

بسمه تعالی
تمرین دوم
دوره یادگیری ماشین بدون نظارت

سوال اول

در این سوال قصد داریم که روش‌های مربوط به کاهش بعد را به صورت دستی پیاده‌سازی کرده و خروجی آن را با استفاده از کتابخانه‌های معروف مقایسه کنیم.

1. ابتدا مجموعه دادگان diabetes.csv که در اختیار شما قرار گرفته است را بارگیری نمایید.
2. ویژگی‌ها و لیبل‌ها را با استفاده از دستورات pandas یا numpy جدا کرده و در متغیرهای متفاوتی ذخیره‌سازی را انجام دهید.
3. حال برای هر دو روش کاهش بعد PCA و LDA به صورت جداگانه مرحله‌ی بعد را پیاده‌سازی کرده و خروجی‌ها را بر هر دو روش به صورت مجزا نمایش دهید.
4. برای هر یک از این دو روش یک‌بار مجموعه دادگان را با استفاده از کتابخانه sklearn کاهش بعد به ۲ انجام دهید. سپس روش کاهش بعد مذکور را به صورت یک تابع پایتون با استفاده از کتابخانه numpy پیاده‌سازی کنید. خروجی روش‌های کاهش بعد دستی و با استفاده از کتابخانه را در نموداری نمایش داده و تحلیل‌های خود را ارائه کنید.

سوال دوم

در این سوال هدف پیاده‌سازی روش بیشینه درست‌نمایی بر روی یک مجموعه دادگان ساختگی است. در این قسمت می‌بایست فایل Q2.ipynb که در اختیار شما قرار گرفته شده است را برای هر بخش سوال پر کنید. توضیحات مربوط به هر بخش در آن آمده است.

سوال سوم

در این سوال می‌خواهیم به مبحث نفرین ابعاد بپردازیم. هر یک از مراحل زیر را به ترتیب انجام داده و در نهایت خروجی مطلوب را گزارش دهید.

1. ابتدا توجه داشته باشید که تنها می‌توانید از کتابخانه‌های numpy و pandas استفاده کنید.
2. دو داده x و y را به صورت تصادفی با استفاده از تابع np.randint تولید کنید.
3. فاصله اقلیدسی این دو داده را برای ابعاد ۲ تا ۵ محاسبه کنید.
4. نمودار فاصله اقلیدسی بر حسب ابعاد داده را رسم کنید و استنباط خود را از خروجی بیان کنید.

سوال چهارم

در این سوال قصد داریم که از مجموعه دادگان مربوط به سرطان سینه که به اسم cancer.csv در اختیار شما قرار گرفته است، به سراغ پیاده سازی مساله forward selection بپردازیم.

- ابتدا این مجموعه دادگان را فراخوانده و ۵ سطر اول آن را نمایش دهید.
- سائز کلی این مجموعه داده را گزارش کرده و بگویید که این مجموعه داده چند ویژگی در بر دارد.
- در بخش بعد به سراغ مرتب سازی ستون ها رفته و مجموعه داده را در صورت وجود دادگان پرت از آن ها عاری کنید.
- حال با استفاده از یک طبقه بند دلخواه این مجموعه داده را طبقه بندی کرده و دقت را در مجموعه داده آموزشی و ارزیابی گزارش کنید.
- با استفاده از کتابخانه sklearn و ماژول feature_selection، ویژگی های مطلوب را بر روی طبقه بند اعمال کرده و گزارش کنید. کدام ویژگی ها انتخاب شدند؟ آیا می توانید علت را توجیه کنید؟ در هر مرحله ماتریس آشفتگی را نیز رسم کنید.