

بنام آنکه عزت از آن اوست



دانشگاه صنعتی امیرکبیر  
(پلی تکنیک تهران)  
دانشکده ریاضی و علوم کامپیوتر

درس یادگیری ماشین  
تمرین شماره دو

استاد محترم  
جناب آقای دکتر شیری قیداری

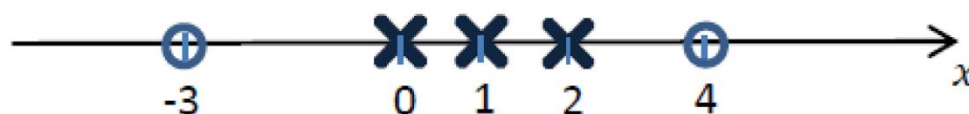
آذر 1399

در این تمرین شما میتوانید از کمک افراد مجرب، اینترنت و یا هر منبع مناسب دیگری بهرمنند شوید، با این شرط که منابع دقیقاً ذکر شده و این استفاده بصورت غیر مستقیم باشد. در این راستا هر گونه کپی برداری یا واگذاری تمرین به اشخاص دیگر تقلب محسوب شده و پیامد آن نه تنها نمره تمرین ذیل، بلکه نمره سایر تمارین و نمره پایانی شما را نیز تحت تاثیر قرار میدهد.

فایلی که برای حل تمرین ارائه میدهید باید شامل گزارش شخصی شما با فرمت PDF بعلاوه کدهای مربوطه باشد. این فایل را با نام خودتان، شماره دانشجویی، شماره تمرین (به عنوان مثال Sheykhi.9326169.HW7) به صورت فایل زیپ شده، در زمان مقرر ارسال نمایید. هر روز دیرکرد شما در ارسال این فایل، موجب از دست دادن 10% از نمره اصلی خواهد شد.

موفق باشید.

۱- در شکل زیر داده های آموزشی مشاهده می شود. به داده های ضربدر خروجی 1- و به داده های دایره خروجی 1 نسبت داده شده است.

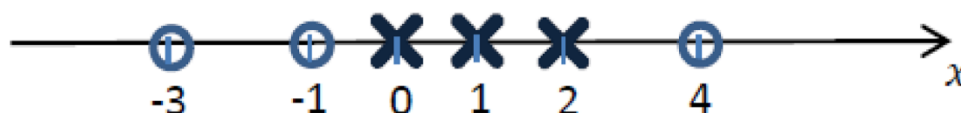


مشخص است داده ها بطور خطی جدا پذیر نیستند، اما با یک تبدیل مناسب (و یا هسته kernel مناسب) قابل دسته بندی می باشند. الف) یک تبدیل مناسب  $\phi(x)$  برای قابل جداسازی شدن داده های مذکور در یک بعد بیابید.

ب) نقاط مذکور را در دستگاه تبدیل (انتقال) یافته مربوط به قسمت الف رسم نمایید. سپس مرز جداسازی SVM و بردارهای پشتیبان را رسم کنید.

ج) با تبدیل معکوس، مرز جداسازی SVM و بردارهای پشتیبان را در دستگاه اولیه (شکل بالا) رسم کنید. (این جدا کننده دیگر خطی نخواهد بود)

د) داده آموزشی جدیدی به مجموعه داده ها طبق شکل زیر اضافه می کنیم:



ه) آیا تبدیل  $\phi(x)$  برای جداسازی این داده ها بطور خطی هنوز کارا است؟

و) حال داده های در فضای دوبعدی جدید  $(x, \phi(x))$  را در نظر بگیرید. این ۶ داده را به همراه خط جدا ساز SVM در فضای دوبعدی رسم نمایید و بردارهای پشتیبان را مشخص کنید.

ز) مرز جداسازی در قسمت (و) به فرم  $w_0 + w_1x + w_2\phi(x) = 0$  خواهد بود. مقادیر  $w_0$ ،  $w_1$  و  $w_2$  را بیابید.

ح) هسته  $K(x, x')$  مربوط به ضرب داخلی در فضای تبدیل یافته را مشخص کنید، بطوریکه  $x$  و  $x'$  دو نقطه از فضای ویژگی اولیه (ورودی) و  $(x, \phi(x))$  و  $(x', \phi(x'))$  داده های فضای تبدیل یافته هستند.

- ۲- (پیاده سازی) یک مجموعه داده دو کلاسی مناسب به انتخاب خود از مجموعه داده‌گان UCI دانلود کنید. داده‌ها را به سه دسته آموزشی، ارزیابی (برای تنظیم پارامترها) و تست تقسیم کنید. با استفاده از مجموعه داده انتخابی، SVM را مورد بررسی قرار دهید.
- الف) خطای SVM را برای داده آموزشی و داده تست و داده‌های ارزیابی بررسی نمایید. (رسم نمودار)
- ب) تعداد بردارهای پشتیبان را مشخص کنید.
- ج) با تغییر اندازه داده آموزشی تاثیر آنرا در نتیجه بررسی و در صورت مشاهده over-fitting آن را گزارش کنید.