



پردیس دانشکده‌های فنی

بسمه تعالی
دانشکده مهندسی برق و کامپیوتر
تمرین سری پنجم درس یادگیری ماشین



دانشگاه تهران

سلام بر تمام دانشجویان عزیز، چند نکته مهم:

۱. حجم گزارش به هیچ عنوان معیار نمره‌دهی نیست، در حد نیاز توضیح دهید.
۲. نکته‌ی مهم در گزارش نویسی روشن بودن پاسخها می‌باشد، اگر فرضی برای حل سوال استفاده می‌کنید حتما آن را ذکر کنید، اگر جواب نهایی عددی است به صورت واضح آن را بیان کنید.
۳. برای سوالات شبیه سازی، فقط از دیتاست داده شده استفاده از کنید. شکل ها به طور واضح و در فرمت درست گزارش شوند.
۴. از بین سوالات **شبیه سازی** حتما به دو مورد پاسخ داده شود.
۵. همچنین دقت شود که نام کلاس‌ها و توابع در فایل ipynb باید رعایت شود اما ورودی و خروجی آن بسته به نیاز می‌تواند تغییر کند و اشکال داده شده مثال هستند و لزومی ندارد عینا همان را در خروجی بدست آورید.
۶. پیشنهاد می‌شود محاسبات خود را در **گوگل کولب** انجام دهید تا حجم محاسبات به سیستم شما فشار وارد نکند.
۷. حداکثر تا نمره ۱۱۰ (۱۰ نمره امتیازی) لحاظ خواهد شد.
۸. هرگونه شباهت در گزارش و کد مربوط به شبیه سازی، به منزله **تقلب** می باشد و کل نمره تمرین **صفر** می‌شود.
۹. در صورت داشتن سوال، از طریق ایمیل banafshehkarimian@ut.ac.ir ، سوال خود را مطرح کنید.

سوال (۱) به سوالات زیر پاسخ دهید: (۱۰ نمره)

- الف) انواع روش‌های جست و جو و انتخاب فیچر در انتخاب ویژگی را نام ببرید. روش‌هایی برای ارزیابی کیفیت ویژگی انتخابی بیان کنید. تفاوت اصلی روش‌های نام برده را بیان کنید.
- ب) مشکلات محاسباتی و عددی که موقع استفاده از PCA و LDA در داده‌های با دایمنشنالیتی بالا و تصاویر ممکن است بوجود بیاید را ذکر کرده و راه حل موجود را ذکر کنید.

سوال (۲) حالتی را در نظر بگیرید که به جای دو کلاس C کلاس داشته باشیم. (۱۵ نمره)

الف) در این حالت نشان دهید:

$$\text{total scatter matrix or } S_T = S_W + S_B$$

ب) درباره رنک ماتریس S_B چه می‌توان گفت؟ استدلال خود را بیان کنید.

سوال (۳) مراحل تشخیص چهره (آیا تصویر چهره است؟) و شناخت چهره (آیا این چهره فرد خاصی است؟) از طریق eigenface ها را به صورت شبه کد بیان کنید. (۱۵ نمره)

سوال (۴) فرض کنید که $P_x(x|w_i)$ بیانگر چگالی‌های احتمالی با میانگین μ_i و واریانس Σ_i برای $i = 1, 2$ باشد که لزوما هم

نرمال نیست. اگر $y = w^t x$ بیانگر projection باشد و چگالی‌های تک بعدی $P(y|w_i)$ دارای میانگین μ_i و واریانس δ_i^2 باشند، w را بیابید طوری که تابع هزینه‌ی $J(w)$ بیشینه گردد. (۱۵ نمره)

$$J(w) = \frac{(\mu_1 - \mu_2)^2}{\delta_1^2 + \delta_2^2}$$

- سوال ۵) برای جدول زیر به موارد زیر پاسخ دهید (کد زدن برای این بخش قابل قبول نمی‌باشد). (۱۵ نمره)
- الف) میزان $\mu + 1$ و $\mu - 1$ را به همراه ماتریس S_B (between-class scatter matrix) محاسبه کنید.
- ب) میزان $S + 1$ و $S - 1$ را به همراه ماتریس S_W (within-class scatter matrix) محاسبه کنید.
- ج) داده‌ها را به همراه بهترین جهتی که آن‌ها را از هم جدا میکند با استفاده از بردارهای ویژه رسم کنید.

i	x	yi
X1	(1,1)	1
X2	(2,1)	1
X3	(1,2)	1
X4	(2,2)	-1
X5	(3,2)	-1

- سوالات پیاده‌سازی: (در کل ۸۰ نمره)
- سوال ۶) در این بخش قصد پیاده‌سازی دو الگوریتم فیچر سلکشن بدون استفاده از کتابخانه برای دیتاست اعداد را داریم. (۲۵ نمره)
- الف) الگوریتم Forwardselection را در کلاس مشخص شده تکمیل کنید (معیار مربوطه را دقت روش NaiveBayes قرار دهید) و نموداری همانند مثال رسم کنید.
- ب) الگوریتم Backwardselection را در کلاس مشخص شده با معیاری همانند معیار بالا تکمیل کنید و نمودار ایندکس فیچرهای حذف شده به دقت را رسم کنید.
- ج) نتایج دو بخش بالا را به همراه مزایا و معایب آن‌ها نسبت به یکدیگر تحلیل کنید.

- سوال ۷) در این بخش قصد داریم الگوریتم LDA را از پایه بدون استفاده از کتابخانه پیاده‌سازی کنیم. (۲۵ نمره)

الف) ابتدا توابع مربوط به تولید ماتریس درون و بین کلاسی اند را کامل کنید.

ب) تابع LDA را تکمیل نمایید و خروجی آن را گزارش دهید.

ج) در یک نمودار مقدار separability measure یا $\text{trace}(S_W^{-1}S_B)$ نسبت به تعداد ویژگیها رسم نمایید و در مورد تاثیر تعداد ویژگیها بر آن بحث کنید.

سوال ۸) در این بخش قصد بررسی PCA را داریم. (۳۰ نمره)

الف) ابتدا تابع PCATransformer را تکمیل کنید و داده ورودی را به ۲ کامپننت کاهش و نمایش دهید (مانند مثال خروجی).

ب) به طور مختصر توضیح دهید که معنای کامپننت در PCA چیست و زمانی که میگوییم داده به ۲ کامپننت کاهش یافته چه معنایی دارد.

ج) برای واریانس خروجی PCA کل فیچر ها نموداری مانند نمودار مثال رسم کنید و ترشلدی که تعداد کامپننت ها واریانسی بالای ۹۰ درصد دارند را مشخص کنید. تحلیل کنید که این ترشلد به چه معناست.

د) ۳۰ کامپننت اول را همانند خروجی نمایش دهید. ایگن وکتورهای یک تصویر بخش هایی رنگ روشن دارد. معنای آن چه می تواند باشد؟

ه) مانند خروجی تصاویر اصلی را با حالت بازگردانده شده ی PCA مقایسه کنید. تفاوت را برای دو مقدار دیگر بجز ترشلد بدست آمده مقایسه و تفاوت را بررسی کنید