



تمرین چهارم یادگیری ماشین

توجه:

- کد باید فقط در زبان پایتون باشد.
- استفاده از کتابخانه‌های از پیش آماده مجاز است، مگر اینکه نقیض آن گفته شود. در گزارش کتابخانه‌ها ذکر شوند.
- لطفاً علاوه بر ضمیمه کردن کد و تصاویر، نتایج را تحلیل و در فایل گزارش خود ضمیمه کنید.
- فایل گزارش به فرمت pdf و به زبان فارسی باشد.

سوال اول

داده‌ها در فضای دو بعدی به صورت زیر داده شده است. با استفاده از روش PCA بهترین برداری را پیدا کنید که واریانس داده‌ها بعد از نگاشت بر روی آن، بیشینه شود. مولفه‌های اصلی^۱ PCA را به صورت تشریحی محاسبه کنید، سپس نگاشت داده‌ها بر روی مولفه اول را به دست آورید و در آخر نگاشت داده‌ها را به فضای دو بعدی برگردانید.

$$\left\{ \begin{pmatrix} 2 \\ -3 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -2 \\ -3 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -2 \\ -2 \end{pmatrix} \right\}$$

سوال دوم

در این تمرین دو مجموعه داده آموزش و تست در اختیار شما قرار گرفته شده است. این دادگان شامل تصاویر دویعدی از صورت است، ولی تصاویر به صورت یک بردار یک بعدی با طول ۴۰۹۶ ذخیره شده است (بعد هر تصویر ۶۴ × ۶۴ است).

۱. از طریق Repeated 5 Fold Cross-Validation با ۱۰ تکرار، خطای MSE بین تصویر بازسازی شده پس از کاهش بعد با تصویر اصلی بر روی داده ارزیابی حساب کنید و در boxplot نمایش دهید. کاهش ابعاد ۵۰۰، ۳۰۰، ۲۰۰، ۱۴۰، ۱۰۰، ۷۰، ۴۵، ۳۰، ۱۵، ۸، ۳ مورد بررسی قرار دهید.

بر اساس boxplot رسم شده، تا چه مقدار کاهش بعد، اطلاعات درون دادگان، حتی با وجود کاهش بعد، به خوبی حفظ شده است.

۲. پس از انتخاب بعد مناسب، داده تست را به آن کاهش دهید و سپس داده را بازسازی کنید. خطای MSE را گزارش دهید.

۳. ۱۰ نمونه اول داده تست به همراه بازسازی شان را نمایش دهید.

¹Principal Components

سوال سوم

مجموعه دادگان BreastCancer را از طریق کد زیر دانلود کنید.

```
from sklearn.datasets import load_breast_cancer  
cancer = load_breast_cancer()
```

۱. ابتدا دادگان را نرمال سازی کنید.

۲. به کمک PCA دادگان را به ابعاد ۲، ۳، ۴، ۵، ۱۰، ۱۵ و ۲۰ کاهش دهید، سپس دسته‌بندی‌های Perceptron متفاوتی را روی خروجی‌های مرحله قبل آموزش دهید. همچنین باید دقت آموزشی و ارزیابی را با استفاده از روش Repeated 5 Fold Cross Validation با تکرار ۱۰ گزارش دهید، سپس نتایج را در Boxplot رسم کنید.

۳. بر اساس Boxplot های رسم شده، کدام یک از مدل‌ها به نظر شما بهترین است ؟