

سوال ۱ (۲۵ نمره): انتخاب ویژگی (feature selection)

۱.۱. درستی یا نادرستی گزاره‌های زیر را با ذکر توضیح مشخص نمایید.

- (۲ نمره) ممکن است ویژگی‌هایی که همبستگی زیادی دارند، افزونه (redundant) نباشند.
- (۲ نمره) یک زوج ویژگی کاملاً همبسته حتماً باعث افزونگی (redundancy) می‌شوند و حذف یکی از آن‌ها منطقی است.
- (۲ نمره) غالب روش‌های چندمتغیره از رویکرد wrapper استفاده می‌نمایند.

۲.۱. مجموعه داده spambase که در فایلی به همین عنوان در اختیار شما قرار داده شده را در نظر بگیرید. در این تمرین هدف آن است که ابتدا از بین ویژگی‌ها تعدادی انتخاب و سپس دسته‌بندی داده‌های کاهش ابعادیافته با استفاده از 1-NN انجام شود. قبل از طرح صورت تمرین، معیارهای انتخاب ویژگی موردنظر مختصراً معرفی خواهند شد.

- معیار بهره اطلاعاتی (Information gain): این معیار برای یک ویژگی می‌تواند بر اساس تغییر مقدار آنتروپی قبل و بعد از شکستن داده‌ها بر اساس مقدار آن ویژگی محاسبه شود:

$$IG(X) = H_D(Y) - \sum_{j \in \text{val}(X)} \frac{|D_j|}{|D|} H_{D_j}(Y)$$

که  $D$  مجموعه داده‌های آموزش قبل از شکستن بر اساس ویژگی  $X$  را نشان می‌دهد و  $D_j$  زیرمجموعه‌ای از نمونه‌های موجود در  $D$  است که مقدار ویژگی  $X$  در آن‌ها برابر  $j$  است. در واقع تفاضل مقدار آنتروپی برجسب‌ها قبل از دیده شدن ویژگی  $X$  با میانگین وزن‌دار آنتروپی‌های بعد از دیده شدن این ویژگی را نشان می‌دهد.

- معیار Relief: این معیار به دنبال یک امتیاز یا وزن برای ویژگی‌ها می‌گردد.

○ امتیازات اولیه را صفر می‌گذارد  $w(j) = 0$  ( $j = 1, \dots, d$ ).

○ در هر دور از  $m$  تکرار، یک داده  $x$  به صورت تصادفی انتخاب می‌شود. نزدیک‌ترین همسایه این داده که شماره‌ی دسته‌اش با  $x$  یکی است (تحت عنوان  $H$ ) و همچنین نزدیک‌ترین همسایه‌ای که شماره‌ی دسته‌اش متفاوت است (تحت عنوان  $M$ )، پیدا می‌شوند. سپس وزن ویژگی‌ها به این صورت به‌روزرسانی می‌شود:

$$w(j) = w(j) - \frac{\text{diff}(j, x, H)}{m} + \frac{\text{diff}(j, x, M)}{m}$$

- مقدار  $\text{diff}(j, x, y)$  درحالتی که ویژگی  $j$ -ام اسمی (nominal) باشند، اگر  $x_j$  و  $y_j$  یکی نباشند یک وگرنه صفر خواهد بود. اما درحالتی که ویژگی مربوطه عددی و ترتیب‌دار باشند به صورت  $\text{diff}(j, x, y) = \frac{|x_j - y_j|}{\max_{i=1, \dots, N} x_j^{(i)} - \min_{i=1, \dots, N} x_j^{(i)}}$  (یعنی قدرمطلق تفاضل مقادیر ویژگی مربوطه که به کل محدوده‌ی مقادیر آن ویژگی نرمال شده است).

در این تمرین هدف آن است که روش‌های زیر برای انتخاب ویژگی پیاده‌سازی شده و نتایج دسته‌بندی حاصل از به‌کارگیری ویژگی‌های انتخاب شده با هم مقایسه شوند. تعداد ویژگی‌هایی که قرار است انتخاب شوند، در مجموعه  $d' = \{1, 5, 10, 20, 30, 40, 50\}$  در نظر گرفته شده‌اند. در هر یک از موارد زیر پس از اعمال روش انتخاب ویژگی از دسته‌بند 1-NN استفاده نمایید و نمودار خطای دسته‌بندی روی داده‌ی آموزش و آزمون را بر اساس تعداد ویژگی‌های انتخاب شده رسم و تحلیل نمایید:

- (۵ نمره) فیلتر تک متغیره (univariate filter) با استفاده از معیار بهره اطلاعاتی (Information Gain)
- (۸ نمره) فیلتر تک متغیره با استفاده از معیار Relief
- (۶ نمره) wrapper چند متغیره (multivariate) با استراتژی جستجوی انتخاب رو به جلوی ترتیبی (sequential forward) که از دقت دسته‌بندی 1-NN به صورت LOOCV برای ارزیابی یک مجموعه ویژگی استفاده می‌کند.

### سوال ۲ (۱۰ نمره): PCA

- ۱.۲ (۶ نمره) فرض کنید پایه‌های orthonormal فضای  $d$ -بعدی باشند. نشان دهید متوسط مجذور خطا (MSE) بین یک نقطه  $d$ -بعدی و تصویر آن در فضای  $d'$ -بعدی حداقل می‌شود اگر
  - پایه‌ها بردارهای ویژه‌ی ماتریس  $R_x$  شوند.
  - زیرفضای  $d'$ -بعدی فضایی باشد که توسط بردارهای ویژه‌ی متناظر با بزرگ‌ترین مقادیر ویژه  $R_x$  پوشش می‌شود. (span)

[راهنمایی: سعی کنید حداقل میانگین مجذورات خطا را با وجود محدودیت  $a_i^T a_i = 1$  ( $i = 1, \dots, d$ ) به دست آورید.]

- ۲.۲ (۴ نمره) روابط مربوط به kernel PCA را به صورت کامل از روی روابط PCA استخراج کنید و گام‌هایی را که برای پیدا کردن تبدیل‌یافته‌ی داده‌ها با داشتن ماتریس هسته‌ی  $K$ ، در روش Kernel-PCA مورد نیاز است، مشخص نمایید.

### سوال ۳ (۲۰ نمره): LDA

- ۱.۳ فرض کنید تابع چگالی احتمال دلخواه با میانگین  $\mu_i$  و ماتریس کوواریانس  $\Sigma_i$  باشد. همچنین تبدیل  $x' = w^T x$  که باعث تصویر کردن نقاط روی بردار  $w$  می‌شود را در نظر بگیرید. میانگین و واریانس داده‌های تصویر شده دسته‌ی  $i$ -ام به ترتیب با  $\mu'_i$  و  $\sigma'^2_i$  نشان داده می‌شوند. اگر احتمال پیشین دسته‌ی  $i$ -ام را مشخص کند:
  - (۳ نمره) نشان دهید بهینه‌ی تابع هدف

$$J(w) = \frac{(\mu'_1 - \mu'_2)^2}{P(\omega_1)\sigma'^2_1 + P(\omega_2)\sigma'^2_2}$$

برابر با  $w = (P(\omega_1)\Sigma_1 + P(\omega_2)\Sigma_2)^{-1}(\mu_1 - \mu_2)$

- (۲ نمره) آیا اگر بخواهیم روش Fisher را با تابع هدف بالا تطبیق دهیم مقادیر احتمالات پیشین را در تابع هدف فوق باید مساوی قرار دهیم یا خیر (توضیح دهید)؟

۲.۳. (۵ نمره) نشان دهید راستای برداری که توسط روش Fisher برای مساله‌ی دو دسته‌ای پیدا می‌شود، جواب مساله‌ی حداقل مجذورات خطا (تابع هزینه‌ی SSE) است در حالتی که برچسب مطلوب برای نمونه‌های دسته‌ی اول  $\frac{N}{N_1}$  و برای نمونه‌های دسته‌ی دوم  $\frac{N}{N_2}$  در نظر گرفته شود  $N_1$  و  $N_2$  به ترتیب تعداد نمونه‌های آموزش موجود برای دسته‌ی اول و دوم را نشان می‌دهند).

[راهنمایی: گام‌های کلی این اثبات در بخش 4.1.5 کتاب Bishop آمده است.]

۳.۳. (۶ نمره) پیدا کردن تبدیل خطی  $\mathbf{x}' = \mathbf{W}^T \mathbf{x}$  از فضای ویژگی  $d$  بعدی به فضای  $d'$  بعدی در مساله کلی چنددسته‌ای (multi-class)، اگر تابع هدفی به دنبال حداکثر کردن مجموع مجذور فواصل زوج داده‌هایی که درون یک دسته قرار می‌گیرند و حداقل کردن مجموع مجذور فواصل زوج داده‌هایی که درون دسته‌های متفاوت قرار می‌گیرند، باشد:

a. نشان دهید تابع هدف مربوطه می‌تواند با استفاده از ماتریس‌های پراکنندگی درون دسته‌ای  $\mathbf{S}_W$  و بین دسته‌ای  $\mathbf{S}_B$  تابع هدف را به صورت زیر بازنویسی کرد:

$$J(\mathbf{W}) = \frac{\text{tr}(\mathbf{W}^T \mathbf{S}_B \mathbf{W})}{\text{tr}(\mathbf{W}^T \mathbf{S}_W \mathbf{W})}$$

b. آیا این تابع هدف همان تابع هدف LDA است؟

۴.۳. (۴ نمره) آیا پیشنهاد کردن تابع هدف MDA (تعمیم چند دسته‌ای LDA) برابر با حل مساله‌ی بهینه‌سازی زیر است؟ در صورت درستی اثبات کنید و گرنه دلایل خود را مطرح نمایید.

$$\begin{aligned} \max_{\mathbf{W}} \text{tr}(\mathbf{W}^T \mathbf{S}_B \mathbf{W}) \\ \text{s. b. } \mathbf{W}^T \mathbf{S}_W \mathbf{W} = \mathbf{I} \end{aligned}$$

#### سوال ۴ (۲۵ نمره): پیاده‌سازی و مقایسه روش‌های کاهش ابعاد PCA و LDA

در این تمرین هدف آن است که برای مجموعه داده MNIST ابتدا داده‌ها توسط روش‌های LDA و PCA از فضای ویژگی اولیه به فضای با ابعاد پایین نگاشت شود و سپس در فضای کاهش ابعاد یافته دسته‌بندی توسط دسته‌بند 1-NN انجام شود. لازم به ذکر است که برای پیدا کردن تبدیلات خطی مربوط به روش‌های PCA و LDA باید تنها از داده‌های آموزش استفاده شود. همان‌گونه که در تمرین سری سوم توضیح داده شد، مجموعه داده‌ی MNIST شامل ارقام دست‌نویس است و هر داده‌ی این مجموعه یک تصویر  $28 \times 28$  است. در فایلی که در اختیار شما قرار گرفته، هر سطر داده حاصل برداری کردن یک تصویر  $28 \times 28$  است که در انتهای آن مشخص شده است که این تصویر مربوط به کدام یک از ارقام 0 تا 9 بوده است (بنابراین از ستون‌های 1 تا 784 برای ویژگی‌های داده و از ستون 785 برای برچسب استفاده نمایید). از ۴۰۰۰ داده‌ی اول برای آموزش و از مابقی داده‌ها برای آزمون استفاده شود.

در پیاده‌سازی این تمرین می‌توانید از این نکته استفاده نمایید که بردارها و مقادیر ویژه یک ماتریس به شکل  $\mathbf{A}\mathbf{A}^T$  را می‌توان از روی بردارها و مقادیر ویژه  $\mathbf{A}^T \mathbf{A}$  پیدا کرد.

#### ۱.۴. به کارگیری PCA:

- a. (۳ نمره) روش PCA را به کارگیرید و تصاویر ۵ بردار ویژه‌ی اول را در گزارش ارائه نمایید.
- b. (۳ نمره) داده‌های مربوطه به ارقام 0، 1 و 2 را در فضای تبدیل‌یافته دو بعدی (که پایه‌های آن را دو مولفه‌ی ویژه‌ی اول تشکیل می‌دهند) به صورت scatter plot با رنگ‌های متفاوت رسم نمایید.
- c. (۴ نمره) میانگین خطای بازسازی تصاویر را به ازای هر یک از تعداد ابعاد فضای تبدیل‌یافته  $d' = \{1,2,5,10,20,50,100,200,500\}$  مشخص نموده و نتایج را در یک نمودار ارائه نمایید.  
[خطای بازسازی هر تصویر را میانگین مجذور خطا در روشی پیکسل‌های بازسازی شده در نظر بگیرید.]
- d. (۵ نمره) نتایج دسته‌بندی در فضای تبدیل‌یافته را به ازای تمام حالتی که استفاده از ۲۰٪، ۴۰٪، ۸۰٪، ۱۰۰٪ اول داده‌های آموزش (برای تعیین تبدیل PCA و همچنین به عنوان داده‌ی آموزش در روش 1-NN) صورت گیرد و در هر کدام از این موارد تعداد ابعاد فضای تبدیل‌یافته  $d' = \{1,2,5,10,20,50,100,200,500\}$  تغییر کند را به دست آورید. کل نتایج را در یک نمودار گزارش نموده و تحلیل نمایید.

#### ۲.۴. به کارگیری LDA:

- a. (۵ نمره) با استفاده از داده‌های آموزش ماتریس تبدیل LDA را پیدا نمایید و داده‌ها را به هر یک از فضاها‌ی ویژگی با ابعاد  $d' = \{1,2,\dots,8,9\}$  نگاشت کنید. برای هر حالت خطای دسته‌بندی 1-NN روی نمونه‌های آزمون را به دست آورید.
- b. (۳ نمره) داده‌های مربوطه به ارقام 0، 1 و 2 را در فضای تبدیل‌یافته دو بعدی  $d' = 2$  به صورت scatter plot با رنگ‌های متفاوت رسم نمایید.

#### ۳.۴. مقایسه PCA و LDA:

- (۲ نمره) نتایج به دست آمده برای PCA و LDA بر روی داده‌های آزمون در حالتی که از کل داده‌های آموزش استفاده شود، به ازای  $d' = \{1,2,\dots,8,9\}$  در یک نمودار نشان داده و مقایسه نمایید.

موفق باشید