

توجه: امتحان به صورت کتاب و جزوه بسته است و مجاز به استفاده از حداکثر دو برگه روابط کمکی هستید.

توجه: امتحان از ۱۱۵ نمره است و برای کسب نمره کامل باید ۱۰۰ نمره بدست آورید.

سوال ۱ خوشه‌بندی ترتیبی (۲۰ نمره)

مجموعه داده‌ی یک بعدی W به شکل زیر را در نظر بگیرید:

$$W = \{x^1, x^2, x^3, x^4, x^5, x^6, x^7\} = \{1, 2, 3, 5, 7, 8, 9\}$$

قصد خوشه‌یابی این ۷ داده یک بعدی به شیوه ترتیبی را داریم. فاصله نقطه از مجموعه را به صورت حداقل در نظر بگیرید:

$$d(x, W) = \min_{y \in W} d(x, y)$$

و d را فاصله‌ی اقلیدسی در نظر بگیرید. آستانه ایجاد کلاستر جدید را th بنامید.

الف) (۱۲ نمره) در حالت‌های زیر کلاستریگ را انجام دهید:

- $th = 2.1$ و ترتیب ورود داده‌ها: $x^1 \rightarrow x^2 \rightarrow x^3 \rightarrow x^4 \rightarrow x^5 \rightarrow x^6 \rightarrow x^7$
- $th = 1.5$ و ترتیب ورود داده‌ها: $x^1 \rightarrow x^2 \rightarrow x^3 \rightarrow x^4 \rightarrow x^5 \rightarrow x^6 \rightarrow x^7$
- $th = 1.5$ و ترتیب ورود داده‌ها: $x^7 \rightarrow x^6 \rightarrow x^5 \rightarrow x^4 \rightarrow x^3 \rightarrow x^2 \rightarrow x^1$
- $th = 2.1$ و ترتیب ورود داده‌ها: $x^1 \rightarrow x^7 \rightarrow x^2 \rightarrow x^6 \rightarrow x^3 \rightarrow x^5 \rightarrow x^4$

ب) (۸ نمره) حال با توجه به نتایج به دست آمده روی تاثیر ترتیب ورود داده‌ها و نیز مقدار آستانه th بحث کنید. به علاوه یک ترتیب ورود و آستانه دیگر به انتخاب خودتان در نظر بگیرید که به تعداد و شکل کلاستر متفاوتی منجر شود.

سوال ۲ خوشه‌بندی سلسله‌مراتبی (۱۵ نمره)

همان داده‌های سوال قبل را به شیوه سلسله‌مراتبی از نوع Agglomerative خوشه‌بندی کنید. فاصله مجموعه از مجموعه را به صورت فاصله‌ی میانگین‌ها در نظر بگیرید:

$$d(w_1, w_2) = d(\text{mean}_{x \in w_1} x, \text{mean}_{y \in w_2} y)$$

در هر مرحله از تصمیم‌گیری برای یافتن حداقل فاصله، اگر فاصله‌ها مساوی شد یکی را به تصادف انتخاب کنید. نتیجه را با خوشه‌بندی ترتیبی مقایسه کنید. (راهنمایی: برای نمایش خوشه‌بندی سلسله‌مراتبی خود از dendrogram استفاده کنید).

سوال ۳ مدل چگالی مخلوط گاوسی (۲۴ نمره)

مدل Gaussian Mixture Density را در نظر بگیرید. هر گاوسی با میانگین μ_k و واریانس σ_k^2 تعریف شده است. تعداد گاوسی‌ها L است. داده‌ها یک بعدی و تعداد آن‌ها n است، یعنی: $X = \{x^q\}_{q=1}^n$.

الف) (۱۵ نمره) روابط حل بسته گام maximization در الگوریتم EM را با ذکر مراحل استخراج روابط بدست آورید. (توجه: روابط بازگشتی برای μ_k ، σ_k^2 و p_k مد نظر است).

ب) (۵ نمره) روشی برای تعیین L پیشنهاد کنید.

ج) (۴ نمره) چرا مهم است که در گام M در الگوریتم EM حل بسته داشته باشیم؟

سوال ۴ انتخاب ویژگی (۱۶ نمره)

مسئله subset selection را در نظر بگیرید. یک راه حل مسئله Exhaustive search است. یعنی وقتی n ویژگی داریم، تمام $2^n - 1$ زیر مجموعه ممکن را بررسی کنیم و هر کدام خطای طبقه‌بندی کمتری ایجاد کرد آن را به عنوان زیر مجموعه بهینه ویژگی‌ها انتخاب کنیم.

الف) (۸ نمره) آیا انتخاب پیشرونده^۱ در گام اول همیشه عضوی از زیر مجموعه‌ی بهینه را انتخاب می‌کند؟ اگر نظرتان موافق است، اثبات کنید و اگر مخالف است مثال نقض بیاورید و یا استدلال منطقی ارائه دهید.

ب) (۸ نمره) به همین سوال با فرض آن که ابعاد بردار ویژگی مستقل از یکدیگر باشند پاسخ دهید.

سوال ۵ L2-SVM (۲۵ نمره)

یک مسئله طبقه‌بندی دو کلاسه با برچسب‌های $\{+1, -1\}$ را در نظر بگیرید. مجموعه‌ی داده‌های آموزشی به صورت $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$ است (داده‌ها در فضای d بعدی هستند). فرض کنید با اندکی تغییر در Soft margin SVM، می‌خواهیم از **مربع** متغیر e_i (که نشان دهنده‌ی میزان خروج نمونه‌ی x_i باشد از محدوده‌ی مجاز است)، در تابع هدف استفاده کنیم. در واقع، مسئله بهینه‌سازی ما به صورت زیر درمی‌آید:

$$\min_{w, b, e} \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^n e_i^2$$

$$s. t. \quad y_i(w^T x_i + b) \geq 1 - e_i, \quad i = 1, \dots, n$$

$$e_i \geq 0, \quad i = 1, \dots, n$$

در روابط بالا، $w \in \mathbb{R}^d$ و $b \in \mathbb{R}$ پارامترهای ابرصفحه‌ی جداکننده و C یک ثابت مثبت است.

¹ Forward Selection

الف) (۵ نمره) نشان دهید که در این مسئله اگر قیدهای آخر (یعنی $e_i \geq 0 \forall i$) را از مسئله حذف کنیم، هیچ تغییری در جواب بهینه‌ی مسئله ایجاد نمی‌شود. (راهنمایی: از برهان خلف استفاده کنید).

حال پس از حذف قیدهای اضافی، مسئله بهینه‌سازی به صورت زیر درمی‌آید:

$$\min_{w,b,e} \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^n e_i^2$$

$$s. t. \quad y_i(w^T x_i + b) \geq 1 - e_i, \quad i = 1, \dots, n$$

ب) (۲ نمره) آیا مسئله بهینه‌سازی فوق، محدب^۲ است؟ توضیح دهید.

ج) (۱۸ نمره) مسئله بهینه‌سازی دوگان^۳ را برای مسئله فوق بدست آورید.

سوال ۶ Ensemble Learning (۱۵ نمره)

یک مسئله طبقه‌بندی دو کلاسه با برچسب‌های $\{+1, -1\}$ را در نظر بگیرید. مجموعه داده‌ی آموزشی $\{(x_1, y_1), \dots, (x_n, y_n)\}$ را در اختیار داریم. تابع خطای طبقه‌بند h را به صورت زیر تعریف می‌کنیم:

$$E = \sum_{i=1}^n (h(x_i) - y_i)^2$$

می‌خواهیم T طبقه‌بند را به صورت زیر با هم ترکیب کنیم:

$$F_T(x) = \sum_{t=1}^T \alpha_t h_t(x_i)$$

که در رابطه بالا، $h_t(x) \in \{+1, -1\}$ و α_t وزن طبقه‌بند t -ام است. اگر بخواهیم مشابه روش AdaBoost ضرایب α_t را به صورت ترتیبی^۴ بدست آوریم، با در نظر گرفتن تابع خطای بالا، مقدار بهینه‌ی α_T را بدست آورید.

موفق و پیروز باشید.

Convex²
Dual Problem³
Sequential⁴