حل تمرین دوم | یادگیری بدون نظارت | مهلت ارسال:

۱. (۰ نمره) مجموعه دادهٔ زیر را در نظر بگیرید.

$$D = \{(۱,۲),(۲,۳),(۳,۵),(۴,۴),(۵,۶)\}$$

(الف) الگوریتم PCA را روی این داده‌ها اجرا کنید. PC1 و PC2 را مشخص کنید.
(ب) داده‌ها را به فضای جدید منتقل کنید.
(پ) اگر فقط از PC1 استفاده شود، چند درصد از واریانس داده‌ها بیان می‌شود؟

۲. (۰ نمره) در این مسئله می‌خواهیم به بررسی الگوریتم خوشه‌بندی K-means بپردازیم. فرض کنید

$$X = \{x_۱, x_۲, \dots, x_n\}$$

داده‌های ما باشند و γ ماتریس indicator باشد به این صورت که $\gamma_{ij} = ۱$ اگر x_i متعلق به خوشهٔ jام باشد و در غیر این صورت برابر ۰ است. فرض کنید $\mu_۱, \mu_۲, \dots, \mu_k$ میانگین خوشه‌ها باشند.
اعوجاج J برای داده‌ها به صورت زیر محاسبه می‌شود:

$$J(\gamma, \mu_۱, \dots, \mu_k) = \sum_{j=۱}^{k} \sum_{i=۱}^{n} \gamma_{ij} \|x_i - \mu_j\|^۲$$

همچنین

$$C = \{۱, ۲, \dots, k\}$$

را به عنوان مجموعه‌ی خوشه‌ها در نظر بگیرید.

(آ) نشان دهید که الگوریتم در تعداد متناهی قدم به پایان می‌رسد. (γ چند مقدار متفاوت می‌تواند بگیرد؟)
(ب) فرض کنید $\hat{x}$ میانگین داده‌های نمونه باشد. مقادیر زیر را در نظر بگیرید:

$$T(X) = \frac{\sum_{i=۱}^{n} \|x_i - \hat{x}\|^۲}{n}$$

$$W_j(X) = \frac{\sum_{i=۱}^{n} \gamma_{ij} \|x_i - \mu_j\|^۲}{\sum_{i=۱}^{n} \gamma_{ij}}$$

$$B(X) = \sum_{j=۱}^{k} \left(\frac{\sum_{i=۱}^{n} \gamma_{ij}}{n} \right) \|\mu_j - \hat{x}\|^۲$$

در اینجا $T(X)$ نشان‌دهنده‌ی انحراف کل، $W_j(X)$ انحراف درون‌خوشه‌ای، و $B(X)$ انحراف بین‌خوشه‌ای است.
رابطه‌ی بین این سه مقدار به چه صورت است؟ نشان دهید که K-means را می‌توان به عنوان کمینه‌کننده‌ی میانگین وزن‌دار مقادیر درون‌خوشه‌ای، و تقریباً بیشینه‌کننده‌ی انحراف بین‌خوشه‌ای دید.

(پ) نشان دهید که کمینه‌ی J تابعی غیرافزایشی بر حسب k (تعداد خوشه‌ها) است. بنابراین چرا انتخاب تعداد خوشه‌ها با کمینه کردن J کاری بی‌معنی است؟

۳. (۰ نمره) دوست شما دو سکه دارد: یک سکه قرمز و یک سکه آبی با بایاس‌های p_r و p_b (سکهٔ قرمز با احتمال p_r و سکهٔ آبی با احتمال p_b رو می‌آید). او همچنین یک ترجیح π برای انتخاب سکهٔ قرمز دارد.

او m پرتاب سکه انجام می‌دهد. برای هر پرتاب، ابتدا سکهٔ قرمز را با احتمال π یا سکهٔ آبی را با احتمال $(1-\pi)$ انتخاب می‌کند. این پروسه برای هر پرتاب مستقل است. ما از انتخاب او اطلاع نداریم و فقط خروج پرتاب‌ها (رو آمدن یا نیامدن) را مشاهده می‌کنیم.

برای هر پرتاب i، متغیر تصادفی X_i به صورت زیر تعریف می‌شود:

$$X_i = \begin{cases} 1 & \text{اگر رو آمده باشد} \\ 0 & \text{otherwise} \end{cases}$$

بنابراین دادهٔ قابل مشاهده برای ما مقادیر

$$x_1, x_2, \dots, x_m$$

هستند که از m متغیر تصادفی به‌دست آمده‌اند. با توجه به این داده، می‌خواهیم پارامترهای

$$\theta = (\pi, p_r, p_b)$$

را تخمین بزنیم.

برای کمک به این موضوع، برای هر پرتاب i متغیر تصادفی مشاهده‌نشدهٔ Z_i را به صورت زیر در نظر می‌گیریم:

$$Z_i = \begin{cases} 1 & \text{اگر در پرتاب } i \text{ از سکهٔ قرمز استفاده شده باشد} \\ 0 & \text{otherwise} \end{cases}$$

(آ) عبارت توزیع احتمال توأم $p(x, z; \theta)$ را به‌دست آورید.

(ب) عبارت complete-data log-likelihood یعنی

$$L_c(\theta) = \sum_{i=1}^{m} \ln p(x_i, z_i; \theta)$$

را بنویسید.

(پ) فرض کنید مقادیر z_i معلوم هستند. تخمین بیشینهٔ درست‌نمایی را برای پارامترهای $\hat{\pi}$، $\hat{p}_r$، $\hat{p}_b$ برحسب x_i و z_i به‌دست آورید.

(ت) با ندانستن مقادیر z_i از الگوریتم EM استفاده می‌کنیم. فرض کنید الگوریتم با مقادیر اولیه $\theta^{(0)}$ شروع شده و $\theta^{(t)}$ پارامترهای شروع تکرار tام است. در گام E-step نیاز به محاسبهٔ

$$P(Z_i = 1 \mid X_i = x_i; \theta^{(t)})$$

داریم. این مقدار را به‌دست آورید.

(ث) برای هر پرتاب i مقدار محاسبه‌شده در بخش قبل را با $\gamma_i^{(t)}$ نمایش می‌دهیم. امید ریاضی complete-data log-likelihood برابر است با:

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

در گام M-step می‌خواهیم $\theta^{(t+1)}$ را طوری تخمین بزنیم که این عبارت بیشینه شود. پارامترهای آپدیت‌شدهٔ $\theta^{(t+1)}$ یعنی $\pi^{(t+1)}$، $p_r^{(t+1)}$، $p_b^{(t+1)}$ را برحسب x_i و $\gamma_i^{(t)}$ مشخص کنید.

۴. (۰ نمره) یک GMM با یک کامپوننت را در فضای $\mathbb{R}^d$ در نظر بگیرید. فرض کنید داده‌های

$$x_1, x_2, \ldots, x_n$$

به صورت .i.i.d مشاهده شده‌اند. تخمینگر MLE برای μ را بیابید.

یک GMM با دو کامپوننت که مشابه هستند را در نظر بگیرید. فرض کنید وزن‌های اختصاص‌یافته به این دو کامپوننت برابر π_1 و π_2 باشند. نشان دهید که ترکیب این دو کامپوننت معادل یک کامپوننت با وزن ترکیبی $\pi_1 + \pi_2$ است.

یک GMM باینری در فضای یک بعدی بگیرید.

$$p(x \mid z = 1) = \mathcal{N}(x \mid \mu_1, \sigma^2), \qquad p(x \mid z = 2) = \mathcal{N}(x \mid \mu_2, \sigma^2)$$

$$p(z = 1) = \pi_1 p(z = 2) = \pi_2 \pi_1 + \pi_2 = 1$$

طبقه بندی بر این اساس است:

$$p(z = 1 \mid x) > p(z = 2 \mid x)$$

مرز طبقه بندی را بیابید