

به نام خدا

## تمرین پنجم یادگیری ماشین

غزل زمانی نژاد

(2)

$$2) \tilde{\mu}_1 = \frac{1}{n} \sum_{i=1}^n w^T x_i = w^T \left( \frac{1}{n} \sum_{i=1}^n x_i \right) = w^T \mu_1$$

همین ترتیب:  $\tilde{\mu}_2 = w^T \mu_2$

$$\tilde{\sigma}_1^2 = \sum_{i=1}^n (w^T x_i)(w^T x_i)^T = \sum_{i=1}^n w^T x_i x_i^T w = w^T \underbrace{\sum_{i=1}^n x_i x_i^T}_{\sigma_1^2} w = w^T \sigma_1^2 w$$

همین ترتیب:  $\sigma_2^2 = w^T \tilde{\sigma}_2^2 w$

$$J(w) = \frac{(w^T \mu_1 - w^T \mu_2)^2}{w^T \sigma_1^2 w + w^T \sigma_2^2 w} = \frac{w^T (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T w}{w^T (\sigma_1^2 + \sigma_2^2) w} = \frac{w^T S_B w}{w^T S_W w}$$

$$w^* = \max w^T S_B w \quad \text{s.t.} \quad w^T S_W w = k$$

$$L(w) = w^T S_B w - \lambda (w^T S_W w - k) \quad \lambda \geq 0$$

$$\nabla_w L(w) = 2 S_B w - 2 \lambda S_W w = 0 \Rightarrow S_B w = \lambda S_W w \Rightarrow S_W^{-1} S_B w = \lambda w$$

$$\lambda w = S_W^{-1} (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T w \Rightarrow w = \frac{\alpha}{\lambda} S_W^{-1} (\mu_1 - \mu_2)$$

$$\Rightarrow w^* = S_W^{-1} (\mu_1 - \mu_2) = (\Sigma_1 + \Sigma_2)^{-1} (\mu_1 - \mu_2)$$

### (3 الف) به چند مشکل اشاره می کنیم:

- مشکل ماتریس کوواریانس: در PCA، باید ماتریس کوواریانس داده‌ها محاسبه شود. این ماتریس معمولاً با ابعاد بالا، بسیار بزرگ می‌شود. می‌توانیم از ترفندهایی مانند محاسبه ماتریس کوواریانس به صورت مستقیم یا استفاده از تقریب‌ها استفاده کنیم.
- برخورد با مسئله انحراف نمونه کوچک: ممکن است در PCA یا LDA با مسئله انحراف نمونه کوچک (Small Sample Size) مواجه شویم که باعث ناپایداری محاسبات می‌شود. می‌توانیم از روش‌های تنظیم انحراف نمونه کوچک مانند تحلیل مؤلفه اصلی با تنظیم (Regularized PCA) یا LDA با استفاده از ماتریس همگرایی استفاده کنیم.
- مشکل تفاضل مقادیر ویژه: ممکن است مقادیر ویژه در محاسبات PCA یا LDA نزدیک به هم باشند که باعث ناپایداری محاسبات می‌شود. می‌توانیم از روش‌های پایداری مانند افزودن یک مقدار مثبت به مقادیر ویژه یا استفاده از روش‌های نرمالی‌سازی مقادیر ویژه استفاده کنیم.
- مصرف منابع محاسباتی زیاد: ممکن است با ابعاد بالا، مصرف حافظه و زمان محاسباتی بسیار زیاد شود. می‌توانیم از روش‌های تقریبی یا تکنیک‌های بهینه‌سازی برای محاسبات استفاده کنیم، مانند استفاده از نمونه‌گیری تصادفی (Randomized PCA) یا تکنیک‌های mini-batch.

### (ب) مزایا:

- استفاده از اطلاعات متقابل می‌تواند بهترین نمایانگر برای تفاوت‌ها و ارتباطات بین داده‌ها باشد. مثلاً اگر اطلاعات لیبل‌ها در دسترس باشد، می‌توان از معیارهایی مانند معیارهای مبتنی بر تفاوت بین دسته‌ها به عنوان مثال، Fisher's discriminant criterion در LDA استفاده کرد.
- در مسائلی که تشخیص الگو از اهمیت بالایی برخوردار است، استفاده از معیارهای جایگزین می‌تواند بهبود مسائل مربوط به تشخیص الگو را فراهم کند.
- در LDA، اگر از معیارهای جایگزین مبتنی بر اطلاعات لیبل‌ها استفاده شود، ممکن است تفاوت بین دسته‌ها افزایش یابد و تبدیلی بهتر از نظر تفکیک‌پذیری دسته‌ها ایجاد شود.

### معایب:

- استفاده از معیارهای دیگر ممکن است نیاز به اطلاعات خاص و مکمل داشته باشد که در برخی موارد ممکن است در دسترس نباشد.
- ممکن است محاسبات مرتبط با معیارهای جایگزین پیچیده‌تر و زمان‌بر باشد.
- هر معیار جایگزینی برای هر مسئله بهینه نیست و ممکن است در برخی حالات به نتایج بهتری نرسد.

(4)

4) الف)  $x = Wz + \mu + n$      $n \sim \mathcal{N}(0, \sigma^2 I)$

$$\text{mean}_x = E[x] = E[Wz + \mu + n] = WE[z] + \mu + E[n] = \mu$$

$$\begin{aligned} \text{Cov}_x &= E[(x - \mu)(x - \mu)^T] = E[(Wz + \mu + n - \mu)(Wz + \mu + n - \mu)^T] = \\ &= E[(Wz + n)(Wz + n)^T] = \\ &= E[Wz z^T W^T + Wz n^T + n z^T W^T + n n^T] = \\ &= WE[z z^T]W^T + WE[z]E[n^T] + E[n]E[z^T]W^T + \sigma^2 I = \\ &= WW^T + \sigma^2 I \end{aligned}$$

(5)

5) الف)  $\mu_1 = \begin{bmatrix} \frac{0+0+5}{3} \\ \frac{2+0+0}{3} \end{bmatrix} = \begin{bmatrix} 1.6 \\ 0.6 \end{bmatrix}$      $\mu_2 = \begin{bmatrix} \frac{1.5+5}{2} \\ \frac{0+2}{2} \end{bmatrix} = \begin{bmatrix} 3.25 \\ 1 \end{bmatrix}$

point	$\mu_1$ dist	$\mu_2$ dist	new cluster	کلام 1:
$x_1$	2.12	3.4	1	
$x_2$	1.7	3.4	1	می دانیم شرط جابجایی داده $\tilde{x}$ از کلاستر $i$
$x_3$	0.61	2.01	1	به کلاستر $j$ :
$x_4$	3.45	2.01	2	$\ \tilde{x} - \mu_i\ ^2 > \ \tilde{x} - \mu_j\ ^2$
$x_5$	3.67	2.01	2	

$$C_1 = \{x_1, x_2, x_3\} \quad C_2 = \{x_4, x_5\}$$

$$\mu_1 = \begin{bmatrix} \frac{0+0+1.5}{3} \\ \frac{2+0+0}{3} \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.6 \end{bmatrix} \quad \mu_2 = \begin{bmatrix} \frac{5+5}{2} \\ \frac{0+2}{2} \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

point	$\mu_1$ dist	$\mu_2$ dist	new cluster	کلام 2:
$x_1$	1.48	5.09	1	
$x_2$	0.78	5.09	1	$\Rightarrow C_1 = \{x_1, x_2, x_3\}$
$x_3$	1.16	3.64	1	$C_2 = \{x_4, x_5\}$
$x_4$	4.53	1	2	
$x_5$	4.71	1	2	خوشه ها تغییر نکردند.

$$b) \mu_1 = \begin{bmatrix} 1.6 \\ 0.6 \end{bmatrix}, \mu_2 = \begin{bmatrix} 3.25 \\ 1 \end{bmatrix}$$

point	$\mu_1$ dist	$\mu_2$ dist	new cluster
$x_1$	3	4.25	1
$x_2$	2.2	4.25	1
$x_3$	0.7	2.75	1
$x_4$	4	2.75	2
$x_5$	4.8	2.75	2

گام 1

$$\Rightarrow C_1 = \{x_1, x_2, x_3\}$$

$$C_2 = \{x_4, x_5\}$$

$$\mu_1 = \begin{bmatrix} 0.5 \\ 0.6 \end{bmatrix}, \mu_2 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

point	$\mu_1$ dist	$\mu_2$ dist	new cluster
$x_1$	1.9	6	1
$x_2$	1.1	6	1
$x_3$	1.6	4.5	1
$x_4$	5.1	1	2
$x_5$	5.9	1	2

گام 2

$$\Rightarrow C_1 = \{x_1, x_2, x_3\}$$

$$C_2 = \{x_4, x_5\}$$

خوبه ها تغییر نکردند