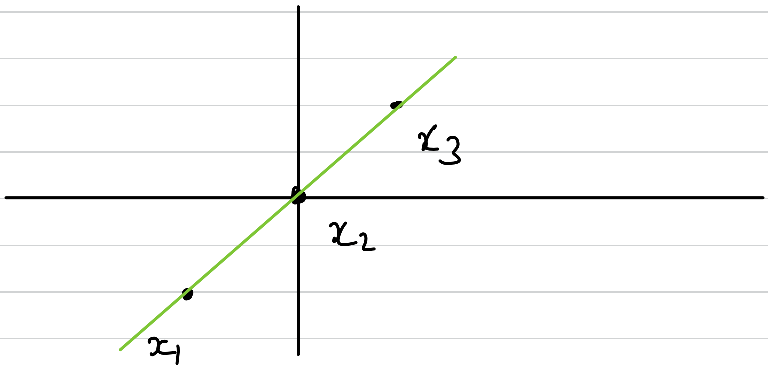
تمرین 6 یادگیری ماشین 99109393 علیرضا حبیب زاده

1.

$x_1 = (-1,-1) \qquad x_2 = (0,0) \qquad x_3 = (1,1)$

(1



پاسخ به صورت شهودی واضح است اما با همان روش کلی حساب میکنیم :

$(\|v\| = 1)$

$$\sigma^2 = \sum (v^T x_i)^2 = (X^T v)^2 = v^T X X^T v \ , \ v^T v = 1$$

$$\nabla L = \nabla_v (\sigma^2 - \lambda v^T v) = 2(XX^T - \lambda I)v = 0$$

$$\Rightarrow |XX^T - \lambda I| = 0$$

$$X^T = \begin{bmatrix} — x_1 — \\ — x_2 — \\ — x_3 — \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}$$

$$XX^T = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 6 & 1 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

$$\begin{vmatrix} 2-\lambda & 2 \\ 2 & 2-\lambda \end{vmatrix} = (2-\lambda)^2 = 2^2 \Rightarrow 2-\lambda = \pm 2$$

$$\lambda = 2 \pm 2$$

$$= 4, 0$$

اما $rank(XX^T)=1$ فقط یک مولفه PCA داریم. $(\lambda = 4)$ به صورت

شهودی هم داده ها با یک محور قابل توصیف هستند.

$$\lambda = 4 \Rightarrow \begin{bmatrix} -2 & 2 \\ & \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ \end{bmatrix} \Longrightarrow x = y$$

$$\text{و } x^2 + y^2 = 1 \Longrightarrow \boxed{v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}} \quad \text{یا } v_1 = \frac{-1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

طبق تعریف درس تنها ویژه‌بردارهای با ویژه‌مقدار ناصفر مولفه‌ی اصلی (لازم) هستند و

تعدادشان هم رتبه‌ی ماتریس XX^T است. اما به هر حال محاسبات آن مولفه‌ی دیگر

را هم که برای توصیف داده‌ها لازم نیست و واریانس صفر دارد بیابیم:

$$\lambda = 0 \Rightarrow \begin{bmatrix} 2 & 2 \\ & \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ \end{bmatrix} \Rightarrow x = -y$$

$$\text{و } x^2 + y^2 = 1 \Longrightarrow \boxed{v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}} \quad \text{یا } v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

(2

$$x_1' = \begin{bmatrix} v_1^T x_1 \\ v_2^T x_2 \end{bmatrix} = \begin{bmatrix} -\sqrt{2} \\ 0 \end{bmatrix}$$

$$x_2' = \sim = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$x_3' = \sim = \begin{bmatrix} \sqrt{2} \\ 0 \end{bmatrix}$$

$$\sigma_1^2 = {x'_{1_1}}^2 + {x'_{2_1}}^2 + {x'_{3_1}}^2 = \sqrt{2}^2 + 0^2 + \sqrt{2}^2 = 4$$

که با $\lambda_1 = 4$ هم خواندن داشتیم چون λ $= v_1^T \underbrace{X X^T v_1}_{\lambda v_1}$.

$$\sigma_2^2 = {x'_{1_2}}^2 + {x'_{2_2}}^2 + {x'_{3_2}}^2 = 0 + 0 + 0 = 0 \quad (\lambda_2 = 0)$$

واریانس های نرمال شده : $\sigma_1^2/n = 4/3$, $\sigma_2^2/n = 0$

(3 خطای بازسازی صفر است چون هر سه نقطه روی خط قرار دارند . به صورت مستقیم :

$$E_1 = \| x_1 - (x_1 \cdot e_1) e_1 \|^2 =$$

$$\| (-1,-1) - (-1,-1) \cdot \frac{1}{\sqrt{2}} (1,1) \cdot \frac{1}{\sqrt{2}} (1,1) \|^2$$

$$= \| (-1,-1) - \frac{1}{2} (-2,-2) \|^2 = 0 \quad \checkmark$$

$$E_2 = \|x_2 - (x_2 \cdot e_1) e_1\|^2 = \| (0,0) - (0,0) \cdot \frac{(1,1)}{\sqrt{2}} \frac{(1,1)}{\sqrt{2}} \|^2$$

$$= 0 \quad \checkmark$$

$$E_3 = \| x_3 - (x_3 \cdot e_1) e_1 \|^2 = \| (1,1) - (1,1) \cdot \frac{(1,1)}{\sqrt{2}} \frac{(1,1)}{\sqrt{2}} \|^2 = 0 \checkmark$$

2.

(1

A به چیزی بستگی ندارد :

$A: \quad P(A)$

$F: \quad P(F)$

$S: \quad P(S \mid F, A)$

$H: \quad P(H \mid S)$

$N: \quad P(N \mid S)$

صرفا در الگوریتم Expected Maximization (EM) باید دقت کنیم A تمام موقعیت ها

مشاهده شده نیست پس باید مقدار آن را در این موقع تخمین بزنیم.

(2

$$P(A=1 \mid F, S, H, N) = \frac{P(A=1, F, S, H, N)}{P(F, S, H, N)}$$

$$P(A=1, F, S, H, N) = P(A=1) \cdot P(F) \cdot P(S \mid F, A=1) \cdot P(H \mid S) \cdot P(N \mid S)$$

$$P(F, S, H, N) = P(F) \cdot P(H \mid S) \cdot P(N \mid S) \times \left[P(A=0) \cdot P(S \mid F, A=0) + P(A=1) \cdot P(S \mid F, A=1) \right]$$

$$\Rightarrow P(A=1 \mid F,S,H,N) =$$

$$\frac{P(A=1).P(F).P(S \mid F,A=1).P(H \mid S).P(N \mid S)}{P(F).P(H \mid S).P(N \mid S) \times \left[P(A=0).P(S \mid F,A=0) + P(A=1).P(S \mid F,A=1)\right]}$$

3) باید طبق تعریف پتوی مارکوف والدین ، فرزندان و همسایه والدین فرزندان

A را لحاظ کنیم که متغیر های $\boxed{F \text{ و } S}$ می‌شود .

4) عبارت های خارج پتو چون $P(N \mid S)$ و $P(H \mid S)$ از صورت و مخرج ساده

می‌شوند . $P(F)$ هم می‌شود اما نیازی نیست .

$$P(A=1 \mid F,S,H,N) = \frac{P(A=1).P(F).P(S \mid F,A=1)}{P(F).\left[P(A=0).P(S \mid F,A=0) + P(A=1).P(S \mid F,A=1)\right]}$$

5) برای توزیعی که آن را نمایندگی کند معمولاً $= 1$ و $= 0$ را طلب

مثلاً طبق تجربه کاربرد اما چون سوال از $1 =$ استفاده کرده این پارامتر ها وابسته

هستند :

$P(A=1)$

$P(S=1, F=0, A=0)$

$P(S=1, F=0, A=1)$ (چهار حالت منطقی ممکن F و A)

$P(S=1, F=1, A=0)$

$P(S=1, F=1, A=1)$

6) جای متغیر غیر قابل مشاهده (؟) مقدار احتمال محاسبه شده را به عنوان تخمین قرار میدهیم و الان 5 مقدار مشاهده شده داریم :

$$P(A=1) = \frac{1}{5}(1+1+0+0.8+0.4) = \boxed{0.64}$$

$$P(S=1 \mid F=0, A=1) = \frac{P(S=1, F=0, A=1)}{P(F=0, A=1)}$$

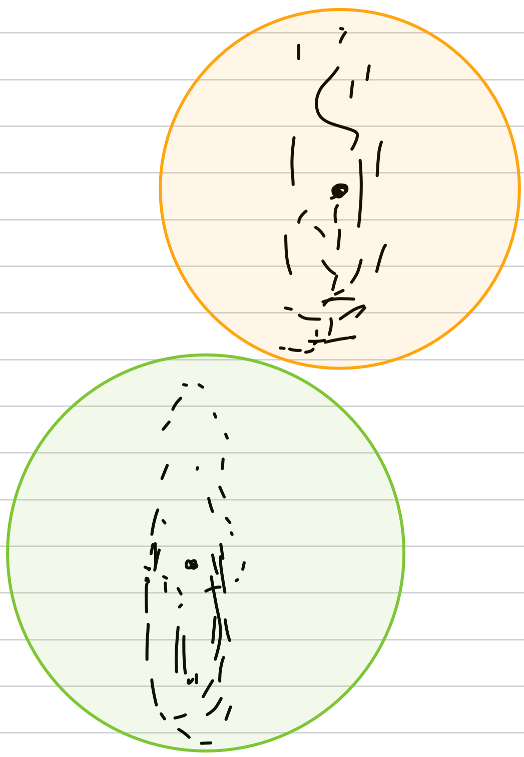
$$= \frac{\cancel{1/5}\,(1+0+0+0.8\times 0+0.4\times 1)}{\cancel{1/5}\,(1+1+0+0.8+0.4)} = \boxed{0.4375}$$

3.

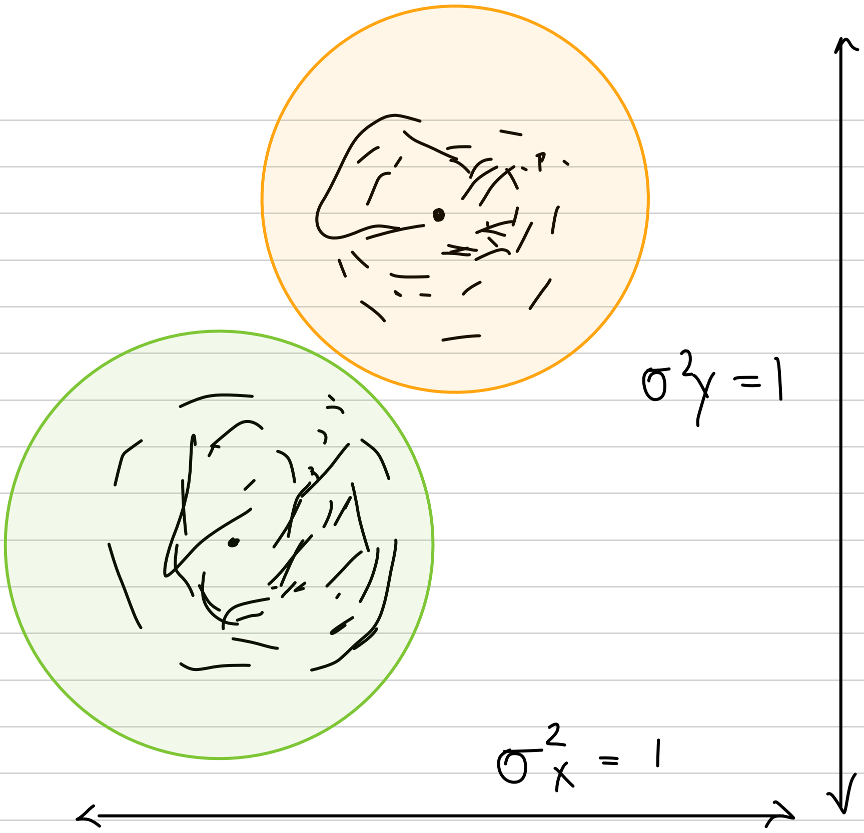
1) به نتیجه‌ی درستی نخواهیم رسید و نمی‌توانیم به درستی دسته‌بندی
کنیم چرا که دسته‌بندی k-means نمی‌تواند اشکال هندسی غیر محدب را دسته‌بندی
کند چرا که دسته‌ها بر اساس فاصله‌ی اقلیدسی از تعدادی نقاط تعیین می‌شوند.
شکل حاصل حتما یک شکل محدب خواهد بود و لذا اشکال غیر محدب
یا شکل در تصویر قابل تفکیک نیستند.

2) نرمال‌سازی مشکل الگوریتم در این حالت خاص را حل نمی‌کند. اما به طور کلی
نرمال کردن داده‌ها باعث می‌شود الگوریتم k-means با همه‌ی ابعاد داده‌ها
به یک اهمیت برخورد کند. چرا که الگوریتم در فضا نقاط و دایره /
(کره) هایی قرار می‌دهد. پس یعنی نمی‌تواند بین ابعاد مختلف و اهمیت
آن‌ها تغییر قائل شود. این گونه بعدی که ذاتا در آن تغییر کم است
کم اهمیت‌تر از بعدی خواهد بود که در آن تغییرات بزرگ است.
ضمنا خطاها و نویزهای ابعاد با اعداد بزرگ هم amplify می‌شود.
مثال: فرض کنید دو بعد حقوق با هم از نظر اهمیت دسته‌بندی تفاوتی ندارند،
اما یک بعد را به سانتی‌متر و دیگری را با متر اندازه‌گیری کنیم و به خورد

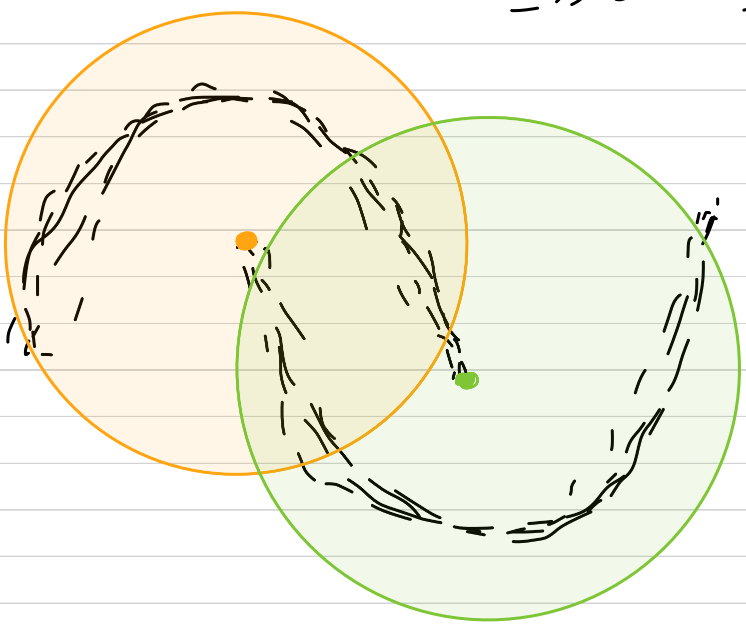
الگوریتم بدهیم. یا فرض کنید نوسانات یا قیمت یک چیز را در یک جهت به ریال بدهیم و چیز دیگر را در جهت دیگر به دلار (که نسبت به هم برابر است بر اساس هردوی این‌ها نمی‌توانند هم دسته‌بندی کنند) حال آن جهتی که ریال بود و نوسانات قبل (۱) آن جهتی که به سانتی‌متر بود خیلی اهمیت بیشتری پیدا خواهند کرد و به تنهایی شعاع دایره‌ها را تعیین خواهند کرد.



تفاوت در جهت x (افقی) دیده نمی‌شود. اما با اعمال نرمال‌سازی حال واقعا دیده وقت خوب به این‌ها خواهد بود که اگر $\sigma_x^2 = \sigma_y^2 = 1$ (پس از نرمال کردن) و تقریبا چنین شکلی خواهند داشت :



۳) احتمالا الگوریتم به چنین نتیجه‌ای خواهد رسید :



عموما امکان تفکیک

اثبات ریاضی : فرض خلف کنیم k-means توانسته باشد شکل سمت چپ بالا را به درستی تفکیک کند. در آن صورت شکل‌ها را داخل دایره‌ای به مرکز بدست آمده است پس دایره‌ای به این صورت وجود دارد :

هر دو نقطه ای که داخل دایره به هم وصل شوند

نیز داخل دایره هستن

این فضای داخلی شکل ∩ که داریم همه داخل دایره‌ی خود ∩ است

برای ∪ هم همینطور. برای همین k-means نمی‌تواند هر یک از داخل

این‌ها را به صورت هم زمان درست تشخیص داده باشد.

الگوریتم k-means با تفکیک اشکال غیر محدب مشکل دارد.