



قبل از اینکه شروع به حل تمرین‌ها کنید، حتماً یک بار فایل **Policies.pdf** (موجود در وبگاه) که حاوی نکات مهم در تحویل تکالیف هست را مطالعه فرمایید.
حل سولاتی که امتیازی هستند، اجباری نبوده و نمره‌ی اضافی خواهد داشت.
طراح هر یک از سولات در زیر مشخص شده است. در صورتی که سالی دارید می‌توانید سوالتان را در گروه مطرح کنید یا از دستیار درس بپرسید.
طراح: گلدانی

۱- به پرسش‌های مفهومی زیر پاسخ کامل دهید:

الف) در روش Feature selection ما عملاً یک سری از داده‌ها را دور می‌ریزیم. اگر فرض کنیم داده‌های ما نویزی نباشند، آیا باز هم این روش مفید خواهد بود.

ب) اضافه نمودن ویژگی اضافه بر روی عملکرد یک دسته بند چه تاثیراتی ممکن است ایجاد نماید.

ج) فرض کنید بخواهیم برای انتخاب ویژگی از معیار correlation تک تک ویژگی‌ها با خروجی استفاده نماییم؛ به این صورت که ویژگی که correlation بیشتری دارد انتخاب شود. آیا این معیار همیشه خوب عمل می‌کند؟

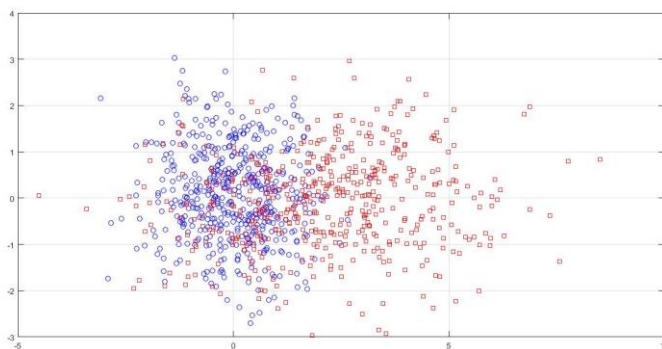
د) redundancy یا افزونگی ویژگی‌ها به معنای این است که دو ویژگی هرچند اثر بالایی داشته باشند کاملاً اثر یکسانی روی خروجی داشته باشند و به خاطر شباهت، وجود یکی از آنها کافی باشد؛ چه روشی برای تشخیص و حذف این ویژگی‌ها پیشنهاد می‌کنید.

۲- فرض کنید داده‌های مساله ما داده‌های "Fisher iris" باشند؛ این داده‌ها شامل ۴ اندازه حقیقی به عنوان ویژگی‌های سه نوع گل زنبق (iris) می‌باشد. می‌خواهیم با تکنیک feature selection دو ویژگی از میان این چهار ویژگی به گونه‌ای انتخاب نماییم که بهترین عملکرد را برای دسته بندی این سه نوع گل داشته باشند. یکی از روش‌هایی که برای انتخاب ویژگی استفاده می‌شود wrapper است در این روش در هر دور پس از انتخاب ویژگی‌ها یک دسته بند train کرده و داده‌ی validation ای را نیز روی آن تست می‌کنیم. در نهایت ویژگی‌هایی را برمی‌گزینیم که کمترین خطا را داشته باشند.

الف) با استفاده از روش wrapper دو ویژگی را انتخاب نمایید.

ب) (امتیازی) روشی دیگر برای انتخاب ویژگی پیشنهاد کنید. و عملکرد آن را بررسی کنید.

۳- در سوال ۷ ام سری نخست تمرین‌ها با استفاده از Bayesian Decision Boundary مرز تصمیمی برای جدا نمودن دو کلاس C_1 و C_2 پیدا نمودید حال سعی کنید با استفاده از داده‌های مربوط به آن سوال به پرسش‌های زیر پاسخ دهید:



شکل ۱

الف) سعی کنید با استفاده از linear SVM این مرز را ترسیم نمایید.
 ب) قسمت قبل را با استفاده از polynomial kernel و radial basis function kernel حل نمایید.
 ج) (امتیازی) آیا می‌توانیم برای حل این مساله از تکنیک LDA استفاده نماییم. مساله را با این تکنیک حل نمایید.

۴- تصویر آرم دانشگاه را به عنوان داده با ابعاد ۱۰۰ در ۱۰۰ در نظر بگیرید؛ قصد داریم با استفاده از تکنیک PCA مولفه‌های اساسی این تصویر را بیابیم. سپس این تصویر را با استفاده از مولفه‌های اساسی محدودی بسازیم (هر سطر تصویر یکی از داده‌ها در نظر گرفته می‌شود).

الف) برنامه‌ای بنویسید که با دریافت تصویر مولفه‌های اساسی تصویر را بیابد. (PCA را پیاده‌سازی کنید)
 ب) با استفاده از برنامه‌ای که نوشته‌اید تصویر را با ۵، ۱۰، ۵۰، ۸۰ مولفه اساسی بازسازی کنید.
 ج) تغییرات ایجاد شده با به کارگیری مولفه‌های اساسی بیشتر در قسمت قبل را بررسی کنید.

۵- ۱۸۹ داده تصویری شامل ۶ حالت مختلف حرکات دست مطابق شکل ۲ داریم در این سوال قصد داریم این حرکات را دسته بندی نماییم.^۱

الف) با استفاده از روش PCA ابتدا ابعاد داده های تصاویر را کاهش داده و سپس با استفاده از SVM این تصاویر را دسته بندی نمایید.

ب) مشخص نمایید با این دسته بندی چند درصد داده ها درست دسته بندی می شوند.

ج) ابعاد را به گونه ای کاهش دهید تا دسته بندی بیشترین دقت را داشته باشد.



شکل ۲

^۱داده های این مساله از Github.com گرفته شده است.