



الگوشناسی آماری (CE-725)

دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف

تمرینات سری دوم - بهار ۱۳۸۹

به نکات زیر توجه فرمائید:


۱. زمان تحویل تمرینات در سایت درس مشخص شده است. دقت نمائید که زمانبندی‌های تعیین شده قابل تغییر نیستند.
۲. تمرینات را با عنوان SPR-HWx-8xxxxxxx (مثلا SPR-HW2-88300785) و در یک فایل فشرده با همین نام به آدرس Muhammadi@ce.sharif.edu ایمیل زده و در اولین جلسه بعد از زمان تحویل، بصورت پرینت شده تحویل استاد درس دهید.
۳. گزارش شما باید مختصر و مفید باشد. برای تمرینات پیاده‌سازی که با لوگوی  مشخص شده‌اند باید کد مطلب نوشته شده ضمیمه گزارش شده و تمامی خروجی‌های برنامه‌ها در گزارش شما ذکر شوند.

 **سوال ۱)** ماتریس FV (فایل mat). حاوی این ماتریس ضمیمه تمرین شده است) را در نظر بگیرید. این ماتریس حاوی ۵۱۹ داده ۱۳۹۹۶ بعدی می‌باشد. می‌خواهیم با استفاده از تحلیل مولفه اصلی به کاهش ابعاد این داده‌ها اقدام کنیم (یک نمونه کد PCA ضمیمه تمرین شده است. از princomp مطلب هم می‌توانید استفاده کنید).

الف) با استفاده از PCA بدون اینکه هیچ data loss داشته باشیم، می‌توانیم این ۱۳۹۹۶ بعد را به چند بعد کاهش دهیم؟
ب) با استفاده از PCA برای اینکه حداکثر ۱ درصد data loss داشته باشیم، می‌توانیم این ۱۳۹۹۶ بعد را به حداقل چند بعد کاهش دهیم؟

پ) با استفاده از PCA برای اینکه حداکثر ۵ درصد data loss داشته باشیم، می‌توانیم این ۱۳۹۹۶ بعد را به حداقل چند بعد کاهش دهیم؟

ت) اگر یکصد مولفه اصلی حاصل از PCA را به عنوان فضای یکصد بعدی جدید انتخاب کنیم، بردن داده‌ها به این فضای جدید، باعث چه مقدار data loss خواهد شد؟ (یا چه مقدار انرژی موجود در داده‌ها حفظ خواهد شد؟)

 **سوال ۲)** ماتریس FV (فایل mat). حاوی این ماتریس ضمیمه تمرین شده است) را در نظر بگیرید. این ماتریس حاوی ۵۱۹ داده ۱۳۹۹۶ بعدی می‌باشد. می‌خواهیم از بین ۱۳۹۹۶ بعد داده‌ها، ۱۰۰۰ بعد را انتخاب نمائیم.

الف) الگوریتم Relief کدام ابعاد را انتخاب می‌کند؟ (نتایج را در یک فایل mat ضمیمه کنید)

ب) روش DTM کدام ابعاد را انتخاب می‌کند؟ (نتایج را در یک فایل mat ضمیمه کنید)

پ) روش LVF کدام ابعاد را انتخاب می‌کند؟ (نتایج را در یک فایل mat ضمیمه کنید)

ت) چه تعداد از ابعاد توسط حداقل دو روش از سه روش فوق انتخاب شده‌اند؟

ث) چه تعداد از ابعاد توسط هر سه روش فوق انتخاب شده‌اند؟

ج) از مراحل (ت) و (ث) چه نتیجه‌ای می‌گیرید؟

سوال ۳ روی مجموعه داده ۱۵۰ تایی iris (فایل mat. حاوی این ماتریس ضمیمه تمرین شده است) PCA و MDA را اعمال کرده و تعداد ابعاد را از چهار بعد به دو بعد کاهش دهید (دو نمونه کد برای PCA و MDA، ضمیمه تمرین شده‌اند). در یک plot دو بعدی، داده‌های کاهش بعد یافته با PCA و در plot دیگر داده‌های کاهش بعد یافته با MDA را نمایش دهید. در هر دو مورد، داده‌های کلاس‌های مختلف را با رنگ‌ها یا علائم متفاوت نمایش دهید (در فایل mat. ضمیمه، لیبل کلاس هر داده نیز به شما داده شده است). با توجه به دو plot بدست آمده کدام یک دو مجموعه داده تغییر یافته را برای کلاسه‌بندی مناسب‌تر می‌دانید؟ آیا نتایج مشاهده شده با انتظارات شما از عملکرد PCA و MDA مطابقت دارند؟ چرا؟

سوال نمره اضافه: در کاربردهایی نظیر پردازش صورت تعداد ابعاد داده‌ها بسیار بالا خواهد شد. مثلاً هر تصویر کوچک شده 256×256 به یک بردار 65536 بعدی تبدیل خواهد شد. اعمال PCA بر روی این داده‌ها مشکلات عملی‌ای را بدنبال خواهد داشت (ماتریس کوواریانس فضای بسیار بالایی را نیاز خواهد داشت). چه راهکاری برای حل این مشکل می‌توانید ارائه دهید؟ دلیل اعتبار راهکار خود را توضیح دهید.

راهنمایی: برای مشاهده یک راهکار مناسب می‌توانید به مقاله Turk و Pentland تحت عنوان Face Recognition Using Eigenfaces مراجعه نمایید.