Subject:

Year. Month. Date. ( )

مثال : طبقه‌بندی در دو دسته

α_1 : deciding ω_1

α_2 : deciding ω_2

$\lambda_{ij} = \lambda(\alpha_i | \omega_j)$

خطر شرطی برای هریک از دسته‌ها :

$$R(\alpha_1|x) = \lambda_{11} P(\omega_1|x) + \lambda_{12} P(\omega_2|x)$$

$$R(\alpha_2|x) = \lambda_{21} P(\omega_1|x) + \lambda_{22} P(\omega_2|x)$$

قاعده‌ی تصمیم‌گیری بیزی برای حداقل کردن خطر :

decide ω_1 if $R(\alpha_1|x) < R(\alpha_2|x)$

$$\Rightarrow \lambda_{11} P(\omega_1|x) + \lambda_{12} P(\omega_2|x) < \lambda_{21} P(\omega_1|x) + \lambda_{22} P(\omega_2|x)$$

$$\Rightarrow (\lambda_{11} - \lambda_{21}) P(\omega_1|x) < (\lambda_{22} - \lambda_{12}) P(\omega_2|x)$$

$$\Rightarrow (\lambda_{11} - \lambda_{21}) p(x|\omega_1) P(\omega_1) < (\lambda_{22} - \lambda_{12}) p(x|\omega_2) P(\omega_2)$$

$$\Rightarrow \frac{p(x|\omega_1)}{p(x|\omega_2)} > \underbrace{\frac{\lambda_{22} - \lambda_{12}}{\lambda_{11} - \lambda_{21}} \frac{P(\omega_2)}{P(\omega_1)}}_{\theta_\lambda} \quad , \lambda_{11} < \lambda_{21}$$

نسبت درست‌نمائی‌ها

یک مقدار آستانه ثابت مستقل از x

Duda_fig2.3

مثال : تابع ضرر صفر و یک $\Lambda = [\lambda(\alpha_i|\omega_j)] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

$$\theta_\lambda = \frac{0-1}{0-1} \frac{P(\omega_2)}{P(\omega_1)} = \frac{P(\omega_2)}{P(\omega_1)}$$

مثال : $\Lambda = \begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix}$

$$\theta_\lambda = \frac{0-2}{0-1} \frac{P(\omega_2)}{P(\omega_1)} = 2 \frac{P(\omega_2)}{P(\omega_1)}$$

Subject:

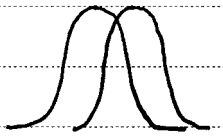
Year. Month. Date. ( )

مثال: مسئله طبقه‌بندی دو کلاسه با ویژگی اسکالر x ω_1, ω_2

$\omega_1: \; p(x|\omega_1) = \frac{1}{\sqrt{\pi}} \exp(-x^2), \quad P(\omega_1) = 1/2$

$\omega_2: \; p(x|\omega_2) = \frac{1}{\sqrt{\pi}} \exp(-(x-1)^2), \quad P(\omega_2) = 1/2$

الف) قاعده‌ی تصمیم‌گیری بیز برای مینیمم‌سازی احتمال خطا:



decide ω_1 if $P(\omega_1|x) > P(\omega_2|x)$

$\Rightarrow P(\omega_1)\,p(x|\omega_1) > P(\omega_2)\,p(x|\omega_2)$

$\Rightarrow \frac{1}{2}\frac{1}{\sqrt{\pi}} \exp(-x^2) > \frac{1}{2}\frac{1}{\sqrt{\pi}} \exp(-(x-1)^2)$

$\Rightarrow \exp(-x^2) > \exp(-(x-1)^2)$

$\Rightarrow -x^2 > -(x-1)^2$

$\Rightarrow x^2 < (x-1)^2$

$\Rightarrow x^2 < x^2 - 2x + 1$

$\Rightarrow x < \frac{1}{2}$

ب) قاعده‌ی تصمیم‌گیری بیز برای مینیمم‌سازی ریسک (خطر) با ماتریس حساسیت $\Lambda = \begin{bmatrix} 0 & 0.5 \\ 1 & 0 \end{bmatrix}$

$$R(\alpha_i|x) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)\,P(\omega_j|x)$$

decide ω_1 if $R(\omega_1|x) < R(\omega_2|x)$

$\Rightarrow \lambda(\omega_1|\omega_1)^{0} P(\omega_1|x) + \lambda(\omega_1|\omega_2) P(\omega_2|x) <$

$\lambda(\omega_2|\omega_1) P(\omega_1|x) + \lambda(\omega_2|\omega_2)^{0} P(\omega_2|x)$

$\Rightarrow 0.5\, P(\omega_2|x) < 1 \times P(\omega_1|x)$

$\Rightarrow 0.5 \exp(-(x-1)^2) < \exp(-x^2)$

$\Rightarrow \exp(-(x-1)^2) < 2\exp(-x^2)$

$\Rightarrow \exp(-(x-1)^2) < \exp(\ln 2 - x^2)$

$\Rightarrow (x-1)^2 > x^2 - \ln 2 \Rightarrow -2x + 1 > -\ln 2 \Rightarrow x < \frac{1+\ln 2}{2}$

Subject:
Year. Month. Date. ( )

مثال : (2-2)

مسئله طبقه بندی دو کلاسه دوبعدی ($x \in \mathbb{R}^2$ بردار ویژگی)

بردارهای ویژگی با دو توزیع نرمال با یک ماتریس کوواریانس مشترک

$$\Sigma = \begin{bmatrix} 1.1 & 0.3 \\ 0.3 & 1.9 \end{bmatrix}$$

و بردارهای میانگین :

$$\omega_1 : \quad \mu_1 = [0, 0]^T$$

$$\omega_2 : \quad \mu_2 = [3, 3]^T$$

(الف) بردار $[1.0, 2.2]^T$ را با استفاده از طبقه‌بندی کننده‌ی بیزی، طبقه‌بندی کنید.

* کافی است فاصله‌ی ماهالونوبیس این بردار را از دو بردار میانگین مربوط به دو طبقه محاسبه کنیم :

$$\text{decide } \omega_1 \text{ if } d_m^2(\mu_1, x) < d_m^2(\mu_2, x)$$

$$\Rightarrow (x-\mu_1)^T \Sigma^{-1} (x-\mu_1) < (x-\mu_2)^T \Sigma^{-1} (x-\mu_2)$$

$$\Rightarrow [1, 2.2] \begin{bmatrix} 0.95 & -0.15 \\ -0.15 & 0.55 \end{bmatrix} \begin{bmatrix} 1 \\ 2.2 \end{bmatrix} < [-2, -.8] \begin{bmatrix} 0.95 & -0.15 \\ -0.15 & 0.55 \end{bmatrix} \begin{bmatrix} -2 \\ -.8 \end{bmatrix}$$

$$\Rightarrow 2.952 < 3.672$$

نکته : اگر از فاصله‌ی اقلیدسی استفاده می‌کردیم، بردار ویژگی به μ_2 نزدیکتر می‌بود.

(ب) کمترین اصلی بیضی با مرکز $[0, 0]^T$ مناظر با فاصله‌ی ماهالونوبیس $d_m = \sqrt{2.952}$ از مرکز را محاسبه کنید.

* برای این هدف مقادیر ویژه‌ی Σ را محاسبه می‌کنیم :

$$\det(\Sigma - \lambda I) = 0 \Rightarrow$$

$$\left| \begin{bmatrix} 1.1-\lambda & 0.3 \\ 0.3 & 1.9-\lambda \end{bmatrix} \right| = 0 \Rightarrow \lambda^2 - 3\lambda + 2 = 0 \Rightarrow \lambda_1 = 1, \lambda_2 = 2$$

برای یافتن بردارهای ویژه‌ی مناظر با این مقادیر، آنها را در معادله‌ی $(\Sigma - \lambda I)v = 0$ قرار می‌دهیم :

$$v_1 = \left[\frac{3}{\sqrt{10}}, -\frac{1}{\sqrt{10}}\right]^T, \quad v_2 = \left[\frac{1}{\sqrt{10}}, \frac{3}{\sqrt{10}}\right]^T$$

v_1 و v_2 بر هم عمودند.

$$\|v_1\| = 3.436 \quad , \quad \|v_2\| = 4.859$$

Subject:

Year. Month. Date. ( )

Receiver Operating Characteristics : ROC مشخصه‌ی

مشخصه‌ی عملکرد گیرنده

Signal Detection Theory and Operating Characteristics

معیاری برای سنجش فاصله میان دو توزیع گاوسی

کاربرد در: روان‌شناسی تجربی، آشکارسازی رادار و ...

$$p(x|\omega_1) \sim N(\mu_1, \sigma^2)$$

$$p(x|\omega_2) \sim N(\mu_2, \sigma^2)$$

- تفکیک پذیری

$$d' = \frac{|\mu_2 - \mu_1|}{\sigma}$$

discriminability

* مطلوب است که d' بزرگ باشد.

وقتی μ_1، μ_2، σ و یا x^* را بدانیم فرض می‌کنیم که حالت طبیعت و تصمیم سیستم را می‌دانیم. این اطلاعات کمک می‌کند d' را بیابیم.

چهار احتمال زیر را در نظر می‌گیریم:

hit $P(x > x^* | x \in \omega_2)$

احتمال اینکه سیگنال داخلی بالاتر از x^* باشد و داشته باشیم سیگنال خارجی موجود است.

false alarm $P(x > x^* | x \in \omega_1)$

احتمال اینکه سیگنال داخلی بالاتر از x^* باشد، علیرغم اینکه سیگنال خارجی حاضر نباشد.

miss $P(x < x^* | x \in \omega_2)$

احتمال اینکه سیگنال داخلی پایین‌تر از x^* باشد، و داشته باشیم سیگنال خارجی موجود است.

correct rejection $P(x < x^* | x \in \omega_1)$

احتمال اینکه سیگنال داخلی پایین‌تر از x^* باشد، علیرغم اینکه سیگنال خارجی حاضر نباشد.

سیگنال داخلی با دو تأثیر x

- با میانگین μ_1 وقتی سیگنال خارجی موجود نباشد (کلاس ω_1)
- با میانگین μ_2 وقتی سیگنال خارجی موجود باشد (کلاس ω_2)
- + نویز تصادفی

x^* مقدار آستانه برای تشخیص وجود/عدم وجود سیگنال خارجی

P4PCO

abscissa کورانتی

ordinate کور عمودی

Subject :

Year . Month . Date . ( )

بقدار زیادی آزمایش انجام می‌دهیم (با فرض x^* ثابت)

تا این احتمالات را به طور تجربی به دست آوریم. به طور خاص hit و false alarm

این نقاط را بر روی یک نمودار رسم می‌کنیم. (نمودار دوبعدی)

اگر چگالی‌ها ثابت باشند (آستانه x^* تغییر باشد)، نرخ های hit و false alarm نیز تغییر می‌کند.

⇐ برای مقدار تفکیک پذیری d' داده شده، نقطه‌ای در راستای یک منحنی هموار (ROC) حرکت می‌کند.

Subject:

Year. Month. Date. ( )

مثال: N داده‌ی تولید شده با pdf گاوسی یک بعدی داریم که میانگین آن μ و واریانس آن مجهول است.

$x_1, x_2, \dots, x_N$

واریانس را به دست آورید: MLE

حل: تابع log-likelihood عبارت است از:

$$L(\sigma^2) = \ln \prod_{k=1}^{N} p(x_k;\sigma^2) = \sum_{k=1}^{N} \ln p(x_k;\sigma^2)$$

$$= \sum_{k=1}^{N} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_k-\mu)^2}{2\sigma^2}\right)\right)$$

$$= -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{k=1}^{N}(x_k-\mu)^2$$

$$\frac{\partial L(\sigma^2)}{\partial \sigma^2} = 0 \Longrightarrow -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{k=1}^{N}(x_k-\mu)^2 = 0 \Rightarrow \sigma^2 = \hat{\sigma}^2_{ML} = \frac{1}{N}\sum_{k=1}^{N}(x_k-\mu)^2$$

این برآورد برای N متناهی، اریب (biased) است؛ زیرا:

$$E\{\hat{\sigma}^2_{ML}\} = \frac{1}{N}\sum_{k=1}^{N} E\{(x_k-\mu)^2\} = \frac{N-1}{N}\sigma^2 \neq \sigma^2 = \text{واریانس واقعی تابع چگالی}$$

با وقتی $N \to \infty$، نااریب می‌شود:

$$\lim_{N\to\infty} E\{\hat{\sigma}^2_{ML}\} = \left(1-\frac{1}{N}\right)\sigma^2 = \sigma^2$$

$N \to \infty$

Subject:

Year. Month. Date. ( )

مثال:

بردارهای $x_1, x_2, \ldots, x_N$ که از یک توزیع نرمال با ماتریس کوواریانس معلوم Σ و میانگین مجهول بیرون کشیده شده‌اند را در نظر بگیرید:

$$p(x_k;\mu) = \frac{1}{(2\pi)^{l/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x_k-\mu)^T\Sigma^{-1}(x_k-\mu)\right)$$

برآورد ML بردار میانگین مجهول را بیابید.

حل: برای N نمونه‌ی موجود داریم:

$$L(\mu) = \ln \prod_{k=1}^{N} p(x_k;\mu) = \sum_{k=1}^{N} \ln p(x_k;\mu)$$
$$= -\frac{N}{2}\ln((2\pi)^l|\Sigma|) - \frac{1}{2}\sum_{k=1}^{N}(x_k-\mu)^T\Sigma^{-1}(x_k-\mu)$$

$$\frac{\partial L(\mu)}{\partial \mu} = \begin{bmatrix} \frac{\partial L}{\partial \mu_1} \\ \frac{\partial L}{\partial \mu_2} \\ \vdots \\ \frac{\partial L}{\partial \mu_l} \end{bmatrix} = \sum_{k=1}^{N}\Sigma^{-1}(x_k-\mu) = 0 \Rightarrow \mu = \frac{1}{N}\sum_{k=1}^{N} x_k = \hat{\mu}_{ML}$$

تذکر: $A = A^T \Rightarrow \frac{\partial(\alpha^T A\alpha)}{\partial\alpha} = 2A\alpha$

پس MLE برای میانگین در حالتی گاوسی، میانگین نمونه‌ای است.
این برآورد طبیعی برای توابع چگالی غیر گاوسی لزوماً همسنی ML نیست.

Subject:

Year. Month. Date. ( )

مثال : بردارهای $x_1, x_2, \ldots, x_N$ از یک توزیع نرمال با ماتریس کوواریانس معلوم Σ و میانگین مجهول μ گرفته شده‌اند. توزیع بردار میانگین مجهول μ به صورت زیر است :

$$p(\mu) = \frac{1}{(2\pi)^{l/2}\sigma_\mu^l} \exp\left(-\frac{1}{2}\frac{\|\mu-\mu_0\|^2}{\sigma_\mu^2}\right)$$

برآورد MAP به صورت زیر است :

$$\frac{\partial}{\partial \mu} \ln\left(\prod_{k=1}^{N} p(x_k|\mu)\, p(\mu)\right) = 0$$

برای $\Sigma = \sigma^2 I$ داریم :

$$\sum_{k=1}^{N} \frac{1}{\sigma^2}(x_k - \hat{\mu}) - \frac{1}{\sigma_\mu^2}(\hat{\mu} - \mu_0) = 0 \Rightarrow$$

$$\hat{\mu} = \hat{\mu}_{MAP} = \frac{\mu_0 + \frac{\sigma_\mu^2}{\sigma^2}\sum_{k=1}^{N} x_k}{1 + \frac{\sigma_\mu^2}{\sigma^2} N}$$

مشاهده می‌شود که :

۱) اگر $\frac{\sigma_\mu^2}{\sigma^2} \gg 1$ باشد، یعنی واریانس σ_μ^2 خیلی زیاد است ⇐ کاوشی تا حد بسیار عریض (با تغییرات کم روی بازه‌ی مورد نظر) می‌باشد ⇐

$$\hat{\mu}_{MAP} \approx \hat{\mu}_{ML} = \frac{1}{N}\sum_{k=1}^{N} x_k$$

۲) برای $N \to \infty$ بدون توجه به مقدار واریانس، برآورد به میانگین تبدیل می‌شود.

۳) برآورد MAP به طور مجانبی به برآورد ML میل می‌کند :

$$N \to \infty \Rightarrow \hat{\theta}_{MAP} \to \hat{\theta}_{ML}$$

۴) اگر $\frac{\sigma_\mu^2}{\sigma^2} \gg 1$ باشد، و N کوچک باشد، پارامتر به سمت پارامتر پیشین میل می‌کند : $\hat{\theta}_{MAP} \to \theta_0$

(مربوط به توزیع پیشین متغیر)

Subject:

Year. Month. Date. ( )

Bayesian Inference (BI) استنتاج بیزی

روش‌های ML و MAP یک برآورد خاص از بردار پارامتر مجهول θ را ارائه می‌کردند.

اما در BI سیر دیگری دنبال می‌شود:

با داشتن مجموعه‌ی X از N بردار آموزشی و

اطلاعات پیشین درمورد pdf : $p(\theta)$

هدف: محاسبه‌ی $p(x|X)$ است.

داریم: (70)
$$p(x|X) = \int p(x|\theta)\, p(\theta|X)\, d\theta$$

(71)
$$p(\theta|X) = \frac{p(X|\theta)p(\theta)}{p(X)} = \frac{p(X|\theta)p(\theta)}{\int p(X|\theta)\, p(\theta)\, d\theta}$$

(72)
$$p(X|\theta) = \prod_{k=1}^{N} p(x_k|\theta)$$

توزیع شرطی $p(\theta|X)$ نیز به عنوان برآورد pdf پسین معلوم است:

(زیرا "دانش" به روزشده در مورد خواص آماری θ پس از مشاهده‌ی مجموعه داده‌ی X است).

محاسبه‌ی $p(x|X)$ نیازبه انتگرال‌گیری دارد (70): عددتا به صورت تقریب عددی قابل محاسبه است.

راهنمائی: اگر $p(\theta|X)$ معلوم باشد، آن‌گاه $p(x|X)$ میانگین $p(x|\theta)$ نسبت به θ است:

$$p(x|X) = E_\theta\{p(x|\theta)\}$$

با فرض بزرگ بودن تعداد نمونه‌های $\{\theta_i\}_{i=1}^{L}$ از بردار تصادفی θ می‌توانیم مقادیر متناظر $p(x|\theta_i)$ را محاسبه کنیم و امید را به عنوان مقدار متوسط تقریب بزنیم:

$$p(x|X) \approx \frac{1}{L}\sum_{i=1}^{L} p(x|\theta_i)$$

اکنون مسئله تبدیل شد به تولید یک مجموعه نمونه $\{\theta_i\}_{i=1}^{L}$.

مثلا اگر $p(\theta|X)$ گاوسی باشد، می‌توان با مولد اعداد شبه تصادفی گاوسی، L نمونه تولید کرد.

مشکل اینجاست که معمولا فرم دقیق $p(\theta|X)$ معلوم نیست. ⇐ حل مشکل با مجموعه روش‌های MCMC

مثال :

$p(x|\mu) \sim N(\mu, \sigma^2)$

$p(\mu) \sim N(\mu_0, \sigma_0^2)$ μ پارامتر مجهول

داریم :

$$p(\mu|X) = \frac{p(X|\mu)p(\mu)}{p(X)} = \frac{1}{\alpha}\prod_{k=1}^{N} p(x_k|\mu)p(\mu)$$

که در آن $\alpha = P(X)$ با توجه به مجموعه داده‌ی آموزشی داده شده ثابت است.

$$p(\mu|X) = \frac{1}{\alpha}\prod_{k=1}^{N}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x_k-\mu)^2}{\sigma^2}\right)\frac{1}{\sigma_0\sqrt{2\pi}}\exp\left(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right)$$

$$= \frac{1}{\sigma_N\sqrt{2\pi}}\exp\left(-\frac{(\mu-\mu_N)^2}{2\sigma_N^2}\right) \qquad \mu_N = \frac{N\sigma_0^2\bar{x}_N + \sigma^2\mu_0}{N\sigma_0^2+\sigma^2}$$

$$\sigma_N^2 = \frac{\sigma^2\sigma_0^2}{N\sigma_0^2+\sigma^2}$$

$$\bar{x}_N = \frac{1}{N}\sum_{k=1}^{N} x_k$$

با داشتن N نمونه، $p(\mu|X)$ نیز گاوسی خواهد بود (نیاز به اثبات دارد).

وقتی $p(\mu|X)$ محاسبه شد، با جایگزینی در (7.0) خواهیم داشت :

$$p(x|X) = \frac{1}{\sqrt{2\pi(\sigma^2+\sigma_N^2)}}\exp\left(-\frac{1}{2}\frac{(x-\mu_N)^2}{\sigma^2+\sigma_N^2}\right) \sim N(\mu_N, \sigma^2+\sigma_N^2)$$

Subject :

Year . Month . Date . ( )

Maximum Entropy Estimation (ME) برآورد ماکزیمم آنتروپی

آنتروپی : معیاری برای عدم اطمینان یک رویداد پیشامد

معیار تصادفی بودن پیام‌ها که در خروجی یک سیستم رخ می‌دهد (پیام = بردار ویژگی)

$$H = -\int_x p(x) \ln p(x)\, dx$$

برآورد ME برای pdf مجهول $p(x)$: آن که آنتروپی را ماکزیمم می‌کند (با در نظر داشتن قیدهای داده شده)

* متناظر با توزیعی که بالاترین تصادفی بودن ممکن را (نسبت به قیدهای موجود) نشان می‌دهد .

مثال : متغیر تصادفی x برای $x_1 \le x \le x_2$ غیرصفر و در سایر موارد صفر است .

برآورد ماکزیمم آنتروپی را برای pdf آن محاسبه کنید :

حل :

$$\max H = \max -\int_x p(x) \ln p(x)\, dx$$

$$\text{subject to } \int_{x_1}^{x_2} p(x)\, dx = 1$$

با استفاده از ضرایب لاگرانژ داریم :

$$H_L = -\int_{x_1}^{x_2} p(x)(\ln p(x) - \lambda)\, dx$$

$$\frac{\partial H_L}{\partial p(x)} = -\int_{x_1}^{x_2} \left[(\ln p(x) - \lambda) + 1\right] dx = 0 \implies p(x) = \exp(\lambda - 1) \triangleq \hat{p}(x)$$

برای محاسبه‌ی λ ، آن را در معادله‌ی قید جایگذاری می‌کنیم :

$$\exp(\lambda - 1) = \frac{1}{x_2 - x_1}$$

پس

$$\hat{p}(x) = \begin{cases} \frac{1}{x_2 - x_1} & , x_1 \le x \le x_2 \\ 0 & , \text{otherwise} \end{cases}$$

* برآورد ماکزیمم آنتروپی برای pdf مجهول ، توزیع یکنواخت است .

نکته : اگر قیدهای دوم و سوم را برای μ و σ^2 مشخص قرار می‌دادیم ، برآورد حداکثر آنتروپی ، تابع گاوسی می‌شد .

Subject:

Year. Month. Date. ( )

fig 2.17 مثال:

$N=100$ نقطه در فضای دو بعدی

از توزیع چندنده استخراج شده‌اند: multimodal

نمونه‌ها با دو مولد تصادفی گاوسی تولید شده‌اند:

$$N(\mu_1, \Sigma_1) \quad N(\mu_2, \Sigma_2) \qquad \mu_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \mu_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \Sigma_1 = \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

در هر زمان یک نمونه‌ی x_k $(k = 1, 2, \ldots, N)$ تولید می‌شود.

یک سکه انداخته می‌شود تا مشخص شود x_k باید توسط کدام گاوسی تولید شود:

$$\begin{cases} P(H) = P = 0.8 \rightarrow ① \\ P(T) = 1 - P = 0.2 \rightarrow ② \end{cases}$$

(پس داده‌ها با مرکزیت $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ چگال‌تر از داده‌ها با مرکزیت $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$ است).

پس:

$$p(x) = g(x; \mu_1, \sigma_1^2) P + g(x; \mu_2, \sigma_2^2)(1 - P)$$

$g(\cdot; \mu, \sigma^2)$: pdf گاوسی با پارامتر میانگین μ و ماتریس کوواریانس $\Sigma = \sigma^2 I = \text{diag}(\sigma^2)$

هدف: یافتن MLE با بردار پارامترهای مجهول

$$\Theta^T = [P, \mu_1^T, \sigma_1^2, \mu_2^T, \sigma_2^2]$$

بر مبنای $N=100$ نقطه $\{(x_k, j_k)\}_{k=1}^{N}$ که در آن $j_k \in \{1, 2\}$

اما j_kها (اطلاعات برچسب) موجود نیستند.

$$x_k = \alpha_k x_k^1 + (1 - \alpha_k) x_k^2$$

x_k^1 از $N(\mu_1, \Sigma_1)$ ، x_k^2 از $N(\mu_2, \Sigma_2)$

$$L(\Theta; \alpha) = \sum_{k=1}^{N} \alpha_k \ln\{g(x_k; \mu_1, \sigma_1^2) P\} + \sum_{k=1}^{N} (1 - \alpha_k) \ln\{g(x_k; \mu_2, \sigma_2^2)(1 - P)\}$$

$$E\{\alpha_k | x_k; \hat{\Theta}\} = 1 \times P(1 | x_k; \hat{\Theta}) + 0 \times (1 - P(1 | x_k; \hat{\Theta})) = P(1 | x_k; \hat{\Theta})$$

انتخاب مقادیر اولیه و استفاده از EM تا همگرا شدن.