تمرین سری پنجم یادگیری ماشین

امیرمحمد عزتی 402212269

سوال ①

الف)

$$\Sigma = \frac{1}{n-1} X^T X = \frac{1}{2} \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\det(\Sigma - \lambda I) = 0 \Rightarrow \lambda^2 - 2\lambda = 0 \Rightarrow \begin{cases} \lambda_1 = 0 \\ \lambda_2 = 2 \end{cases} \checkmark$$

$$(\Sigma - \lambda I) v = 0 \Rightarrow \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0 \Rightarrow v_1 = v_2$$

$$\text{eigen vector for } \lambda = 2 \Rightarrow v = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

ب)

$$z_i = x_i^T v$$

$$z_1 = \begin{bmatrix} -1 & -1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = -\sqrt{2}$$

$$z_2 = 0$$

$$z_3 = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \sqrt{2}$$

$$\mathrm{Var}(z) = \frac{1}{3} \sum_{i=1}^{3} (z_i - \mu)^2$$

$$\mu(z) = \frac{-\sqrt{2} + \sqrt{2}}{3} = 0$$

$$\Rightarrow \mathrm{Var}(z) = \frac{4}{3}$$

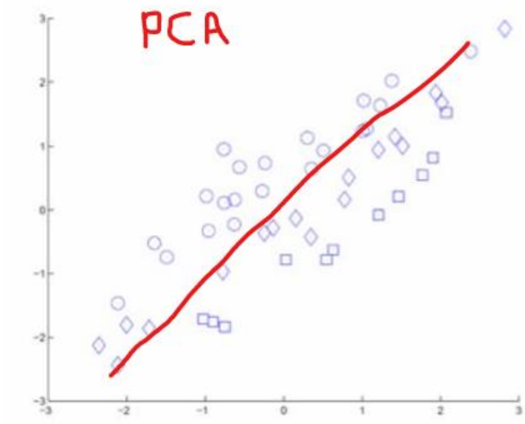
پ)

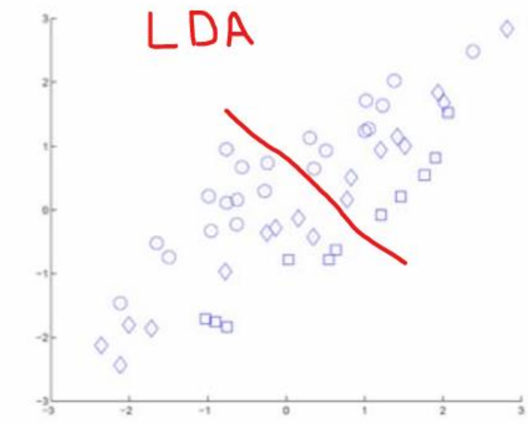
$$\hat{x}_1 = z_1 \cdot v^T = -\sqrt{2} \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} -1 & -1 \end{bmatrix} \checkmark$$

$$\hat{x}_2 = z_2 \cdot v^T = 0 \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

$$\hat{x}_3 = z_3 \cdot v^T = \sqrt{2} \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \end{bmatrix}$$

سوال2:





سوال 3

الف) می‌دانیم که در هر iteration، چون تابع J اکیداً نزولی است، کاهش می‌یابد. این نشان می‌دهد که یک state یکسان از γ در iteration ها نداریم و چون نهایتاً $n \times k$ درایه در ماتریس γ داریم و تعدادش محدود است، پس در تعداد step محدود به پایان می‌رسد.

ب)

$$\sum_{j=1}^{K} \sum_{i=1}^{n} \gamma_{ij} W_j(X) + n B(X) =$$

$$= \sum_{j=1}^{K} \sum_{i=1}^{n} \gamma_{ij} \|x_i - \mu_j\|^2 + \gamma_{ij} \|\mu_j - \hat{x}\|^2 = \sum_{j=1}^{K} \sum_{i=1}^{K} \gamma_{ij} \Big(x_i^2 + \hat{x}^2 - 2x_i\hat{x}$$

$$+ 2x_i\hat{x} + \mu_j^2 - 2x_i\mu_j - 2\hat{x}\mu_j \Big)$$

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

$$= \underbrace{n^2 T(X) + K}_{\text{ثابت}}$$

چون مقدار $\sum_{j=1}^{K} \sum_{i=1}^{n} \gamma_{ij} W_j(X)$ در طول K-means کاهش می‌یابد پس باید مقدار $nB(X)$ افزایش یافته و maximized شود.

پ) فرض می‌کنیم مقدار تابع J با مقدار K کلاستر دیگر بدون کاهش است. آنگاه اگر یک کلاستر اضافه کنیم $(K+1)$ حداقل آن نقطه‌ای که کلاستر جدید شده است $J=0$ دارد پس مقدار J کلی کاهش می‌یابد. پس تعداد کلاستر را اگر تا n افزایش دهیم، همواره میزان J کاهش می‌یابد تا به صفر برسد. اما این کار باعث overfit روی داده است پس بی‌معنی می‌باشد.

Scanned with CamScanner

سوال ④

الف) توزیع برنولی برای پرتاب یک سکه اگر P_0 احتمال رو آمدن باشد:

$$P(x|P_0) = \prod_{i=1}^{m} P_0^{x_i} (1-P_0)^{1-x_i}$$

در این حالت با دو سکه آبی و قرمز و با احتمال رو آمدن P_r ، P_b

با احتمال π سکه قرمز و با احتمال $1-\pi$ سکه آبی انتخاب می‌شود.

$$P(x|P_r, P_b) = \prod_{i=1}^{m} \left[\pi \left(P_r^{x_i} (1-P_r)^{1-x_i} \right) + (1-\pi) \left(P_b^{x_i} (1-P_b)^{1-x_i} \right) \right]$$

اگر متغیر تصادفی مشاهده نشده هم در نظر بگیریم در آخر به عبارت زیر می‌رسیم:

$$P(x, Z; \theta) = L = \prod_{i=1}^{m} \left[\pi \left(P_r^{x_i} (1-P_r)^{1-x_i} \right) \right]^{Z_i} \left[(1-\pi) \left(P_b^{x_i} (1-P_b)^{1-x_i} \right) \right]^{1-Z_i}$$

ب)

$$\ln p(x, z; \theta) = \sum_{i=1}^{m} \left[Z_i \ln \left(\pi P_r^{x_i} (1-P_r)^{1-x_i} \right) \right] + \left[(1-Z_i) \ln \left((1-\pi) P_b^{x_i} (1-P_b)^{1-x_i} \right) \right]$$

$$\Rightarrow \ln p(x, z; \theta) = \sum_{i=1}^{m} Z_i \left[\ln \pi + x_i \ln P_r + (1-x_i) \ln (1-P_r) \right] + (1-Z_i) \left[\ln (1-\pi) + x_i \ln P_b + (1-x_i) \ln (1-P_b) \right]$$

پ) از log-likelihood نسبت به π مشتق می‌گیریم:

$$\frac{\partial \ln L}{\partial \pi} = \sum_{i=1}^{m} Z_i \left(\frac{1}{\hat{\pi}} \right) - \frac{1-Z_i}{1-\hat{\pi}} = 0$$

$$\sum_{i=1}^{m} Z_i (1-\hat{\pi}) - \hat{\pi} (1-Z_i) = 0 \quad \Rightarrow \quad \hat{\pi} = \frac{1}{m} \sum_{i=1}^{m} Z_i$$

نسبت به P_r مشتق می‌گیریم:

$$\frac{\partial \ln L}{\partial P_r} = \sum_{i=1}^{m} Z_i\left(\frac{x_i}{\hat{P}_r^n}\right) - Z_i\left(\frac{1-x_i}{1-\hat{P}_r}\right) = 0$$

$$\sum_{i=1}^{m} Z_i x_i - Z_i x_i \hat{P}_r^n + Z_i x_i \hat{P}_r - P_r Z_i = 0$$

$$\Rightarrow \hat{P}_r^n = \frac{\sum_{i=1}^{m} Z_i x_i}{\sum_{i=1}^{m} Z_i}$$

نسبت به P_b مشتق می‌گیریم:

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

$$\sum_{i=1}^{m} (1-Z_i)x_i - (1-Z_i)x_i\hat{P}_b + (1-Z_i)x_i\hat{P}_b^n - \hat{P}_b(1-Z_i) = 0$$

$$\Rightarrow \hat{P}_b = \frac{\sum_{i=1}^{m}(1-Z_i)x_i}{\sum_{i=1}^{m}(1-Z_i)}$$

ت) θ_t پارامتر در لحظه t باشد برای تخمین توزیع Z در لحظه $t+1$ داریم:

$$P(Z_i=1 \mid x_i, \theta_t) = \frac{P(x_i \mid Z_i=1, \theta_t)\, P(Z_i=1 \mid \theta_t)}{P(x_i \mid \theta_t)}$$

$$P(Z_i=1 \mid x_i, \theta_t) = \frac{P(x_i \mid Z_i=1, \theta_t)\, P(Z_i=1 \mid \theta_t)}{P(x_i \mid Z_i=1, \theta_t)\, P(Z_i=1 \mid \theta_t) + P(x_i \mid Z_i=0, \theta_t)\, P(Z_i=0 \mid \theta_t)}$$

$$\begin{cases} P(Z_i=1 \mid \theta_t) = \pi^t \\ P(x_i \mid Z_i=1, \theta_t) = (P_r^t)^{x_i}(1-P_r^t)^{1-x_i} \\ P(x_i \mid Z_i=0, \theta_t) = (P_b^t)^{x_i}(1-P_b^t)^{1-x_i} \end{cases} \Rightarrow$$

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

ث) امید ریاضی log-likelihood :

$$E[L] = \sum_{i=1}^{m} \left[\gamma_i^t \ln p(x_i, 1; \theta) + (1-\gamma_i^t) \ln p(x_i, 0; \theta) \right]$$

مشتق نسبت به پارامترها را مشتق حساب می‌کنیم. امّا چون در بخش "پ" محاسبه شده
صرفاً بازنویسی می‌کنیم:

$$\pi^{t+1} = \frac{1}{m} \sum_{i=1}^{m} \gamma_i^t$$

$$P_r^{t+1} = \frac{\sum_{i=1}^{m} \gamma_i^t x_i}{}$$

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