

Output:

k = 3, gamma = 0.00

intra-cluster variance = 7.26e+07

centers =

Center1	Center2	Center3
59.3898	100.6459	141.5072
60.4516	106.0896	117.9699
36.7264	80.3168	206.3199
0	0	0
0	0	0



k = 3, gamma = 0.20

intra-cluster variance = 1.99e+08

centers =

Center1	Center2	Center3
78.4967	80.7338	137.7399
80.5963	87.4997	118.6896
49.9512	59.7337	189.4200
53.3976	162.7566	123.8301
66.8637	194.6094	176.2907



k = 3, gamma = 0.50

intra-cluster variance = 3.81e+08

centers =

Center1	Center2	Center3
118.0897	108.8453	111.3502
103.3630	100.3070	107.5439
153.9392	119.3006	123.5182
316.4397	259.4618	326.6168
391.2046	148.1497	653.7691



k = 5, gamma = 0.00

intra-cluster variance = 5.99e+07

centers =

Center1	Center2	Center3	Center4	Center5
174.5180	151.8989	67.8860	111.4877	132.0335
168.1562	125.0388	69.8405	119.7999	104.4159
221.3916	247.4979	45.4489	88.3609	198.9429

0	0	0	0	0
0	0	0	0	0



k = 5, gamma = 0.20

intra-cluster variance = 1.66e+08

centers =

Center1	Center2	Center3	Center4	Center5
74.9356	80.9186	138.9643	138.6632	136.9392
77.9777	87.7597	136.0696	115.8237	115.8471
48.3182	58.4944	136.5568	197.6109	197.2371
76.5122	152.2872	102.5669	194.8894	106.2825
70.9007	245.0822	93.0651	64.8407	215.5749



k = 5, gamma = 0.50

intra-cluster variance = 3.18e+08

centers =

Center1	Center2	Center3	Center4	Center5
111.2670	102.4147	144.8414	107.0779	121.7268
100.2315	100.5224	107.8325	102.6129	108.6940

138.3051	97.3029	223.3703	109.4313	158.4691
466.5242	80.4528	569.2745	266.4643	342.8152
340.8080	479.3574	63.3363	117.7213	614.3944



Questions:

1. which k gives lower intra-cluster variance?

比起 $k=3$, $k=5$ 能給予更低的 intra-cluster variance。

2. what happens as γ goes up, and why?

當 γ 增加時

1. intra-cluster variance 將越來越大
2. 圖像上有越來越明顯的區塊化，甚至到了 $\gamma = 0.5$ 的時候根本就看不出原圖了，只剩圖片的簡單分割

γ 代表的是 xy 的權重。

我先解釋第 2 點

我們考慮 $k = 3$, $\gamma = 0.5$ 的例子

考慮 centers 中的 RGB，計算過後發現他們 3 個之間的距離分別為：8.7427, 31.4378, 35.9810，取 max 大約為 36

考慮 centers 中的 xy （有乘上 γ ），計算過後發現他們 3 個之間的距離分別為：510.0596, 262.7617, 249.6441，取 min 大約為 250

也就是說， xy 的權重在這個例子中真的太大了，以至於你的分群根本就跟你的 RGB 值無關。

反觀

考慮 $k = 3$, $\gamma = 0.2$ 的例子

考慮 centers 中的 RGB，計算過後發現他們 3 個之間的距離分別為：12.1803, 145.0553, 156.2447

考慮 centers 中的 xy （有乘上 γ ），計算過後發現他們 3 個之間的距離分別為：43.0215, 130.1346, 168.1617

這個例子中很明顯 RGB 跟 xy 的權重是差不多的。

再來解釋第 1 點

當 γ 從 0.0 變到 0.2 時，就是增加考慮位置的因素，也就擴增了維度，intra-cluster variance 的概念本來就是距離的概念，而你在 3D 上的距離本來就會比 5D 上的距離來的小，就好像空間中兩點投影到平面上一樣，所以 intra-cluster variance 增加是合理的。

而當 γ 從 0.2 變到 0.5 時，就等於將 xy 的質乘上 2.5，你算出來的距離放大也是合理的，所以 intra-cluster variance 增加是合理的。