

## تکلیف شماره ۲-۲ یادگیری ماشین: بخش عملی Bayes و naïve Bayes

در این بخش با دادگان Iris باید الگوریتم Bayes و naïve Bayes را پیاده‌سازی کنید. برای پیاده‌سازی می‌توانید از نرم‌افزار MATLAB یا پایتون استفاده کنید، ولی نباید از دستورات آماده برای طراحی طبقه‌بند استفاده کنید. دادگان را می‌توانید از لینک زیر دانلود کنید:  
<https://archive.ics.uci.edu/ml/datasets/Iris>

دادگان داده شده شامل ۱۵۰ داده با بعد ۳ (۳ ویژگی) در ۳ کلاس است، که در هر کلاس ۵۰ داده وجود دارد.

حال مراحل زیر را انجام دهید:

الف) یک PDF گوسی برای متغیرهای چند بعدی با استفاده از دادگان بسازید. برای این کار، دادگان را بصورت تصادفی به ۲ بخش آموزش شامل ۷۰ درصد داده‌ها و تست شامل ۳۰ درصد باقیمانده تقسیم کنید. طبقه‌بندی بیز را با در نظر گرفتن توزیع پیشینه یکنواخت (مساوی) برای تمام کلاسها تولید کنید (تخمین ML). این انتخاب تصادفی و آموزش را ۳ بار تکرار کنید. میزان صحت طبقه‌بندی متوسط روی داده‌های آموزش و تست (متوسط روی کل کلاسها) را هم در حالت هر یک از ۳ تکرار، و هم متوسط هر ۳ حالت، گزارش کنید. همچنین ماتریس confusion را پیدا کنید.

ب) بخش الف) را این بار با 4-fold cross validation تکرار کنید، یعنی هر بار ۷۵ درصد داده‌ها در آموزش دخالت دارند، و ۲۵ درصد در تست. نتیجه صحت متوسط طبقه‌بندی نهایی را گزارش کنید، و با حالت قبل مقایسه کنید. فکر می‌کنید با افزایش K در K-fold تغییری حاصل خواهد شد؟ همچنین ماتریس confusion را پیدا کنید.

پ) Gaussian naïve Bayes: الگوریتم Gaussian naïve bayes را روی دادگان آموزش بخش الف) پیاده کنید. میزان صحت متوسط طبقه‌بندی را روی مجموعه‌های آموزش و تست محاسبه کنید، و با بخش الف) مقایسه انجام دهید. همچنین ماتریس confusion را پیدا کنید.

ت) Gaussian naïve Bayes: الگوریتم Gaussian naïve bayes را روی دادگان بخش ب) با 4-fold cross validation پیاده کنید. میزان صحت متوسط طبقه‌بندی را روی مجموعه‌های آموزش و تست محاسبه کنید، و با بیز در ب) و حالت قبل مقایسه انجام دهید. همچنین ماتریس confusion را پیدا کنید.

ث) نمره تشویقی: ماتریس confidence را در بهترین نتیجه حاصل در ۴ بخش گذشته محاسبه کنید.

ج) نمره تشویقی: با کتابخانه‌های موجود، صحت الگوریتم درخت تصمیم مناسب و ماتریس confusion را روی این دادگان در حالت ب) پیدا کنید و با نتایج این تمرین مقایسه کنید.

گزارش: برنامه‌های نوشته شده در MATLAB یا پایتون به همراه نتایج خواسته شده در هر مرحله را ضمیمه کنید. همچنین یک بحث و بررسی در مورد نتایج و مشخص کردن طبقه‌بند بهتر انجام دهید. گزارش شامل تحلیل و مقایسات و توضیحات کافی در مورد نحوه پیاده‌سازی‌ها، لیست برنامه‌ها و پارامترهای بکار رفته می‌باشد. لطفاً از نوشتن گزارش‌های طولانی اجتناب کنید.

توجه: کل تمرین و گزارش باید بصورت انفرادی نوشته شود، و به برنامه‌ها و گزارش‌های کپی شده نمره‌ای تعلق نمی‌گیرد.

توجه: به هیچ وجه از کدهای آماده یا کتابخانه‌های آماده برای طراحی طبقه‌بند (بجز بخش ج برای درخت تصمیم) استفاده نکنید. ولی برای محاسبه توزیع گوسی در حالت تک‌متغیره یا چندمتغیره مجاز به استفاده از دستورات موجود در کتابخانه‌ها هستید.

نمره تشویقی اضافه: بخش (ث)، بخش (ج)، تحلیل درست نتایج و خصوصاً ماتریس confusion و اینکه آیا نتایج در همه کلاسها یکسان هستند یا خیر، نمودارهای کامل، کد صحیح با انتخاب پارامترهای درست برای آموزش، مقایسه با نتایج تمرین قبلی

موعد تحویل: جمعه ۹ خرداد ۱۳۹۹

روش تحویل: تحویل در سایت courses

موفق باشید

سیدین