

توجه: استفاده از کتاب، جزوه، اسلایدهای درس، موبایل و تبلت در حین امتحان غیر مجاز است و تقلب محسوب می شود.

توجه: امتحان از ۱۱۰ نمره است و برای کامل شدن باید ۱۰۰ نمره کسب نمایید.

سوال ۱ سوالات چند گزینه‌ای (۲۱ نمره)

در هر یک از سوالات زیر، گزینه‌های صحیح را مشخص کنید. توجه داشته باشید که ممکن است چند گزینه صحیح باشد و باید همه‌ی آن‌ها را مشخص کنید. نیازی به هیچ توضیحی نیست. (هر سوال ۳ نمره)

الف) یک مدل دسته‌بند دودویی آموزش داده‌اید که دقت بسیار بالایی در داده‌های آموزش دارد، اما در داده‌های اعتبارسنجی^۱ دقت بسیار کمتری دارد. از میان موارد زیر، کدام گزینه‌ها ممکن است درست باشد:

- (۱) این یک نمونه از بیش‌برازش^۲ است
 - (۲) این یک نمونه از کم‌برازش^۳ است.
 - (۳) آموزش به خوبی regularized نشده است.
 - (۴) نمونه‌های آموزش و آزمون از توزیع‌های متفاوتی نمونه‌برداری شده‌اند.
- ب) بردارهای پشتیبان چیست؟

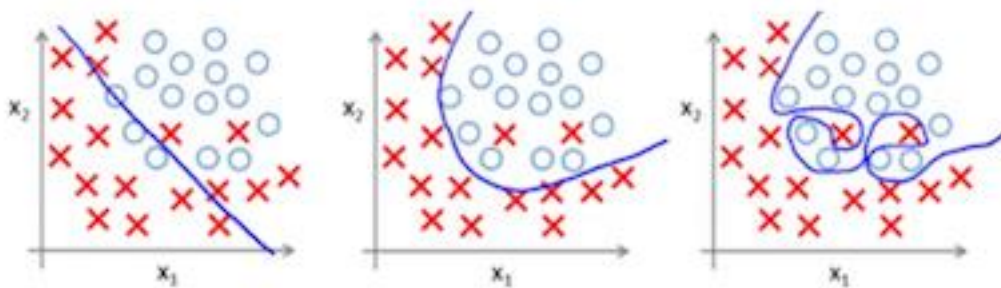
- (۱) نمونه‌هایی هستند که دورترین فاصله را از مرز تصمیم مشخص می‌کنند.
- (۲) تنها نمونه‌های لازم برای محاسبه $f(x)$ در SVM هستند.
- (۳) مراکز کلاس‌ها هستند.
- (۴) تمام نمونه‌هایی هستند که وزن غیر صفر (α_k) در SVM دارند.

ج) کدام یک از موارد زیر تنها در صورت جداپذیری خطی داده‌های آموزش قابل استفاده است؟

- (۱) Linear hard-margin SVM
- (۲) Linear Logistic Regression
- (۳) Linear Soft margin SVM
- (۴) The centroid method
- (۵) Parzen windows

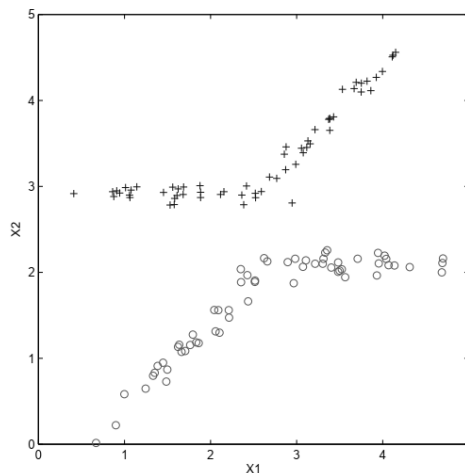
^۱ validation
^۲ overfitting
^۳ underfitting

د) سه دسته‌بند مختلف بر روی داده‌های یکسان آموزش داده شده‌اند. مرز تصمیم آن‌ها در زیر نشان داده شده است. کدامیک از گزاره‌های زیر درست است؟



- ۱) دسته‌بند سمت چپ با دقت بالا و انعطاف پذیری کمی آموزش داده شده است.
- ۲) دسته‌بند سمت چپ با دقت پایین و انعطاف پذیری بالا آموزش داده شده است.
- ۳) دسته‌بند سمت راست با دقت پایین و انعطاف پذیری بالا آموزش داده شده است.
- ۴) دسته‌بند سمت راست با دقت بالا و انعطاف پذیری کمی آموزش داده شده است.

ه) مجموعه‌ی داده‌ی زیر را برای یک مسأله طبقه‌بندی دو کلاسه در نظر بگیرید. اگر بخواهیم از مدل مخلوط گاوسی (GMM) برای هر یک از دو کلاس استفاده کنیم، چه تعداد مولفه را برای هر کلاس مناسب می‌دانید؟



- ۱) یک مولفه با ماتریس کوواریانس قطری
- ۲) یک مولفه با ماتریس کوواریانس دلخواه
- ۳) دو مولفه با ماتریس کوواریانس قطری
- ۴) دو مولفه با ماتریس کوواریانس دلخواه

و) اگر $k_1(x, y)$ و $k_2(x, y)$ دو تابع هسته (kernel) معتبر برای روش $kernel-SVM$ باشند، آنگاه کدامیک از موارد زیر نمی‌تواند یک تابع هسته معتبر باشد؟

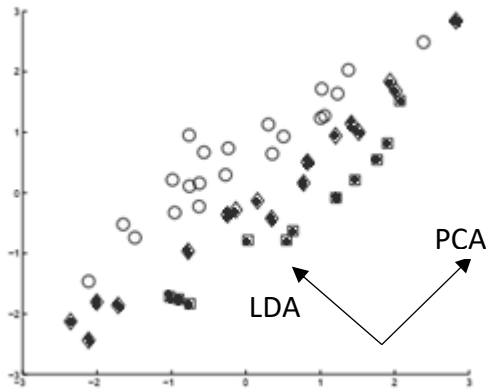
$$k_1(x, y) + k_2(x, y) \quad (۱)$$

(۲) $a > 0$ به ازای $k_1(x, y)$

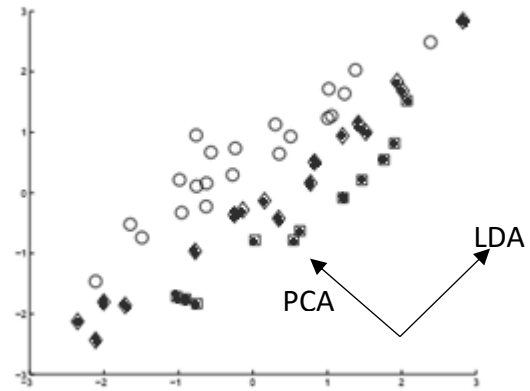
(۳) $k_1(x, y)k_2(x, y)$

(۴) $k_1(x, x) - k_2(y, y)$

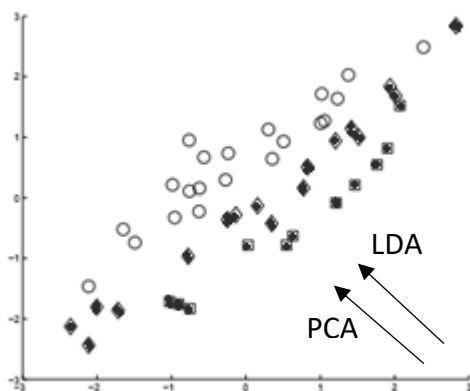
ز) فرض کنید داده‌های دو کلاس در فضای دو بعدی باشند. کدام گزینه جهت درست مولفه اول PCA و LDA را نشان می‌دهد.



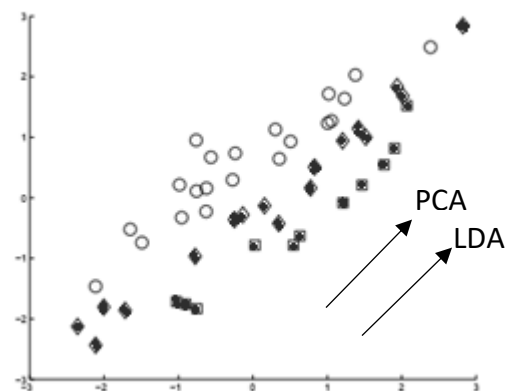
(۲)



(۱)



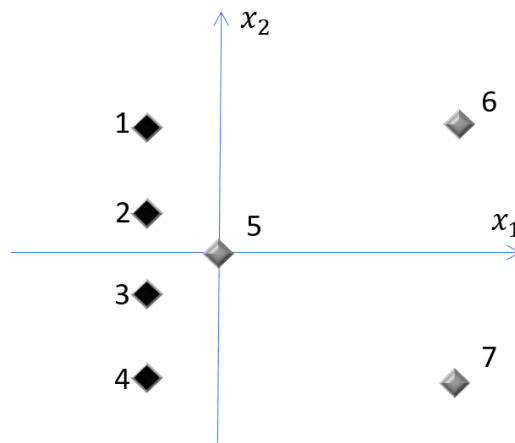
(۴)



(۳)

سوال ۲ دسته‌بندی به روش مرکز کلاس^۱ (۱۰ نمره)

یک مسئله دسته‌بندی دو کلاسه در یک فضای ویژگی دو بعدی $x = [x_1, x_2]$ با برچسب $y = \pm 1$ را در نظر بگیرید. داده‌های آموزشی شامل ۷ نمونه است که در شکل زیر نشان داده شده است. ۴ الماس سیاه برای کلاس مثبت (نمونه‌های ۱ تا ۴) و ۳ الماس سفید برای کلاس منفی (نمونه‌های ۵ تا ۷).



الف (۴ نمره) در شکل بالا، مراکز^۱ دو کلاس را رسم کنید (آن‌ها را با یک دایره با علامت (+) داخل آن، برای کلاس مثبت و یک دایره با علامت (-) برای کلاس منفی مشخص کنید). سپس مراکز را با یک خط چین ضخیم به هم وصل کنید. در نهایت مرز تصمیم روش مرکز کلاس را با یک خط پررنگ صلب رسم کنید.

ب (۲ نمره) نرخ خطای آموزش چند است؟

ج (۲ نمره) آیا نمونه‌ای وجود دارد که با حذف آن، مرز تصمیم به گونه‌ای تغییر کند که نمونه حذف شده به سمت دیگر مرز جداکننده برود؟ (با "بله" یا "خیر" پاسخ دهید)

د (۲ نمره) نرخ خطای leave-one-out چند است؟

سوال ۳ SVM (ماشین بردار پشتیبان) (۱۵ نمره)

مجددا داده‌های آموزشی موجود در شکل قبل را در نظر بگیرید.

الف (۳ نمره) مرز تصمیم حاصل از روش SVM خطی hard margin را با یک خط صلب پهن رسم کنید. حاشیه‌ها را به هر دو طرف با خط‌چین‌های نازک مشخص کنید. بردارهای پشتیبان را با دایره مشخص کنید.

ب (۲ نمره) نرخ خطای آموزش چند است؟

ج (۲ نمره) حذف کدام نمونه منجر به تغییر مرز تصمیم می‌شود؟

د (۲ نمره) نرخ خطای leave-one-out چند است؟

ه (۲ نمره) انعطاف پذیرتری روش مرکز کلاس (سوال قبل) را با SVM مقایسه کنید.

و (۴ نمره) قسمت (الف) را در حالتی که از soft-margin SVM استفاده کنیم، پاسخ دهید.

سوال ۴ Ensemble Learning (۱۹ نمره)

برای یک مسأله رگرشن، K تا مدل از پیش آموزش دیده در اختیار داریم. می‌خواهیم با ترکیب این مدل‌ها به صورت زیر یک مدل جدید بسازیم:

$$g(x) = \sum_{k=1}^K w_k f_k(x)$$

که در آن f_k ها مدل‌های رگرسور از پیش آموزش دیده و w_k ها وزن هر یک از مدل‌ها هستند. همه w_k ها مثبت و حاصل جمعشان ۱ است.

الف) (۱۵ نمره) اگر خطای MSE مدل f_k برابر با e_k باشد و خطای MSE مدل g را با e_g نشان دهیم، ثابت کنید:

$$e_g \leq \sum_{k=1}^K w_k e_k$$

که در رابطه‌ی بالا داریم:

$$e_k = \frac{1}{n} \sum_{i=1}^n (f_k(x_i) - y_i)^2$$

راهنمایی: از نامساوی زیر می‌توانید استفاده کنید:

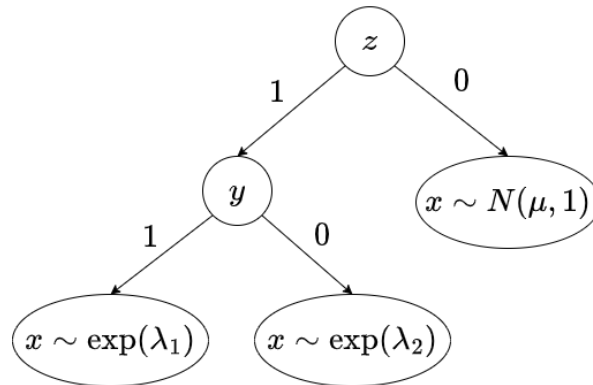
اگر w_k ها مثبت و حاصل جمعشان ۱ باشد، آنگاه داریم:

$$\left(\sum_{k=1}^K w_k a_k \right)^2 \leq \sum_{k=1}^K w_k a_k^2$$

ب) (۴ نمره) در قسمت قبل نشان دادیم که خطای مدل حاصل کمتر از میانگین خطای مدل‌های اولیه است. به نظر شما این کاهش خطا ناشی از کاهش خطای واریانس است یا خطای بایاس یا هر دو؟ توضیح دهید.

سوال ۵ (۳۰ نمره)

فرض کنید متغیر تصادفی $x \in \mathbb{R}$ مطابق درخت زیر تولید می‌شود:



که در شکل بالا z و y متغیرهای باینری و مستقل از یکدیگر هستند. اگر احتمال ۱ بودن متغیرهای z و y به ترتیب برابر با α و β باشد، آنگاه توزیع احتمال توأم هر سه متغیر x ، y و z به صورت زیر خواهد بود:

$$P(x, y, z) = \left((1 - \alpha) \frac{1}{2\pi} \exp\left(-\frac{(x - \mu)^2}{2}\right) \right)^{1-z} \left(\alpha((1 - \beta)\lambda_2 \exp(-\lambda_2 x))^{1-y} (\beta\lambda_1 \exp(-\lambda_1 x))^y \right)^z$$

فرض کنید مجموعه داده‌ی $\{x_1, \dots, x_n\}$ را که به صورت i.i.d. هستند، در اختیار داریم. حال می‌خواهیم با استفاده از روش EM، پارامترهای توزیع بالا را تخمین بزنیم.

الف) (۵ نمره) تابع لگاریتم complete likelihood را بدست آورید.

ب) (۱۵ نمره) امید ریاضی^۱ تابع لگاریتم complete likelihood را نسبت به متغیرهای پنهان (یعنی z_i و y_i) بدست آورید. توجه داشته باشید که امید ریاضی را باید به شرط متغیر مشاهده شده (یعنی x_i) محاسبه کنید.

$$\text{راهنمایی: } \mathbb{E}[y_i z_i | x_i] = P(y_i z_i = 1 | x_i)$$

ج) (۱۰ نمره) با استفاده از امید ریاضی بدست آمده در مرحله قبل، مقدار بهینه‌ی پارامترهای β و μ و λ_1 را بدست آورید.

سوال ۶ کاهش بعد (۱۵ نمره)

در روش کاهش بعد LDA، می‌خواهیم یک ترم منظم‌ساز^۲ نرم $\ell_2 - norm$ روی بردار جهت w اعمال کنیم. به عبارت دیگر، می‌خواهیم مسأله بهینه‌سازی زیر را حل نماییم:

$$\begin{aligned} \max_w \quad & w^T S_B w - \alpha \|w\|_2^2 \\ \text{s. t.} \quad & w^T S_W w = K \end{aligned}$$

که در رابطه‌ی بالا، α یک ثابت مثبت است.

الف) (۱۰ نمره) مقدار بهینه‌ی بردار w را بدست آورید.

ب) (۵ نمره) تحت چه شرایطی، مسأله‌ی بهینه‌سازی بالا جواب دارد ولی مسأله بهینه‌سازی LDA معمولی جواب ندارد؟ به عبارت دیگر، توضیح دهید که ترم منظم‌ساز اضافه شده، چه مشکل احتمالی را از LDA معمولی برطرف می‌کند.

موفق و پیروز باشید.

^۱ Expectation
^۲ Regularization term