یادگیری ماشین برای بیوانفورماتیک

نیم‌سال دوم ۹۸ـ۹۹

دانشکده‌ی مهندسی کامپیوتر

گردآورنده: پری‌شاد بهنام قادر ــ امین میرزایی

تمرین دوم — پرسپترون و ماشین بردار پشتیبان

مسأله‌ی ۱.

یک دسته بند پرسپترون در حال پیمایش و آموزش روی داده هایی با ترتیب مشخص می باشد. پس از دیدن داده i ام پارامتر های آن به این شکل درآمده است: $(w_۰, w_۱, w_۲) = (۲, ۱, ۱)$

الف) مرز تصمیم این دسته بند را رسم کنید.

ب) فرض کنید داده $i+۱$ دارای مختصات $(x = ۲, y = ۱)$ و برچسب $+۱$ می باشد و داده ی $i+۲$ دارای مختصات $(x = ۳, y = ۲)$ و برچسب -۱ می باشد. مراحل به روز رسانی پارامترهای مدل و تغییرات مرز دسته بند را پس از دیدن هر کدام از داده های $i+۱$ ام و $i+۲$ ام با رسم شکل نشان دهید.

مسأله‌ی ۲.

نشان دهید که ترتیب داده ها در بردار وزن حاصل از نسخه ی تک نمونه ی الگوریتم پرسپترون می‌تواند اثرگذار باشد. با مثال نشان دهید اگر الگوریتم پرسپترون را روی داده‌ها اجرا کنیم و در هر چرخه یک داده را بررسی کنیم و بردار وزن را بروزرسانی کنیم، ترتیب بررسی داده‌ها در زمان آموزش، در بردار نهایی می‌تواند اثر بگذارد.

مسأله‌ی ۳.

فرض کنید مسأله‌ی دسته‌بندی با الگوریتم پرسپترون را برای داده‌هایی از دو کلاس حل کرده‌ایم و بردار w^* نتیجه شده‌است به طوری که همه‌ی داده‌ها را به درستی با حاشیه‌ی γ دسته‌بندی می‌کند یعنی داریم

$$\forall i,\ w^{*T} x_i y_i > \gamma$$

با دانستن این حقیقت که تمام داده‌ها در ابرکره‌ای با شعاع R قرار دارند، ثابت کنید تعداد گام‌های لازم برای رسیدن به این بردار نهایی حداکثر $\frac{R^۲\|w^*\|_۲^۲}{\gamma^۲}$ گام بوده است. (از استقرا روی بردار در هر گام استفاده کنید و فرض کنید بردار وزن اولیه بردار تماما صفر باشد.)

مسأله‌ی ۴.

مسأله‌ی SVM حاشیه نرم با نُرم ۲ به صورت مسأله‌ی بهینه‌سازی زیر تعریف می‌شود.

$$\min_{w,b,\zeta} \frac{۱}{۲}\|w\|^۲ + \frac{C}{۲}\sum_{i=۱}^{N} \zeta_i^۲$$
$$s.t.\ y^{(i)}(w^T x^{(i)} + b) \geqslant ۱ - \zeta_i,\ i = ۱, \ldots, N$$

آ) نشان دهید شرط $\zeta_i \geqslant ۰$ در این مسأله در جواب مسأله‌ی بهینه‌سازی بی‌تأثیر است.

ب) این عبارت را با استفاده از ضرایب لاگرانژ بازنویسی و سپس با استفاده از مشتق‌گیری نسبت به پارامترهای w، b و ζ_i عبارت لاگرانژی را بهینه کنید.

ج) حال با استفاده از بخش قبل فرم دوگان این نوع از مسأله‌ی SVM را به دست آورید.

مسأله‌ی ۵.

در فرم دوگان مسأله‌ی SVM دیدید می‌توان از بردارهای ویژگی پیچیده‌تر x به جای خود x استفاده کرد. هم‌چنین با مفهوم کرنل آشنا شدید و مشاهده کردید که برای دسته‌بندی‌های غیرخطی به جای استفاده از ضرب داخلی بردارهای ویژگی داده‌ها، مستقیما از تابع کرنل استفاده می‌کنیم. به چنین تابعی یک کرنل معتبر گفته می‌شود.

یکی از روش هایی که برای اثبات معتبر بودن یک کرنل به کار می‌رود استفاده از قانون زیر است:
یک کرنل معتبر است اگر و تنها اگر ماتریس گرم[1] آن مثبت نیمه معین باشد.
و منظور از کرنل معتبر کرنلی است که بتوان آن را به شکل رو به رو نوشت

$$k(x,y) = \phi(x)^T \phi(y)$$

حال با استفاده از قانون بالا ثابت کنید که کرنل RBF یک کرنل معتبر است.

مسأله‌ی ۶.

برای هر مجموعه دادگانی که در شکل ۱ به تصویر کشیده شده است مشخص کنید کدام کرنل‌ها می‌توانند داده‌ها را بدون خطا دسته‌بندی کنند.

۱. $k(x_۱, x_۲) = x_۱^T x_۲$
۲. $k(x_۱, x_۲) = e^{-۱۰\|x_۱ - x_۲\|^۲}$
۳. $k(x_۱, x_۲) = (۱ + x_۱^T x_۲)^۲$
۴. $k(x_۱, x_۲) = e^{-۰/۱\|x_۱ - x_۲\|^۲}$

مسأله‌ی ۷.

با فرض معتبر بودن کرنل‌های $k_۱$ و $k_۲$، معتبر بودن کرنل‌های زیر را ثابت کنید.

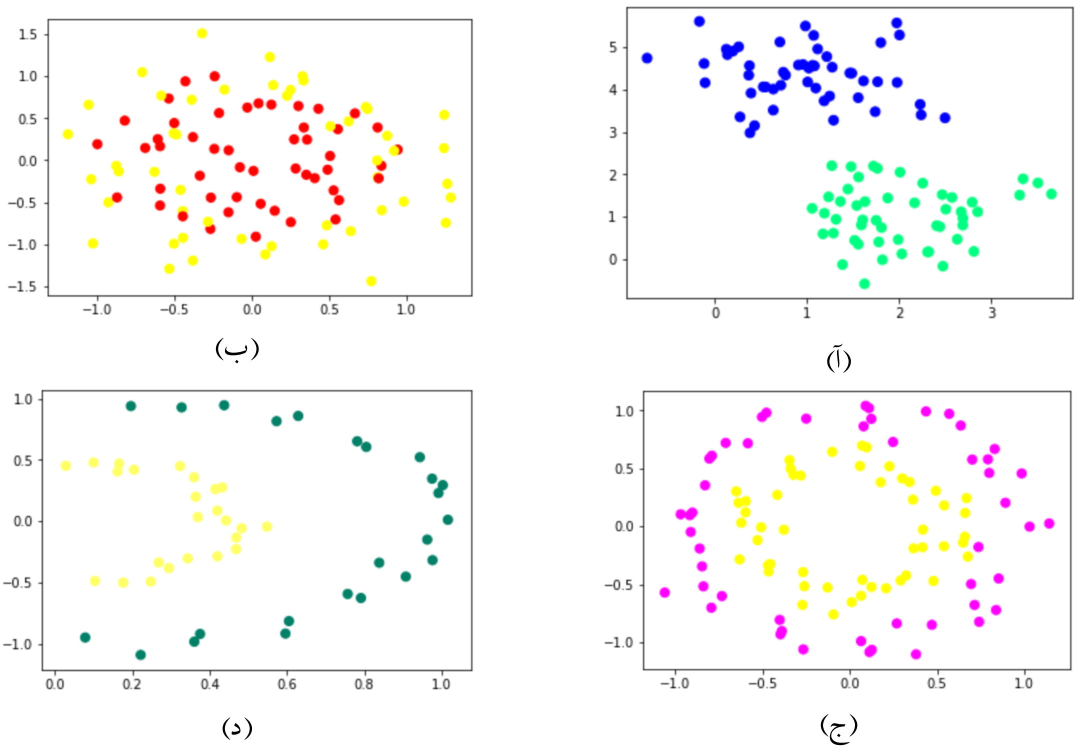
۱. $k_۳(x_۱, x_۲) = k_۲(x_۱, x_۲) + k_۲(x_۱, x_۲)$
۲. $k_۴(x_۱, x_۲) = k_۲(x_۱, x_۲) k_۲(x_۱, x_۲)$
۳. $k_۵(x_۱, x_۲) = e^{k_۱(x_۱, x_۲)}$
۴. $k_۶(x_۱, x_۲) = \frac{۱}{۱ - x_۱^T x_۲}$

مسأله‌ی ۸.

بسیاری از الگوریتم‌های دسته‌بندی دارای فرم مبتنی بر کرنل هستند. در این سوال به بررسی روش پرسپترون می‌پردازیم.

آ) با فرض این‌که از بردار تماما صفر برای بردار وزن w آغاز کرده باشیم، اگر α_i تعداد دفعاتی باشد که داده‌ی iـ‌ام اشتباه دسته‌بندی شده است، رابطه‌ی بروزرسانی وزن‌ها را با استفاده از α_i بازنویسی کنید.

[1] Gram matrix



شکل ۱: مجموعه دادگان مسأله‌ی ۶

ب) می‌دانیم در این الگوریتم دسته ی هر داده با محاسبه‌ی $sign(w^T x)$ به دست می‌آید. این رابطه را با استفاده از بخش قبل بازنویسی کنید و نشان دهید با استفاده از روش مبتنی بر کرنل دیگر نیازی به محاسبه‌ی فضای ویژگی $\phi(x)$ نیست.

مسأله‌ی ۹.

کرنل‌ها تنها در فضای بردارها استفاده ندارند.

۱. ثابت کنید در فضای مجموعه‌ها $k(A, B) = 2^{|A\cap B|}$ یک کرنل معتبر است.

۲. $\hat{X}$ را مجموعه‌ی تمام زیرمجموعه‌های متناهی X در نظر بگیرید. ثابت کنید اگر k یک کرنل معتبر روی $X \times X$ باشد، $\hat{k}(A, B) = \sum_{x_1 \in A, x_2 \in B} k(x_1, x_2)$ یک کرنل معتبر روی $\hat{X} \times \hat{X}$ است.

مسأله‌ی ۱۰ (امتیازی).

ثابت کنید هر دادگان حاوی تعداد متناهی داده با کرنل گاوسی $k(x_1, x_2) = e^{-\frac{\|x_1 - x_2\|^2}{2\sigma^2}}$ تفکیک‌پذیر است اگر برای مقدار σ حق انتخاب داشته باشیم. همچنین در ارتباط با تعداد بردارهای پشتیبان وقتی $\sigma \to 0$ صحبت کنید.

نکات مهم

- بخش تئوری را در قالب یک فایل pdf با اسم HW2_STD-Num آپلود کنید.
- ددلاین تمرین ساعت ۲۳:۵۹ روز ۱۱ فروردین می‌باشد.