



پردیس دانشکده های فنی

به نام خدا
دانشکده‌ی مهندسی برق و کامپیوتر
تمرین سری چهارم یادگیری ماشین



دانشگاه تهران

سلام بر دانشجویان عزیز، چند نکته مهم:

۱. حجم گزارش به هیچ عنوان معیار نمره دهی نیست، در حد نیاز توضیح دهید.
۲. نکته‌ی مهم در گزارش نویسی روشن بودن پاسخ‌ها می‌باشد، اگر فرضی برای حل سوال استفاده می‌کنید حتماً آن را ذکر کنید، اگر جواب نهایی عددی است به صورت واضح آن را بیان کنید.
۳. کد های نوشته شده برای هر سوال شبیه سازی را در فایل ipynb متناظر آن سوال بنویسید.
۴. کدهای ارسال شده بدون گزارش و یا کامنت گذاری دقیق در کد فاقد نمره می‌باشند.
۵. برای سوالات شبیه سازی، فقط از دیتاست داده شده استفاده کنید.
۶. نمره تمرین از ۱۰۰ نمره می‌باشد
۷. هرگونه شباهت در گزارش و کد مربوط به شبیه سازی، به منزله تقلب می‌باشد و کل تمرین برای طرفین ۱۰۰- خواهد شد.
۸. در صورتی که تشخیص داده شود از چت بات ها به صورت مستقیم برای پاسخ سوال های تئوری و شبیه سازی استفاده شده است، نمره ۱۰۰- در نظر گرفته خواهد شد.
۹. فایل نهایی خود را در یک فایل زیپ شامل، pdf گزارش و فایل کدها آپلود کنید. نام فایل زیپ ارسالی الگوی ML_HW#_StudentNumber داشته باشد.
۱۰. در صورت داشتن سوال، از طریق گروه درس یا ایمیل‌های زیر با تدریسار مربوطه سوال‌های خود را مطرح کنید.

سوال ۱ و ۲ و ۳ : k.hooshanfar@ut.ac.ir
سوال ۴ و ۵ و ۶ : amirh.bonakdar@ut.ac.ir

سوال ۱: (۱۲ نمره)

به سوالات زیر پاسخ دهید.

الف) یکی از مزیت‌های تابع فعال‌ساز relu این می‌باشد که برخلاف دو تابع فعال‌ساز sigmoid و tanh ، دچار مشکل $\text{vanishing gradient}$ نمی‌شود. این مشکل را توضیح دهید و استدلال کنید چرا sigmoid و tanh دچار این مشکل می‌شوند ولی relu دچار این مشکل نمی‌شود.

ب) در یک مساله طبقه‌بندی دوکلاسه از بین دو تابع فعال‌ساز relu یا sigmoid کدام را انتخاب خواهید کرد؟ چرا؟

ج) فرق بین تابع فعال‌ساز relu با leaky relu در چیست؟ در چه مواردی بهتر است که به جای relu از leaky relu استفاده شود؟

ه) توضیح دهید مقداردهی اولیه وزن‌ها و بایاس‌ها، تحت چه شرایطی بر یادگیری شبکه عصبی و عملکرد نهایی آن تاثیر دارد.

ی) توضیح دهید تحت چه شرایطی اضافه کردن تعداد لایه‌های نهان در یک شبکه عصبی (مانند MLP) می‌تواند عملکرد مدل شبکه عصبی را افزایش دهد، کاهش دهد یا بر آن تاثیری نداشته باشد.

سوال ۲: (۱۳ نمره)

لطفاً به سؤال‌های موجود در فایل Q2.ipynb پاسخ دهید.

نیازی به نوشتن هیچ کد برنامه‌نویسی نیست؛ تنها کافی است به سوالات مشخص شده پاسخ دهید. می‌توانید

پاسخ سوال‌ها را کنار پاسخ دیگر سوال‌ها قرار دهید یا اینکه در همان فایل ipynb پاسخ را نوشته و خروجی

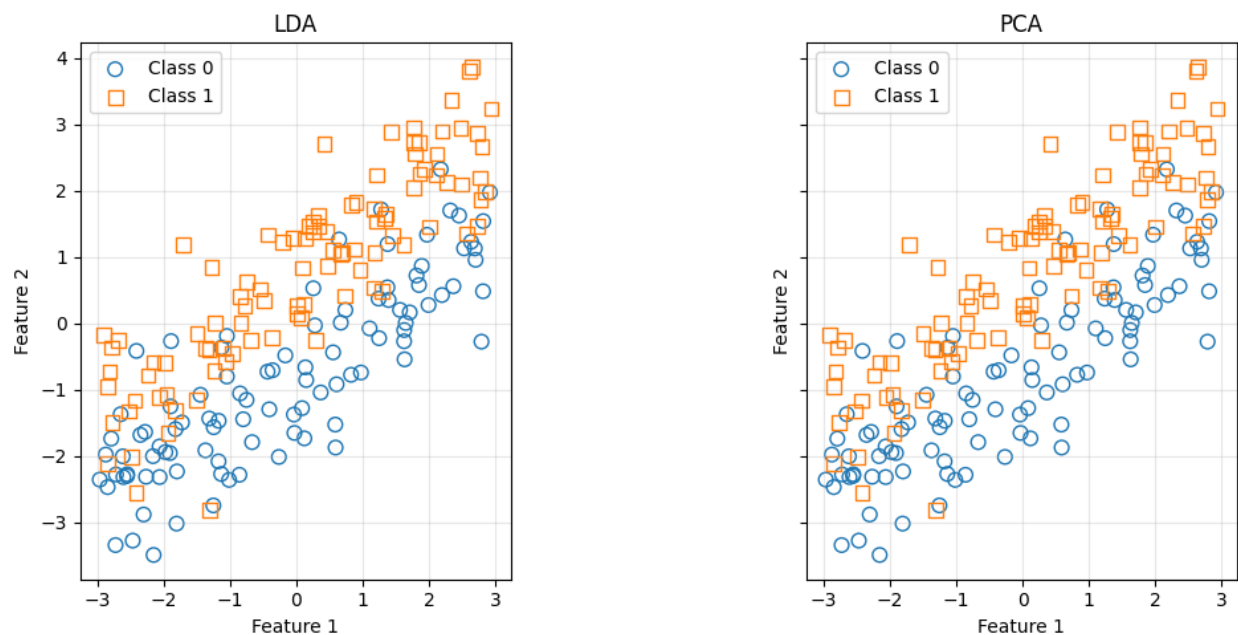
pdf مربوط به این سوال نیز را قرار دهید.

سوال ۳: (۱۵ نمره)

به هر کدام از موارد زیر پاسخ دهید.

الف) در شکل ۱ زیر، جهت مؤلفه‌ی اصلی^۱ اول را در نمودار سمت راست رسم کنید و جهت تفکیک‌کننده‌ی خطی فیشر اول^۲ را در نمودار سمت راست رسم نمایید. (منظور جهت اولین مؤلفه‌ی LDA می‌باشد) (جهت تقریبی و شهودی مدنظر است، نیازی به محاسبات نیست).

نکته: برای PCA، برچسب‌گذاری نقاط را در نظر نگیرید.



شکل ۱

ب) بدست آوردن مؤلفه‌های LDA:

این دیتاست با ۳ کلاس $(\omega_1, \omega_2, \omega_3)$ در فضای دو بعدی در نظر بگیرید:

$$\text{class } \omega_1: x_1 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, x_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\text{class } \omega_2: x_1 = \begin{bmatrix} -3 \\ 0 \end{bmatrix}, x_2 = \begin{bmatrix} -3 \\ -2 \end{bmatrix}$$

^۱ the first principal component

^۲ the first Fisher's linear discriminant

$$class \omega_3: x_1 = \begin{bmatrix} 3 \\ 0 \end{bmatrix}, x_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

به موارد زیر پاسخ دهید:

الف) بردارهای میانگین را برای هر کلاس و برای کل داده‌ها محاسبه کنید.

ب) Within-Class Scatter Matrix (S_W) را محاسبه کنید.

ج) Between-Class Scatter Matrix (S_B) را محاسبه کنید.

د) مسئله بهینه‌سازی برای بدست آوردن مولفه‌های LDA را بنویسید.

ه) مولفه‌های LDA را بدست آورید.

سوال ۴: (۲۰ نمره)

به ترتیب به سوالات زیر مورد Kernel PCA پاسخ دهید.

الف) در PCA استاندارد، مؤلفه‌های اصلی^۳ برای یک مجموعه داده‌ی با میانگین^۴ صفر $X \in \mathbb{R}^{N \times D}$ با تجزیه مقادیر ویژه ماتریس کوواریانس آن پیدا می‌شوند.

فرض کنید $x_i^T \in \mathbb{R}^D$ سطر i ام از ماتریس X باشد. ماتریس کوواریانس به صورت زیر نمایش داده می‌شود.

$$C = \frac{1}{N} \sum_{i=1}^N x_i x_i^T$$

j امین بردار ویژه، u_j ، و مقدار ویژه، λ_j در رابطه زیر صدق می‌کنند.

$$C u_j = \lambda_j u_j$$

حال یک نگاشت $\phi: \mathbb{R}^D \rightarrow \mathbb{R}^M$ به یک فضای بعد بالا $D \ll M$ را در نظر بگیرید. همچنین فرض کنید $\phi(x_i)$ میانگین^۵ صفر دارد.

$$\sum_{i=1}^N \phi(x_i) = 0$$

همچنین ماتریس کواریانس S به صورت زیر محاسبه می‌شود.

$$S = \frac{1}{N} \sum_{i=1}^N \phi(x_i) \phi(x_i)^T$$

اگر ما به صورت مستقیم با حل مسئله تجزیه مقادیر ویژه زیر در فضای نگاشت شده بخواهیم PCA انجام دهیم،

$$S v_i = \lambda_i v_i$$

احتمالا با چه مشکلی مواجه می‌شویم؟ به طور خلاصه پاسخ دهید.

³ principal components

⁴ mean

⁵ mean

در ادامه در قسمت‌های ب تا د، می‌خواهیم با استفاده از کرنل‌ها، PCA در فضای ویژگی بعد بالا را بدست آوریم. در این راه حل ما به جای آن که بر روی ماتریس S با سایز $M \times M$ تجزیه مقادیر ویژه را محاسبه کنیم، بر روی ماتریسی با سایز $N \times N$ تجزیه مقادیر ویژه را محاسبه خواهیم کرد.

ب) نشان دهید j امین principal component، v_j ، می‌تواند به صورت ترکیب خطی نگاشت نقاط در فضای ویژگی نمایش داده شود. به بیان دیگر یک بردار N بعدی، $w_j = (w_{j1}, \dots, w_{jn}, \dots, w_{jN})^T$ ، وجود دارد که در رابطه زیر صدق می‌کند.

$$v_j = \sum_{i=1}^N w_{ji} \phi(x_i) = \phi(X)^T w_j$$

دقت شود تنها مقادیر ویژه غیر ۰ مورد نظر هستند.

ج) ماتریس کرنل را به صورت زیر در نظر بگیرید.

$$K \in \mathbb{R}^{N \times N}: K_{ij} = k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$$

ثابت کنید هر w_j ، که متناظر با مقدار ویژه مخالف صفر λ_j است، از حل رابطه زیر بدست می‌آید.

$$K w_j = N \lambda_j w_j$$

راهنمایی: از مسئله تجزیه مقادیر ویژه اصلی بر حسب S ، v_j شروع کنید. و از نتیجه‌ی بخش قبلی استفاده کنید. می‌توانید فرض کنید که K معکوس پذیر است.

د) بدست آوردن بردار principal component، v_j ، از روی بردار وزن w_j نیازمند محاسبه $\phi(X)$ است که ممکن است مشکل‌زا باشد. در kernel PCA ما از محاسبه بردارهای principal component اجتناب می‌کنیم و به صورت مستقیم از تصویر نقاط بر روی principal component ها استفاده می‌کنیم.

نشان دهید تصویر هر نقطه‌ی نگاشت شده، $\phi(x)$ بر روی بردار principal component، v_j ، می‌تواند تنها با استفاده از کرنل و w_j محاسبه شود.

$$\phi(x)^T v_j = [k(x, x_1), k(x, x_2), \dots, k(x, x_N)] w_j$$

سوال ۵: (شبیه سازی، ۲۵ نمره)

پیاده سازی های در همان فایل Q5.ipynb انجام دهید. هم چنین بعضی از سوال ها حل کردنی هستند، آن ها را میتوانید در همان فایل ipynb پاسخ دهید و یا اینکه پاسخ خود را کنار پاسخ دیگر سوال ها قرار دهید. در صورتی پاسخ خود را در فایل ipynb می نویسید لطفا خروجی pdf آن را برای این سوال قرار دهید.

سوال ۶: (شبیه سازی، ۱۵ نمره)

الف) بارگیری داده: دیتاست Olivetti faces را بوسیله تابع `sklearn.fetch_olivetti_faces` دریافت کنید.

ب) تقسیم داده: ۸۰٪ داده را برای آموزش و ۲۰٪ را برای تست (با حفظ توزیع کلاس) در نظر بگیرید.

ج) ایجاد نویز: روی تصاویر بخش تست، نویز گاوسی با $\sigma = 0.4$ اعمال کنید.

د) کاهش نویز با PCA: روی داده‌های تمیز آموزش، PCA را آموزش دهید.

ه) برای مقادیر $k \in [2, 4, 10, 20, 50, 100, 200, 320]$ تصاویر نویزی را فشرده و بازسازی کنید.

و) به ازای مقادیر مختلف k در قسمت ه mean squared error را محاسبه کنید. (بین داده بازسازی شده و داده تمریز بدون نویز)

ح) کاهش نویز با kernel PCA: موارد د، ه، و را برای kernel PCA با کرنل چندجمله‌ای درجه ۵ تکرار کنید.

ط) منحنی mean squared error بر حسب k را برای PCA و kernel PCA در یک نمودار رسم کنید و نتایج تا تحلیل کنید.

ی) به ازای هر کدام از روش‌های PCA و kernel PCA، تصویر اصلی، تصویر نویزی و تصاویر بازسازی‌شده به ازای k های مختلف را در کنار هم نمایش دهید و تحلیل کنید.

ک) موارد ب تا ی را به ازای دیتاست digits تکرار کنید. در این قسمت از مقادیر $k \in [2, 4, 10, 20, 64]$ استفاده کنید. این دیتاست را می‌توانید با تابع `sklearn.datasets.load_digits` دریافت کنید.

استفاده از کتابخانه‌هایی مانند `sklearn` و ... در پیاده سازی کد مجاز است.