DATE / / SUBJECT:

«به نام خدا»

امیررضا آذری

۹۹۱۰۱۰۸۷

یادگیری ماشین

تمرین پنجم

۱

الف)

با حل SVD مقدار $v=\left[\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]^T$ بدست می‌آید. البته به طور واضح نیازی به SVD هم نیست و با توجه به نقاط، مولفه اصلی اول مشخص است. ($\left[-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right]^T$ نیز درست است.)

ب)

$$z_1 = [-1, -1]\begin{bmatrix}\frac{\sqrt{2}}{2}\\ \frac{\sqrt{2}}{2}\end{bmatrix} = -\sqrt{2}$$

$$z_2 = [0, 0]\begin{bmatrix}\frac{\sqrt{2}}{2}\\ \frac{\sqrt{2}}{2}\end{bmatrix} = 0$$

$$z_3 = [1, 1]\begin{bmatrix}\frac{\sqrt{2}}{2}\\ \frac{\sqrt{2}}{2}\end{bmatrix} = \sqrt{2}$$

$$\Rightarrow \text{sample mean} = 0, \quad \text{variance} = \frac{1}{3}\sum_{i=1}^{3}(z_i - 0)^2 = \frac{1}{3}(2+2) = \frac{4}{3}$$

$$\frac{1}{3-1}(2+2) = 2 \leftarrow \text{unbiased estimation} \ *$$

پ)

$$\hat{x}_1 = z_1 . v = -\sqrt{2} . \begin{bmatrix}\frac{\sqrt{2}}{2}\\ \frac{\sqrt{2}}{2}\end{bmatrix} = [-1, -1]$$

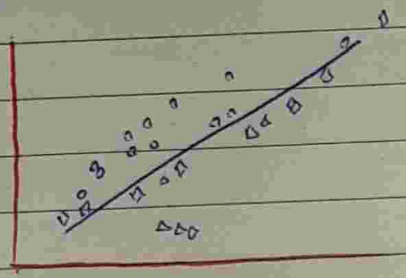
$$\hat{x}_2 = z_2 . v = 0$$

$$\hat{x}_3 = z_3 . v = \sqrt{2} . \begin{bmatrix}\frac{\sqrt{2}}{2}\\ \frac{\sqrt{2}}{2}\end{bmatrix} = [1, 1]$$

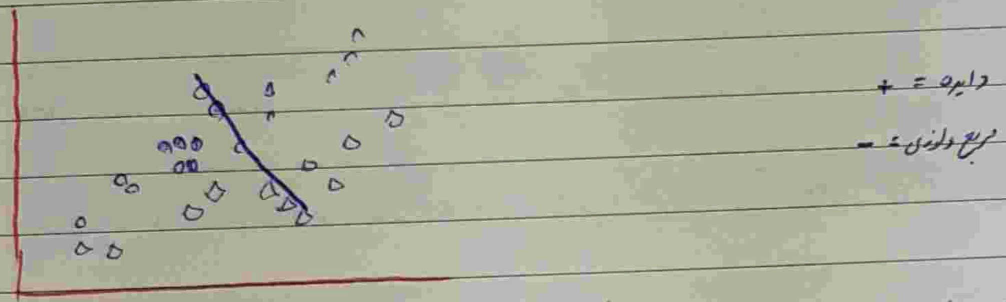
$$\hat{x}_1 = x_1, \ \hat{x}_2 = x_2, \ \hat{x}_3 = x_3 \longrightarrow \text{reconstruction error} = 0$$

۲

الف) First PCA component :



ب) First LDA component :



* تصویرها ممکن است نادقیق باشد اما سعی کردم منظور کامل را برسانم.

۳

الف)

از مطالب تدریس شده در کلاس می‌دانیم که L با هر پیمایش kmeans، یک monotonic decreasing است. این به این معناست که ما نمی‌توانیم یک استیت برای γ را ۲ بار مشاهده کنیم. از آنجایی که γ، $n^۲$ تا entry دارد (۰ یا ۱)، تعداد متناهی از مقادیر ممکن دارد. از این نتیجه می‌شود که kmeans حتی باید در نهایت به پایان برسد.

ب)

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

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

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

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

$$= n^2 T(X) + K$$

$$\Rightarrow \sum_{j=1}^{k} \sum_{i=1}^{n} \gamma_{ij} W_j(X) + nB(X) = n^2 T(X) + K$$

دقت کنید $\sum_{j=1}^{k}\sum_{i=1}^{n}\gamma_{ij}W_j(X)$ همان ل است بنابراین در حین kmeans ، minimized می‌شود. از آنجایی که $n^2T(X)$ یک ثابت و constant است، نتیجه می‌شود $nB(X)$ به طور تقریبی maximized می‌شود.

پ)

با یک استدلال استقرایی ساده نشان می‌دهیم. فرض کنید که تا مقداری از k ، ل در k غیرکاهشی باشد. حال فرض کنید یک cluster دیگر به بعضی location دلخواه اضافه کنیم. از آنجایی که این اجرای تازه kmeans هنوز کانورج نکرده است، می‌دانیم که در نقطه مینیمم ل برای k+1 کلاستر نیستیم. اگر نشان دهیم که ل همچنان غیرکاهشی است، آنگاه گام استقرا را تکمیل کرده‌ایم.

توجه داشته باشید که حداقل برای نقطه‌ای که الان مرکز خوشه است، عبارت ل برابر ۰ خواهد بود.

این بدان معناست که ل کاهش یافته است هنگامی که یک کلاستر جدید اضافه شد.

اگر بخواهیم k را انتخاب کنیم که ل را به حداقل برساند، ما در نهایت k=n را برمی‌گزینیم زیرا ل برابر صفر می‌شود.

از اینجا که ما نمی‌خواهیم خیلی اورفیت کنیم، نمی‌توانیم به سادگی argmax را برای k انتخاب کنیم.

۴

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)

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

پ)

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

$$\Rightarrow \hat{p}_b = \frac{\sum_{i=1}^{m} \left((1-z_i) x_i\right)}{\sum_{i=1}^{m} (1-z_i)}$$

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

$$\Rightarrow \hat{\pi} = \frac{\sum_{i=1}^{m} z_i}{m}$$

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

ت)

$$p(z_i=1 \mid X_i=x_i ; \theta^t) = \frac{p(X_i=x_i \mid z_i=1, \theta)\, p(z_i=1 \mid \theta)}{p(X_i=x_i \mid \theta)}$$

$$= \frac{\pi p_r^{x_i} (1-p_r)^{1-x_i}}{\pi p_r^{x_i} (1-p_r)^{1-x_i} + (1-\pi) p_b^{x_i} (1-p_b)^{1-x_i}}$$

ث)

معادل با پاسخ بخش ب خواهد بود. تنها z_i را با γ_i^t جایگزین می‌کنیم.

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

$$\hat{p}_r^{t+1} = \frac{\sum_{i=1}^m (\gamma_i^t x_i)}{\sum_{i=1}^m \gamma_i^t}$$

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