

پروژه درس یادگیری ماشین

هدف:

در این پروژه قصد داریم به دسته بندی دو مجموعه داده با استفاده از الگوریتم‌های مختلف یادگیری ماشین و مقایسه نتایج آن‌ها بپردازیم.

اولین مجموعه داده تحت عنوان Olivetti Faces مشتمل بر ۴۰۰ تصویر با ابعاد 64×64 می‌باشد که از ۴۰ فرد در ۱۰ موقعیت مختلف با توجه به تغییر تابش نور، حالات چهره (چشم‌های باز/ بسته، در حال لبخند زدن/ بدون لبخند) و جزئیات صورت (با عینک/ بدون عینک) ثبت شده است. بخشی از این مجموعه داده را در شکل ۱ مشاهده می‌نمایید.



شکل ۱. بخشی از مجموعه داده Olivetti faces

انتظار می‌رود پس از اتمام دسته بندی هر یک از تصاویر آزمایشی، در کلاس‌های مختلفی از ۰ تا ۳۹، که نشان دهنده هویت فرد است، قرار گیرند.

مراحل کار بر روی این مجموعه داده به شرح زیر می‌باشد:

- ۱) بارگذاری مجموعه داده
- ۲) تقسیم مجموعه داده به دو بخش آموزش و آزمایش
- ۳) دسته بندی داده با استفاده از الگوریتم‌های SVM, KNN, Random forest
- ۴) مقایسه معیارهای دقت و زمان اجرای الگوریتم‌های فوق
- ۵) کاهش ابعاد مجموعه داده با روش PCA
- ۶) تکرار مراحل ۳ و ۴ پس از اعمال PCA

مجموعه داده دوم تحت عنوان iPhone شامل ۴۰۰ رکورد متعلق به افراد مختلف است که انتظار می‌رود پس از اتمام دسته بندی، بتوانیم تشخیص دهیم هریک از افراد مجموعه آزمایش با توجه ویژگی‌های خود اعم از سن، جنسیت و سطح درآمد مایل به خرید گوشی آیفون می‌باشند یا خیر.

مراحل کار بر روی این مجموعه داده به شرح زیر می‌باشد:

- (۱) بارگذاری مجموعه داده
- (۲) تبدیل ویژگی جنسیت از حالت categorical به حالت numeric
- (۳) نرمالسازی داده‌ها
- (۴) تقسیم مجموعه داده به دو بخش آموزش و آزمایش
- (۵) دسته بندی داده با استفاده از الگوریتم‌های SVM, KNN, Naive Bayes
- (۶) مقایسه معیار دقت

فایل‌های مربوط به کدهای پایتون پروژه و یک فایل گزارش کار که شامل هر گونه نمودار و تحلیل می‌باشد؛ در قالب یک فایل زیپ با فرمت Fname_Lname_project ارسال نمایید.