

تمرین سری سوم داده کاوی - مهلت تحویل: ۱۵ خرداد ۱۴۰۰

در این سری تمرین شما بایستی با استفاده از روش کاهش ابعاد PCA دادگان مصور را به فضای جدید منتقل کرده و بر روی آنها یک دسته بند (classifier) (یکبار از نوع شبکه عصبی و بار دیگر از نوع ماشین بردار پشتیبان) اعمال کنید.

کد خود به همراه فایل گزارش را در سامانه آموزش ایلرن بارگزاری کنید. در توضیحات خود مشخص کنید دسته بندی مورد استفاده چه ساختاری دارد و پارامترها را چگونه مقدار دهی کرده اید و نتایج را گزارش کنید.

توضیحاتی درباره مسئله:

دسته بندی و تشخیص تصاویر با استفاده از الگوریتم PCA یک مسئله قدیمی و نسبتاً معروف در حوزه علوم داده میباشد. فرض کنید هر تصویر یک ماتریس با سایز مشخص $n*n$ است که مقدار هر درایه آن عددی بین صفر و یک است (gray scale). حال اگر هر تصویر را به صورت یک ماتریس $1*n^2$ بعدی در نظر بگیریم میتوانیم m تصویر را با یک ماتریس $m*n^2$ نشان دهیم.

حال بایستی با استفاده از PCA داده های فعلی را به داده هایی با تعداد ابعاد ویژگی هایی از n^2 تبدیل کنیم. نهایتاً با استفاده از دو دسته بند مذکور به دسته بندی داده ها موجود پردازید. (در بخش لینک های کمکی میتوانید منابعی برای کسب اطلاعات بیشتر پیدا کنید)

توضیحات مربوط به مجموعه داده:

تصاویری با سایز (اندازه ماتریس) یکسان از افراد مشهور در مجموعه داده موجود است. از هر فرد حداقل ۲۰۰ تصویر موجود است و هر فرد با یک شماره مشخص شده است. برای ساده سازی تعداد افراد برابر ۲ در نظر گرفته شده است و ابعاد هر تصویر $41*41$ است.

بارگذاری داده:

روش اول (ترجیحاً از این روش استفاده کنید) با استفاده از دستور زیر میتوانید داده ها را به صورت مستقیم دانلود و در کد خود استفاده کنید. (اگر اینترنت کند دارید میتوانید از google colaboratory استفاده کنید).

```
from sklearn.datasets import fetch_lfw_people
lfw_people = fetch_lfw_people(min_faces_per_person=200, resize=0.4)

x = lfw_people.data
y = lfw_people.target
```

روش دوم فایل ex3.csv را از کانال درس دانلود کنید.

در صورت وجود هر گونه ابهام سوالات خود را از استاد یا دستیاران آموزشی پرسید.

لینک های کمکی :

[what is Google colaboratory \(colab\) ?](#)

[recommended video](#)

[PCA](#)