۱ـ الف) برای این کار ابتدا مقدارهای ویژه ماتریس کوواریانس را پیدا می‌کنیم:

$$x_1 = (-1,-1),\quad x_2 = (0,0),\quad x_3 = (1,1) \Rightarrow X = \begin{bmatrix} -1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}$$

$$\Rightarrow Cov = \frac{1}{3} X^T X = \frac{1}{3} \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 2/3 & 2/3 \\ 2/3 & 2/3 \end{bmatrix} = S$$

حال که ماتریس کوواریانس بدست آمد مقادیر ویژه‌ی آن را به کمک رابطه $|S - \lambda I| = 0$ می‌یابیم:

$$\begin{vmatrix} \frac{2}{3} - \lambda & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} - \lambda \end{vmatrix} = \left(\frac{2}{3} - \lambda\right)^2 - \frac{4}{9} = 0 \Rightarrow \begin{cases} \frac{2}{3} - \lambda = \frac{2}{3} \Rightarrow \lambda = 0 \\ \frac{2}{3} - \lambda = -\frac{2}{3} \Rightarrow \boxed{\lambda = \frac{4}{3}} \end{cases}$$

$\leftarrow$ مقدار ویژه مورد نظر

حال نیز $\lambda = \frac{4}{3}$، مولفه اصلی اول را بدست می‌آوریم:

$$Sv = \lambda v = \begin{bmatrix} \frac{2}{3} v_1 + \frac{2}{3} v_2 \\ \frac{2}{3} v_1 + \frac{2}{3} v_2 \end{bmatrix} = \begin{bmatrix} \frac{4}{3} v_1 \\ \frac{4}{3} v_2 \end{bmatrix} \to v_1 = v_2 ,\quad v_1^2 + v_2^2 = 1 \Rightarrow v_1 = v_2 = \frac{\sqrt{2}}{2}$$

مولفه اصلی اول برابر $\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$ است

ب) برای تصویر کردن داده‌ها از رابطهٔ $x' = xA$ استفاده می‌کنیم که ماتریس A ستون‌هایش بردارهای ویژه ماتریس کوواریانس هستند (همان مولفه‌های اصلی):

$$x' = \begin{bmatrix} -1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} -\sqrt{2} \\ 0 \\ \sqrt{2} \end{bmatrix}$$

مختصات جدید در فضای یک بعدی:

$(-1,-1) \to -\sqrt{2}$

$(0,0) \to 0$, $\boxed{var = \lambda_1 = \frac{4}{3}}$

$(1,1) \to \sqrt{2}$

Subject :
Date

پ) خطای بازسازی، معادل جمع مقادیر ویژه‌هایی است که از آن‌ها استفاده نشده و دور انداختیم؛ در این مسئله مقدار ویژه‌ای که در نظر نگرفته‌ایم، صفر است که معادل خطای بازسازی صفر می‌شود.

۲- چون همانطور که گفته شد، PCA به برچسب توجهی نمی‌کند، جهت آن معادل جهتی است که واریانس تمامی داده‌ها روی آن افزایش می‌یابد.

در LDA، به دنبال خطی هستیم که در صورت تصویر شدن نمونه‌ها روی آن، کلاس‌ها به بهترین شکل ممکن قابل تفکیک باشند.

۳- در این الگوریتم بصورت کلی دو گام داریم؛ آپدیت کردن γ_{ij} و محاسبه centroid های جدید یا μ_j ها، که بصورت زیر هستند:

$$\gamma_{ij} = \begin{cases} 1 & \|x_i-\mu_j\|^2 < \|x_i-\mu_{j'}\|^2 \ \forall j' \neq j \\ 0 & o.w \end{cases}$$

$$\mu_j = \frac{\sum_i \gamma_{ij} x_i}{\sum_i \gamma_i}$$

الف) می‌دانیم برای مقادیر γ محدودیت داریم، n^2 درایه که هرکدام ۲ حالت دارند که می‌شود 2^{n^2} حالت در کل، و درصورت monotonic بودن اعوجاج (J) می‌توان ادعا کرد هر γ ای که مشاهده می‌شود فقط در مسیر کاهش J بوده، در غیر این صورت monotonic نیست. در گام اول الگوریتم، فاصله هر نمونه را کمینه می‌کنیم (یعنی عبارت مربوط به x_i) و در گام دوم، با میانگین در نظر گرفتن مرکز خوشه j ام با توجه به x_i هایی که در آن قرار گرفته‌اند عبارت مربوط به خوشه‌ها کم می‌شود؛ حال نشان دادیم J بصورت monotonic کاهش می‌یابد، پس هر γ در مسیر کاهش آن دیده خواهد شد، تعداد حالات γ محدود است پس الگوریتم پایان می‌یابد.

ب) میانگین وزن‌دار مقادیر درون خوشه‌ای را بصورت زیر می‌نویسیم:

$$\sum_{j=1}^{K} \left(\sum_{i=1}^{n} \gamma_{ij}\right) w_j(X) = \sum_{j=1}^{K}\sum_{i=1}^{n} \gamma_{ij} \|x_i-\mu_j\|^2 = J/n$$

پس کمینه کردن J معادل کمینه کردن میانگین وزن‌دار واریانس‌هاست.

$$nB(x) = \sum_j \sum_i \gamma_{ij} \|\mu_j - \bar{x}\|^2$$

$$\Rightarrow \sum_j \Big(\sum_i \gamma_{ij}\Big) W_j(x) + nB(x) = \sum_j \sum_i \gamma_{ij} \left(\|x_{ij} - \mu_j\|^2 + \|\mu_j - \bar{x}\|^2\right)$$ اضافه می‌کنیم و کم می‌کنیم

$$= \sum_j \sum_i \gamma_{ij} \left(x_i^T x_i - 2x_i^T \mu_j + \mu_j^T \mu_j + \mu_j^T \mu_j + \bar{x}^T \bar{x} - 2\mu_j^T \bar{x} - 2x_i^T \bar{x} + 2x_i^T \bar{x}\right)$$

$$Z$$

$$= \sum_j \sum_i \gamma_{ij} \left(x_i^T x_i + \bar{x}^T \bar{x} - 2x_i^T \bar{x} + Z\right)$$

$$= n \sum_i \|x_i - \bar{x}\|^2 + Z = n(nT(x)) + Z = n^2 T(x) + Z$$

حال رابطه‌ی بین T، W_j و B بدست می‌آید:

$$\Rightarrow \sum_j \Big(\sum_i \gamma_{ij}\Big) W_j(x) + nB(x) = n^2 T(x) + Z$$

پ) چون مدل‌های بر مبنای K-means همیشه با افزایش K، به داده‌های موجود overfit می‌شوند و مقدار J کاهش می‌یابد، این امر که با توجه به K مقدار J را کاهش دهیم بی‌معنی است؛ کاهش یافتن (یا ثابت ماندن) J را هم می‌توان با استقرا نشان داد:

اگر یک خوشه به K خوشه‌ی ما اضافه شود، درصورت وجود داشتن نقاطی در خوشه‌ی جدید مقدار J کاهش می‌یابد و اگر نه، ثابت خواهد ماند که البته فرض استقرا می‌گوید در حالت K خوشه‌ای J یک تابع غیرافزایشی است. پس هیچوقت با افزودن خوشه، مقدار J افزایش نخواهد یافت.

Subject:
Date

۴-الف) برای هر دسته به صورت زیر یک توزیع برنولی در نظر می‌گیریم:

$$P(x;p_r) = p_r^{x}(1-p_r)^{1-x} \quad , \quad P(x;p_b) = p_b^{x}(1-p_b)^{1-x}$$

می‌دانیم که هر دسته با احتمال π یا $1-\pi$ انتخاب می‌شود، و z_i مشخص کننده قرمز بودن آن است! با توجه به این موارد می‌توان نوشت:

$$P(x,z;\theta) = \left(\pi p_r^{x}(1-p_r)^{1-x}\right)^{z}\left((1-\pi)p_b^{x}(1-p_b)^{1-x}\right)^{1-z}$$

ب)

$$L_c(\theta) = \sum_i \ln P(x^{(i)},z^{(i)};\theta) = \sum_i z^{(i)}\left(\ln\pi + x^{(i)}\ln p_r + (1-x^{(i)})\ln(1-p_r)\right)$$
$$+ (1-z^{(i)})\left(\ln(1-\pi) + x^{(i)}\ln p_b + (1-x^{(i)})\ln(1-p_b)\right)$$

ج) با فرض دانستن z_iها کافی است از $L_c(\theta)$ نسبت به هر متغیر مشتق بگیریم:

$$\frac{\partial L_c}{\partial \pi} = \sum_1^m \frac{z^{(i)}}{\pi} - \frac{(1-z^{(i)})}{1-\pi} = \frac{1}{\pi}\sum_i z^{(i)} - \frac{1}{1-\pi}\left(m - \sum_i z^{(i)}\right) = 0$$

$$\Rightarrow \pi^* = \frac{\sum_1^m z^{(i)}}{m}$$

$$\frac{\partial L_c}{\partial p_r} = \sum_1^m z^{(i)}\left(\frac{x^{(i)}}{p_r} - \frac{(1-x^{(i)})}{1-p_r}\right) = \frac{1}{p_r}\sum_1^m z^{(i)}x^{(i)} - \frac{1}{1-p_r}\left(\sum_1^m z^{(i)} - \sum_1^m z^{(i)}x^{(i)}\right) = 0$$

$$\Rightarrow p_r^* = \frac{\sum_1^m z^{(i)}x^{(i)}}{\sum_1^m z^{(i)}}$$

$$\frac{\partial L_c}{\partial p_b} = \sum_1^m (1-z^{(i)})\left(\frac{x^{(i)}}{p_b} - \frac{1-x^{(i)}}{p_b}\right) = \frac{1}{p_b}\sum_1^m (1-z^{(i)})x^{(i)} - \frac{1}{1-p_b}\left(\sum_1^m (1-z^{(i)}) - \sum_1^m (1-z^{(i)})x^{(i)}\right)$$

$$\Rightarrow p_b^* = \frac{\sum_1^m (1-z_i)x^{(i)}}{\sum_1^m (1-z^{(i)})}$$

Subject:
Date

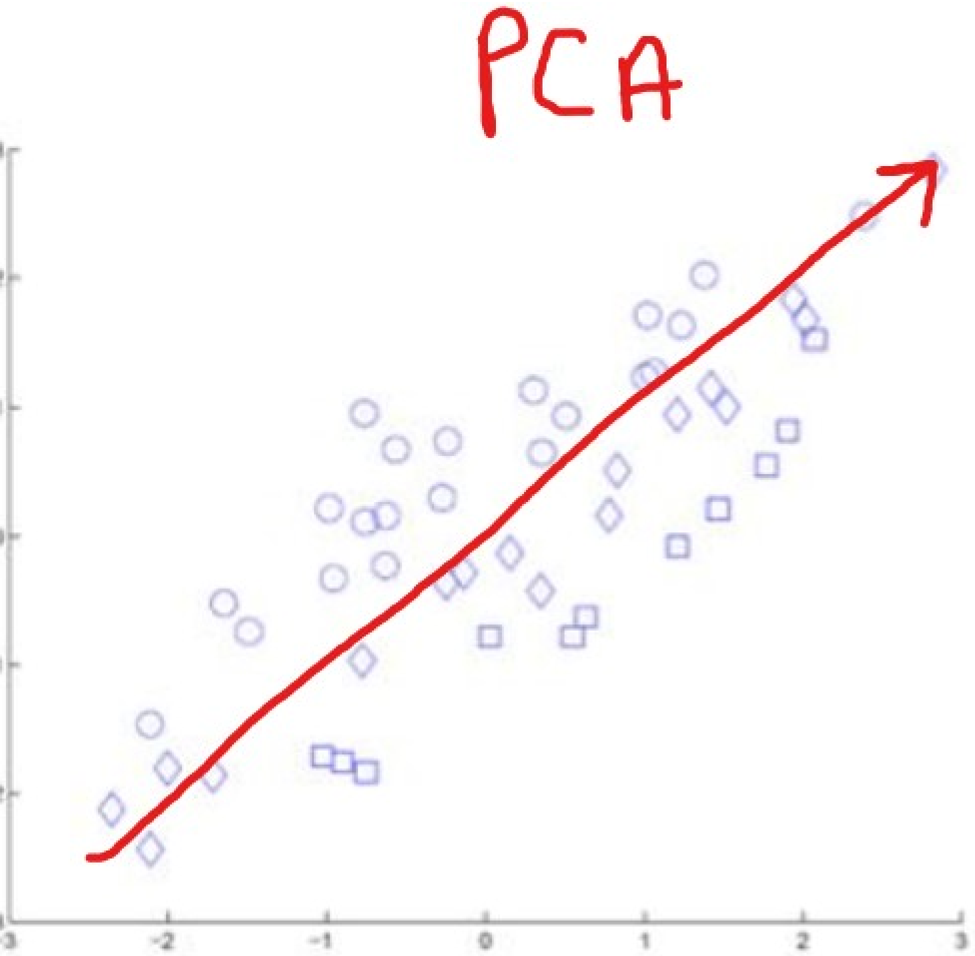
ت)

$$P(z^{(i)}=1|x^{(i)},\theta^t) = \frac{P(x^{(i)}|z^{(i)}=1;\theta^t)\,P(z^{(i)}=1|\theta^t)}{\sum_{j=1}^{n} P(x^{(i)}|z^{(i)}=j;\theta^t)\,P(z^{(i)}=j|\theta^t)}$$

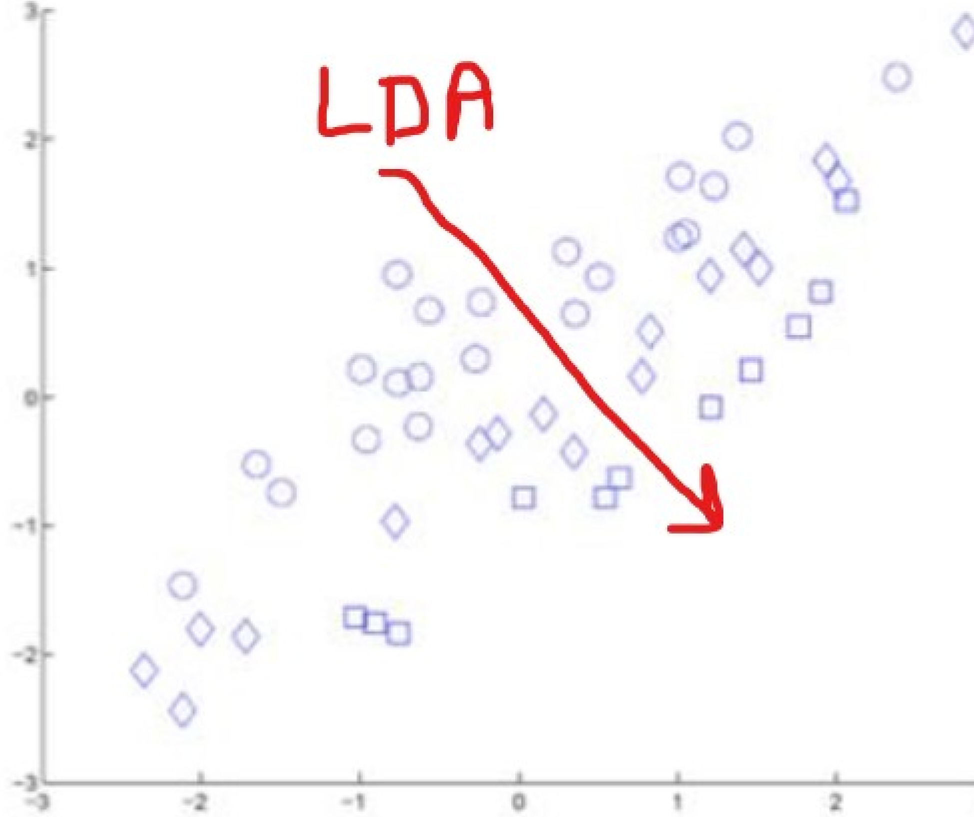
$$= \frac{p_r^{x^{(i)}}(1-p_r)^{1-x^{(i)}}\cdot\pi}{p_r^{x^{(i)}}(1-p_r)^{1-x^{(i)}}\times\pi + p_b^{x^{(i)}}(1-p_b)^{1-x^{(i)}}(1-\pi)}$$

ث) به جای z_i می خواهیم از γ_i^t استفاده کنیم که حال روابط بخش پ را خواهیم داشت:

$$\pi^{t+1} = \frac{\sum_{1}^{m}\gamma_i^t}{m}, \quad p_r^{t+1} = \frac{\sum_{1}^{m}\gamma_i^t x_i}{\sum_{1}^{m}\gamma_i^t}, \quad p_b^{t+1} = \frac{\sum_{1}^{m}(1-\gamma_i^t)x^{(i)}}{\sum_{1}^{m}(1-\gamma_i^t)}$$



PCA
-3
-2
-1
0
1
2
3



LDA
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2