Support Document for Experiment 1

Stage 3:Synthesis analysis of separability and learning rate

Combining thinking the influence of separability and learning rate, when separability is bigger, which means data are more distributed, the nice learning rate is bigger.

I do experiment with varying learning rate in the range of [0.01,3] and number of data point, which represent separability, in the range of [100,1000]. The number of iteration is listed in Table 2. Because of the error brought by random data, the changing tendency is not quite obvious, but we can still see that when separability get smaller (number of data points increase), the learning rate of the least range of iteration get smaller, which match my assumption.

m\n	0.01	0.04	0.07	0.1	0.4	0.7	1	1.5	2	2.5	3
100	61.4	69.8	38.02	42.12	37.3	36.78	33.66	29.63	31.54	28.79	28.76
200	84.78	70.4	72.24	71.37	75.7	71.44	65.47	70.96	70.01	57.06	72.5
300	124.41	61.66	75.46	68.78	68.39	72.57	56.78	52.11	67.72	69.33	62.67
400	176.19	127.74	134.38	142.43	146.89	144.6	122.48	121.97	129.99	108.96	118.05
500	224.36	206.19	160.94	153.25	156.09	139.92	127.05	162.91	196.9	143.22	163.55
600	224.36	206.19	160.94	153.25	156.09	139.92	127.05	162.91	196.9	143.22	163.55
800	352.84	343.98	418.57	278.92	326.36	257.98	368.53	334.54	307.68	284.1	313.16
1000	303.62	289.34	280.95	282.25	256.89	205.95	274.21	184.89	338.09	235.31	264.9

Table 1 training parameter

For another problem mentioned in last paper:

Why the training time is less when learning rate is 0.01 when number of iteration is high? I did more experiment and found that in most of the time, the tendency of iteration is the same as the tendency of learning rate, so that should be an error.