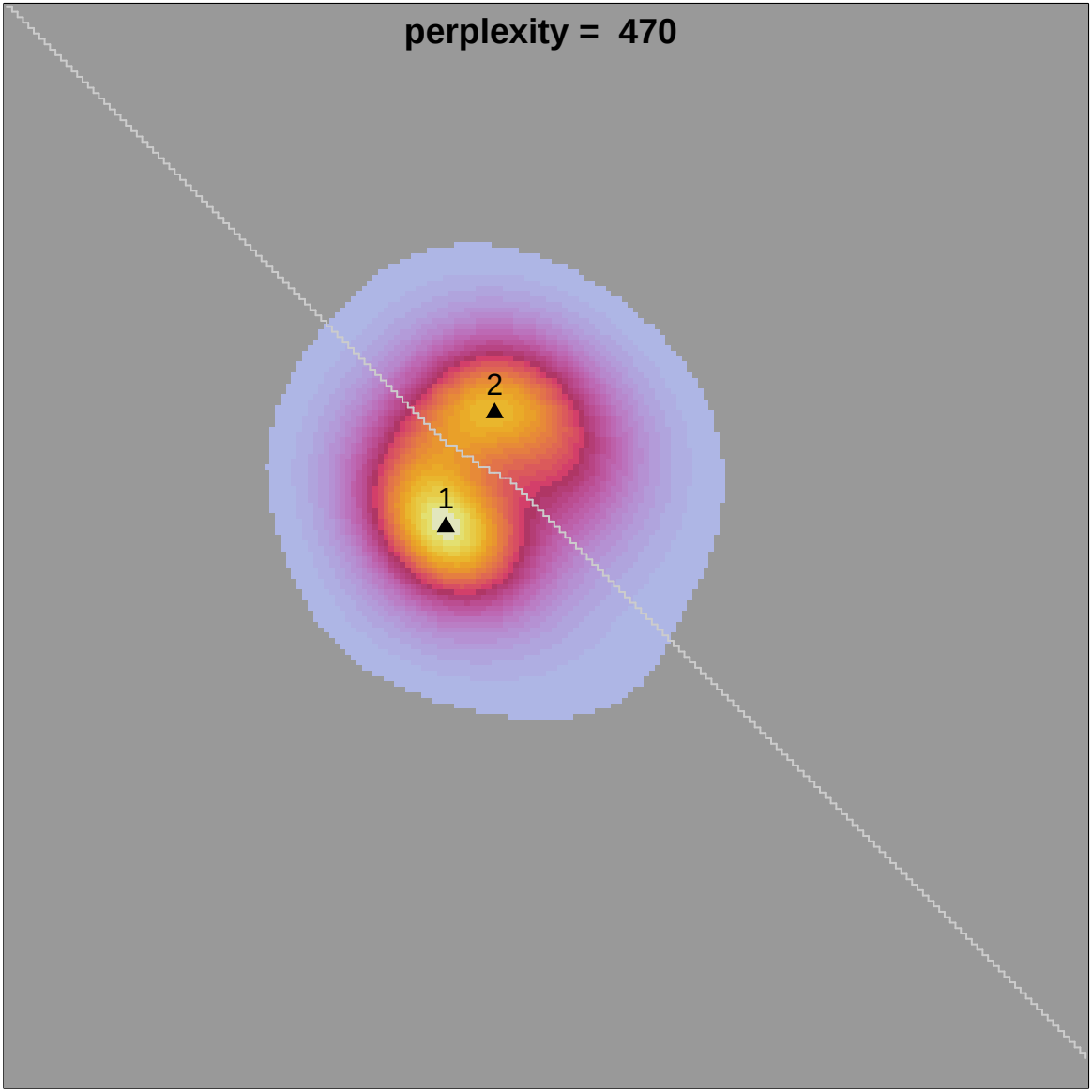

perplexity = 460
Cost
Size
epochs
0.98
0.96
0.94
0.92
0.90
4.0
3.5
3.0
2.5
0
20
40
60
80
100

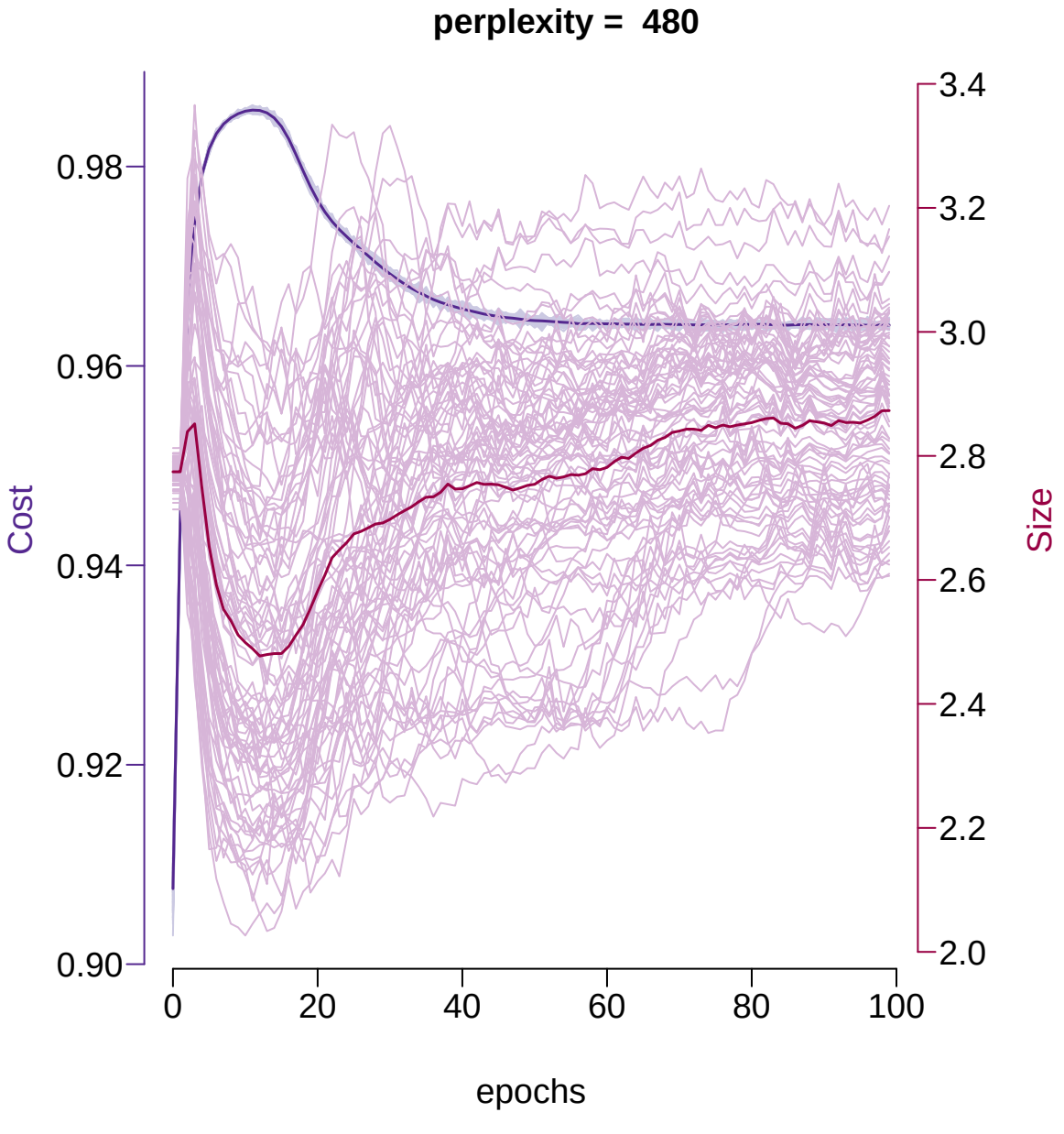

perplexity = 460
Y2
Y1
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0

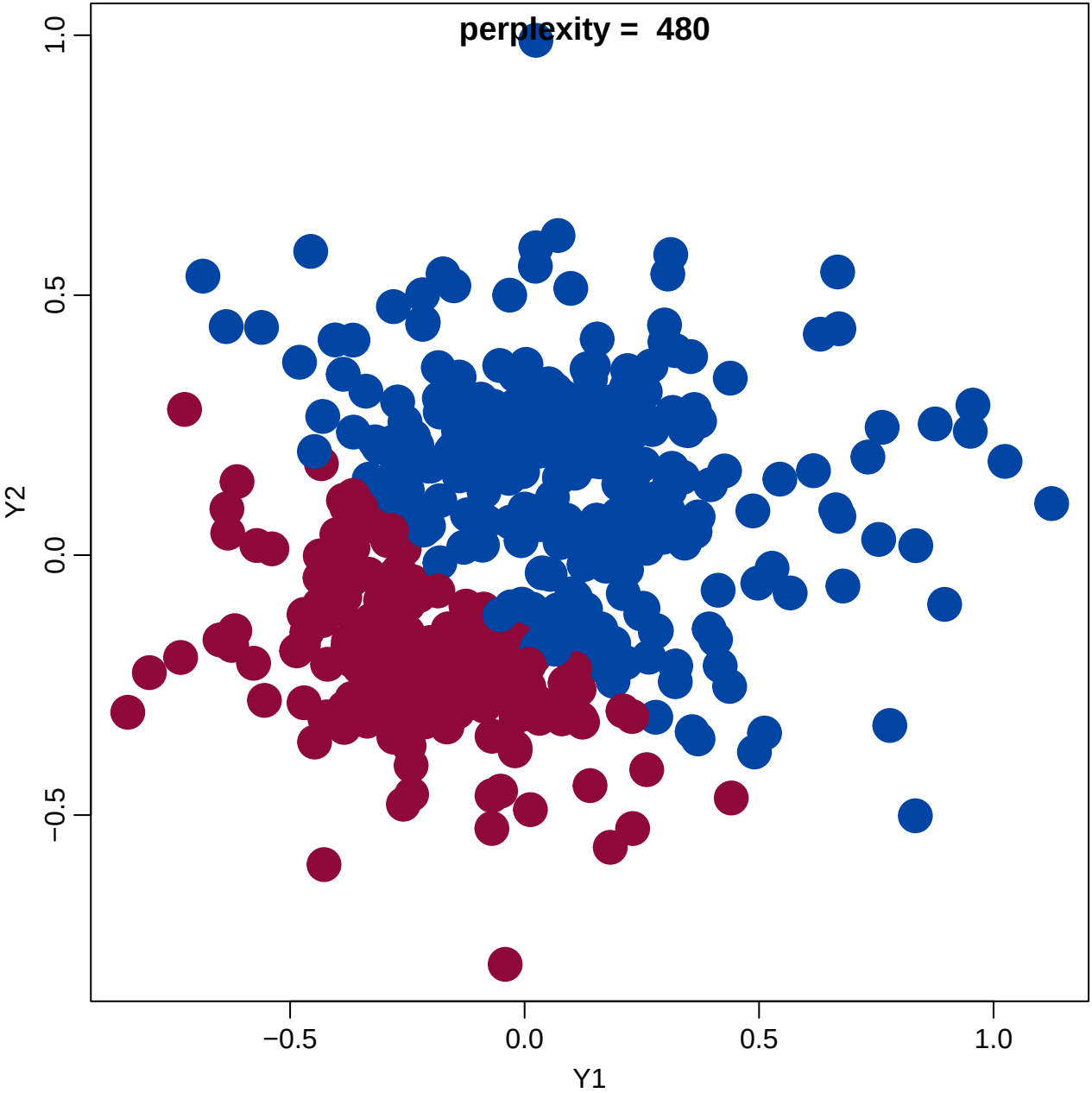

perplexity = 460
1
2

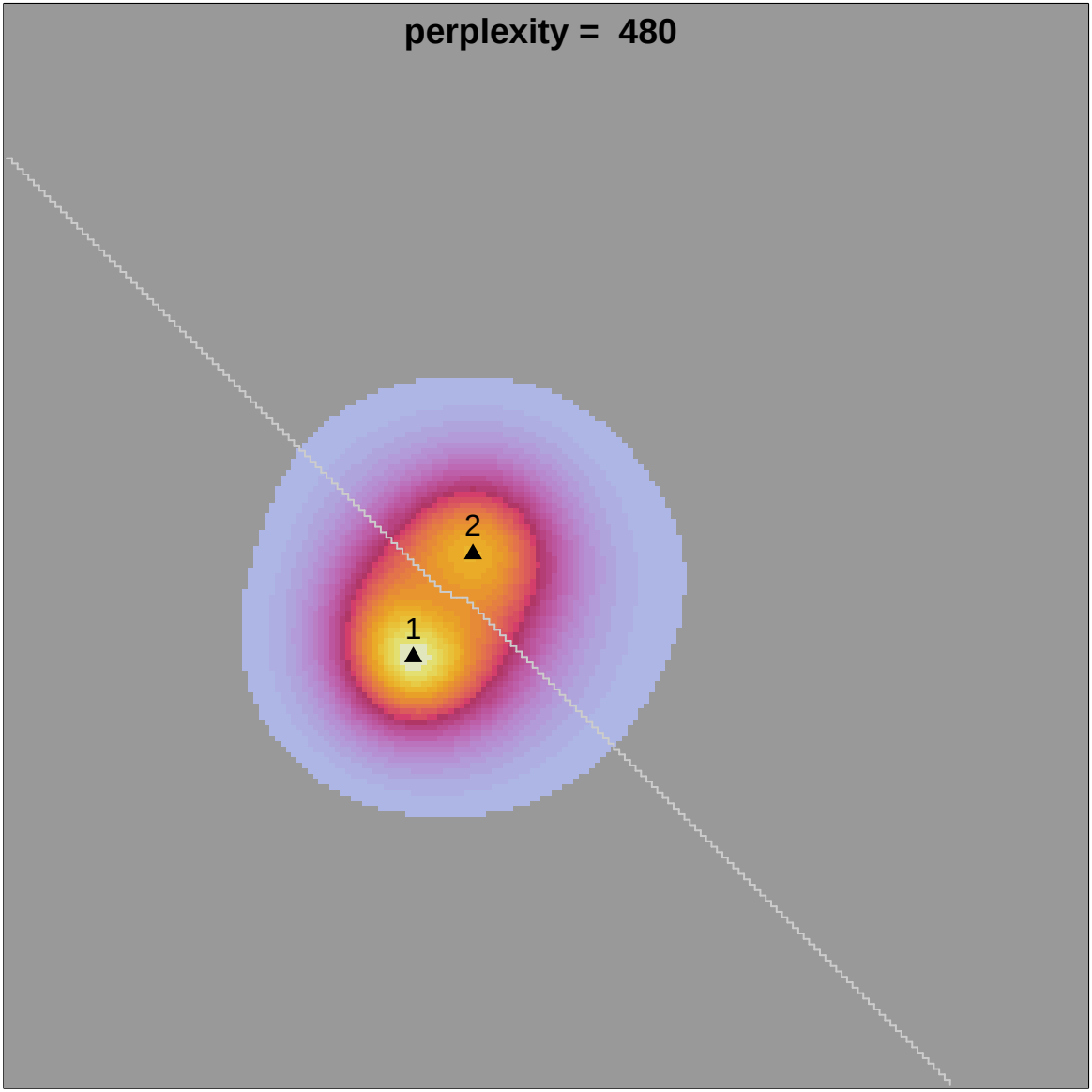

perplexity = 470
Cost
0.98
0.96
0.94
0.92
0.90
Size
3.5
3.0
2.5
0
20
40
60
80
100
epochs

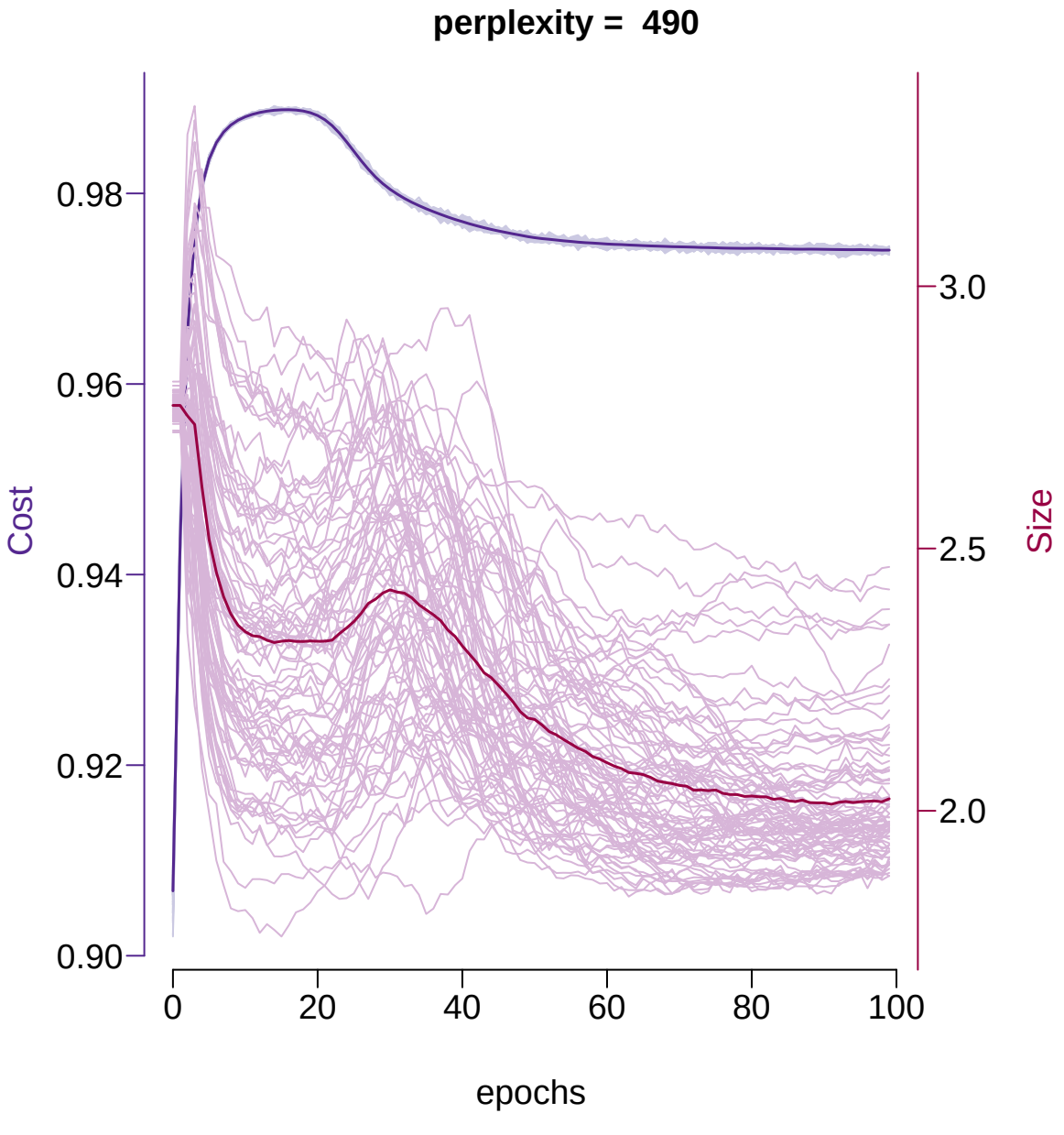

perplexity = 470
1.0
0.5
0.0
−0.5
−1.0
Y2
−0.5
0.0
0.5
Y1

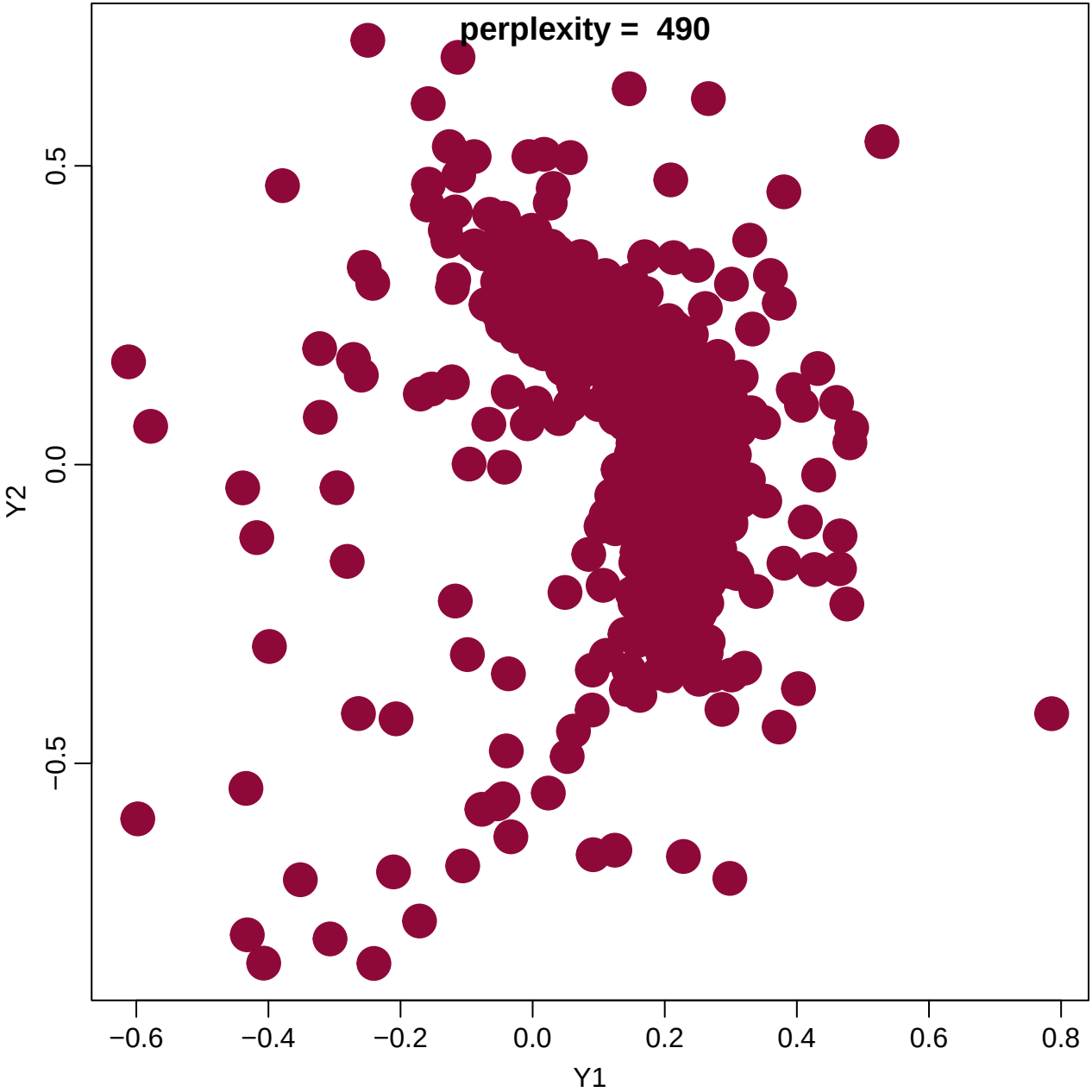

perplexity = 470
2
1

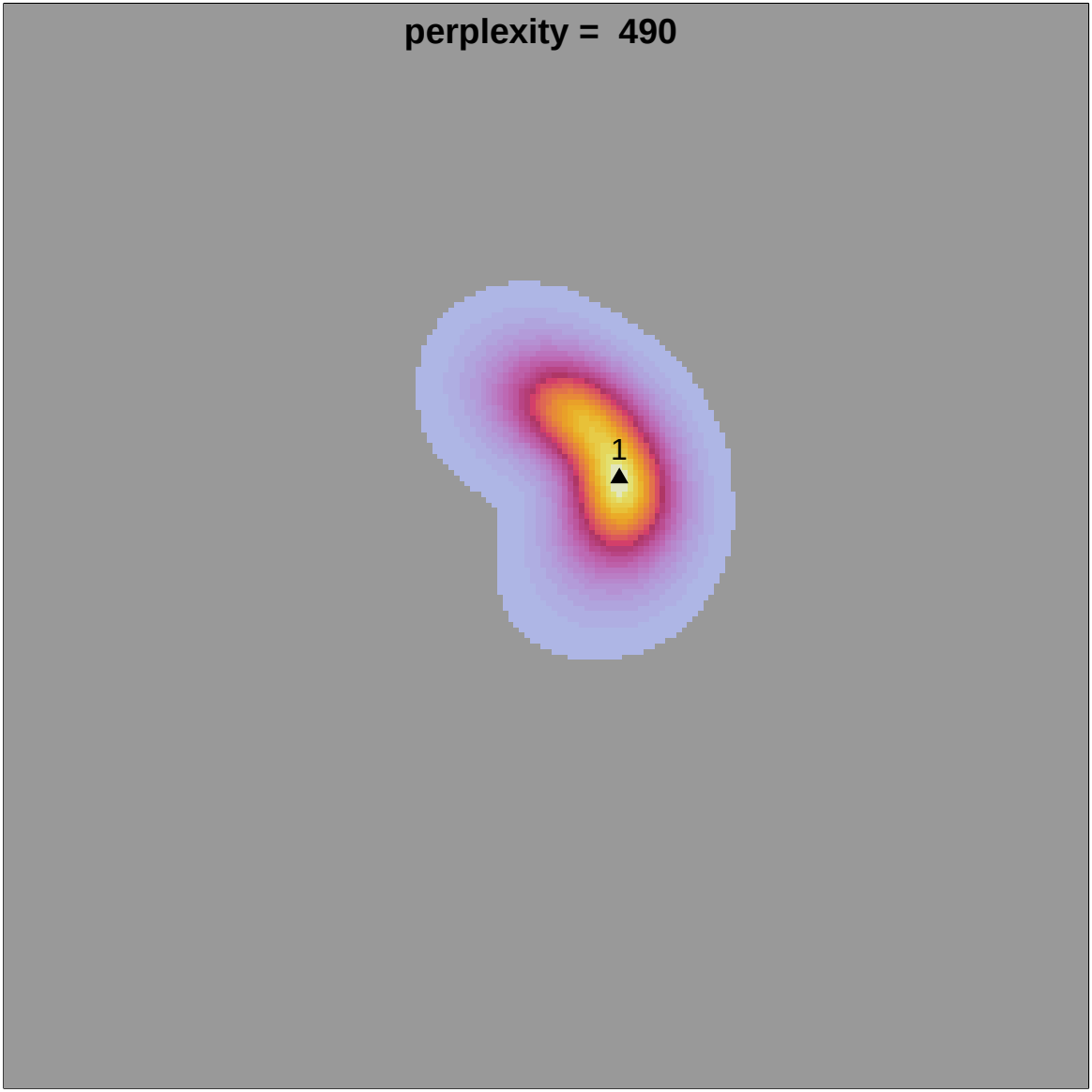

perplexity = 480
Cost
0.98
0.96
0.94
0.92
0.90
Size
3.4
3.2
3.0
2.8
2.6
2.4
2.2
2.0
0
20
40
60
80
100
epochs

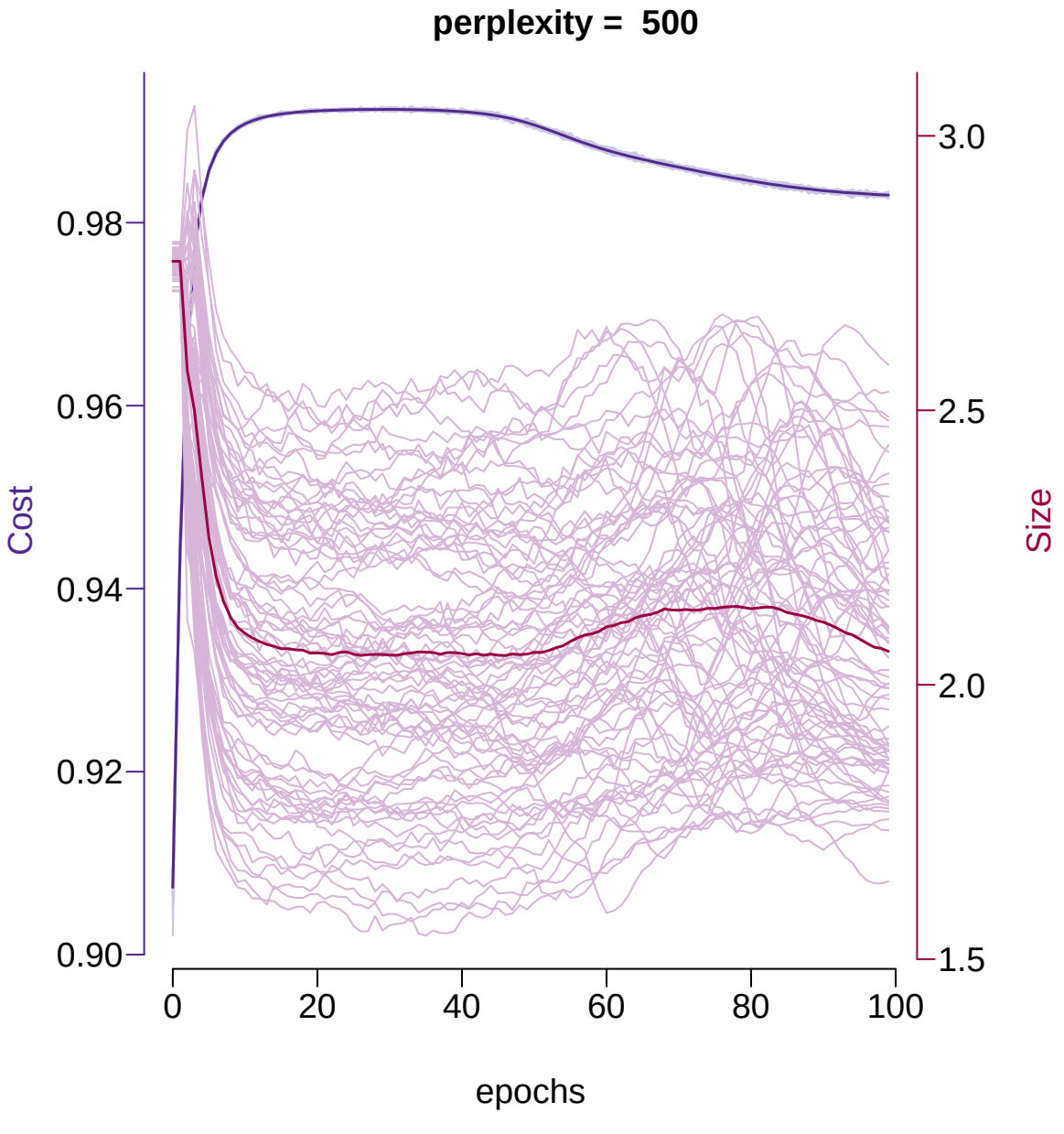

perplexity = 480
1.0
0.5
0.0
−0.5
Y2
−0.5
0.0
0.5
1.0
Y1

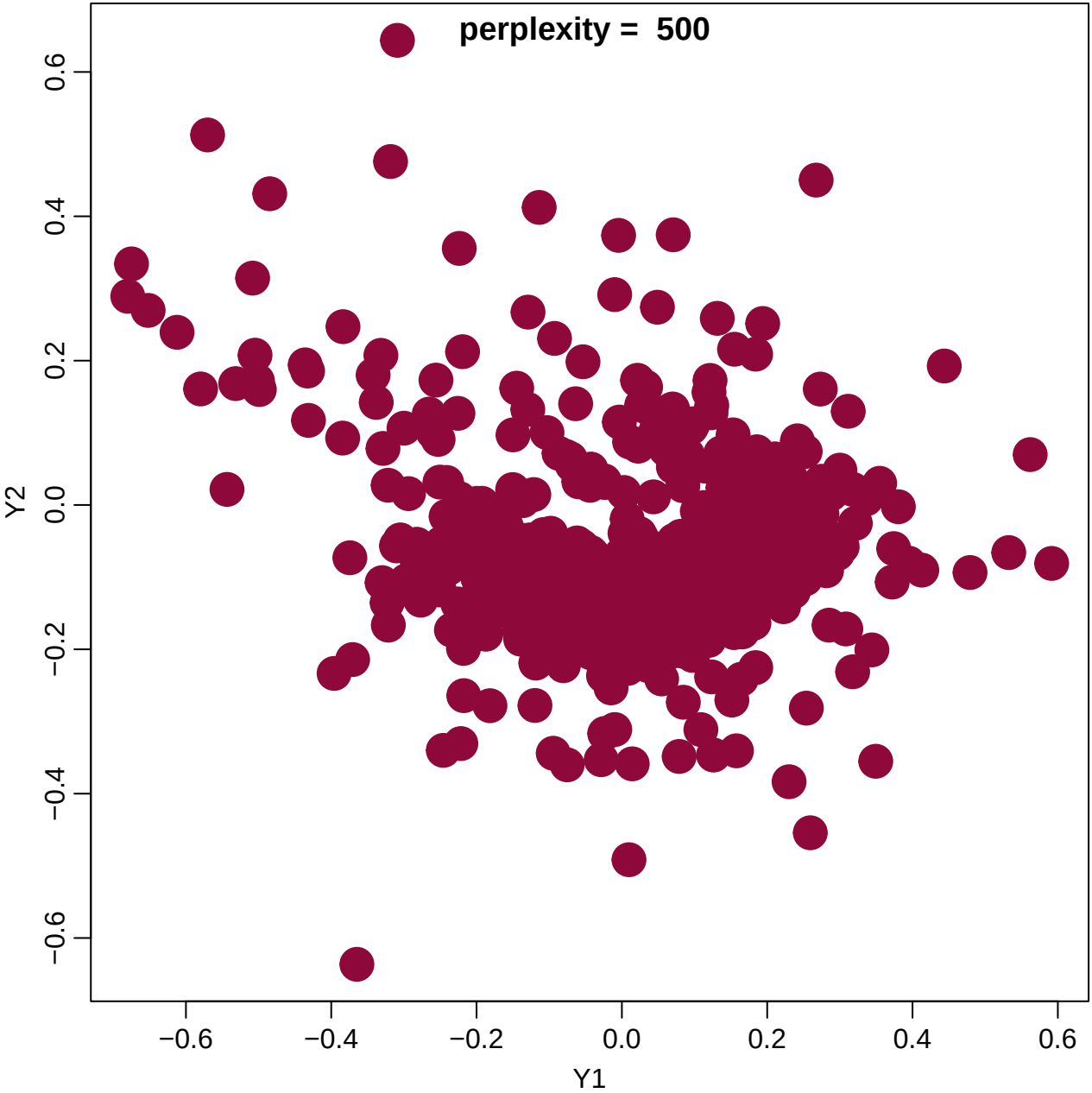

perplexity = 480
2
1

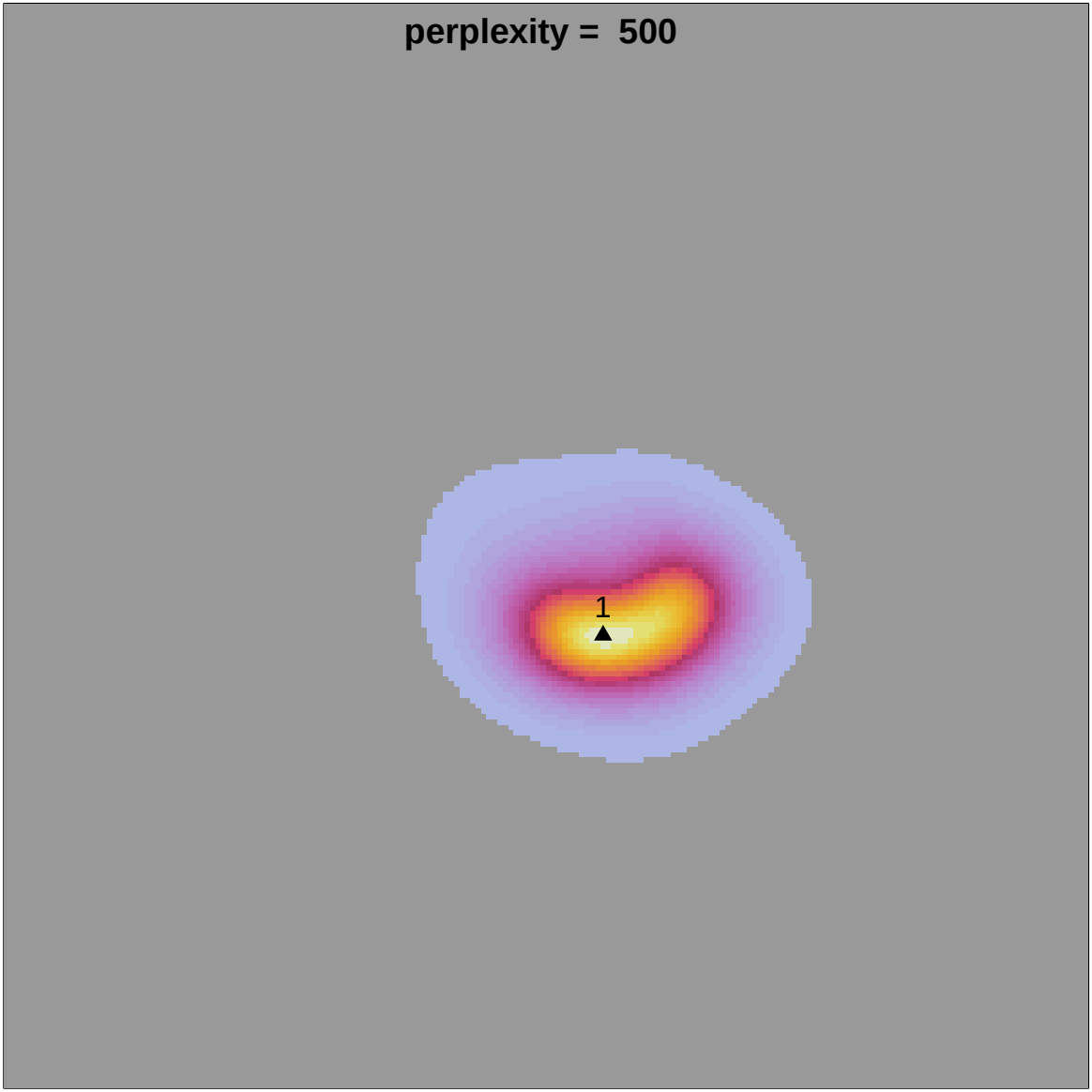

perplexity = 490
Cost
0.98
0.96
0.94
0.92
0.90
Size
3.0
2.5
2.0
0
20
40
60
80
100
epochs

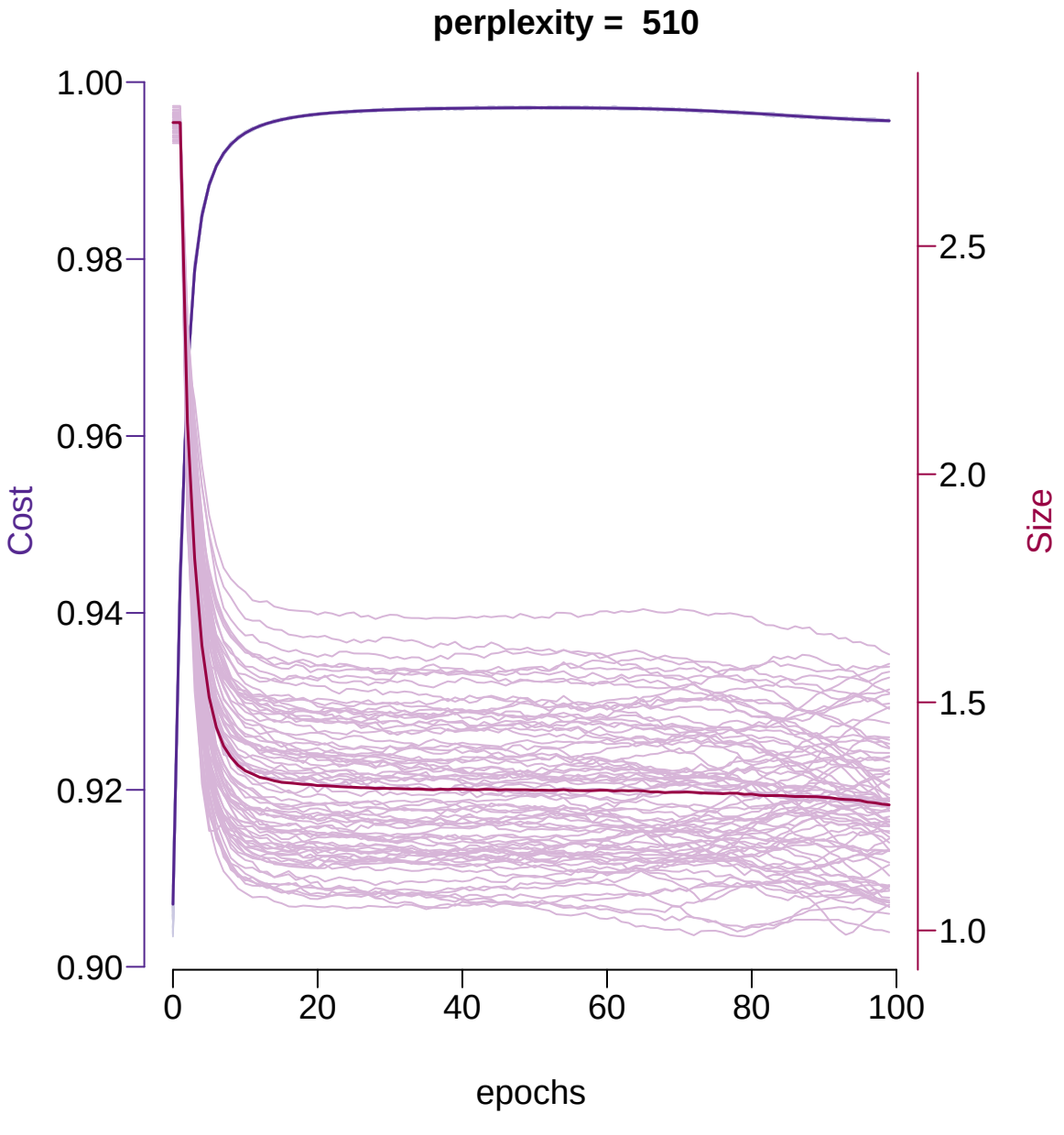

perplexity = 490
0.5
0.0
−0.5
Y2
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
Y1

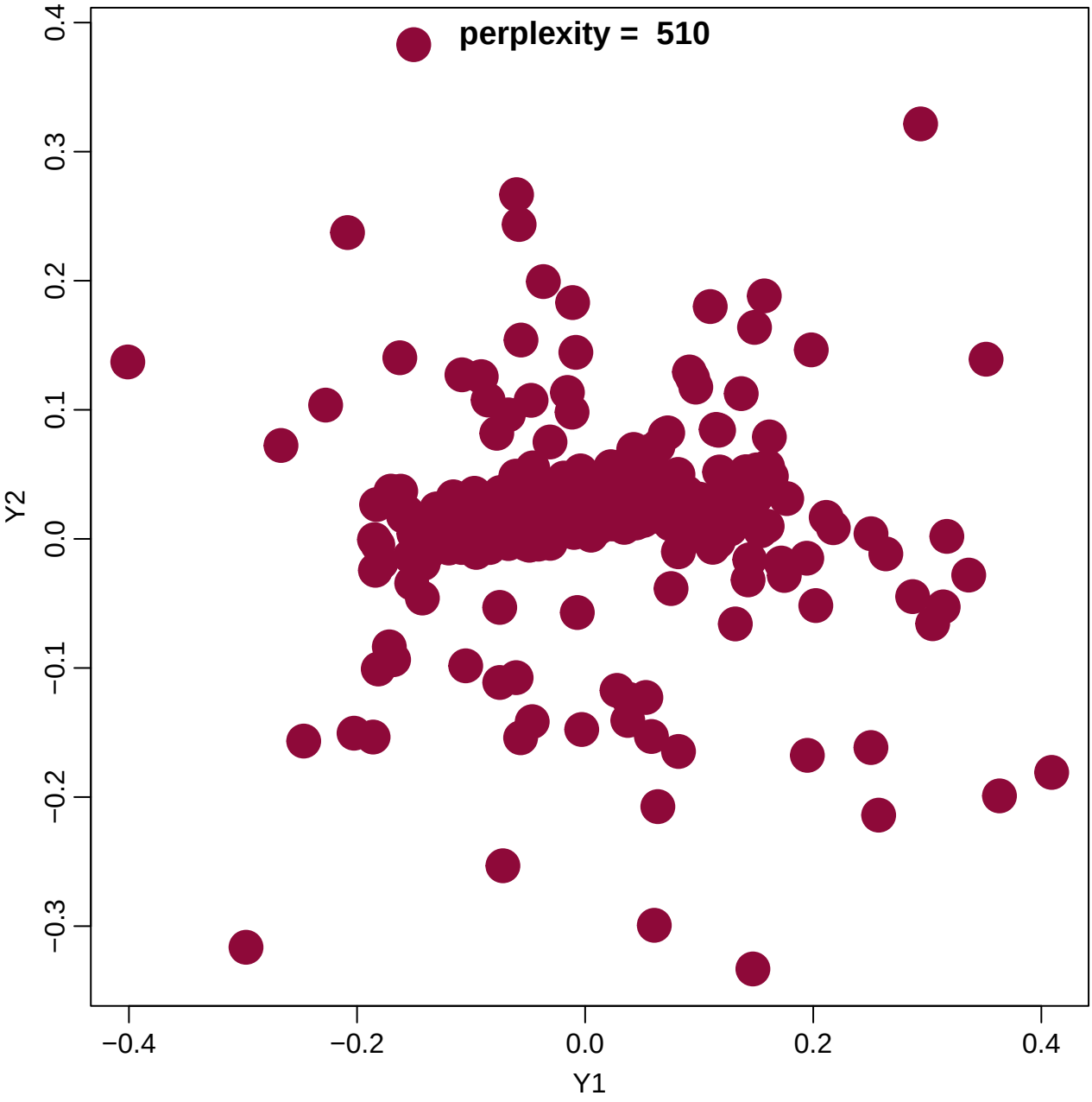

perplexity = 490
1

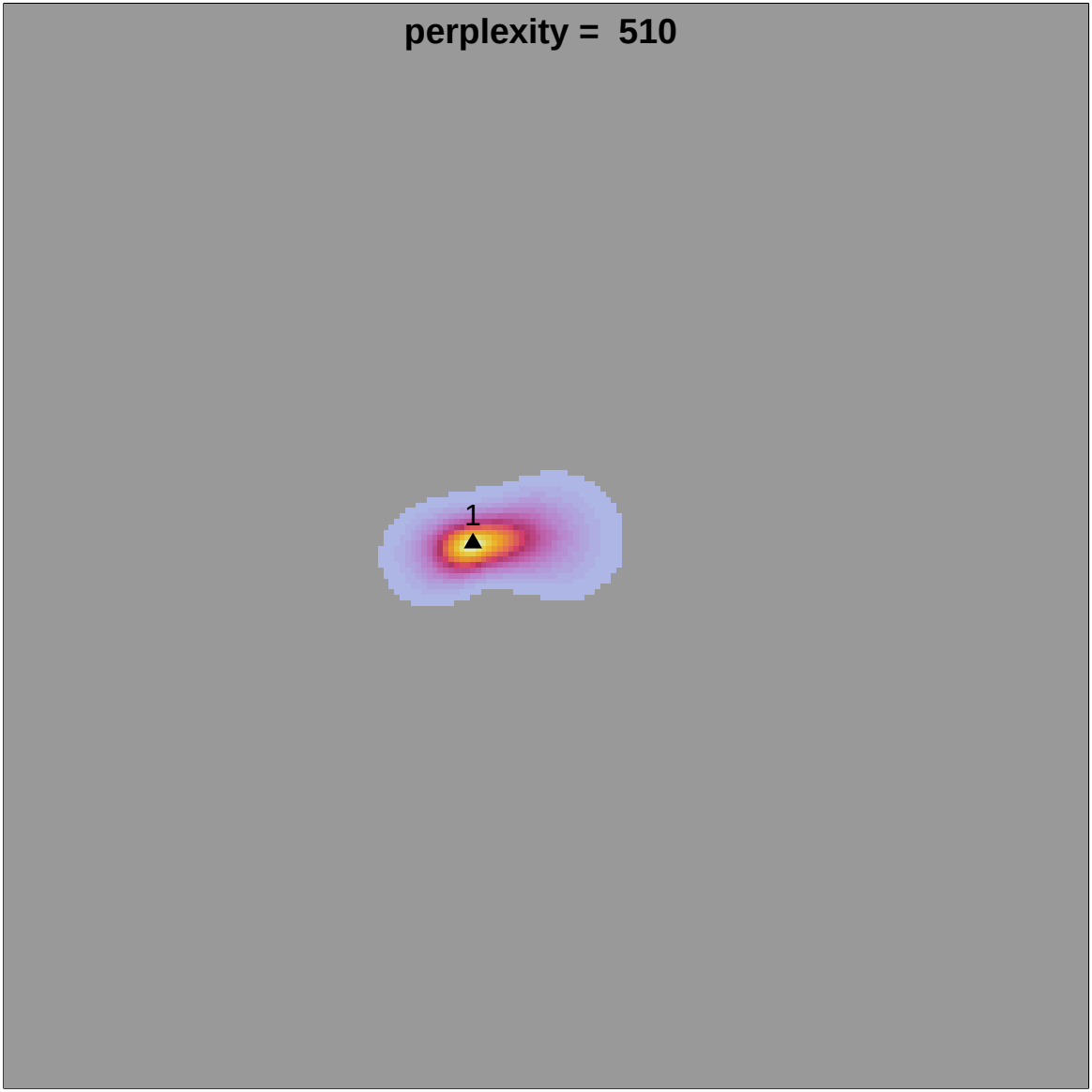

perplexity = 500
Cost
0.90
0.92
0.94
0.96
0.98
Size
1.5
2.0
2.5
3.0
0
20
40
60
80
100
epochs

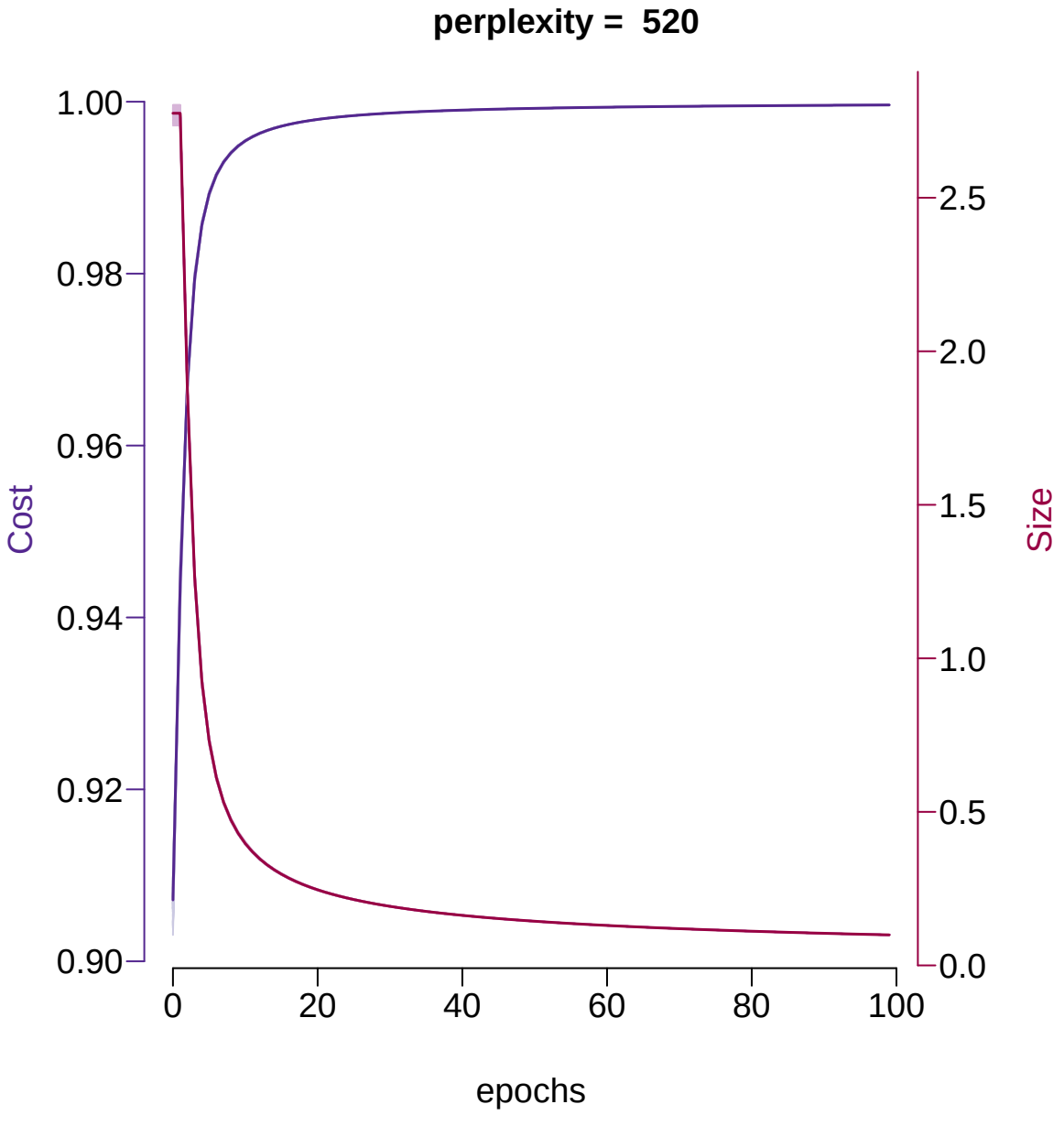

perplexity = 500
Y2
Y1
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6

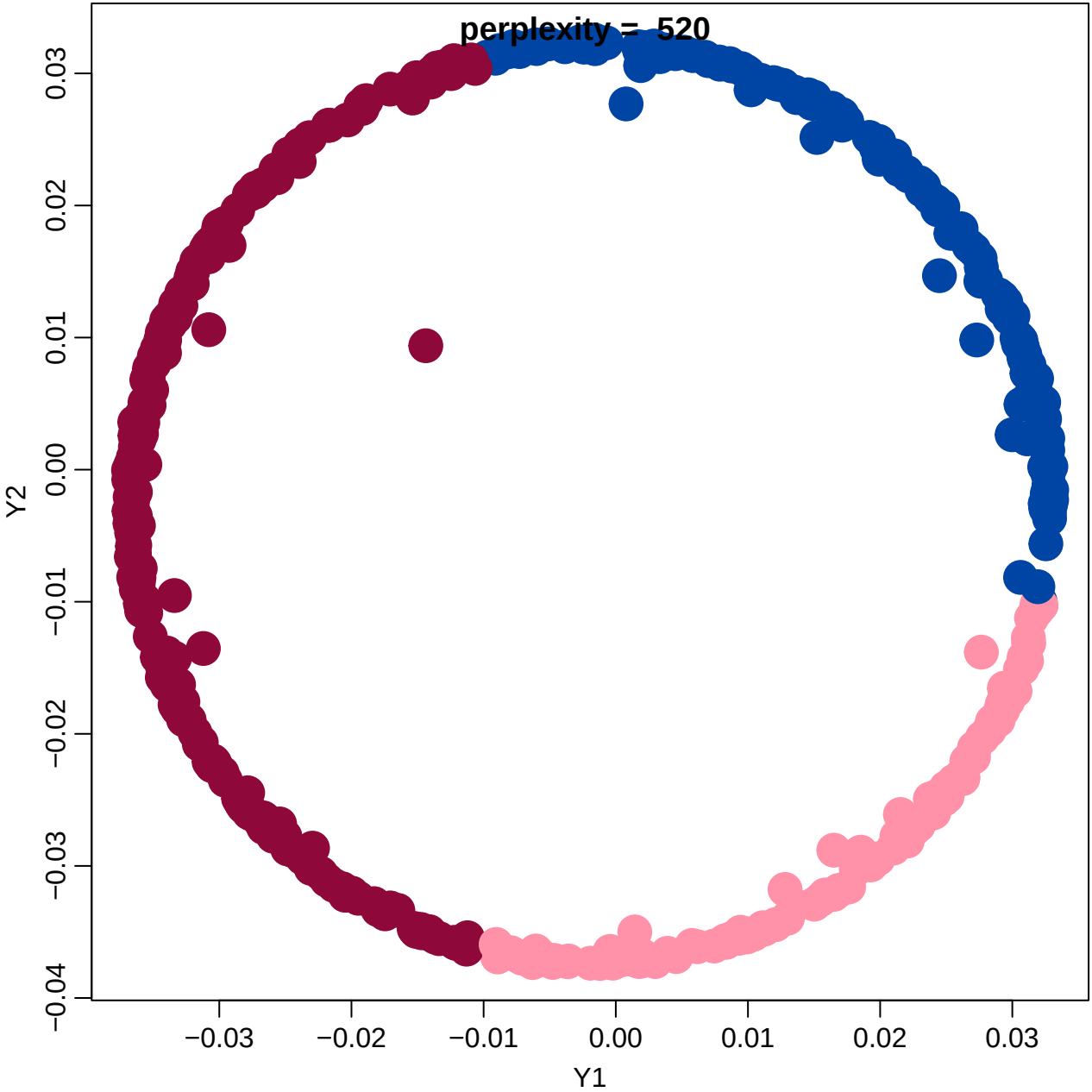

perplexity = 500
1

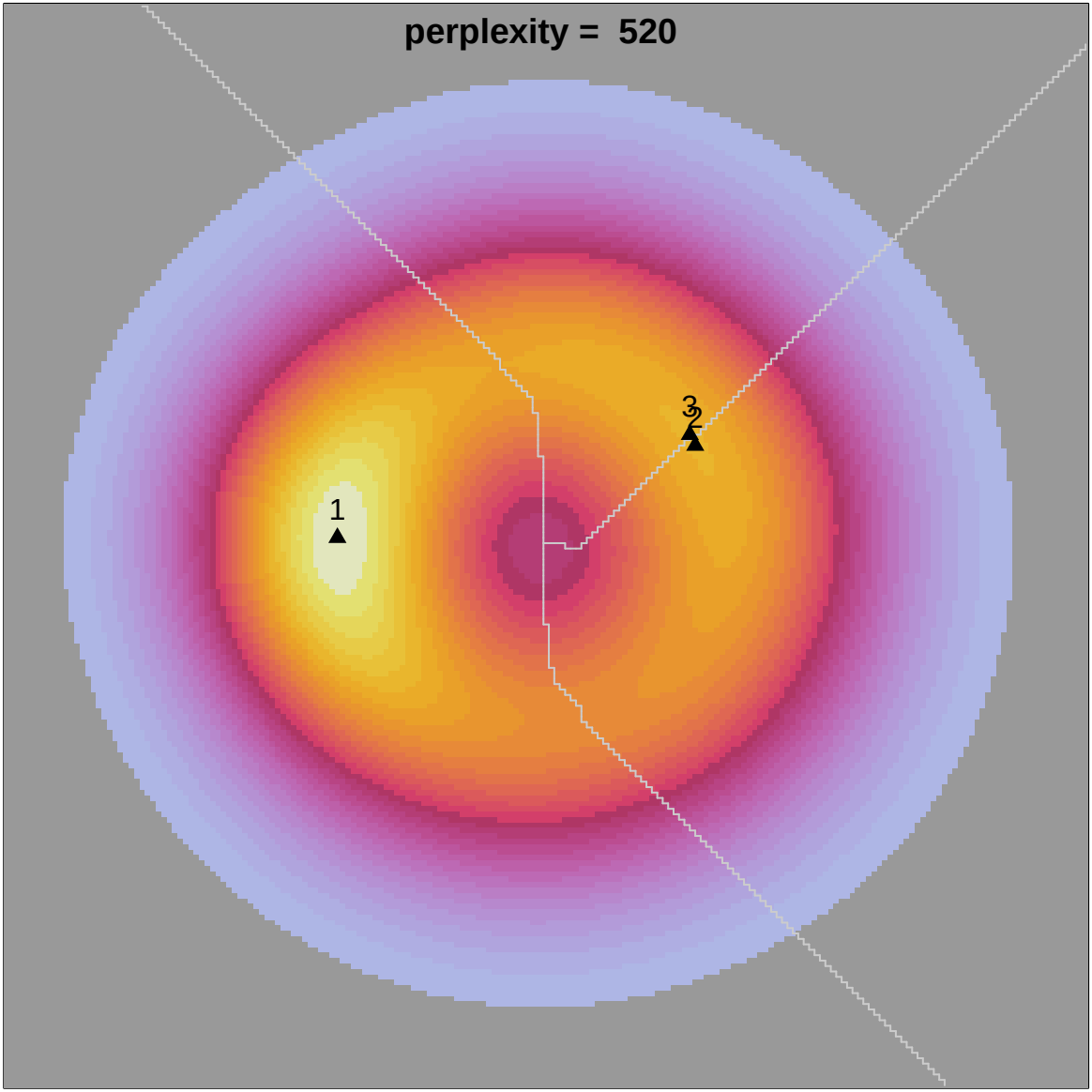

perplexity = 510
1.00
0.98
0.96
0.94
0.92
0.90
Cost
2.5
2.0
1.5
1.0
Size
0
20
40
60
80
100
epochs

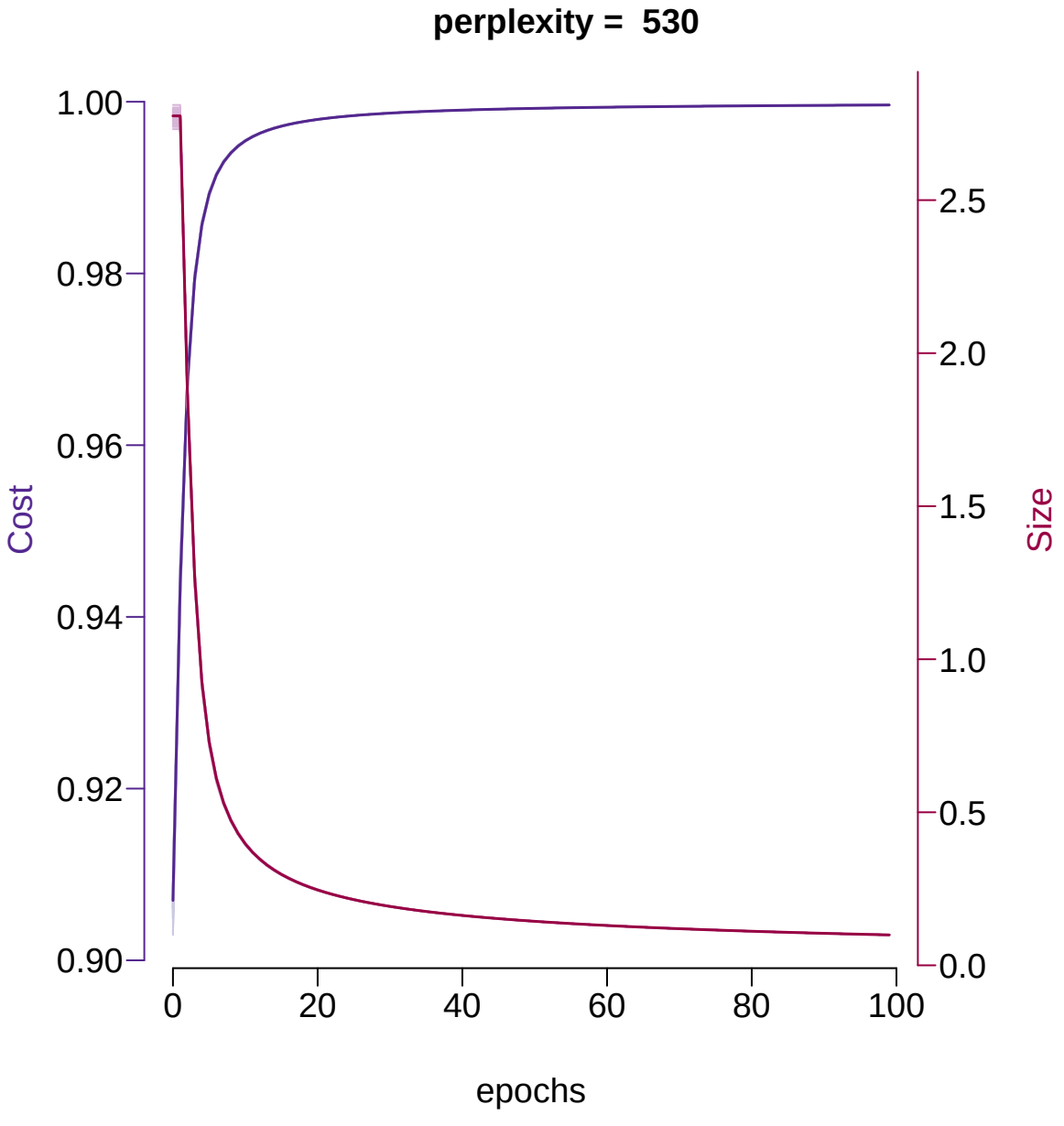

perplexity = 510
Y2
Y1
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.2
0.0
0.2
0.4

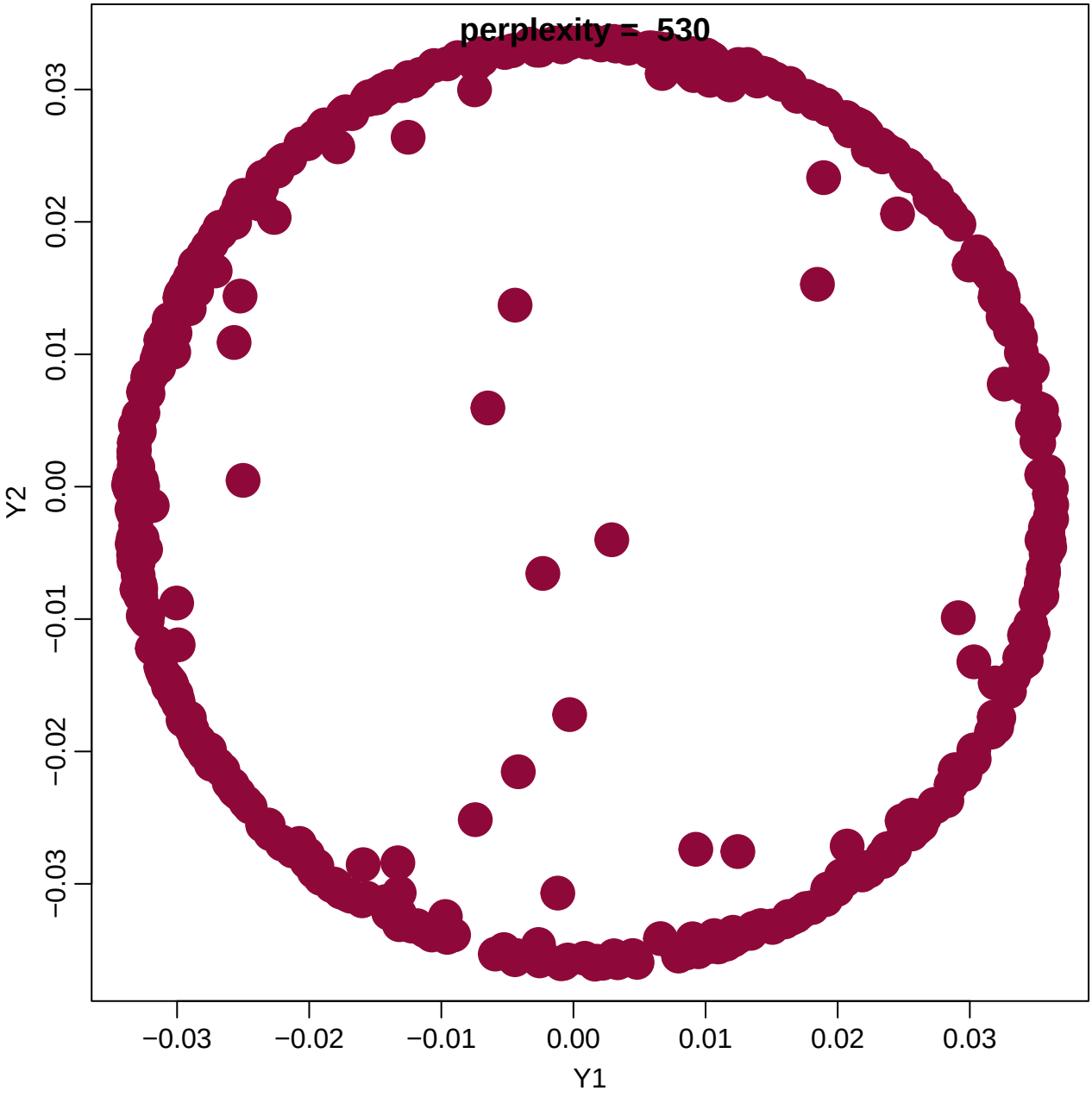

perplexity = 510
1

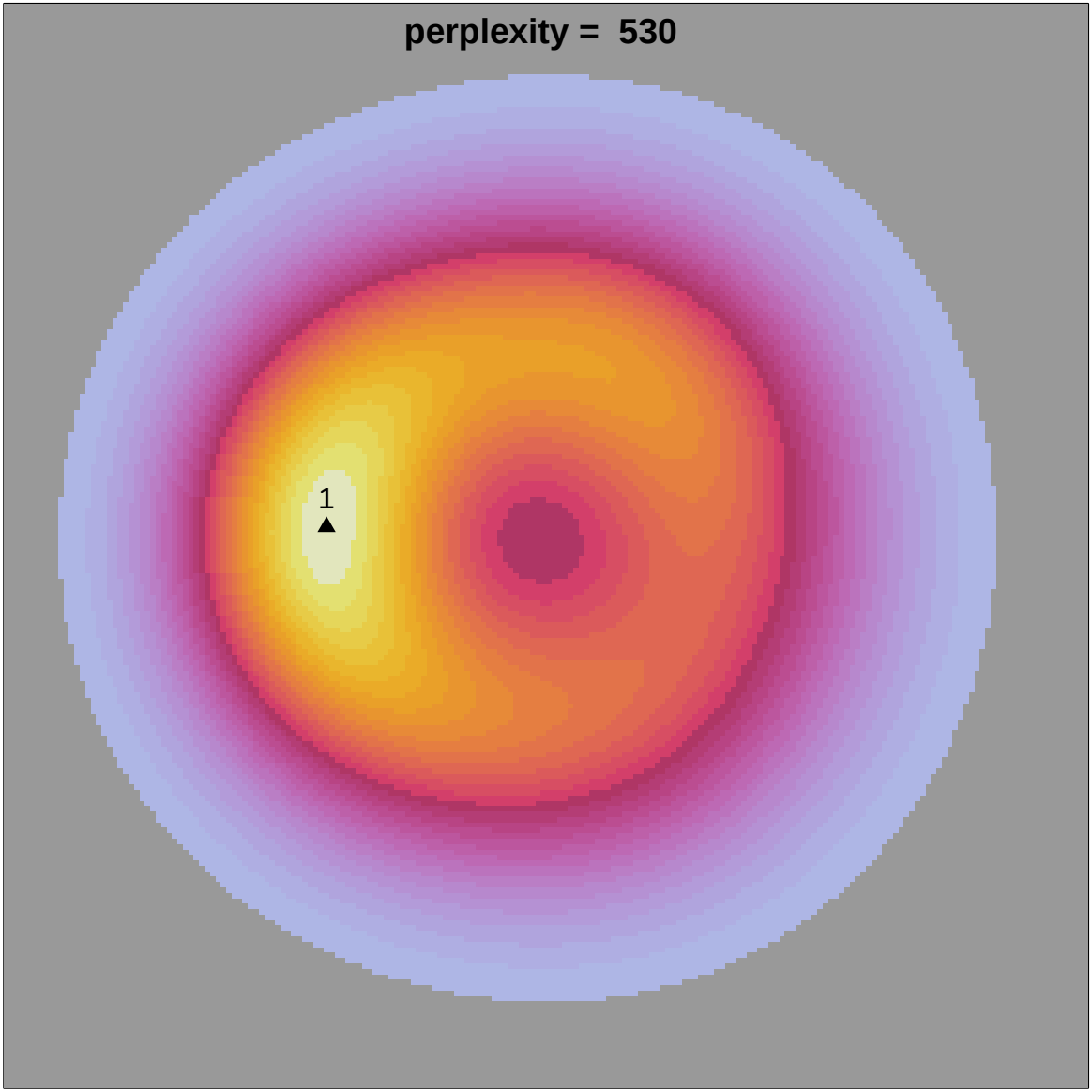

perplexity = 520
1.00
0.98
0.96
0.94
0.92
0.90
Cost
2.5
2.0
1.5
1.0
0.5
0.0
Size
0
20
40
60
80
100
epochs

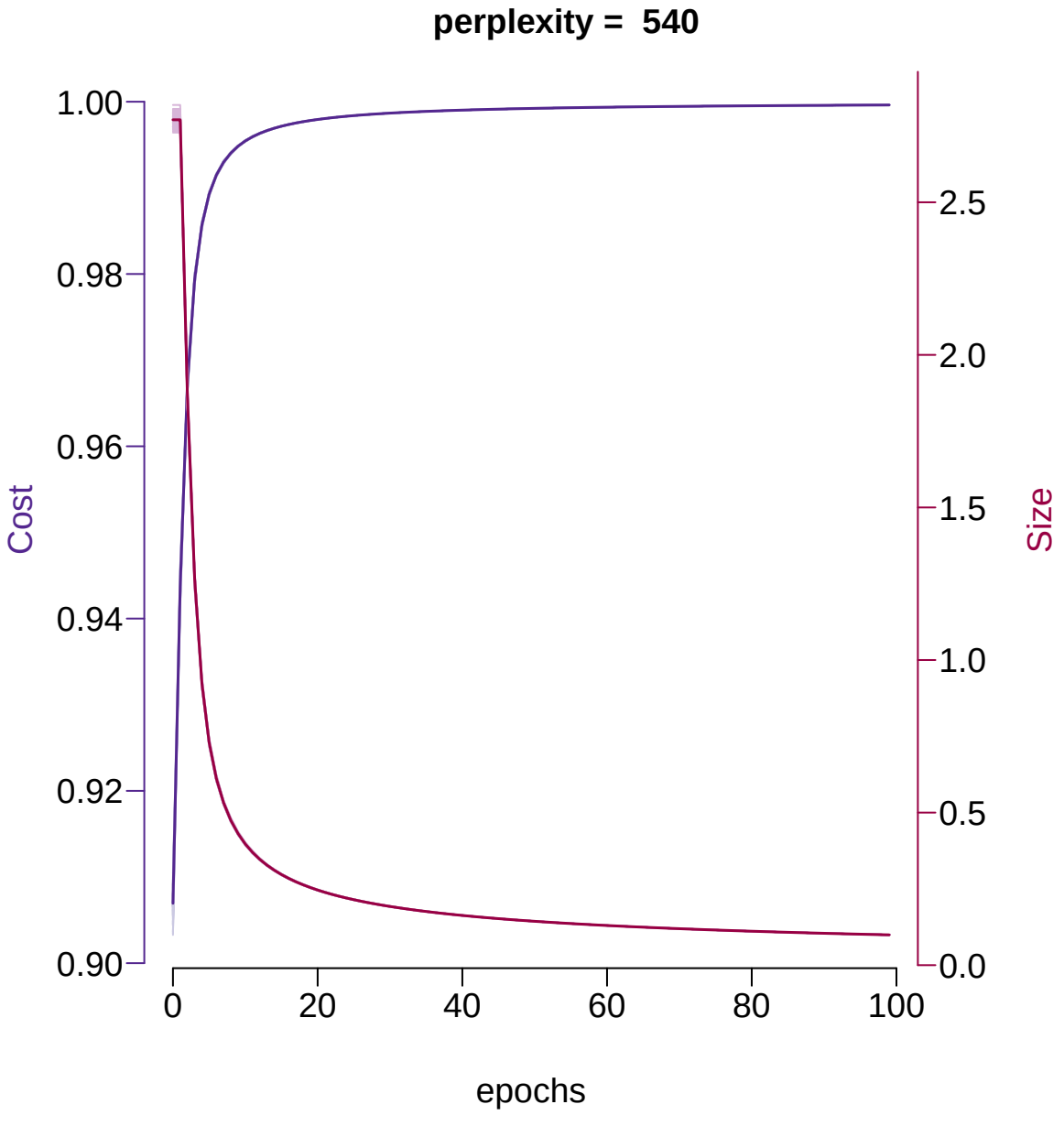

perplexity = 520
Y2
Y1
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
−0.04
−0.03
−0.02
−0.01
0.00
0.01
0.02
0.03

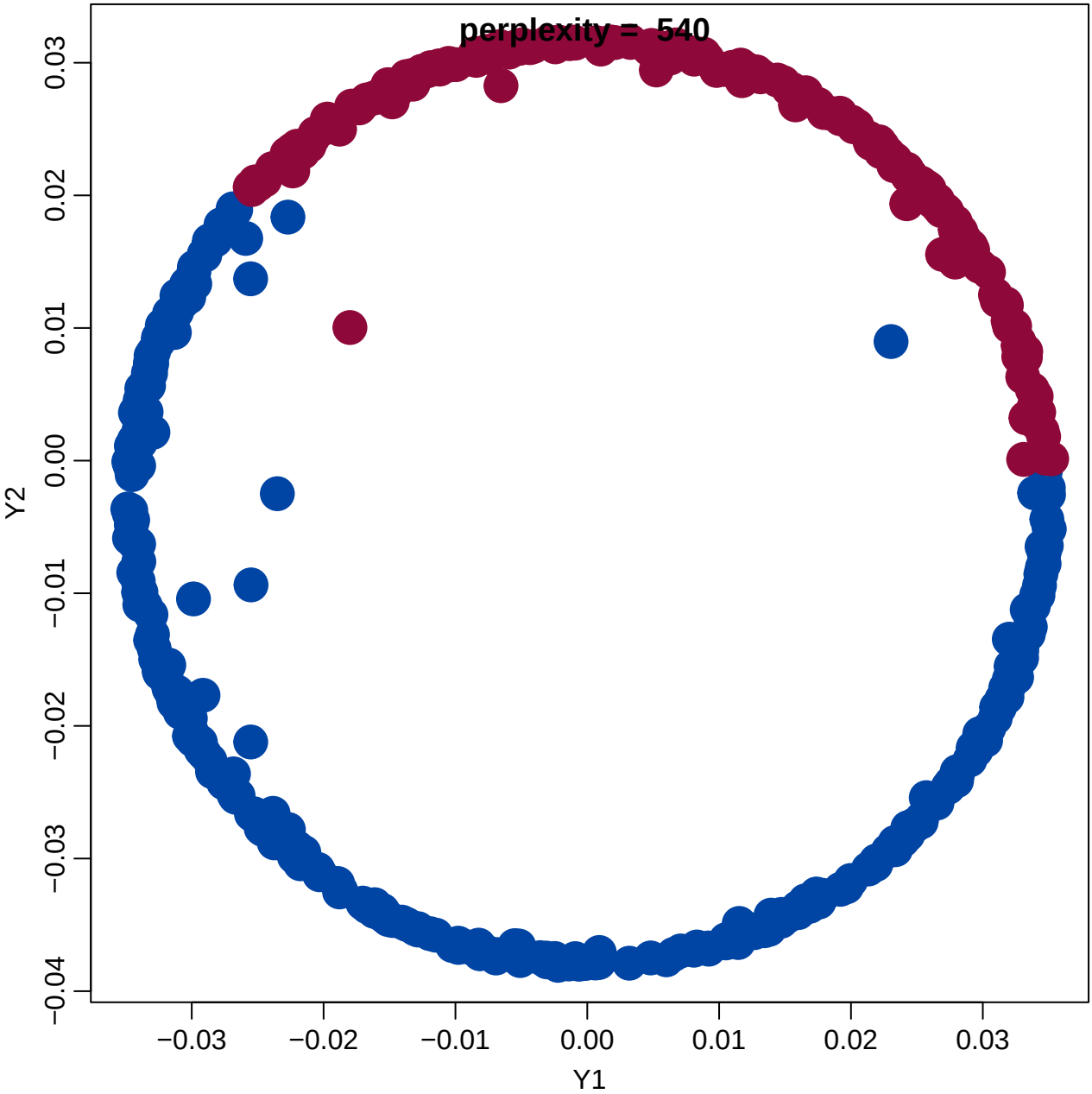

perplexity = 520
1
2
3

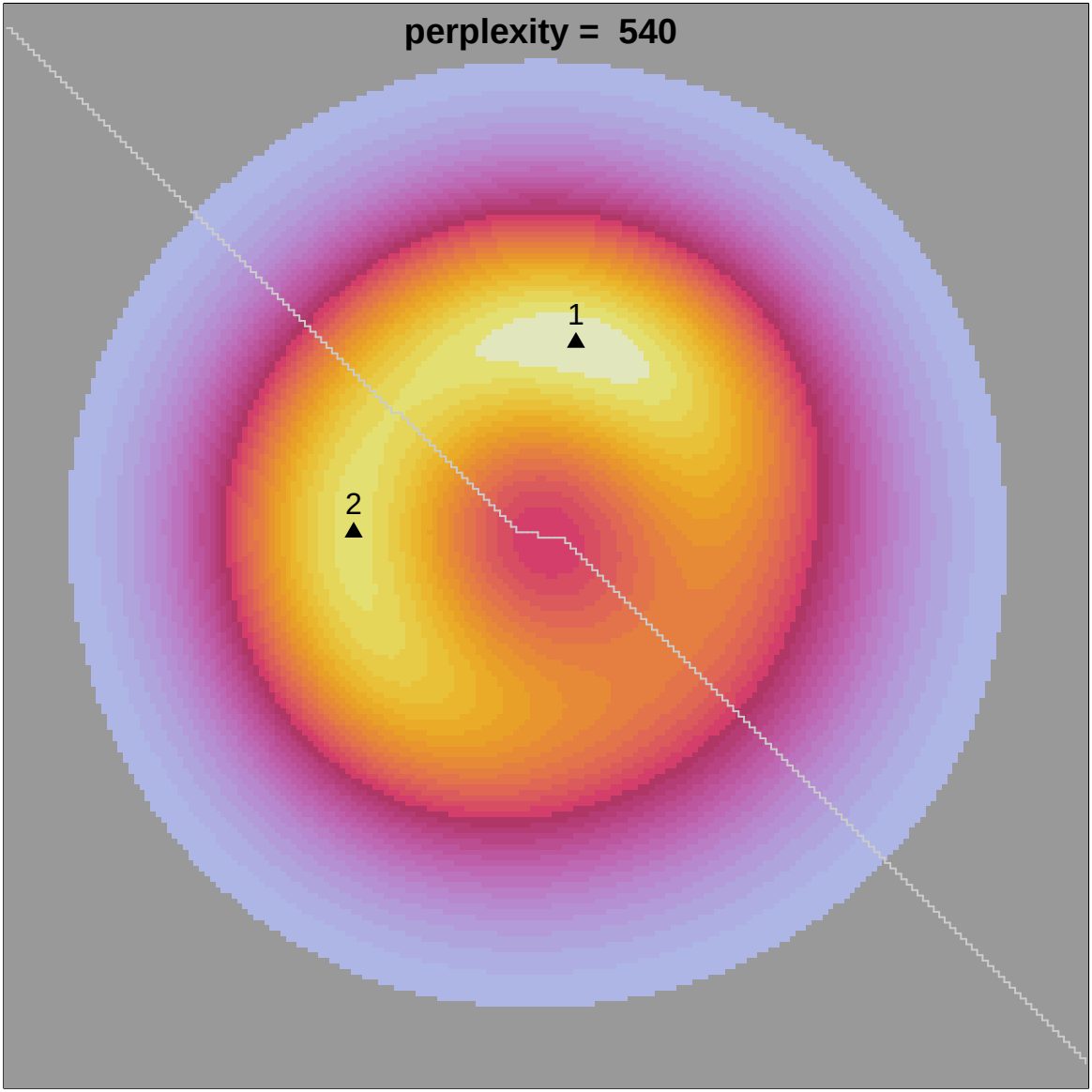

perplexity = 530
1.00
0.98
0.96
0.94
0.92
0.90
Cost
2.5
2.0
1.5
1.0
0.5
0.0
Size
0
20
40
60
80
100
epochs

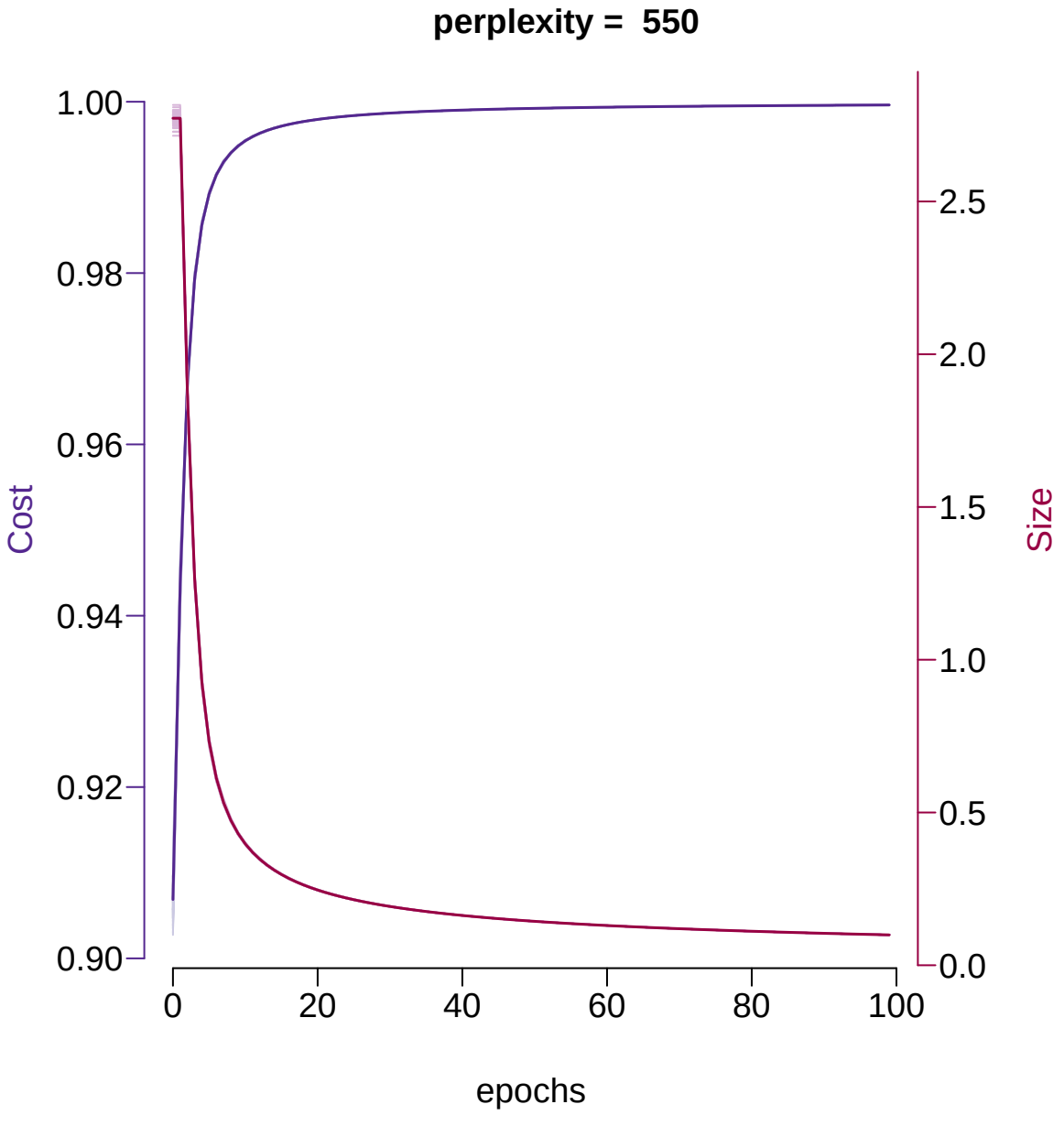

perplexity = 530
Y2
Y1
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
−0.03
−0.02
−0.01
0.00
0.01
0.02
0.03

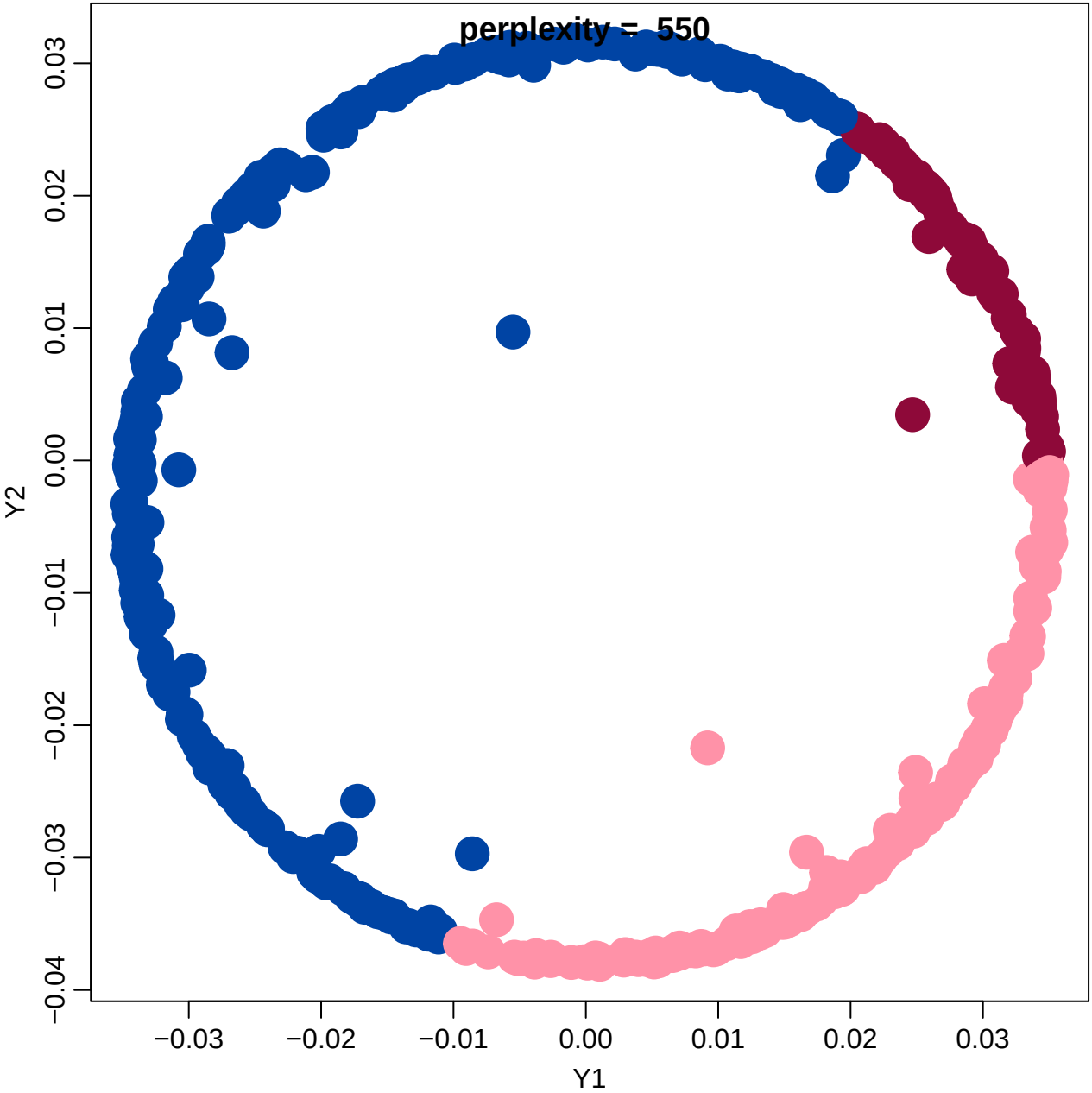

perplexity = 530
1

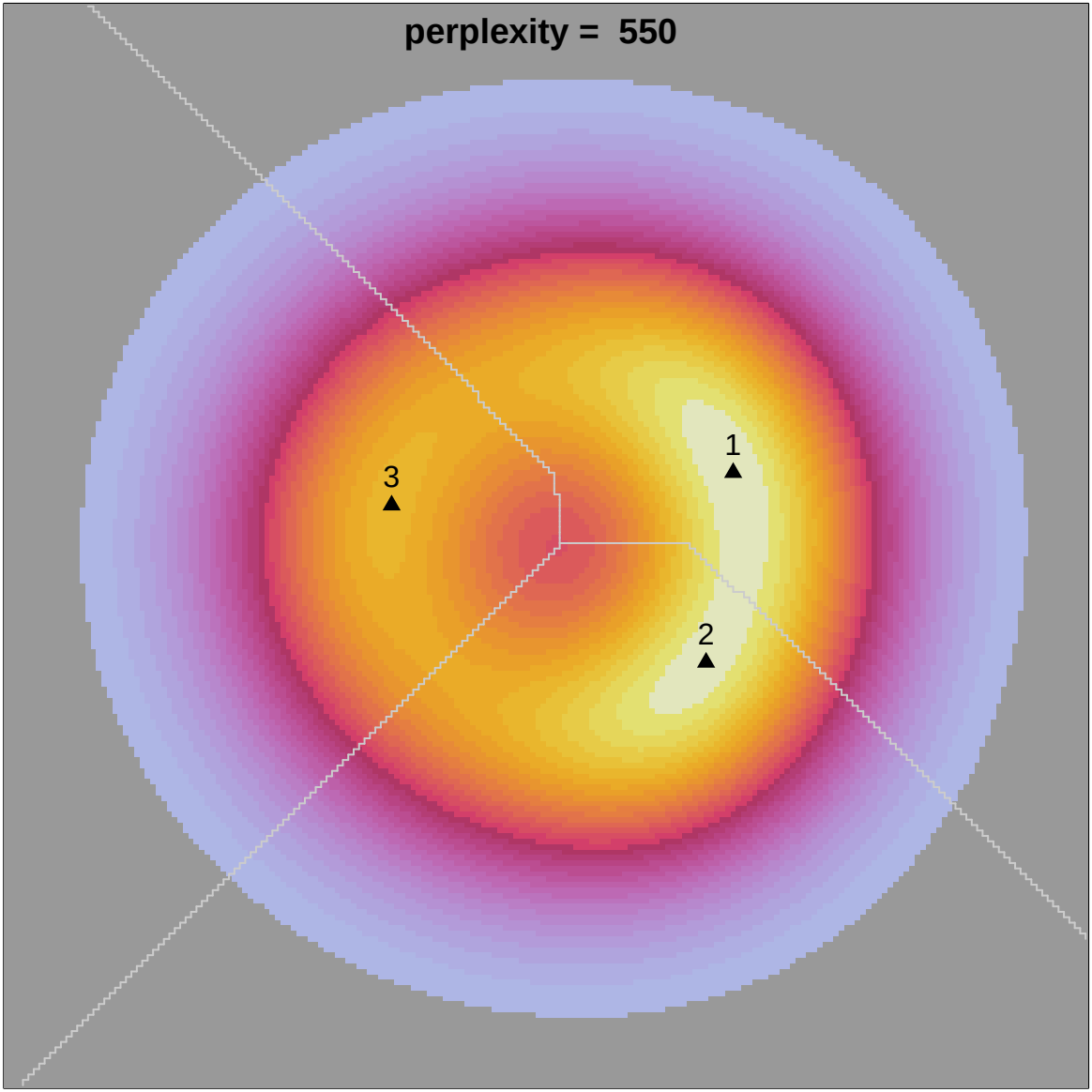

perplexity = 540
1.00
0.98
0.96
0.94
0.92
0.90
Cost
2.5
2.0
1.5
1.0
0.5
0.0
Size
0
20
40
60
80
100
epochs

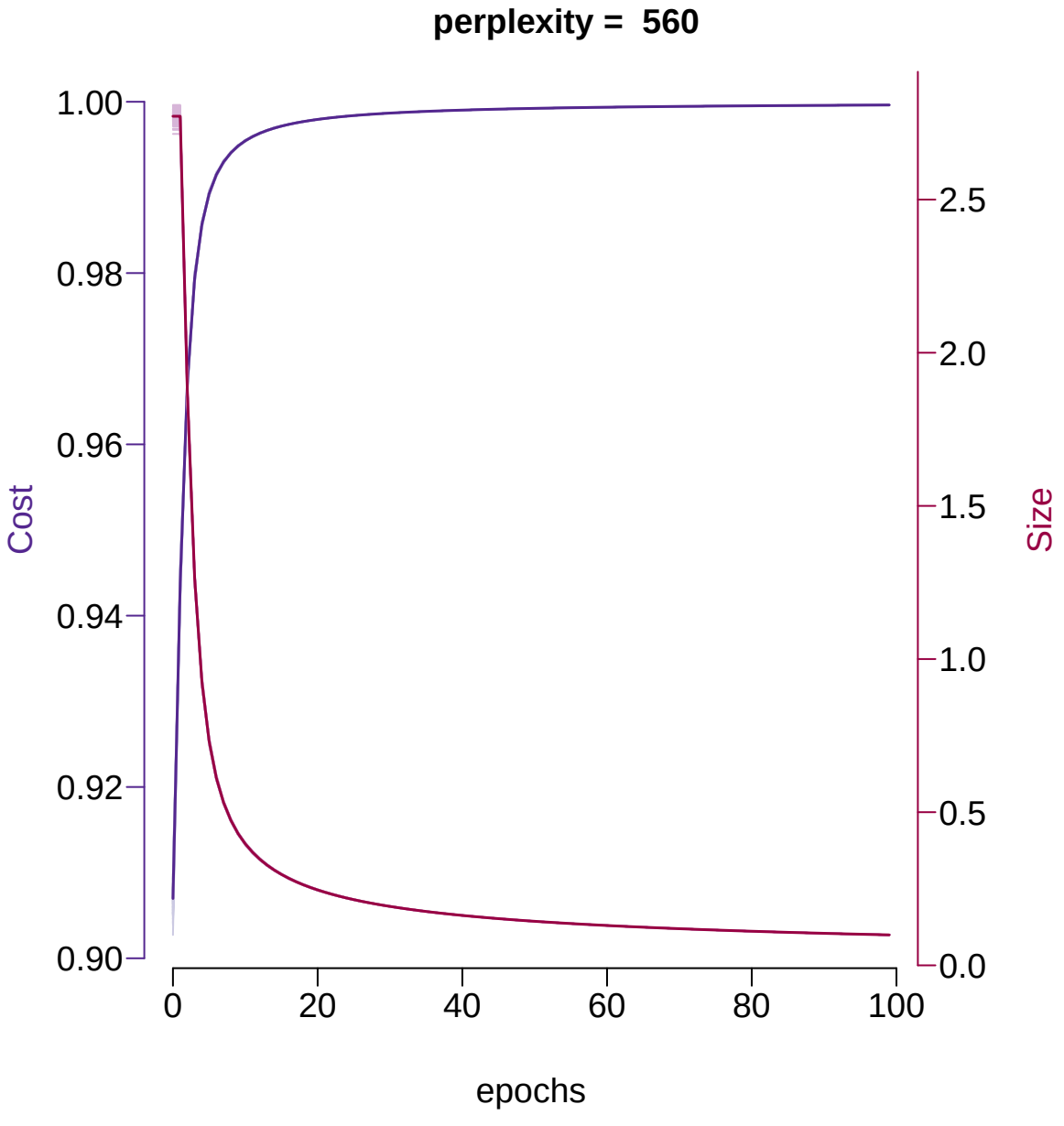

perplexity = 540
Y2
Y1
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
−0.04
−0.03
−0.02
−0.01
0.00
0.01
0.02
0.03

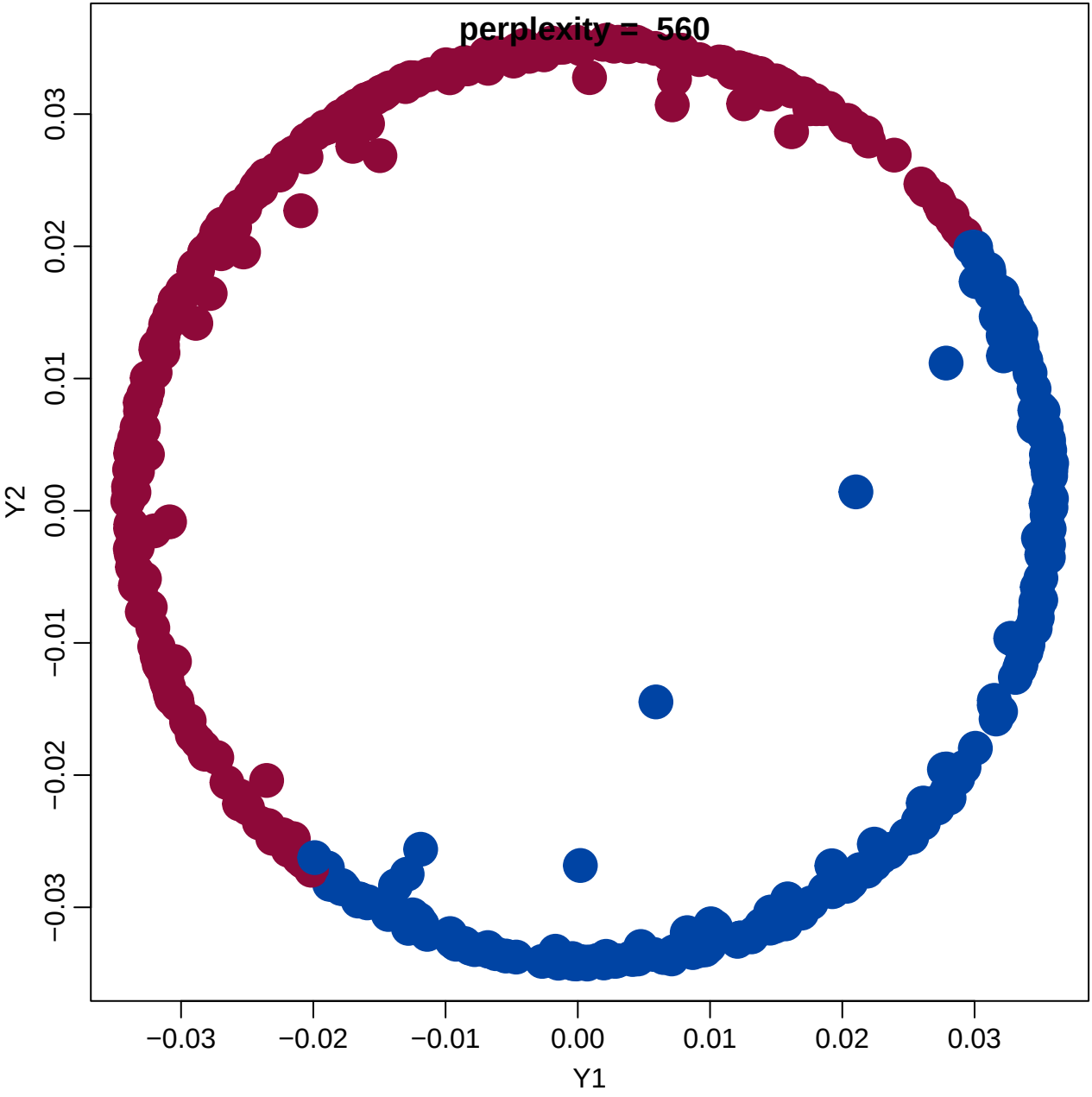

perplexity = 540
1
2

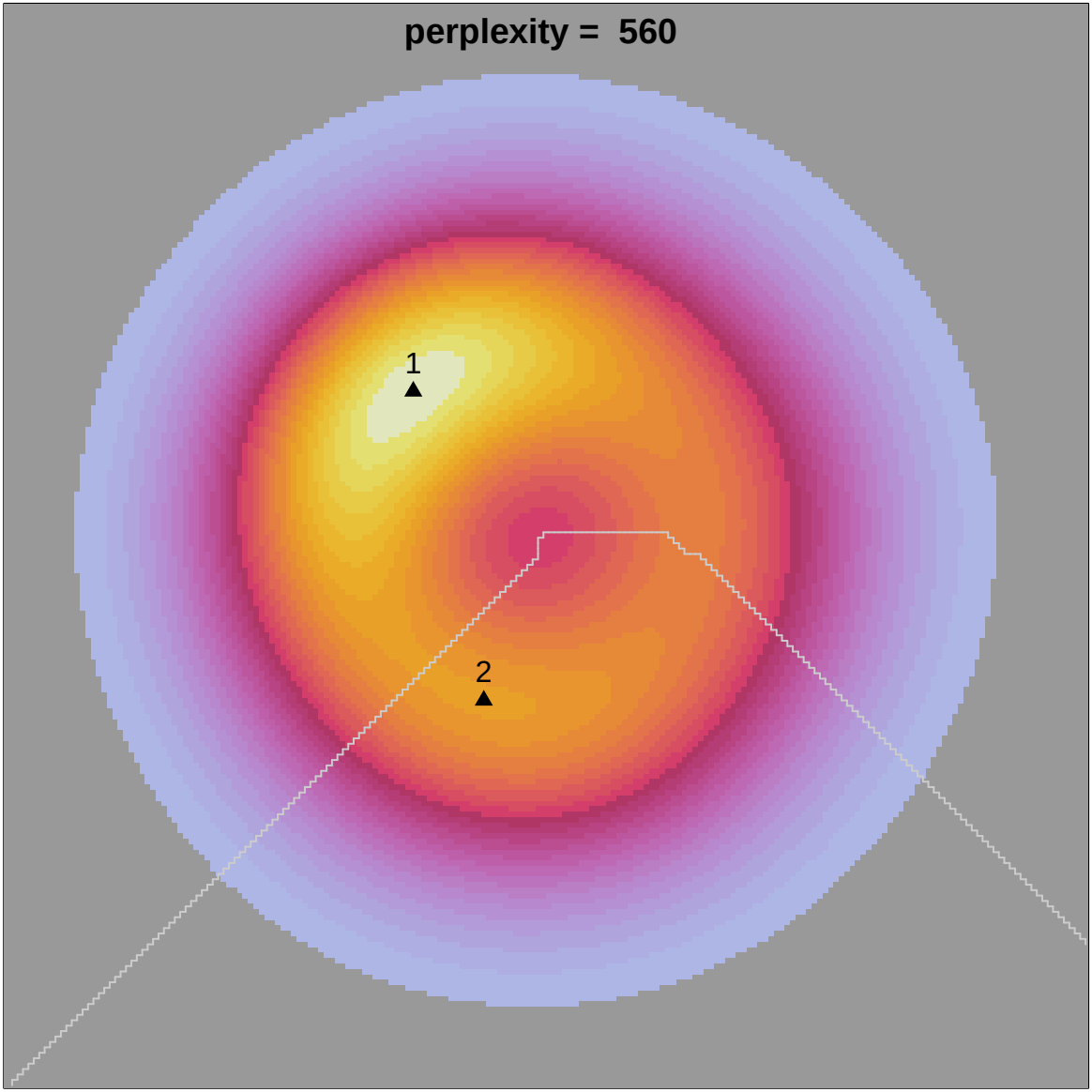

perplexity = 550
1.00
0.98
0.96
0.94
0.92
0.90
Cost
2.5
2.0
1.5
1.0
0.5
0.0
Size
0
20
40
60
80
100
epochs

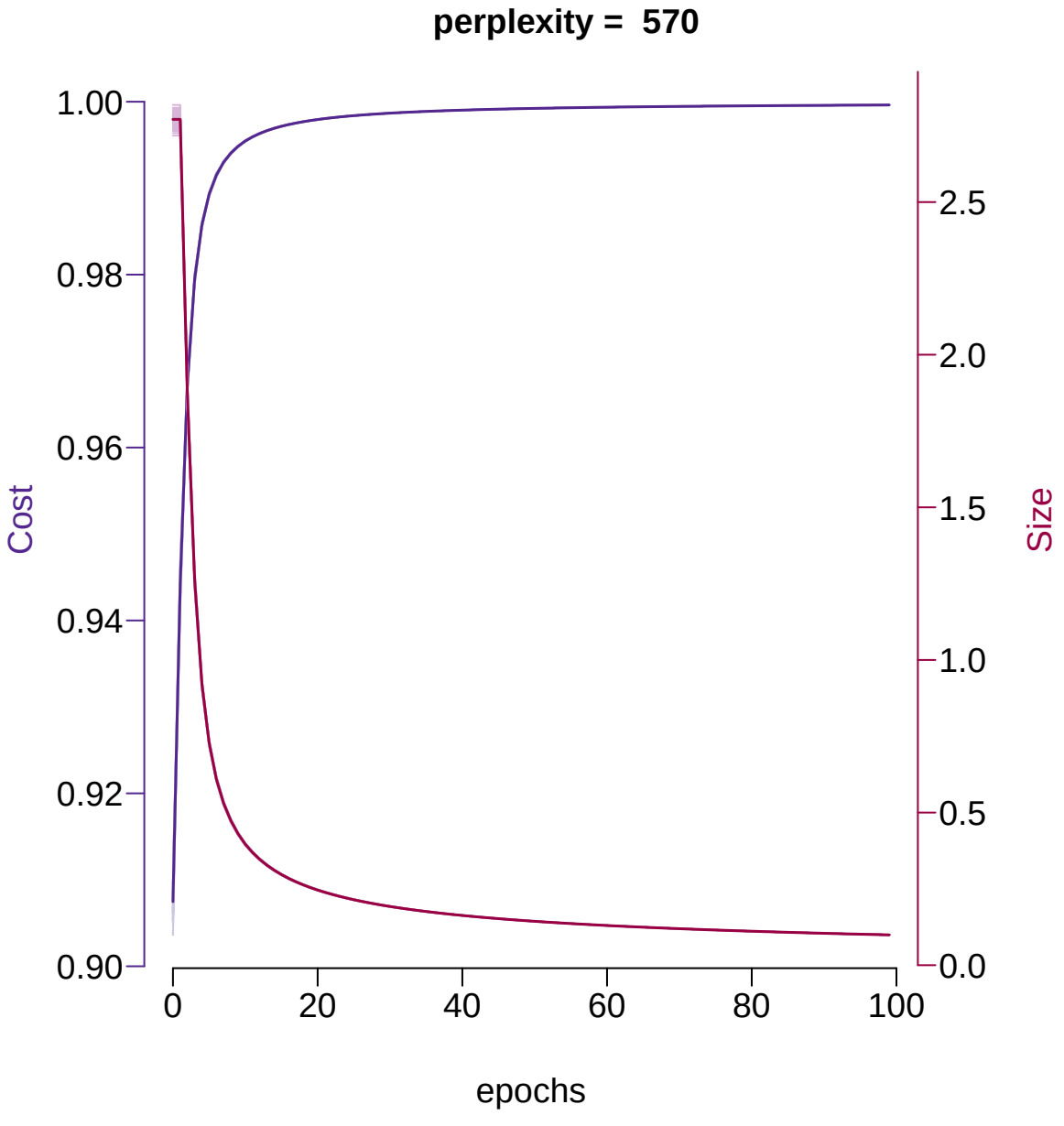

perplexity = 550
Y2
Y1
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
−0.04
−0.03
−0.02
−0.01
0.00
0.01
0.02
0.03

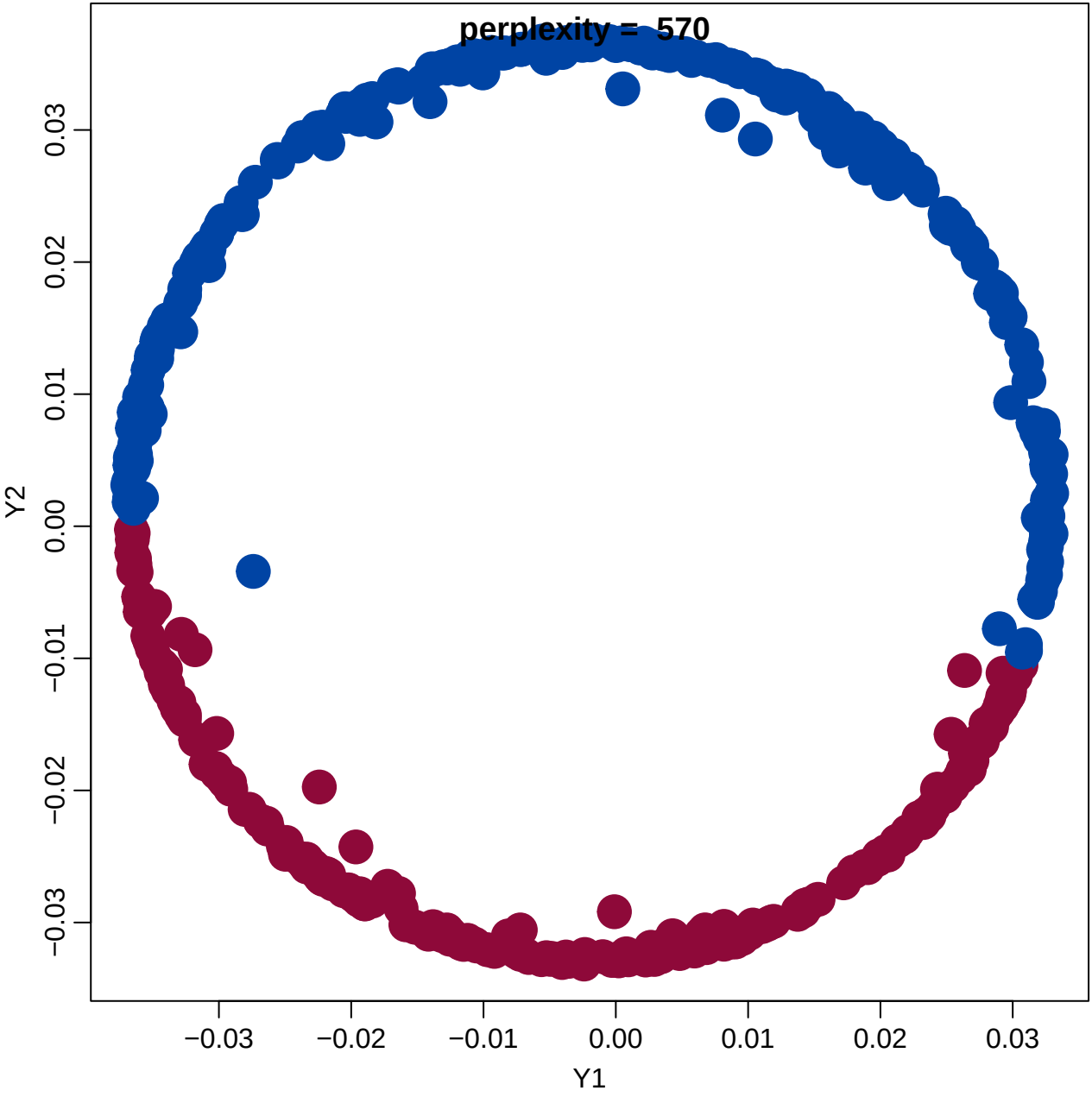

perplexity = 550
1
2
3

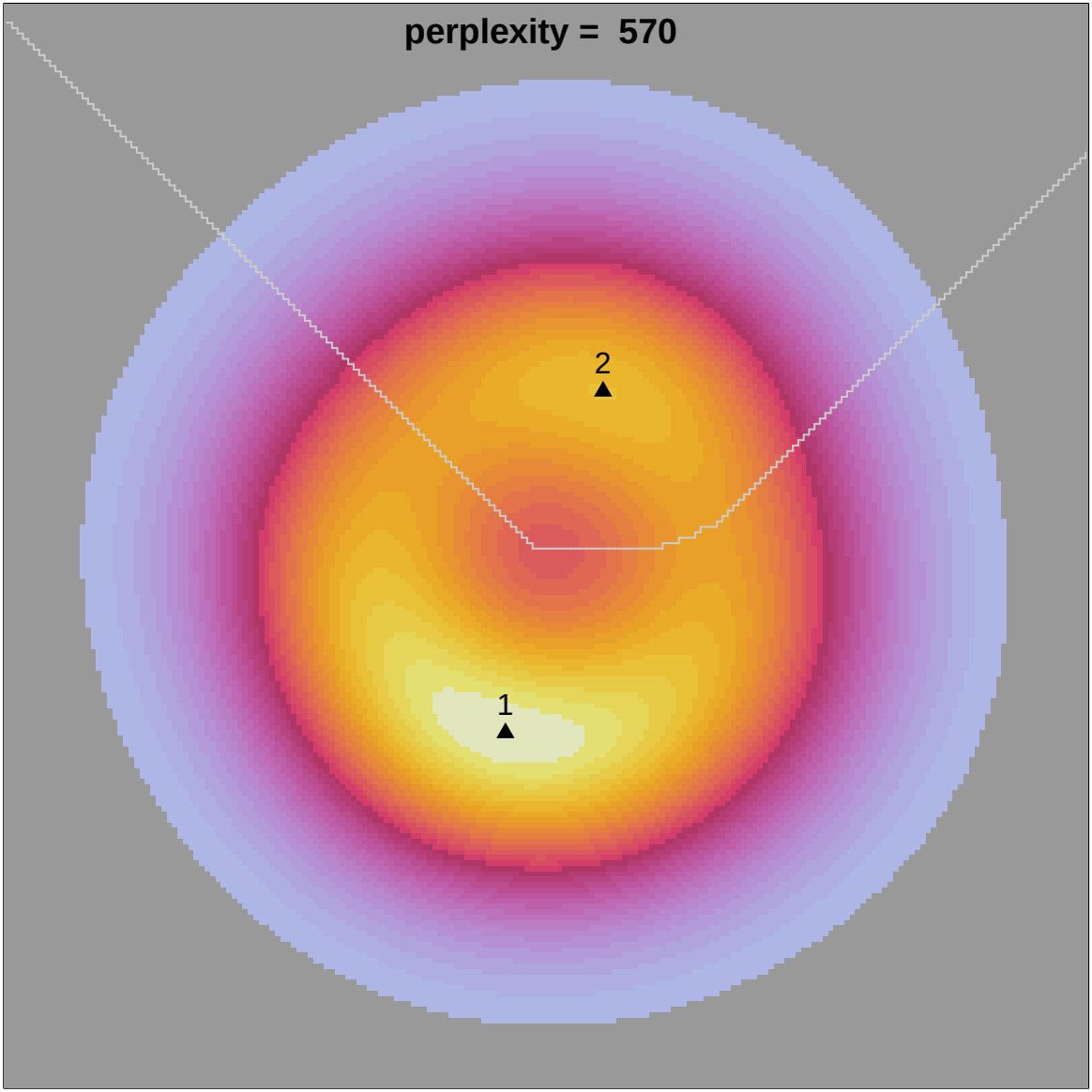

perplexity = 560
1.00
0.98
0.96
0.94
0.92
0.90
Cost
2.5
2.0
1.5
1.0
0.5
0.0
Size
0
20
40
60
80
100
epochs

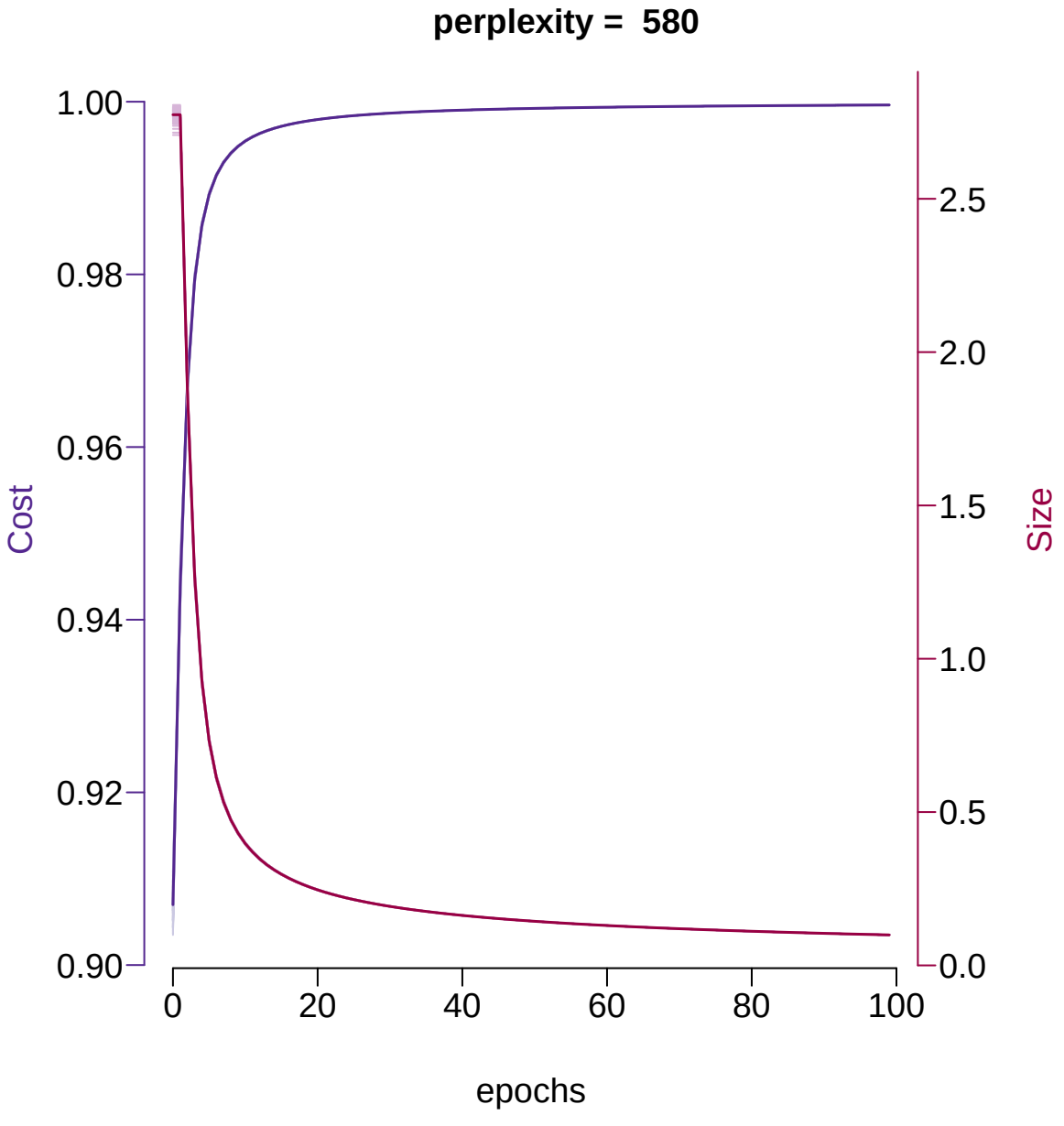

perplexity = 560
Y2
Y1
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
−0.03
−0.02
−0.01
0.00
0.01
0.02
0.03

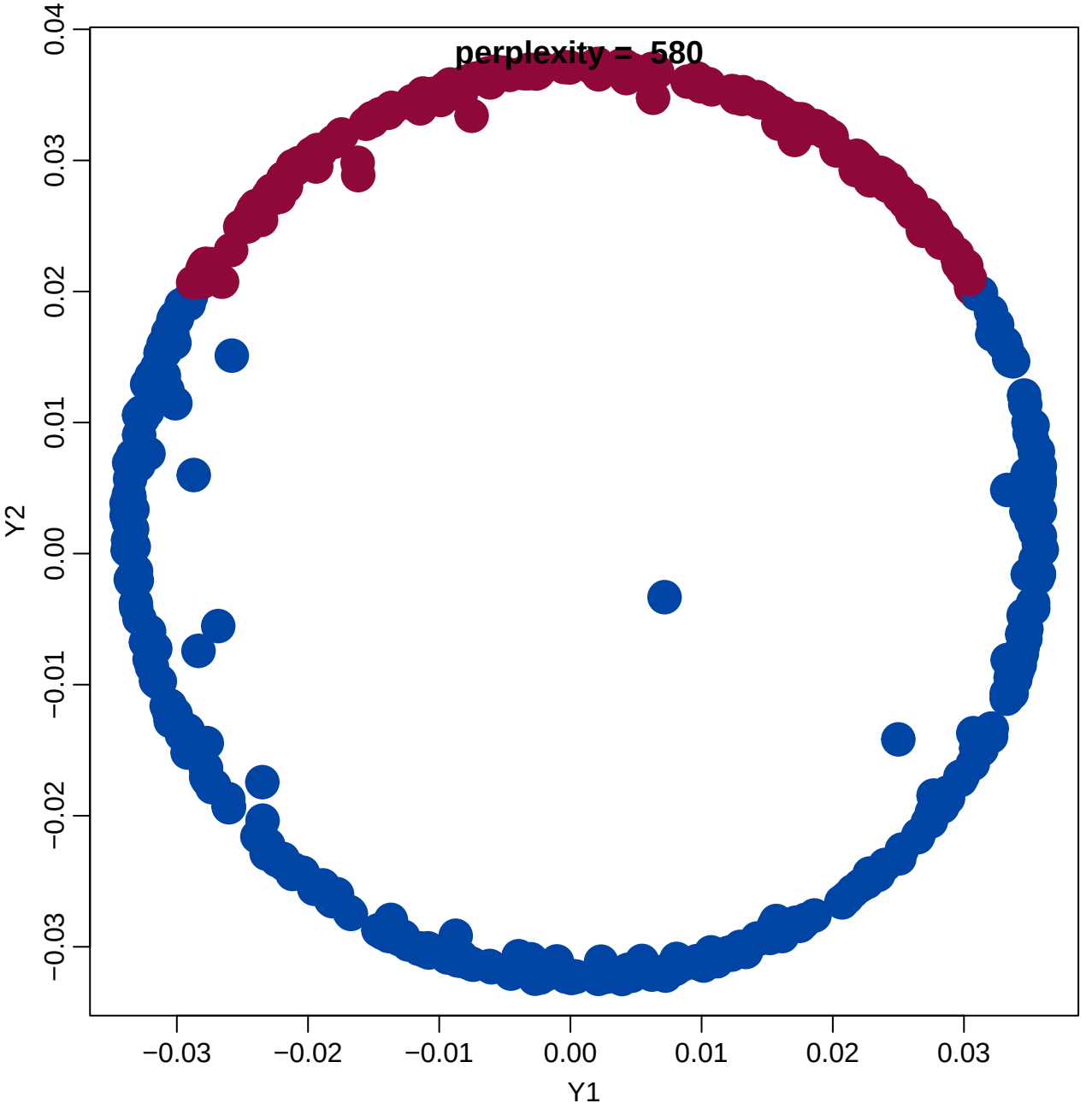

perplexity = 560
1
2

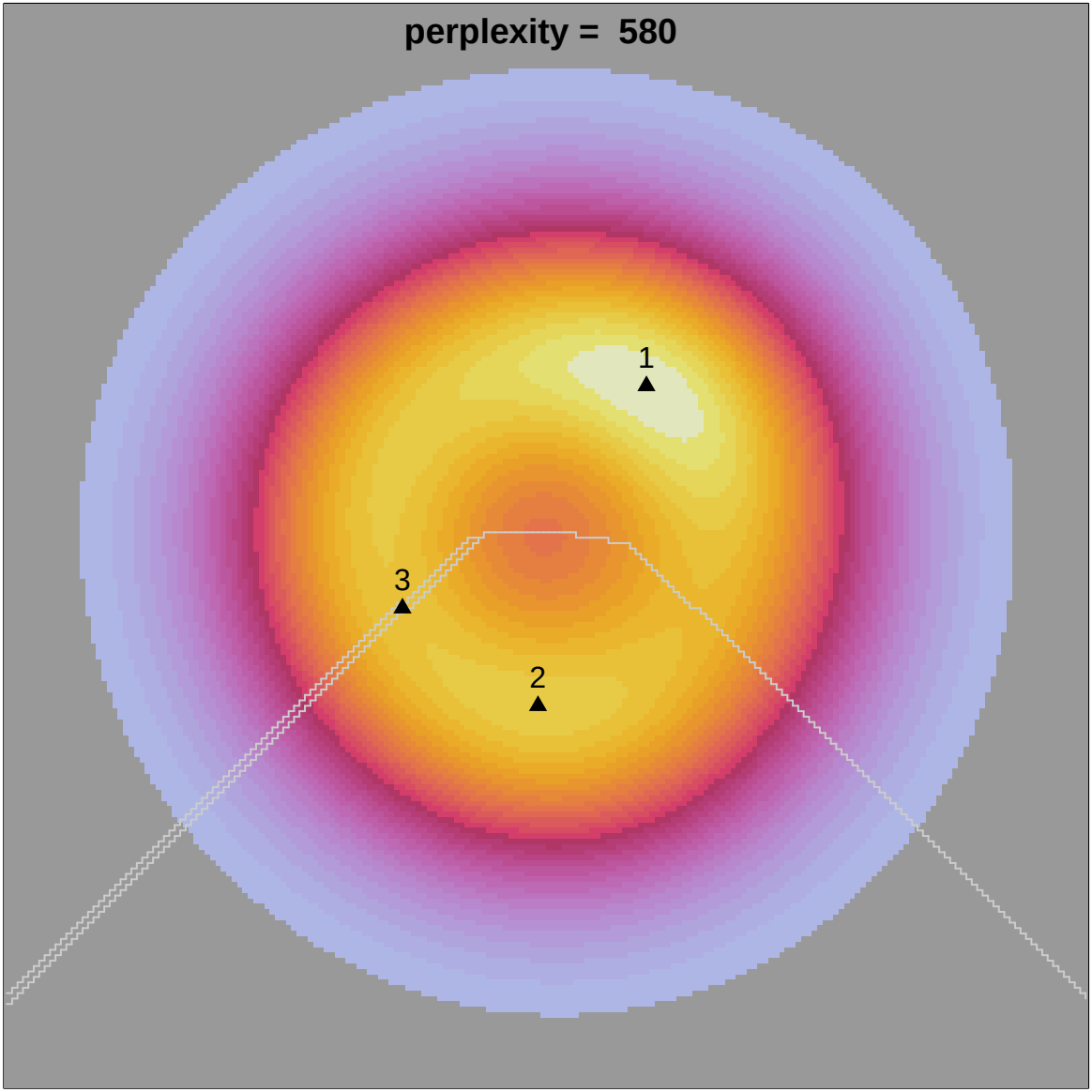

perplexity = 570
1.00
0.98
0.96
0.94
0.92
0.90
Cost
2.5
2.0
1.5
1.0
0.5
0.0
Size
0
20
40
60
80
100
epochs

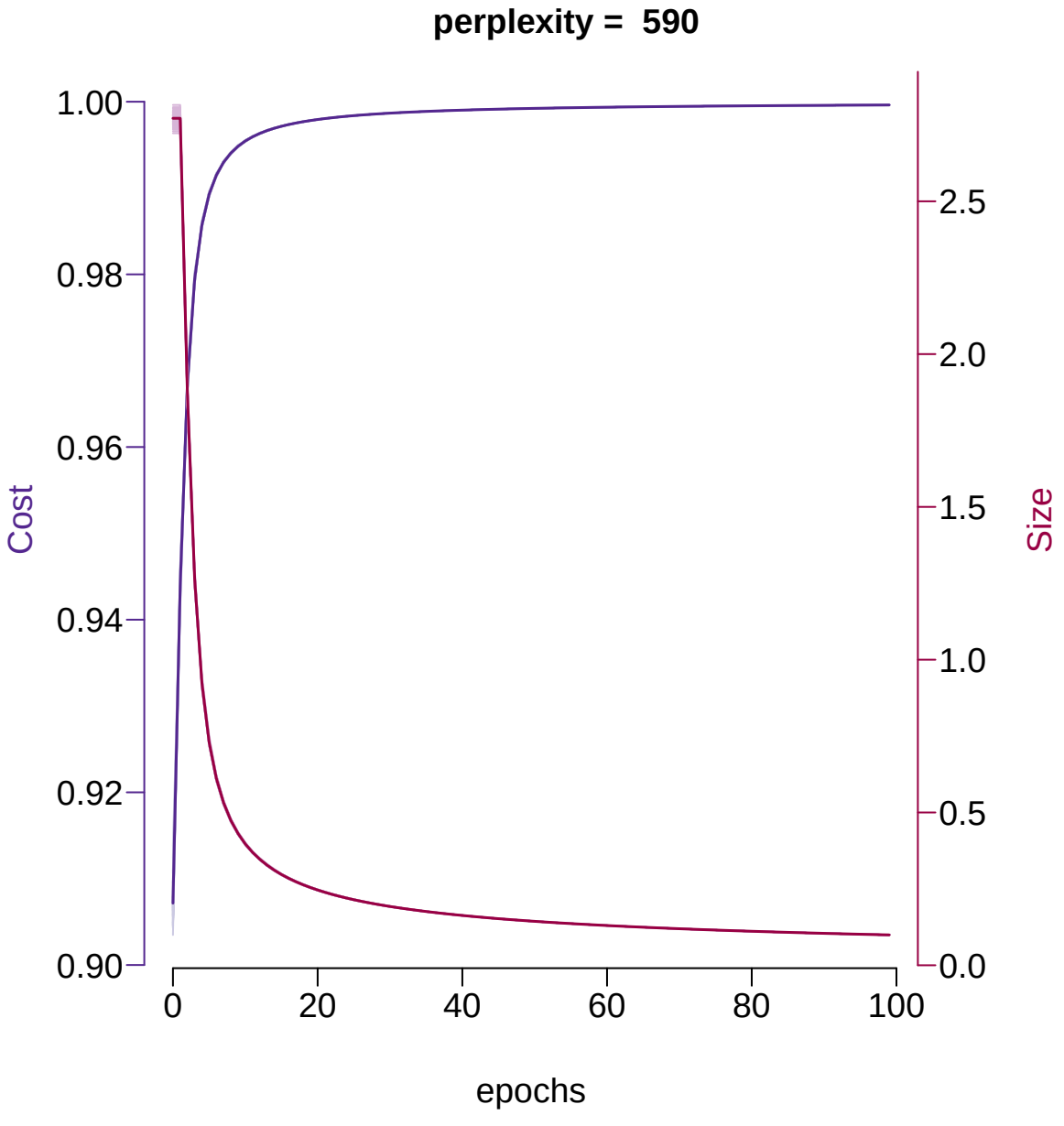

perplexity = 570
Y2
Y1
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
−0.03
−0.02
−0.01
0.00
0.01
0.02
0.03

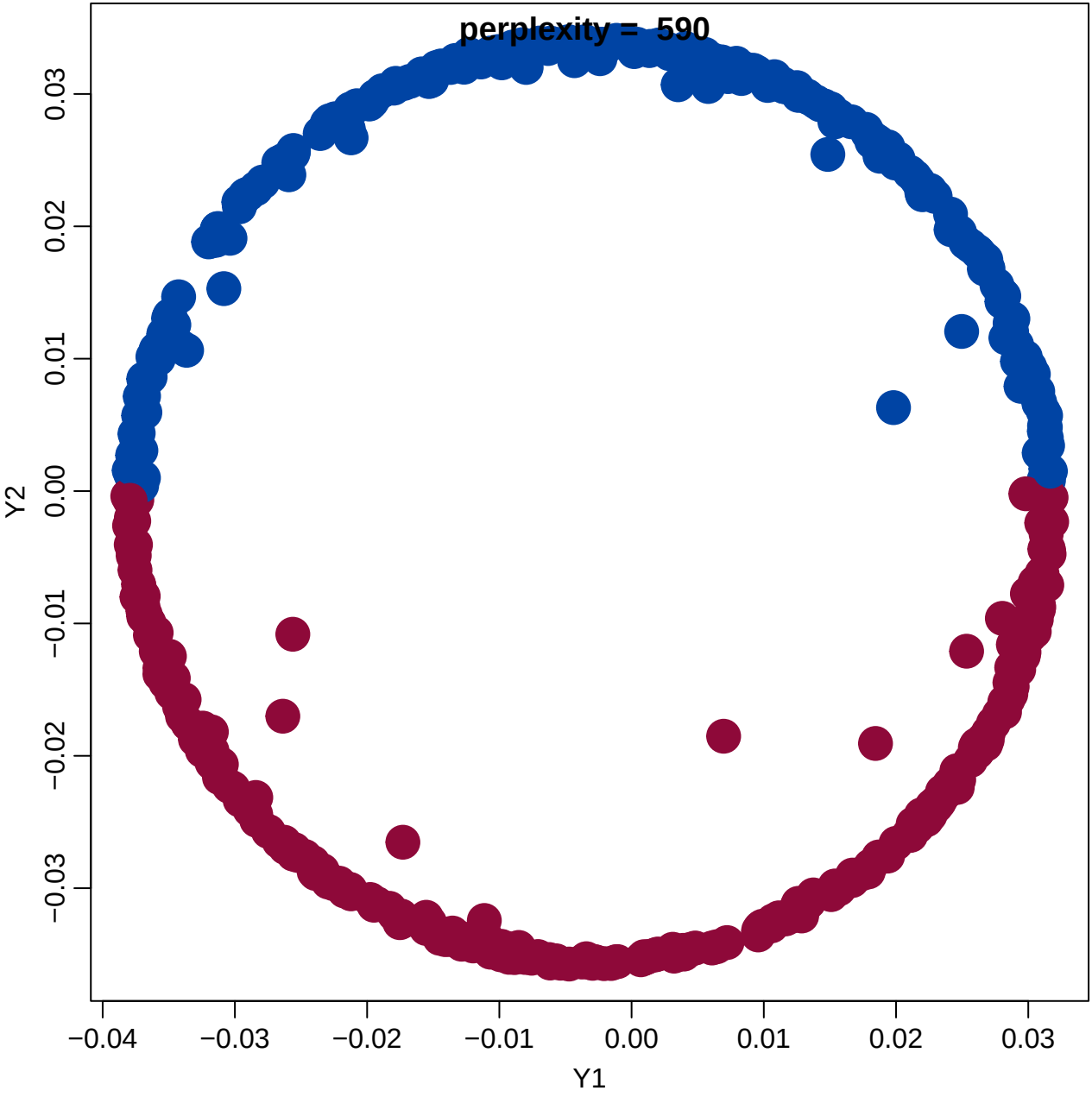

perplexity = 570
2
1

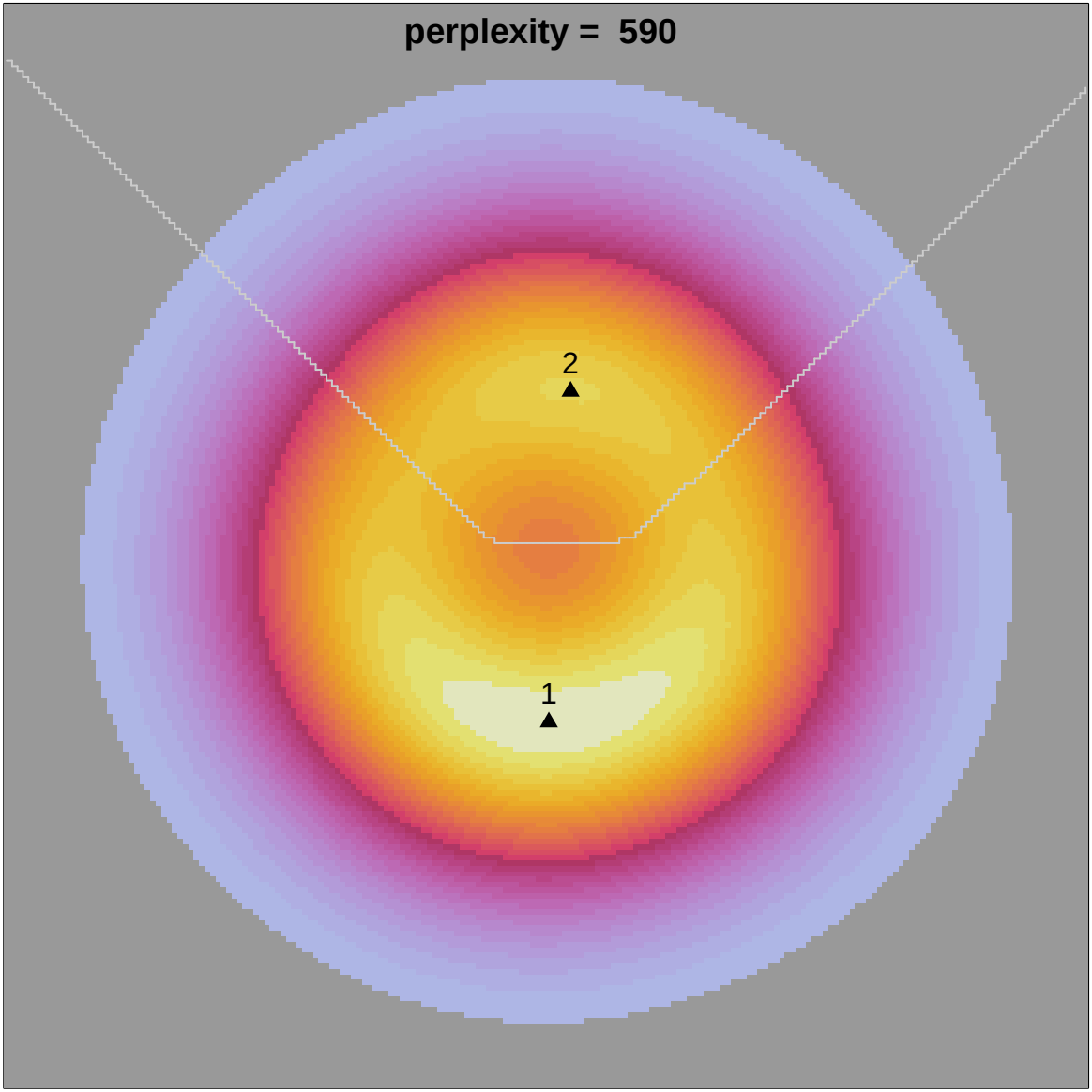

perplexity = 580
1.00
0.98
0.96
0.94
0.92
0.90
Cost
2.5
2.0
1.5
1.0
0.5
0.0
Size
0
20
40
60
80
100
epochs

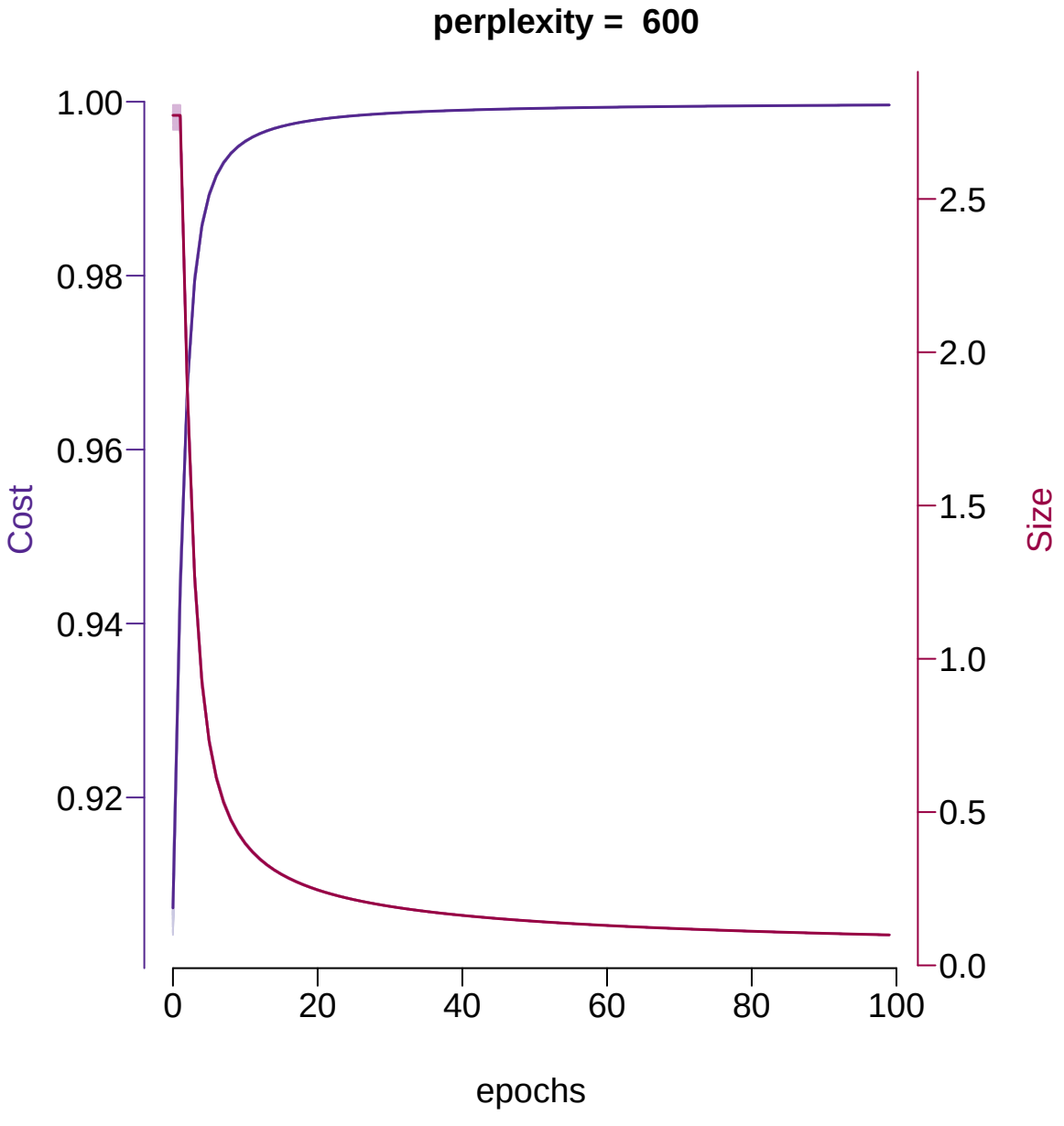

perplexity = 580
Y2
Y1
0.04
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
−0.03
−0.02
−0.01
0.00
0.01
0.02
0.03

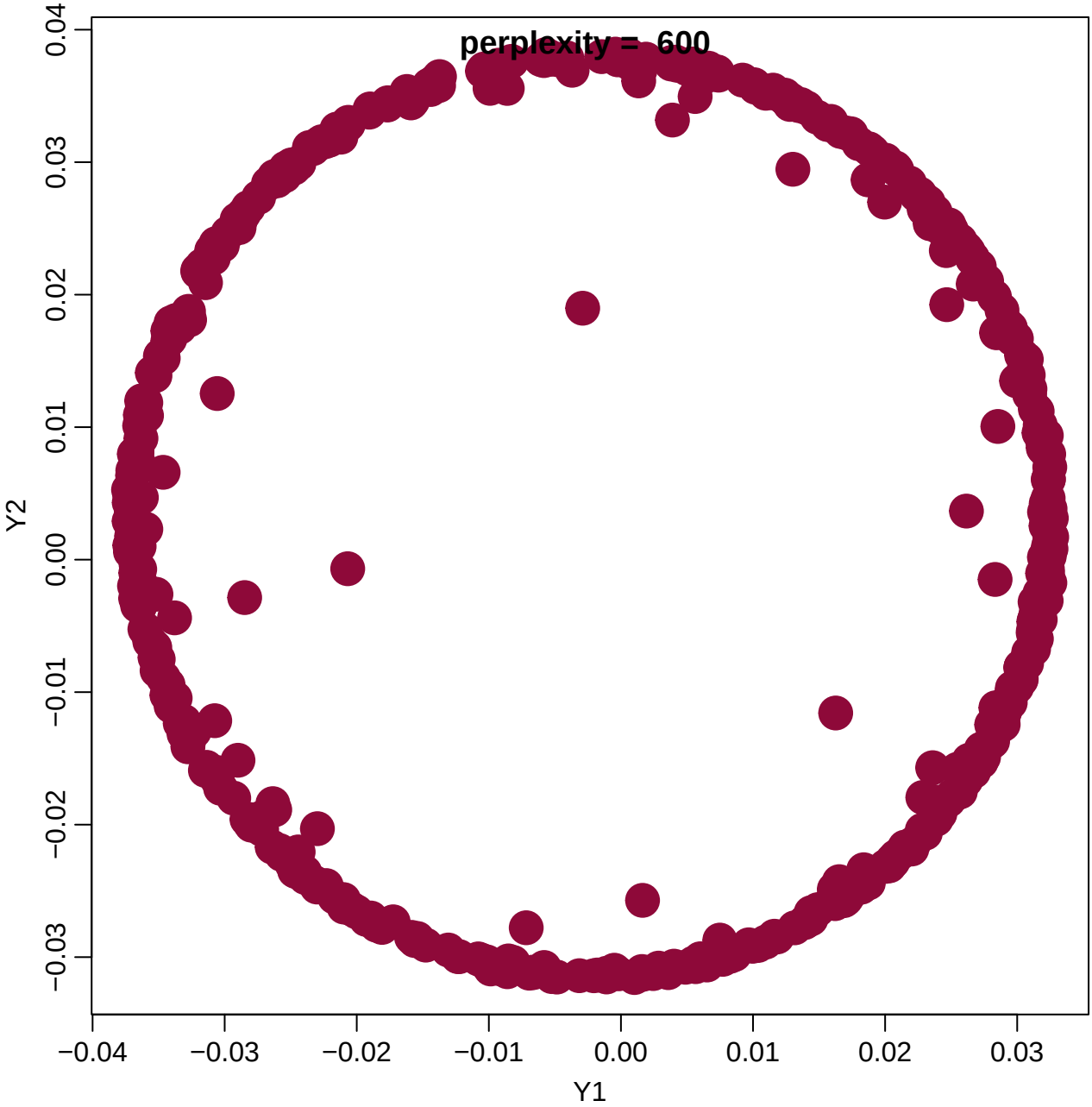

perplexity = 580
1
2
3

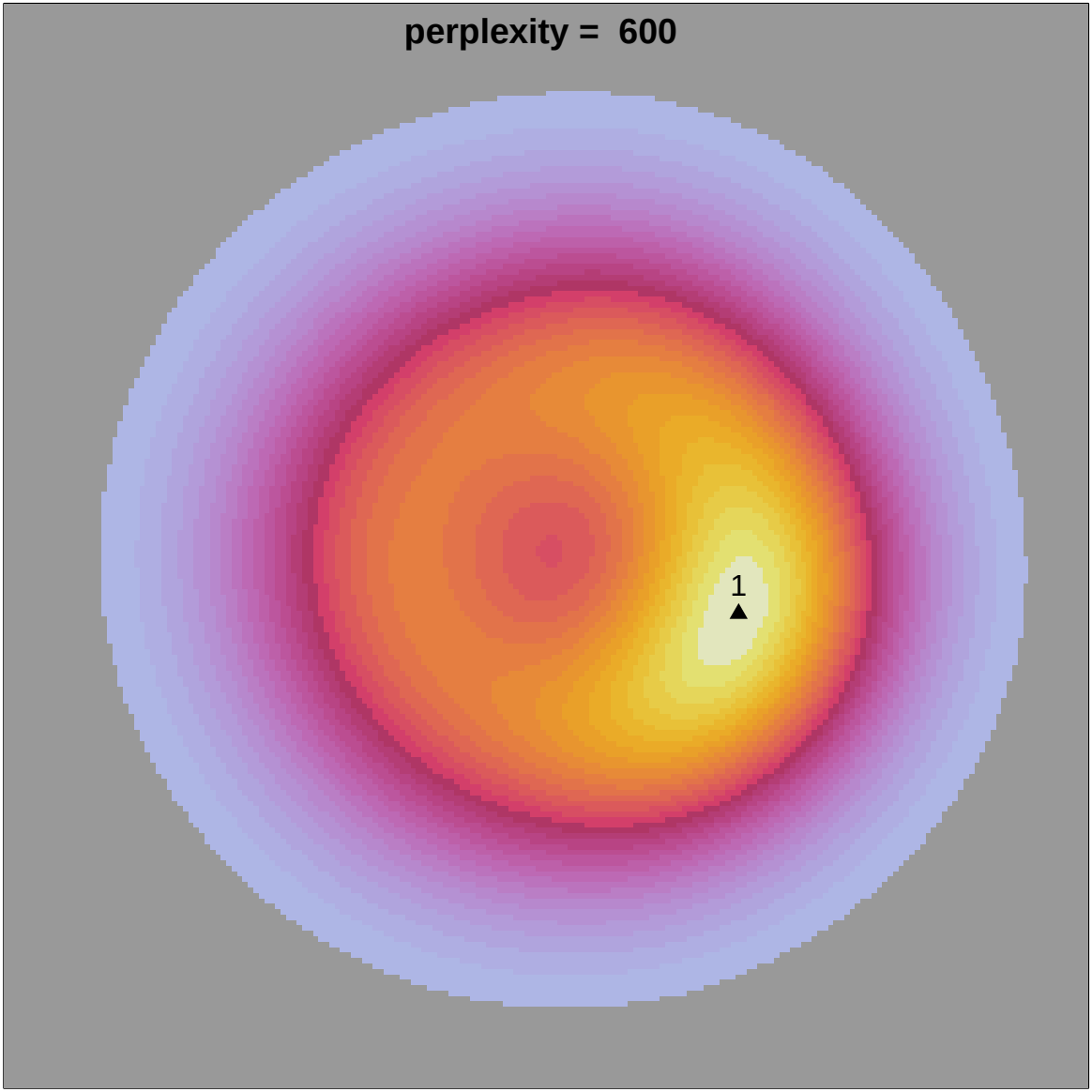

perplexity = 590
1.00
0.98
0.96
0.94
0.92
0.90
Cost
2.5
2.0
1.5
1.0
0.5
0.0
Size
0
20
40
60
80
100
epochs


perplexity = 590
Y2
Y1
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
−0.04
−0.03
−0.02
−0.01
0.00
0.01
0.02
0.03


perplexity = 590
2
1


perplexity = 600
1.00
0.98
0.96
0.94
0.92
Cost
2.5
2.0
1.5
1.0
0.5
0.0
Size
0
20
40
60
80
100
epochs


perplexity = 600
Y2
Y1
0.04
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
−0.04
−0.03
−0.02
−0.01
0.00
0.01
0.02
0.03


perplexity = 600
1