

پروژه شماره ۲: پیاده سازی یک الگوریتم ساده با هدف خلاصه سازی

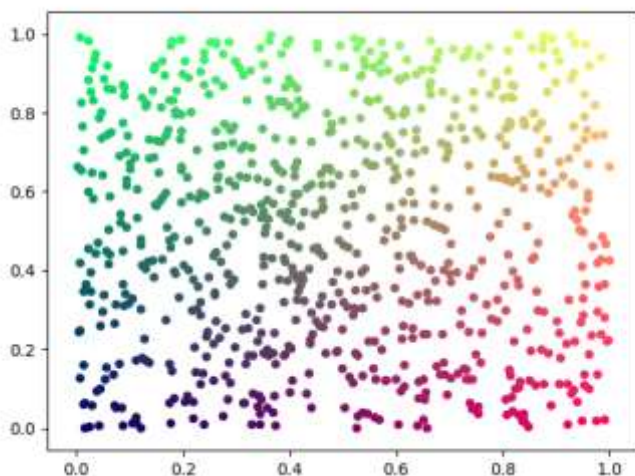
داده ها با بهره بردن از الگوریتم خوشه بندی

مراحل پیاده سازی الگوریتم: هدف پروژه پیاده سازی (شی گرا) قابلیت های زیر بروی یک مجموع ای از نقاط داده شده می باشد.

- قابلیت خواندن فایل CSV نقاط (در فرمت مشخص شده در شکل زیر). یک فایل CSV به عنوان مثال به همراه پروژه داده شده است. این فایل شامل اطلاعات RGB مربوط به هزار نقطه میباشد.

pointid	red	green	blue
0	0.31378129356345197	0.7173301854886083	0.4
1	0.9311206368223862	0.766477359434707	0.4
2	0.4782768105355699	0.316651044731469	0.4
3	0.47110090088727874	0.9257300716463256	0.4
4	0.4510074466580104	0.9657846428615328	0.4
5	0.9570391706246054	0.7581552588569668	0.4
6	0.7782570875902526	0.6396524171944227	0.4
7	0.3579831714596484	0.8857094838468258	0.4

- قابلیت ترسیم نقاط رنگی (با RGB داده شده). به منظور ترسیم ستون اول (red) به عنوان مختصات x و ستون دوم (green) به عنوان مختصات y نقاط در نظر گرفته شود.



- خلاصه سازی اطلاعات به k نقطه. (یافتن k نقطه که تمامی نقاط را نمایندگی کنند).
برای این منظور، با در نظر گرفتن فاصله اقلیدسی بین نقاط به عنوان معیار شباهت نقاط رنگی، با استفاده از الگوریتم Lloyd نقاط را به k خوشه دسته بندی کنید. توجه کنید که centroid هر خوشه میانگین RGB نقاط هر خوشه می باشد و عملاً می تواند به عنوان یک نماینده (و یک خلاصه) مناسب برای اعضای آن خوشه باشد.
- امکان ترسیم نقاط در کنار نماینده های آنها. نماینده ها با سایزی بزرگتر نمایش داده شود.
(در شکل زیر دو خروجی الگوریتم برای $k=4$ و $k=8$ نمایش داده شده است)

