



هدف پروژه

در این پروژه، هدف این است که تصاویر مرتبط با طبیعت را بر اساس ویژگی‌های مشخصی که در بخش‌های بعدی توضیح داده می‌شوند، خوشه‌بندی کنیم. داده‌های تصویری انتخاب شده شامل مجموعه‌ای متنوع از کلاس‌های مرتبط با طبیعت از جمله جنگل، بیابان، ساحل و سایر دسته‌بندی‌های محیطی است. در نهایت، هدف ما این است که نمونه‌های دارای ویژگی‌های مشابه در یک خوشه قرار گیرند.

دیتاست

این دیتاست شامل شش کلاس و هر کلاس شامل 600 نمونه از تصاویر ماهواره‌ای مربوط به طبیعت است. لینک دسترسی به دیتاست در سامانه ویو در اختیار شما قرار گرفته است.

فاز اول: Feature Extraction

در این فاز از شما می‌خواهیم که تصاویر را پردازش کرده و ویژگی‌های مفیدی را از آن‌ها استخراج کنید. هدف این است که بتوانید داده‌ها را به گونه‌ای آماده کنید که امکان تفکیک بهتر کلاس‌های مختلف فراهم شود. برخی از ویژگی‌های پیشنهادی که می‌توانید استخراج کنید شامل موارد زیر هستند:

- ویژگی‌های رنگی: شناسایی رنگ‌های غالب در تصویر
 - ویژگی‌های بافتی: بررسی الگوهای تکرارشونده در تصاویر
 - ویژگی‌های لبه‌ای: محاسبه تراکم لبه‌ها در تصویر
 - ویژگی‌های آماری: محاسبه میانگین، واریانس و توزیع شدت روشنایی در تصویر و ...
 - همچنین، شما می‌توانید پیش‌پردازش‌هایی مانند تغییر اندازه تصاویر، تبدیل به سطح خاکستری، افزایش کنتراست و حذف نویز را برای بهبود داده‌ها انجام دهید.
 - ویژگی‌های ذکر شده کاملاً اختیاری هستند اما تنها به این ویژگی‌ها اکتفا نکنید. حداقل به استخراج 6 ویژگی بپردازید.
 - استفاده از مدل‌های آماده برای استخراج ویژگی مجاز نمی‌باشد.
- در نهایت، ویژگی‌های استخراج‌شده را در یک فایل مناسب ذخیره کنید تا در مراحل بعدی مورد استفاده قرار گیرند.

فاز دوم: Feature Selection

در این فاز از شما می‌خواهیم که از بین ویژگی‌های موجود، حداقل 3 ویژگی را انتخاب کنید تا بر اساس آن‌ها خوشه‌بندی را انجام دهید. برای این کار باید همبستگی (correlation) بین تک‌تک ویژگی‌ها را حساب کرده و سپس یک ماتریس همبستگی (correlation matrix) ایجاد کنید. پس از آن باید ویژگی‌ها را به گونه‌ای انتخاب کنید که خوشه‌بندی بهتری داشته باشید.

- محاسبه correlation بر عهده خودتان است و استفاده از کتابخانه‌ها و توابع عمومی مجاز نیست.

- باتوجه به حداقل تعداد ویژگی‌ای که قید شد، بهینه‌ترین threshold را انتخاب کنید.

فاز سوم: Clustering

در این فاز باید با استفاده از الگوریتم‌های DBSCAN، KMeans، Agglomerative و MeanShift دیتاست موجود را بر اساس ویژگی‌های انتخاب شده در فاز قبلی، خوشه بندی کنید.

- برای پارامترهای هر الگوریتم باید hyperparameter tuning انجام دهید و بهترین مقادیر را انتخاب کنید.

- جهت مقایسه ویژگی‌های متمایز کننده هر خوشه و پیدا کردن ویژگی‌های بارز آن‌ها، یک heatmap باتوجه به اعضای خوشه‌ها plot کنید.

فاز چهارم: Visualization

در این بخش ابتدا با استفاده از کتابخانه‌های مناسب ابعاد دیتا را کاهش داده و سپس نتیجه خوشه‌بندی را visualize کنید. نتایج این خوشه‌بندی را در داکيومنت خود تحلیل کنید.

فاز پنجم: Evaluation

در این فاز از شما می‌خواهیم که عملکرد خوشه‌بندی انجام‌شده را ارزیابی کنید. معیارهای تدریس شده در کلاس درس شامل precision، recall و f1-score را پیاده سازی کرده و نتایج را گزارش کنید. پیاده سازی این معیارها بر عهده خودتان است. معیار دیگری به نام Silhouette Score را نیز محاسبه کنید. علت تفاوت اعداد خروجی این معیارها را تحلیل کرده و در داکيومنت خود ذکر کنید.

فاز ششم: Prediction

در این فاز، داده‌هایی که به‌عنوان داده‌ی تست در اختیار شما قرار داده شده‌اند، باید توسط مدل خوشه‌بندی‌شده بررسی شوند و مشخص شود که به کدام خوشه تعلق دارند. با استفاده از ویژگی‌های انتخاب‌شده، داده‌های تست را پیش‌پردازش کرده و نرمال‌سازی کنید.

- مدل KMeans از قبل آموزش‌دیده‌شده را روی این داده‌ها اعمال کنید و برچسب خوشه‌ی متناظر را به هر نمونه اختصاص دهید.

- برای ۱۰ داده‌ی تست، نتایج خوشه‌بندی را visualize کنید. برای هر یک از این ۱۰ داده‌ی تست، ۵ داده‌ی دیگر از خوشه‌ی مرتبط با آن نیز نمایش داده شود.

- نتایج این فاز را در قالب یک فایل CSV ذخیره کرده و تحلیل خود را در مستندات پروژه ارائه دهید.

نکات تکمیلی

- علاوه بر سورس کد پروژه، فایل مستندات نیز باید آپلود شود.
- نام اعضای گروه در فایل مستندات ذکر شود و فقط یکی از اعضا پروژه را آپلود کند.
- هر گونه شباهت نامتعارف بین کد شما و کد سایر گروه ها تقلب محسوب می شود و نمره ای برای این پروژه دریافت نخواهید کرد.
- در صورت نوشتن داکيومنت تمیز (برای مثال با LATEX) نمره اضافه برای شما در نظر گرفته خواهد شد.
- فایل شامل سورس کد پروژه و مستندات را در قالب فایل zip و با نام شماره دانشجویی خود ذخیره و ارسال نمایید.
- در صورت داشتن هرگونه سوال می توانید با [SetareKahnamuee](#) و [fatemeh_dehbashii](#) و [Kourosh_Hsz](#) در ارتباط باشید یا در گروه درسی مطرح کنید.

موفق باشید؛
تیم حل تمرین