بسمه‌تعالی

یادگیری ماشین **نمره امتحان: ۱۳۰**

امتحان پایان‌ترم- نیمسال دوم ۱۴۰۰-۱۴۰۱ **وقت آزمون: ۲ ساعت و ۳۰ دقیقه**

توجه: استفاده از کتاب، جزوه، اسلایدهای درس، اینترنت، گوشی و تبلت در حین امتحان غیر مجاز است و تقلب محسوب می‌شود. در صورت تشخیص تقلب، نمره کل امتحان **صفر** منظور خواهد شد.

سوال ۱ سوالات پاسخ کوتاه (۲۱ نمره)

در هر یک از موارد زیر درست یا نادرست بودن آن را مشخص کنید و به صورت مختصر علت را توضیح دهید. (هر مورد ۳ نمره)

الف) اگر از روش رای اکثریت برای ترکیب چند طبقه‌بند استفاده شود، طبقه‌بند حاصل همیشه خطای کمتری نسبت به تک‌تک طبقه‌بندهای اصلی دارد.

ب) خطای بایاس برای طبقه‌بند نزدیک‌ترین همسایه (1-NN) صفر است.

ج) در دسته‌بند SVM اگر نمونه‌هایی که support vector نیستند را از مجموعه‌ی داده‌های آموزشی حذف کنیم، هیچ تاثیری بر روی مرز جداکننده دو کلاس نمی‌گذارد.

د) اگر یک مسأله بهینه‌سازی مقید[1] محدب باشد، آن‌گاه strong duality برقرار است.

ه) یک مجموعه داده با اندازه k در اختیار داریم. یک دسته‌بندی وجود دارد که نمی‌تواند این مجموعه‌ی داده را shatter کند. بُعد VC این دسته‌بند حتما کمتر از k است.

و) اگر برای دسته‌بند kernel-SVM، تابع‌های k_1 و k_2 دو کرنل معتبر[2] باشند، آن‌گاه تابع $k_1 + k_2$ نیز یک کرنل معتبر خواهد بود. به عبارت دیگر، حاصلجمع دو کرنل معتبر، خود یک کرنل معتبر است.

ز) اگر داده‌ها جدایی‌پذیر خطی باشند، آن‌گاه hard-margin linear SVM و لاجیستیک رگرشن هر دو یک مرز جداساز بدست می‌آورند.

سوال ۲ خوشه‌بندی مبتنی بر بهینه‌سازی (۲۴ نمره)

می‌خواهیم با استفاده از یک روش مبتنی بر بهینه‌سازی، داده‌ها را خوشه‌بندی کنیم. برای این منظور، می‌خواهیم کوچک‌ترین ابرکره‌ای که کل داده‌ها را درون خود جای می‌دهد، بیابیم. برای یافتن این ابرکره، مسأله بهینه‌سازی زیر را باید حل کنیم:

$$\min_{r,c} r^2$$

$$subject\ to\colon\ \|x_i - c\|^2 \le r^2, \quad i = 1, \dots, n$$

که در رابطه‌ی بالا، r و c به ترتیب شعاع و مرکز ابرکره است. x_i نمونه‌ی i-ام و n تعداد کل نمونه‌ها می‌باشد.

[1] Constrained optimization problem

[2] valid

الف) (۵ نمره) مسأله‌ی بهینه‌سازی بالا را به گونه‌ای تغییر دهید تا این امکان وجود داشته باشد که بعضی از نمونه‌ها خارج از ابرکره قرار بگیرند. (**راهنمایی**: مشابه soft-margin SVM عمل کنید).

ب) (۵ نمره) تابع لاگرانژین مسأله‌ی بهینه‌سازی قسمت (الف) را تشکیل دهید.

ج) (۱۰ نمره) مسأله دوگان را برای مسأله‌ی قسمت (الف) بدست آورید.

د) (۴ نمره) توضیح دهید که در این مدل، ضرایب لاگرانژ برای کدام نمونه‌ها صفر و برای کدام نمونه‌ها بزرگتر از صفر می‌شود.

سوال ۳ الگوریتم EM (۲۵ نمره)

توزیع مخلوط نمایی[۳] زیر را در نظر بگیرید:

$$p(x) = \alpha \lambda_1 e^{-\lambda_1 x} + (1-\alpha)\lambda_2 e^{-\lambda_2 x}$$

با فرض داشتن مجموعه داده‌ی $\{x_1, \dots, x_n\}$ می‌خواهیم با استفاده از روش EM، پارامترهای مدل بالا را که شامل α، λ_1 و λ_2 است، بدست آوریم (نمونه‌ها .i.i.d هستند).

الف) (۵ نمره) اگر متغیر تصادفی پنهان $\mathrm{z_i}$ را به صورت زیر تعریف می‌کنیم:

$$z_i = \begin{cases} 0 & x_i \ comes\ from\ the\ first\ component \\ 1 & otherwise \end{cases}$$

تابع *log-complete likelihood* را تشکیل دهید:

$$\log p(x_1, z_1, \dots, x_n, z_n | \lambda_1, \lambda_2, \alpha) = ?$$

ب) (۱۰ نمره) ***مرحله E***: امید ریاضی تابع *log-complete likelihood* نسبت به متغیرهای پنهان را بدست آورید. برای این منظور، مقدار $\mathbb{E}_{z_i|x_i,\alpha^t,\lambda_1^t,\lambda_2^t}[z_i]$ را نیز باید محاسبه کنید.

ج) (۱۰ نمره) ***مرحله M***: روابط به روز رسانی پارامتر α و λ_1 را بدست آورید. برای سادگی در روابط، امید ریاضی $\mathbb{E}_{z_i|x_i,\alpha^t,\lambda_1^t,\lambda_2^t}[z_i]$ را که در قسمت قبل بدست آمده با نماد γ_i^t نشان دهید.

سوال ۴ یادگیری جمعی (*Ensemble Learning*) (۱۵+۱۰ نمره)

الف) (۱۰ نمره) برای یک مسأله رگرسیون، دو مدل $y = h_1(x)$ و $y = h_2(x)$ را با استفاده از کمینه کردن مجموع مجذور خطا آموزش داده‌ایم. حال می‌خواهیم با ترکیب این دو مدل، یک مدل قوی‌تر بسازیم. فرض کنید مجموعه داده‌ی $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$ را در اختیار داریم، اگر مدل جدید را به صورت $y = \alpha h_1(x) + \beta h_2(x)$ در نظر بگیریم، مقدار بهینه‌ی ضرایب α و β را بدست آورید. توجه داشته باشید که $\alpha + \beta$ لزوما ۱ نیست. بدست آوردن دستگاه دو معادله و دو مجهول برای α و β کفایت می‌کند و نیازی به حل دستگاه نیست.

[۳] Mixture of Exponentials

ب) (**اختیاری ۱۵ نمره**) اگر خطای دو مدل $h_1(x)$ و $h_2(x)$ به ترتیب e_1 و e_2 باشد و از رابطه‌ی $h(x) = \frac{1}{2}h_1(x) + \frac{1}{2}h_2(x)$ برای ترکیب این دو مدل استفاده کنیم. نشان دهید که خطای (مجموع مجذور خطا) مدل $h(x)$ در رابطه‌ی زیر صدق می‌کند:

$$e \leq \frac{1}{4}(e_1 + e_2) + \frac{1}{2}\sqrt{e_1 e_2}$$

راهنمایی: نامساوی *Cauchy-Schwarz* به صورت زیر است:

$$\left(\sum_{i=1}^{n} u_i v_i\right)^2 \leq \left(\sum_{i=1}^{n} u_i^2\right)\left(\sum_{i=1}^{n} v_i^2\right)$$

سوال ۵ (۵ نمره)

نشان دهید اگر توزیع $P(\theta|\alpha)$ یک conjugate prior برای پارامتر θ باشد، آن‌گاه توزیع مخلوط $\sum_{d=1}^{D} \lambda_d P(\theta|\alpha_d)$ هم یک conjugate prior برای θ خواهد بود.

سوال ۶ کاهش بُعد (۱۵ نمره)

مجوعه داده‌ی زیر را در نظر بگیرید. با استفاده از *PCA* می‌خواهیم داده‌ها را به فضای یک بُعدی ببریم. جهت برداری که روش *PCA* برای نگاشت به فضای یک بُعدی بدست می‌آورد را محاسبه کنید.

$$D = \left\{\begin{bmatrix}10\\8\end{bmatrix}, \begin{bmatrix}8\\2\end{bmatrix}, \begin{bmatrix}2\\0\end{bmatrix}, \begin{bmatrix}4\\6\end{bmatrix}\right\}$$

سوال ۷ خوشه‌بندی (۱۵ نمره)

می‌خواهیم مجموعه داده‌های $\mathcal{X}$ را به k خوشه تقسیم کنیم. همانطور که می‌دانید، الگوریتم k-means با انتخاب k مرکز دسته به صورت تصادفی آغاز می‌شود. فرض کنید که راهبرد زیر را برای انتخاب k مرکز دسته اولیه داشته باشیم:

مرکز دسته اول را تصادفی انتخاب کرده و مرکز دسته‌های بعدی را به گونه‌ای انتخاب می‌کنیم که از مرکز دسته‌هایی که قبلا انتخاب شده‌اند بیشترین فاصله را داشته باشد.

به کمک یک مثال نشان دهید که این راهبرد الزاما به دسته‌بندی بهینه منجر نمی‌شود. به عبارت دیگر نشان دهید که ممکن است هزینه راهبرد پیشنهادی از هزینه الگوریتم بهینه خوشه‌بندی k-means که در عبارت زیر نمایش داده شده، بیشتر است:

$$\phi = \sum_{x \in \mathcal{X}} \min_{c \in \mathcal{C}} \|x - c\|^2$$

موفق و پیروز باشید.