



## درس یادگیری ماشین

مدرس: دکتر سامان هراتی زاده

الگوریتم های K-means و HR<sup>۱</sup>

تاریخ تحویل حضوری: ۱۳۹۵/۹/۲۱

نیم سال اول ۹۶-۱۳۹۵

تمرین شماره ی چهار

مهلت ارسال تمرین: ۱۳۹۵/۹/۲۰

### ۱. قطعه بندی عکس<sup>۲</sup>

قطعه بندی یک تصویر بر تفکیک و جدا سازی تصویر به نواحی، به طوری که پیکسل های هر ناحیه دارای یک ویژگی خاص (که می تواند متعلق به یک شیء باشد) مشترک می باشند، دلالت دارد. اساسی ترین ویژگی های مورد استفاده در قطعه بندی یک تصویر تک رنگ<sup>۳</sup>، میزان درخشندگی تصویر و در قطعه بندی یک تصویر رنگی مولفه های رنگی اش می باشند. قطعه بندی تصویر در زمینه هایی مختلفی از قبیل بینایی ماشین<sup>۴</sup> و پردازش تصویر<sup>۵</sup> مورد استفاده قرار می گیرد. در حوزه ی بینایی ماشین از الگوریتم K-means معمولاً برای قطعه بندی تصاویر استفاده می کنند. همان گونه که در شکل ۱ می بینید، هر پیکسل در عکس های رنگی با توجه به رنگ حاصل از بردار سه بعدی RGB آن مشخص می شود. (نا گفته پیداست که هر عکس مجموعه ای از بردارهای ویژگی سه بعدی است). در مورد عکس های تک رنگ، روال یکسان است با این تفاوت که هر پیکسل با یک بردار ۱ بعدی نشان داده می شود. هدف از این تمرین، نوشتن برنامه ای برای ساختن یک مدل به منظور قطعه بندی عکس ها با استفاده از الگوریتم K-means است. (۱۵۰ نمره)

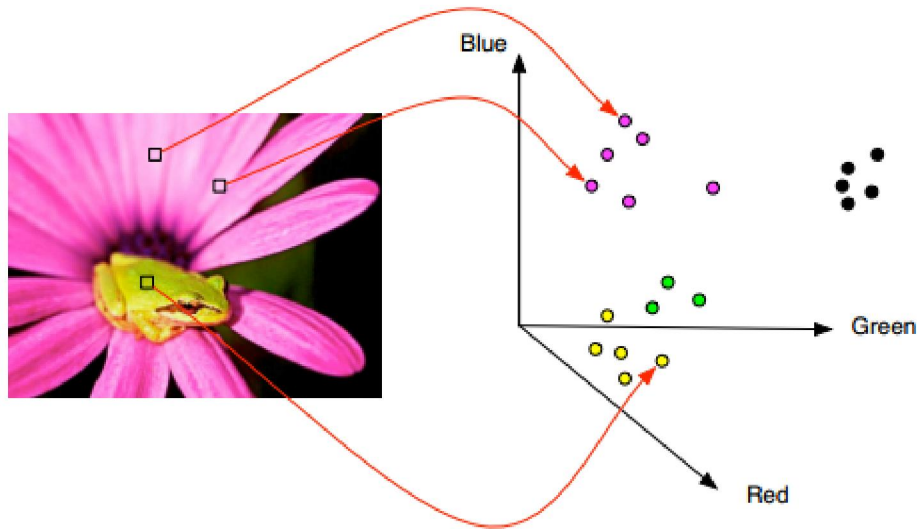
<sup>۱</sup> Hierarchical Clustering

<sup>۲</sup> Image Segmentation

<sup>۳</sup> Gray-Level

<sup>۴</sup> Computer vision

<sup>۵</sup> Image processing



شکل ۱. نمایش یک عکس رنگی در یک فضای برداری سه بعدی

## داده ها :

مجموعه داده پیوست شده ، شامل ۳۰۰ عکس منحصربه فرد است که به ازای هر عکس یک نسخه ی رنگی و یک نسخه تک-رنگ وجود دارد. مجموعه داده ی مورد نظر را می توانید از آدرس زیر بارگیری کنید.

<https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/BSDS۳۰۰-images.tgz>

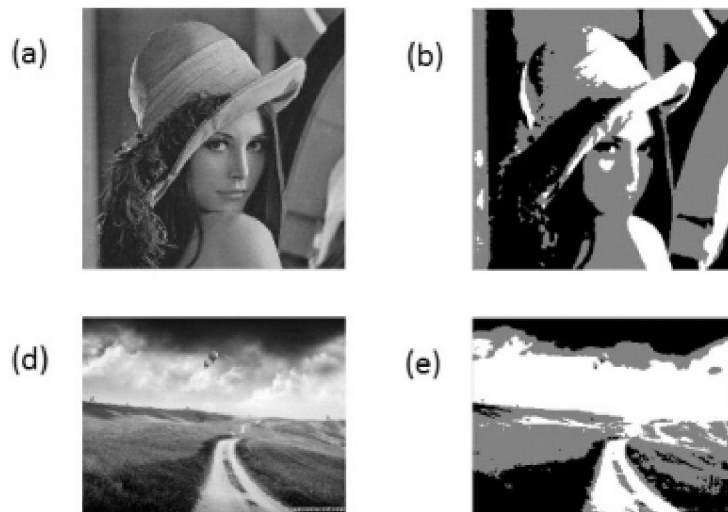
## ۱-۱. پیاده سازی (۹۰ نمره)

(آ) با استفاده از الگوریتم K-means مدلی پیاده سازی کنید و عکس های رنگی را قطعه بندی نمایید. ۱۰ عکس را به صورت تصادفی از مجموعه عکس ها انتخاب کنید و نتایج حاصل از خوشه بندی را مانند آنچه در شکل ۲ می بینید در گزارش خود نمایش دهید. (پیشنهاد : یک انتخاب مناسب برای رنگ هر خوشه ، مقدار مرکز<sup>۱</sup> خوشه می باشد).

(ب) با استفاده از الگوریتم K-means مدلی پیاده سازی کنید و عکس های تک رنگ را قطعه بندی نمایید. ۱۰ عکس را به صورت تصادفی از مجموعه عکس ها انتخاب کنید و نتایج حاصل از خوشه بندی را مانند آنچه در شکل ۲ می بینید در گزارش خود نمایش دهید.

(پ) برای بهبود نتایج حاصل از الگوریتم یک پیشنهاد جالب استفاده از موقعیت هر پیکسل در تصویر علاوه بر ویژگی های رنگی آن است. مدل پیاده سازی کنید که برای قطعه بندی عکس های رنگی از ۵ تایی، (R,G,B, line, column) استفاده کند.

<sup>۱</sup>centroid



(a), (d) Original image; (b), (e) K-means algorithm;

شکل ۲. نحوه‌ی نمایش مطلوب خروجی مدل در گزارش تمرین

## ۲-۱. تحلیل پارمتر (۶۰ نمره)

آ) آیا موقعیت‌های اولیه مراکز خوشه‌ها بر روی نتایج حاصل از مدل‌هایتان تاثیر گذارند؟ (استدلال خود را با ارائه‌ی نتایج و هم به صورت تئوریک بیان کنید)

ب) بهترین مقدار K را برای مدل‌تان بدست آورید. (روش مورد استفاده‌تان را شرح دهید)

پ) تغییر مقدار آستانه چه اثری بر نتایج حاصل از مدل‌تان دارد؟ (استدلال خود را با ارائه‌ی نتایج و هم به صورت تئوریک بیان کنید)

ت) روش‌های متعددی برای ارزیابی الگوریتم‌های با ناظر ارائه شده است و استفاده از آن‌ها برای ارزیابی نتایج مدل‌های با ناظر معمول است. با این حال تا به حال به ارزیابی الگوریتم‌های بدون ناظر فکر کرده اید؟

## ۲. خوشه‌بندی شهرها و استان‌های مربوط به خودپردازها (۱۰۰ نمره)

مجموعه داده‌ای مشتمل بر تراکشهای بانکی تعدادی دستگاه خودپرداز در اماکن مختلف شهرهای چند استان مختلف موجود است. اما اطلاعاتی از محل استقرار دستگاهها خودپرداز در دست نیست. با استفاده از الگوریتم خوشه بندی سلسله مراتبی مدلی جهت شناسایی دستگاههایی که در شهر هستند ارائه دهید. مدلی نیز جهت شناسایی دستگاههایی که در یک استان قرار گرفته اند ارائه دهید. برای هر شهر و هر استان یک کد جداگانه در نظر بگیرید.

### داده‌ها :

مجموعه داده ی مورد نظر این سوال در فایل با عنوان transaction-data.xlsx پیوست شده است. گفتنی است در شیت دوم فایل گسترده مذکور اطلاعات لازم در مورد فیلدهای مختلف داده آورده شده است.

### خروجی مورد نظر :

آ) با توجه به مجموعه دادگانی که در اختیارتان قرار گرفته است ، نحوه‌ی ساختن مدل خود با استفاده از الگوریتم های خوشه‌بندی سلسله مراتبی را به طور کامل شرح دهید.(تمام مراحل شامل جمع‌آوری داده، پیش‌پردازش، ساخت مدل ، انتخاب یکی از روش‌های خوشه بندی سلسله مراتبی و نیز دلیل انتخاب و هم چنین نحوه‌ی ارزیابی‌تان را شرح دهید.)

ب) (تشویقی) مدل موردنظرتان را پیاده‌سازی کنید.(۱۰۰ نمره)

### نکات مهم:

۱.نمره سوالات پیاده سازی و تحلیلی با کد به صورت زیر در نظر گرفته میشود:

۱.کدها ۴۰٪

۲.گزارش ۳۰٪

۳.تحویل حضوری ۳۰٪

II.نمره سوالات تحلیلی بدون کد به صورت زیر در نظر گرفته میشود:

۱.گزارش ۶۰٪

۲.تحویل حضوری ۳۰٪

III.نمرات کدها، گزارش و تحویل حضوری منوط به ارسال به موقع کدها و گزارش است