

تمرین چهارم درس یادگیری ماشین

برای پیاده‌سازی این تمرین باید از ابزار LIBSVM استفاده کنید. این ابزار را می‌توانید از این [لینک](#) دریافت کنید.

سوال ۱

در این بخش از مجموعه داده‌ی موجود در فایل data1.csv استفاده می‌شود. این مجموعه داده از دو ویژگی x_1 و x_2 و کلاس y تشکیل شده است. برای گزارش دقت (Accuracy) مدل در این سوال از 5-fold Cross Validation استفاده کنید (از امکانات LIBSVM می‌توانید استفاده کنید).

(الف) مجموعه داده را بارگذاری کرده و داده‌ها را با رنگ‌های مختلف برای هر کلاس نمایش دهید.

(ب) این داده‌ها به صورت خطی جداپذیر نیستند. تابع نگاشتی ارائه دهید که داده‌ها را از ۲ بعد به ۱ بعد برده تا به صورت خطی جداپذیر باشند. داده‌های نگاشت شده را نمایش دهید.

(ج) تابع نگاشتی ارائه دهید که داده‌ها را از ۲ بعد به ۲ بعد برده تا به صورت خطی جداپذیر باشند. داده‌های نگاشت شده را نمایش دهید.

(د) تابع نگاشتی ارائه دهید که داده‌ها را از ۲ بعد به ۳ بعد برده تا به صورت خطی جداپذیر باشند. داده‌های نگاشت شده را نمایش دهید (در ۳ بعد).

(ه) با در نظر داشتن توابع نگاشت در قسمت‌های قبل، برای هر کدام تابع کرنل را به صورت دستی به دست آورده و ماتریس کرنل را برای داده‌ها در هر قسمت محاسبه کنید (پیاده سازی).

(و) با استفاده از LIBSVM یک SVM خطی (بدون نگاشت) بر روی داده‌های اصلی آموزش داده و دقت مدل را گزارش کنید. چه نتیجه‌ای می‌گیرید؟

(ز) یک SVM خطی برای هر کدام از داده‌های نگاشت شده در قسمت‌های ب، ج و د آموزش داده و دقت مدل را گزارش کنید (۳ مدل). چه نتیجه‌ای می‌گیرید؟

(ح) یک SVM با کرنل از قبل محاسبه شده برای هر کدام از داده‌های اصلی با توجه به قسمت‌های ب، ج و د آموزش داده و دقت مدل را گزارش کنید. چه ارتباطی بین قسمت ز و ح وجود دارد؟ (از precomputed kernel در LIBSVM استفاده کنید).

سوال ۲

مجموعه داده‌ی Parkinsons را از این [لینک](#) دریافت کنید. این مجموعه داده را به سه بخش آموزش، ارزیابی و آزمایش تقسیم کنید. در صورت shuffle کردن داده‌ها این کار را یک بار انجام داده، ذخیره کرده و هربار همان داده‌ها را بارگذاری کنید تا مقایسه‌ی یکسان بین حالت‌های مختلف در قسمت‌های بعد صورت گرفته شود.

(الف) از ابزار LIBSVM برای آموزش SVM با کرنل Radial Basis Function استفاده کنید.

۱- تنظیم هایپرپارامترهای SVM برای هر کدام از کرنل‌ها نقش مهمی در آموزش مدل مناسب دارد. فضای این هایپرپارامترها را به صورت یک grid برای کرنل RBF جست و جو کرده و پس از انتخاب بهترین ترکیب با توجه به دقت حاصل بر روی مجموعه داده‌ی

ارزیابی، دقت مدل را بر روی مجموعه داده‌ی آزمایش به دست آورده و هر دو را گزارش کنید (برای این کرنل مقادیر γ و C را می‌توانید با تغییر توان در بازه‌ای مثل 10^{-10} تا 10^{10} برای هر یک از γ و C یا بازه‌های مختلف برای هر کدام مشخص کنید).

- ۲- تعداد ۲۰ نقطه را به طور تصادفی در فضای هایپرپارامترها انتخاب کرده و نتیجه‌ی آموزش را همانند بخش ۱ گزارش کنید.
- ۳- آیا می‌توانید یک روش هوشمندانه‌تر برای جست و جوی فضای هایپرپارامترها ارائه دهید؟ (یا از روش‌های هوشمندانه‌ی موجود استفاده کنید).

ب) قسمت الف را برای کرنل polynomial تکرار کنید. توجه کنید که هایپرپارامترهای این مدل می‌توانند متفاوت باشند. می‌توانید به توضیحات داخل کد LIBSVM مراجعه کنید.

ج) قسمت الف را برای کرنل sigmoid تکرار کنید.

د) مقدار پارامتر C چه تاثیری در مدل دارد؟ با توجه به فرمول SVM این تاثیر را شرح دهید. مقادیر بسیار بزرگ و بسیار کوچک (نزدیک صفر) برای این مقدار قرار داده و با یکی از کرنل‌ها آموزش داده و نتیجه را تحلیل کنید.

ه) تعداد بردارهای پشتیبان را برای مدل‌های آموزش داده شده توسط پارامترهای بهینه‌ای که یافته‌اید، گزارش کنید (در قسمت‌های الف، ب و ج).

سوال ۳

الف) آیا الگوریتم آداپوست در برابر داده نویز حساس است؟ چرا؟

ب) مجموعه داده‌ی دو بعدی زیر را در نظر بگیرید. می‌خواهیم یک دسته بند به کمک الگوریتم آداپوست بر روی این مجموعه داده آموزش دهیم. برای این کار الگوریتم آداپوست را با سه تکرار ($t=1,2,3$) انجام داده و پارامترهای $i = 1,2,..,8$ $w_t(i)$, α_t , ϵ_t را در هر تکرار بدست آورید. همچنین decision stump هر مرحله را نیز رسم نمایید.

Features			class
X1	-1	0	+
X2	-0.5	0.5	+
X3	0	1	-
X4	0.5	1	-
X5	1	0	+
X6	1	-1	+
X7	0	-1	-
X8	0	0	-

فایل گزارش را به فرمت HW4_Report_95131062.pdf آماده کرده و به همراه بقیه فایل‌های مورد نیاز (کد و ...) در یک فایل فشرده به فرمت HW4_95131062.zip قرار داده و در سایت درس بارگذاری کنید. فرصت بارگذاری این تمرین، تا ساعت ۲۳:۵۵ دوشنبه ۱۱ دی می‌باشد.

ایمیل درس: ceitml17@gmail

موفق باشید