

تمرین پیاده سازی

این تمرین با استفاده از پایگاه داده AT&T انجام می گیرد این پایگاه داده تصاویر 40 نفر است که از هر کدام 10 تصویر در پایگاه داده موجود است (مجموعاً 400 تصویر). این پایگاه داده قبلاً ارسال شده است. همچنین در آدرس زیر اطلاعات بیشتری درباره این پایگاه چهره وجود دارد.

<http://cam-orl.co.uk/facedatabase.html>

در این تمرین، در ابتدا از تصاویر ویژگی های مناسب استخراج کرده و سپس با استفاده از الگوریتم های خوشه بندی به ارزیابی خوشه بندی های مختلف می پردازیم. در این پایگاه داده 40 نفر (کلاس) وجود دارد. از تمامی داده ها برای خوشه بندی استفاده می کنیم (داده های آموزشی و تست نداریم). اندازه تصاویر را به 32 در 32 تبدیل کنید. با استفاده از روش PCA از تصاویر ویژگی استخراج نمائید و داده ها را به فضای دوبعدی نگاشت کنید. موارد زیر در این تمرین مد نظر می باشد:

- ۱) اجرای الگوریتم k میانگین (kmeans) و محاسبه معیار دقت به عنوان یک روش ارزیابی با ناظر
- ۲) اجرای الگوریتم k میانگین (kmeans) و محاسبه معیار Silhouette به عنوان یک روش ارزیابی بدون ناظر
- ۳) اجرای الگوریتم FCM و محاسبه معیار دقت به عنوان یک روش ارزیابی با ناظر
- ۴) اجرای الگوریتم FCM و محاسبه معیار Silhouette به عنوان یک روش ارزیابی بدون ناظر

نکات)

- هر الگوریتم را 20 بار تکرار کنید و نتیجه خوشه بندی را به صورت متوسط این 20 بار گزارش کنید.
- تعداد خوشه ها را 40 در نظر بگیرید
- برای اجرای الگوریتم pca از scikit-learn استفاده کنید.
- برای اجرای الگوریتم kmeans و FCM از scikit-learn استفاده کنید.
- همچنین برای ارزیابی به کمک هر دو معیار گفته شده از sklearn.metrics استفاده کنید.
- پس از تبدیل داده ها به فضای دوبعدی، داده های نگاشت یافته را ترسیم و در یک شکل کلاس های آنها را با رنگ های مختلف مشخص کنید. سپس نتیجه الگوریتم kmeans و FCM را در شکل های جداگانه ترسیم کنید. در شکل های جدید خوشه های یاد گرفته شده را با رنگ های مختلف مشخص کنید. در این بخش، هدف آن است که بتوان بررسی کرد خوشه های یاد گرفته شده تا چه اندازه بر کلاس های اولیه منطبق هستند.
- این تمرین 3 نمره از نمره پایانی شما را به خود اختصاص می دهد.
- مهلت تحویل تمرین شنبه ۱۷ خرداد می باشد و زمان اعلام شده به هیچ عنوان تمدید نخواهد شد.

با آرزوی موفقیت

زهره مصلحی