



تمرین سوم 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

داده‌های دو کلاس در فضای دو بعدی به صورت زیر داده شده است. با استفاده از روش LDA بهترین برداری را پیدا کنید که داده‌های دو کلاس بعد از نگاشت بر روی آن، بیشترین فاصله را مطابق معیار LDA داشته باشند. معیار LDA، پارامترهای بهینه مدل و نتایج نهایی بدست آمده را گزارش دهید. تمام روند بایستی به صورت تشریحی انجام شود.

$$C_0 = \left\{ \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}$$
$$C_1 = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\}$$

سوال دوم

مجموعه دادگاه MNIST را از طریق کد زیر دانلود کنید. ابتدا ۲۰ درصد دادگان را به عنوان مجموعه تست و مابقی را به عنوان مجموعه آموزش در نظر بگیرید. معیار ارزیابی در این مسئله دقت^۱ می‌باشد.

```
from sklearn.datasets import fetch_openml
mnist = fetch_openml('mnist_784')
```

۱. به کمک LDA دادگان را به ابعاد ۲، ۳، ۴، ...، ۸ و ۹ کاهش دهید. سپس دسته بند های Perceptron متفاوتی را بر روی هر یک از خروجی‌های مرحله قبل آموزش دهید. همچنین بایستی خطای 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 (از مجموعه داده تست استفاده نکنید).

۲. با توجه به Boxplot، مدل‌ها را از دیدگاه Overfit و Underfit بررسی کنید. همچنین با توجه به نتایج، آیا می‌توان تنظیمات بهتری ارائه داد؟ توضیح دهید.

۳. تنظیم بهینه را تعیین کنید و مدل با تنظیم بهینه را بر روی داده تست اجرا کنید، سپس دقت مدل را محاسبه کنید و همچنین ماتریس درهم ریختگی را رسم کنید.

¹Accuracy

سوال سوم (امتیازی)

۱. ماتریس پراکندگی درون کلاسی و بین کلاسی را برای یک مجموعه داده با بیشتر از دو کلاس تعریف کنید.
۲. با توجه به ماتریس های تعریف شده در بخش قبل، یک تابع هزینه برای LDA با بیشتر از دو کلاس ارائه دهید.
۳. حال با کمینه کردن تابع هزینه بخش قبل و نوشتن محاسبات مربوطه، بردار جهت نمایی را بیابید.