سینا نمازی شماره دانشجویی: ۴۰۰۱۱۰۱۵۴

۱_

الف)

$$X = \begin{bmatrix} -1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} \qquad \vec{\mu} = \frac{\vec{x}_1 + \vec{x}_2 + \vec{x}_3}{3} = [0 \;\; 0] \Rightarrow \tilde{X} = X$$

$$\Rightarrow S = \frac{\tilde{X}\tilde{X}^T}{N} = \begin{bmatrix} \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} \end{bmatrix} \qquad \text{eigenvalues of } S = \{0, \frac{4}{3}\} \hookrightarrow \lambda$$

$$\begin{bmatrix} \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \frac{4}{3} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \Rightarrow \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \text{first Principal component}$$

ب)

$$x_1 = \begin{bmatrix} -1 & -1 \end{bmatrix} \Longrightarrow x'_1 = \begin{bmatrix} -1 & -1 \end{bmatrix} \times \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \underset{\sim}{\begin{bmatrix} -\sqrt{2} \end{bmatrix}}$$

$$x_2 = \begin{bmatrix} 0 & 0 \end{bmatrix} \Longrightarrow x'_2 = \begin{bmatrix} 0 & 0 \end{bmatrix} \times \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \underset{\sim}{\begin{bmatrix} 0 \end{bmatrix}}$$

$$x_3 = \begin{bmatrix} 1 & 1 \end{bmatrix} \Longrightarrow x'_3 = \begin{bmatrix} 1 & 1 \end{bmatrix} \times \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \underset{\sim}{\begin{bmatrix} \sqrt{2} \end{bmatrix}}$$

$$\Longrightarrow \text{Variance} = \frac{(-\sqrt{2}-0)^2 + (0-0)^2 + (\sqrt{2}-0)^2}{3} = \underset{\sim}{\frac{4}{3}}$$

پ)

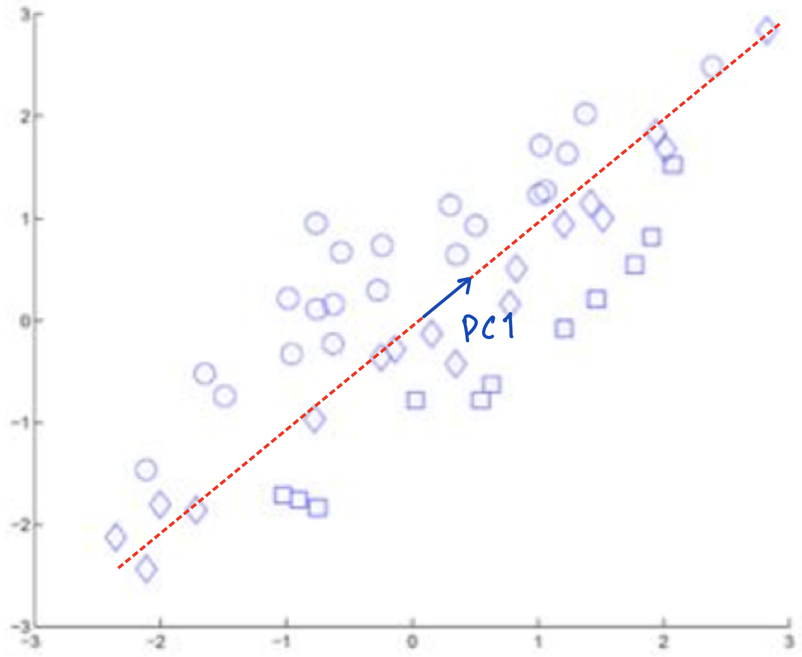
$$\hat{x}_1 = \begin{bmatrix} \sqrt{2} \end{bmatrix} \times \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} -1 & -1 \end{bmatrix} = x_1$$

$$\hat{x}_2 = \begin{bmatrix} 0 \end{bmatrix} \times \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix} = x_2 \Longrightarrow \underset{\sim}{\text{reconstruction error} = 0}$$

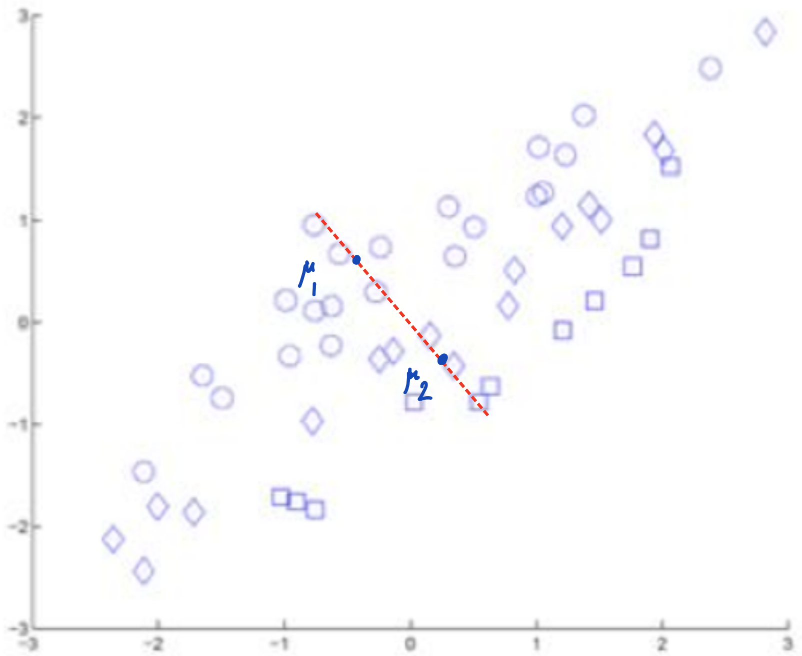
$$\hat{x}_3 = \begin{bmatrix} \sqrt{2} \end{bmatrix} \times \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} 1 & 1 \end{bmatrix} = x_3$$

-2

الف)



ب)



الف) از آن جایی که تابع هزینه الگوریتم K-means ، با گذشت iteration ها به صورت monotonically decreasing است، ماتریس γ یکسان را بیش از یک بار مشاهده نمی‌کنیم. با این که ماتریس γ ، 2^{mk} خانه دارد، ولی می‌دانیم هر x_i به یک خوشه assign می‌شود، تعداد حالت‌های مختلف γ برابر با k^n است. بنابراین الگوریتم در حداکثر k^n قدم به پایان می‌رسد.

ب)

$$\sum_{j=1}^{k}\sum_{i=1}^{n} \gamma_{ij}\, w_j(X) + nB(X) = \sum_{j=1}^{k}\sum_{i=1}^{n} \gamma_{ij}\, \|x_i-\mu_j\|^2 + \gamma_{ij}\, \|\mu_j-\hat{x}\|^2$$

$$= \sum_{j=1}^{k}\sum_{i=1}^{n} \gamma_{ij}\left(\|x_i-\mu_j\|^2 + \|\mu_j-\hat{x}\|^2\right) = \sum_{j=1}^{k}\sum_{i=1}^{n} \gamma_{ij}\left(x_i^T x_i - \hat{x}^T\hat{x} - 2x_i^T\hat{x} + 2x_i^T\hat{x} + \mu_j^T\mu_j - 2x_i^T\mu_j - 2\hat{x}\mu_j\right)$$

$$= \sum_{j=1}^{k}\sum_{i=1}^{n} \gamma_{ij}\left[\left(x_i^T x_i + \hat{x}^T\hat{x} - 2x_i^T\hat{x}\right) + \left(\mu_j^T\mu_j - 2\mu_j^T x_i - 2\mu_j^T\hat{x} + 2x_i^T\hat{x}\right)\right]$$

$$= nT(X) + \underbrace{\sum_{j=1}^{k}\sum_{i=1}^{n} \gamma_{ij}\left(\mu_j^T\mu_j - 2\mu_j^T x_i - 2\mu_j^T\hat{x} + 2x_i^T\hat{x}\right)}_{A}$$

دقت کنید که $\left[\sum_{j=1}^{k}\sum_{i=1}^{n} \gamma_{ij}\, w_j(X)\right]$ که در ابتدا استفاده کردیم، برابر با $J \times n$ است. بنابراین با کمینه کردن J ، nJ نیز کمینه می‌شود (چون n ثابت است). می‌دانیم $T(X)$ ثابت است و بنابراین $nT(X)$ نیز ثابت است. و با فرض ثابت بودن A در فرمول بالا، می‌توان گفت کمینه کردن $\left[\sum_{j=1}^{k}\sum_{i=1}^{n} \gamma_{ij}\, w_j(X)\right]$ معادل با بیشینه کردن $nB(X)$ و در نتیجه بیشینه کردن $B(X)$ است.

بنابراین الگوریتم K-means باعث بیشینه شدن انحراف بین خوشه‌ای (به شکل تقریبی) و کمینه شدن میانگین وزن‌دار مقادیر درون خوشه‌ای می‌شود.

ب) با استقرا نشان می‌دهیم که $J_{min}(k+1)$ کمتر یا مساوی با $J_{min}(k)$ است. فرض کنید با انتخاب k دلخواه الگوریتم را اجرا می‌کنیم و بعد از همگرایی به $J_{min}(k)$ می‌رسیم. حال فرض کنید یک centroid بیش اضافه می‌کنیم و آن را بر روی یکی از k تا centroid قبلی قرار می‌دهیم و تعدادی از اعضای آن خوشه را به صورت تصادفی به خوشه جدید اضافه می‌کنیم. حال $J(k+1)$ برابر با $J_{min}(k)$ شده و با اجرای الگوریتم، این مقدار نیز به شکل monotonic کاهش می‌یابد، بنابراین $J_{min}(k+1) \le J_{min}(k)$. بنابراین انتخاب k براساس کمینه شدن J بی‌معنیست چون با این روش همواره kهای بزرگ‌تر انتخاب می‌شوند و با افزایش k می‌توان J را کاهش داد.

-4

الف)

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

ب)

$$\ln L_c(\theta) = \sum_{i=1}^{m}\left[z_i\left(\ln(\pi) + x_i \ln(P_r) + (1-x_i)\ln(1-P_r)\right) + (1-z_i)\left(\ln(1-\pi) + x_i \ln(P_b) + (1-x_i)\ln(1-P_b)\right)\right]$$

پ)

$$\frac{\partial \ln L_c(\theta)}{\partial P_b} = 0 \Rightarrow \sum_{i=1}^{m}(1-z_i)\left(\frac{x_i}{P_b} - \frac{1-x_i}{1-P_b}\right) = 0 \Rightarrow \hat{P}_b = \frac{\sum_{i=1}^{m}((1-z_i)x_i)}{\sum_{i=1}^{m}(1-z_i)}$$

$$\frac{\partial \ln L_c(\theta)}{\partial P_r} = 0 \Rightarrow \sum_{i=1}^{m} z_i\left(\frac{x_i}{P_r} - \frac{1-x_i}{1-P_r}\right) = 0 \Rightarrow \hat{P}_r = \frac{\sum_{i=1}^{m}(z_i x_i)}{\sum_{i=1}^{m} z_i}$$

$$\frac{\partial \ln L_c(\theta)}{\partial \pi} = 0 \Rightarrow \sum_{i=1}^{m}\left(\frac{z_i}{\pi} - \frac{1-z_i}{1-\pi}\right) = 0 \Rightarrow \hat{\pi} = \frac{\sum_{i=1}^{m} z_i}{m}$$

ت)

$$P(z_i=1 \mid X_i=x_i;\theta^t) = \frac{P(X_i=x_i \mid z_i=1;\theta)\,P(z_i=1;\theta)}{P(X_i=x_i;\theta)} = \frac{\pi P_r^{x_i}(1-P_r)^{1-x_i}}{\pi P_r^{x_i}(1-P_r)^{1-x_i} + (1-\pi)P_b^{x_i}(1-P_b)^{1-x_i}}$$

ث)

با توجه به بخش پ :

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