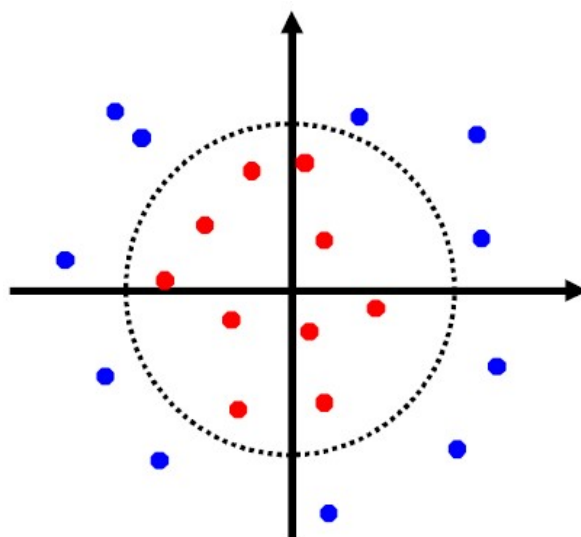


## تکلیف شماره ۴ یادگیری ماشین: SVM

۱- دادگان در ۲ دسته 1 و -1 با استفاده از مختصات قطبی طوری بسازید که منحنی جداساز یک دایره باشد، مثلاً:



توجه کنید که خودتان دادگان آموزش و تست را با همین ساختار بسازید.

الف) با الگوریتم SVM خطی و مجموعه دادگان train یک یادگیر SVM بسازید. میزان صحت طبقه‌بندی را روی مجموعه‌های train و test محاسبه کنید. سپس منحنی حاشیه‌ها و مرز تصمیم را رسم کنید، و روی آن support vector ها را مشخص نمایید.

ب) با الگوریتم SVM غیرخطی در حالت soft SVM یک یادگیر SVM بسازید. پارامتر soft SVM را با استفاده از روش cross-validation تنظیم کنید تا بهترین میزان صحت را روی مجموعه validation بدهد. میزان صحت طبقه‌بندی را روی مجموعه‌های train و test در این بهترین مقدار انتخابی محاسبه کنید. سپس منحنی حاشیه‌ها و مرز تصمیم را رسم کنید، و روی آن support vector ها را مشخص نمایید.

پ) با الگوریتم SVM غیرخطی در حالت کرنل یک یادگیر SVM بسازید. کرنل را RBF و کرنل چندجمله‌ای درجه ۲ و ۳ انتخاب کنید. می‌توانید انتخاب بهترین کرنل را براساس میزان صحت روی مجموعه validation به روش cross-validation انجام دهید. میزان صحت طبقه‌بندی را روی مجموعه‌های train و test در این بهترین کرنل انتخابی محاسبه کنید. سپس منحنی حاشیه‌ها و مرز تصمیم را رسم کنید، و روی آن support vector ها را مشخص نمایید.

۲- در این بخش با دادگان داده شده در تمرین ۱ (Adult) باید الگوریتم SVM را در چند حالت پیاده سازی کنید. برای پیاده سازی می‌توانید از نرم افزار متلب یا پایتون استفاده کنید:

مراحل زیر را انجام دهید:

الف) با الگوریتم SVM خطی و مجموعه دادگان train داده شده یک یادگیر SVM بسازید. میزان صحت طبقه‌بندی را روی مجموعه‌های train و test محاسبه کنید. سپس تعداد support vector ها را مشخص نمایید.

ب) با الگوریتم SVM غیرخطی در حالت soft SVM یک یادگیر SVM بسازید. پارامتر soft SVM را با استفاده از دادگان validation که خودتان از مجموعه آموزش می‌سازید، تنظیم کنید تا بهترین میزان صحت را روی مجموعه validation بدهد. منحنی تغییرات صحت طبقه‌بندی روی دادگان validation براساس مقدار C را بکشید. میزان صحت طبقه‌بندی را روی مجموعه‌های train و test در این بهترین مقدار انتخابی محاسبه کنید. سپس تعداد support vector ها را مشخص نمایید و با حالت قبل مقایسه کنید.

پ) با الگوریتم SVM غیرخطی در حالت کرنل یک یادگیر SVM بسازید. کرنل را RBF و کرنل چندجمله‌ای انتخاب کنید. می‌توانید انتخاب بهترین کرنل را براساس میزان صحت روی مجموعه validation انجام دهید. میزان صحت طبقه‌بندی را روی مجموعه‌های train و test در این بهترین کرنل انتخابی محاسبه کنید. سپس تعداد support vector ها را مشخص نمایید.

ت) با الگوریتم SVM غیرخطی در حالت ترکیبی کرنل و soft SVM یک یادگیر SVM بسازید. کرنل را بهترین کرنل مرحله قبل انتخاب کنید. پارامتر soft SVM را با استفاده از دادگان validation تنظیم کنید تا بهترین میزان صحت را روی مجموعه validation بدهد. میزان صحت طبقه‌بندی را روی مجموعه‌های train و test در این بهترین مقدار محاسبه کنید. سپس تعداد support vector ها را مشخص نمایید.

گزارش: برنامه های نوشته شده به همراه نتایج و شکل های خواسته شده در هر مرحله را ضمیمه کنید. همچنین یک بحث و بررسی در مورد نتایج و اثر استفاده از SVM غیرخطی انجام دهید. گزارش شامل تحلیل و مقایسات و توضیحات کافی در مورد نحوه پیاده سازی ها، لیست برنامه ها و پارامترهای بکار رفته می‌باشد. لطفاً از نوشتن گزارشهای طولانی اجتناب کنید.

توجه: می‌توانید از دستورات و کتابخانه مرتبط در متلب یا پایتون استفاده کنید.

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

نمره مثبت اضافی: تحلیل کامل نتایج، مقایسه با نتایج تمرین شماره ۱ یعنی درخت تصمیم.

موعد تحویل: ۱۸ مرداد ۱۳۹۹

موفق باشید

سیدین