STATISTICAL PATTERN RECOGNITION

باز شناسایی آماری الگو

عباس برزگری نژاد

بهار ۱۴۰۴

ابر صفحه‌های تصمیم گیری(*Decision Hyper planes*)

تنها عاملی که باعث غیر خطی شدن معادله

$$g_i(x) = -\frac{1}{2}x^T[\Sigma_i]^{-1}x + \frac{1}{2}x^T[\Sigma_i]^{-1}\mu_i + \frac{1}{2}\mu_i{}^T[\Sigma_i]^{-1}x - \frac{1}{2}\mu_i{}^T[\Sigma_i]^{-1}\mu_i + Ln(P(w_i)) + C_i$$

گردید عبارت $x^T[\Sigma_i]^{-1}x$ می‌باشد. حال اگر فرض کنیم که ماتریس کوواریانس در تمام کلاسها یکسان می‌باشد یعنی

$$[\Sigma_i] = [\Sigma]$$

لذا فرم درجه دوم $x^T[\Sigma_i]^{-1}x$ در تمام توابع جدا کننده یکسان خواهد بود. از این رو این عبارت در محاسبه ماکزیمم بی اثر خواهد بود و در معادلات صفحه تصمیم گیری عاملی خنثی می گردد. همانطور که قبلا نیز اشاره کردیم $g_i(x)$ یک تابع جدا کننده می‌باشد که با استفاده از آزمون تصمیم گیری زیر ما را در انتخاب یاری می‌رساند:

x در کلاس W_i قرار دارد اگر $g_i(x) > g_j\ (x) \qquad \forall\ i \neq j$

بنابراین ابر صفحه‌های تصمیم گیری با استفاده از معادلات زیر بدست می‌آید:

$$g_i(x) = -\frac{1}{2}x^T[\Sigma_i]^{-1}x + \frac{1}{2}x^T[\Sigma_i]^{-1}\mu_i + \frac{1}{2}{\mu_i}^T[\Sigma_i]^{-1}x - \frac{1}{2}{\mu_i}^T[\Sigma_i]^{-1}\mu_i + Ln(P(w_i)) + C_i \qquad C_i = -\frac{l}{2}Ln\,2\pi - \frac{1}{2}Ln|[\Sigma_i]^{-1}|$$

$$g_{ij}(x) \equiv g_i(x) - g_j(x) = 0$$

با ثابت فرض کردن ماتریس کوواریانس در تمام کلاسها به وضوح مقدار C_i نیز در تمام توابع جدا کننده یکسان خواهد بود
بنابراین می‌توان عبارتهای $x^T[\Sigma_i]^{-1}x$ و C_i را از این توابع حذف نمود و به تابع اصطلاح یافته زیر رسید:

$$g_i(x) = {W_i}^T x + W_{iO}$$

که در آن

$$W_i = [\Sigma_i]^{-1}\mu_i$$

$$W_{iO} = -\frac{1}{2}{\mu_i}^T[\Sigma_i]^{-1}\mu_i + Ln(P(w_i))$$

توابع جداکننده برای چگالی نرمال

تابع چگالی احتمال نرمال را مجدداً در نظر بگیرید:

$$P_i(x) = \frac{1}{(2\pi)^{\frac{L}{2}} \ |[\Sigma_i]|^{\frac{1}{2}}} \exp(-\frac{1}{2}[x-\mu_i]^T[\Sigma_i]^{-1}(x-\mu_i)), \ \ i=1,...,M$$

می‌توان تابع جدا کننده را به صورت زیر در نظر گرفت:

$$g_i(x) = \ln P(X|w_i) + LnP(w_i)$$

اگر $P(X|w_i)\sim N(\mu_i,[\Sigma_i])$ باشد در این صورت خواهیم داشت:

$$g_i(x) = -\frac{1}{2}(X-\mu_i)^T|[\Sigma_i]^{-1}|(X-\mu_i) - \frac{L}{2}Ln\ 2\pi - \frac{1}{2}Ln|[\Sigma_i]| + Ln(P(w_i))$$

در این بخش قصد داریم تابع جدا کننده و دسته بندی‌های حاصله را به ازای تعدادی از حالت‌های خاص شرح می‌دهیم:

حالت اول:

$$[\Sigma_i] = \sigma^2 I$$

در همه کلاسها ماتریس کوواریانس ثابت و قطری می‌باشد

حالت دوم:

$$[\Sigma_i] = [\Sigma]$$

در همه کلاسها ماتریس کوواریانس ثابت می‌باشد

حالت سوم:

$$[\Sigma_i] = Arbitrary$$

در هر کلاس ماتریس کوواریانس خاص خود آن کلاس می باشد.

حالت اول: $[\Sigma_i]=\sigma^2 I$

وقتی که ماتریس کوواریانس ثابت است ولی ماتریس قطری است. ماتریس قطری روی قطر اصلی همه واریانس و بقیه جاها صفر می‌باشد:

$$[\boldsymbol{\Sigma}_{\boldsymbol{i}}] = \begin{bmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{bmatrix}$$

$$[\Sigma_i]^{-1} = \frac{1}{\sigma^2} I \qquad\qquad [\Sigma_i][\Sigma_i]^{-1} = I$$

$$\left\| \begin{bmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{bmatrix} \right\| = (\sigma^2)^n$$

از آنجا که هر دو مقدار $|[\Sigma_i]|$ و $\frac{L}{2} Ln\ 2\pi$ مستقل از مقدار i می باشند، لذا می توان آن ها را نادیده گرفت. بنابراین توابع جدا کننده ساده زیر بدست می آیند:

$$g_i(x) = -\frac{1}{2\sigma^2}(X-\mu_i)^T(X-\mu_i) + Ln\big(P(w_i)\big)$$

$$g_i(x) = -\frac{1}{2\sigma^2}\|X-\mu_i\|^2 + Ln\big(P(w_i)\big)$$

یادآوری:

$$X = (x_1, x_2, \dots\dots, x_n)$$

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \qquad Or \qquad \|x\|_p = \sqrt{|x_1|^p + |x_2|^p + \cdots + |x_n|^p}$$

$$if \quad p = 2 \quad => \quad \|x\| = \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2}$$

بسط فرم درجه دوم $(X-\mu_i)^T(X-\mu_i)$ منجر به تابع زیر می شود

$$g_i(x) = -\frac{1}{2\sigma^2}\left(X^T\mathrm{X} - 2{\mu_i}^T X + {\mu_i}^T\mu_i\right) + Ln\big(P(w_i)\big)$$

X^TX چون اندیس i ندارد در همه کلاسها ثابت هست و می‌توان از آن صرف نظر کرد. بنابراین توابع جداکننده خطی در حالت اول بدین صورت در می آیند:

$$g_i(x) = -\frac{1}{2\sigma^2}\left(-2{\mu_i}^T X + {\mu_i}^T\mu_i\right) + Ln\big(P(w_i)\big)$$

پس:

$$g_i(x) = {W_i}^T x + W_{iO}$$

که در آن

$$W_i = \frac{1}{\sigma^2}\mu_i$$

$$W_{iO} = -\frac{1}{2\sigma^2}{\mu_i}^T\mu_i + Ln(P(w_i))$$

W_{iO} را آستانه تحمل($Threshold$) یا اریب($bias$) در جهت i ام می نامیم.

کلاسیفایر یا دسته بندهایی که از توابع جداکننده خطی استفاده می کنند یک ماشین خطی می نامیم. در این مرحله صرفا توجه داشته باشید که سطوح تصمیم برای یک ماشین خطی، قطعاتی از ابرصفحه های تعریف شده بوسیله معادلات خطی

$$g_i(x) = g_j(x)$$

با حداکثر احتمالات ثانویه، می باشند. برای حالت خاص در نظر گرفته شده این معادله را می توان بدین صورت در نظر گرفت:

$$g_{ij}(x) \equiv g_i(x) - g_j(x) = 0$$

$$\frac{1}{\sigma^2}{\mu_i}^T X - \frac{1}{2\sigma^2}{\mu_i}^T \mu_i + Ln\big(P(w_i)\big) - [\frac{1}{\sigma^2}{\mu_j}^T X - \frac{1}{2\sigma^2}{\mu_j}^T \mu_j + Ln\left(P\left(w_j\right)\right)]=۰$$

$$\frac{1}{\sigma^2}(\mu_i - \mu_j)^T X - \frac{1}{2\sigma^2}\left({\mu_i}^T \mu_i - {\mu_j}^T \mu_j\right) + Ln\left(\frac{P(w_i)}{P(w_j)}\right) =۰$$

چون

$${\mu_i}^T \mu_i - {\mu_j}^T \mu_j=(\mu_i - \mu_j)^T(\mu_i + \mu_j)$$

$$(\mu_i - \mu_j)^T X - \frac{1}{2}(\mu_i - \mu_j)^T(\mu_i + \mu_j) + \sigma^2 Ln\left(\frac{P(w_i)}{P(w_j)}\right) = ٠$$

$$(\mu_i - \mu_j)^T\left[X - \left(\frac{1}{2}(\mu_i + \mu_j) - \frac{\sigma^2}{(\mu_i-\mu_j)^T(\mu_i-\mu_j)}(\mu_i - \mu_j)Ln\left(\frac{P(w_i)}{P(w_j)}\right)\right)\right] = ٠$$

بنابراین

$$W^T(X - X_0) = 0$$

که در آن

$$W=(\mu_i-\mu_j)$$

,

$$X_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\sigma^2}{\|\mu_i-\mu_j\|^2}(\mu_i - \mu_j)Ln\left(\frac{P(w_i)}{P(w_j)}\right)$$

حالت دوم: $[\Sigma_i]=[\Sigma]$

یکی دیگر از حالات ساده زمانی رخ می دهد که ماتریس کوواریانس به ازای تمام کلاس ها یکسان باشد. از آنجا که هر دو مقدار $|[\Sigma_i]|$و $\frac{L}{2}Ln\ 2\pi$ مستقل از مقدار i می باشند، ، لذا می توان آن ها را نادیده گرفت. این ساده سازی منجر به تابع جدا کننده زیر می شود:

$$g_i(x) = -\frac{1}{2}(X-\mu_i)^T[\Sigma]^{-1}(X-\mu_i) + Ln\big(P(w_i)\big)$$

$$= -\frac{1}{2}[x^T[\Sigma]^{-1}x - x^T[\Sigma]^{-1}\mu_i - {\mu_i}^T[\Sigma]^{-1}x + {\mu_i}^T[\Sigma]^{-1}\mu_i] + Ln(P(w_i))$$

در رابطه فوق با توجه به اینکه عبارت درجه دوم $x^T[\Sigma]^{-1}x$ مستقل از مقدار i *می باشد*، لذا می توان آن را نادیده گرفت.
پس خواهیم داشت:

$$g_i(x) = {\mu_i}^T[\Sigma]^{-1}x - \frac{1}{2}{\mu_i}^T[\Sigma]^{-1}\mu_i + Ln(P(w_i))$$

با فرض

$$W_i = [\Sigma]^{-1}\mu_i$$

$$W_{iO} = -\frac{1}{2}{\mu_i}^T[\Sigma]^{-1}\mu_i + Ln(P(w_i))$$

خواهیم داشت:

$$g_i(x) = {W_i}^T x + W_{iO}$$

چون جدا کننده ها خطی می باشند، لذا مرزهای تصمیم گیری باز هم ابرصفحه ها می باشند. اگر R_i و R_j پیوسته باشند، مرز بین آن ها معادله ای به صورت زیر می باشد:

$$g_{ij}(x) \equiv g_i(x) - g_j(x) = 0$$

$${\mu_i}^T[\Sigma]^{-1}x - \frac{1}{2}{\mu_i}^T[\Sigma]^{-1}\mu_i + Ln(P(w_i)) - {\mu_j}^T[\Sigma]^{-1}x + \frac{1}{2}{\mu_j}^T[\Sigma]^{-1}\mu_j - Ln\left(P\left(w_j\right)\right) = 0$$

$$(\mu_i-\mu_j)^T[\Sigma]^{-1}X-\frac{1}{2}(\mu_i-\mu_j)^{\,T}[\Sigma]^{-1}(\mu_i+\mu_j)+Ln\left(\frac{P(w_i)}{P(w_j)}\right)=٠$$

$$(\mu_i-\mu_j)^T[\Sigma]^{-1}\left[X-\left(\frac{1}{2}(\mu_i+\mu_j)-\frac{Ln\left(\frac{P(w_i)}{P(w_j)}\right)}{(\mu_i-\mu_j)^T[\Sigma]^{-1}(\mu_i-\mu_j)}(\mu_i-\mu_j)\right)\right]=٠$$

بنابراین

$$W^T(X-X_0)=0$$

که در آن

$$W=[\Sigma]^{-1}(\mu_i-\mu_j)$$

,

$$X_0=\frac{1}{2}(\mu_i+\mu_j)-\frac{Ln\left(\frac{P(w_i)}{P(w_j)}\right)}{(\mu_i-\mu_j)^T[\Sigma]^{-1}(\mu_i-\mu_j)}(\mu_i-\mu_j)$$

حالت سوم: $[\Sigma_i]=Arbitrary$

در این حالت ماتریس های کوواریانس در هر دسته متفاوت می باشند. تنها عبارتی که مستقل از i می باشد،عبارت $\frac{L}{2}Ln\,2\pi$ است که می توان آن را نادیده گرفت و لذا توابع جداکننده بدست آمده ذاتاً درجه دوم می باشند:

$$g_i(x) = -\frac{1}{2}(X-\mu_i)^T[\Sigma_i]^{-1}(X-\mu_i) + Ln\big(P(w_i)\big) - \frac{1}{2}Ln|[\Sigma_i]|$$
$$= -\frac{1}{2}[x^T[\Sigma_i]^{-1}x - x^T[\Sigma_i]^{-1}\mu_i - {\mu_i}^T[\Sigma_i]^{-1}x + {\mu_i}^T[\Sigma_i]^{-1}\mu_i] + Ln(P(w_i)) - \frac{1}{2}Ln\big|[\Sigma_i]\big|$$
$$= -\frac{1}{2}x^T[\Sigma_i]^{-1}x + {\mu_i}^T[\Sigma_i]^{-1}x - \frac{1}{2}{\mu_i}^T[\Sigma_i]^{-1}\mu_i - \frac{1}{2}Ln\big|[\Sigma_i]\big| + Ln(P(w_i))$$

بنابراین:

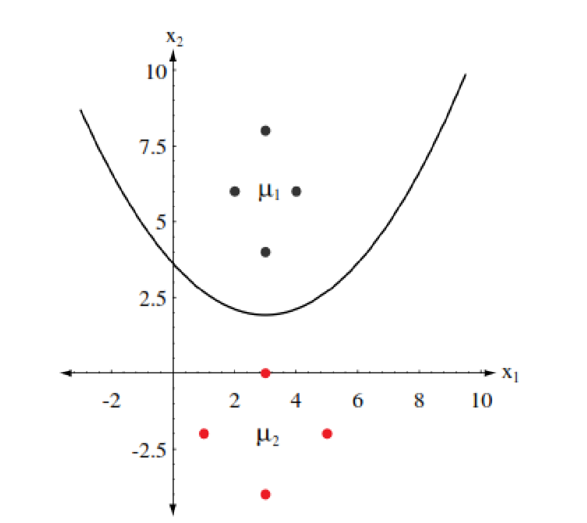
$$g_i(x) = x^T[W_i]x + {w_i}^T x + w_{iO}$$

که در آن

$$[W_i] = -\frac{1}{2}[\Sigma_i]^{-1}$$
$$w_i = [\Sigma_i]^{-1}\mu_i$$
$$w_{iO} = -\frac{1}{2}{\mu_i}^T[\Sigma_i]^{-1}\mu_i - \frac{1}{2}Ln|[\Sigma_i]| + Ln(P(w_i))$$

مثال(نواحی تصمیم برای داده های نرمال دو بعدی):

همانطور که در شکل نیز مشاهده می کنید، فرض کنید w_1 مجموعه شامل چهار نقطه سیاه و w_2 نقاط قرمز باشند. بردار های امید ریاضی و ماتریس های کوواریانس را بدین صورت در نظر بگیرید:



$$\boldsymbol{\mu}_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}; \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix} \text{ and } \boldsymbol{\mu}_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}; \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

برای بدست آوردن ناحیه های تصمیم ابتدا عکس ماتریس های کوواریانس را محاسبه می کنیم:

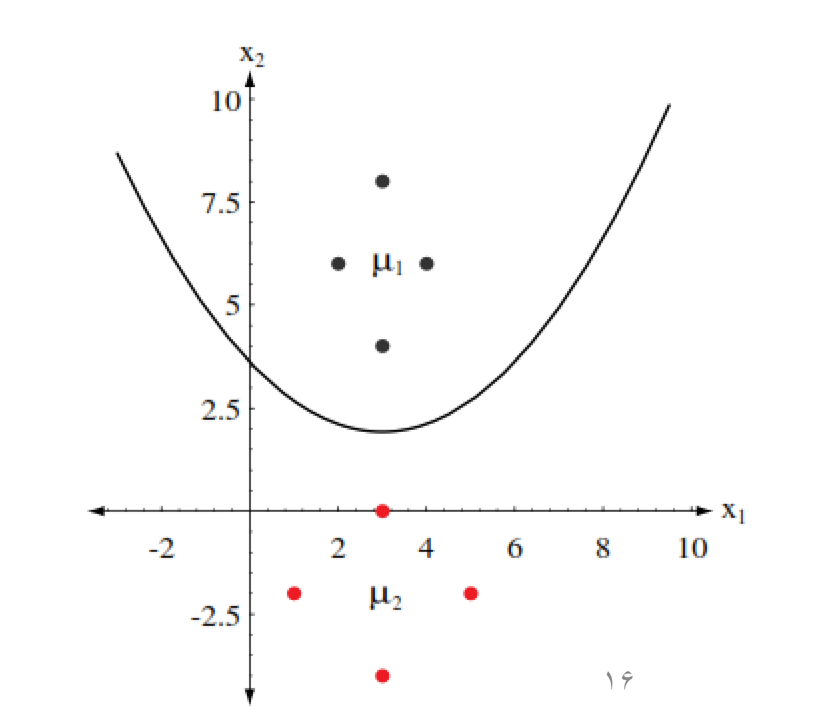
$$\boldsymbol{\Sigma}_1^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_2^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

با فرض برابری احتمالات اولیه($P(w_1)$=$P(w_2)$=۰/۵)و جایگذاری در $g_1(x)$ و $g_2(x)$ خواهیم داشت:

$$g_1(x) = -x_1^2 - \frac{1}{4}x_2^2 + 6x_1 + 3x_2 - 18 - ln2$$
$$g_2(x) = -\frac{1}{4}x_1^2 - \frac{1}{4}x_2^2 + \frac{3}{2}x_1 - x_2 - \frac{13}{4} - \frac{1}{2}ln4 - ln2$$

برای بدست آوردن مرز تصمیم داریم:

$$g_{12}(x) \equiv g_1(x) - g_2(x) = 0$$
$$\Rightarrow -3x_1^2 + 18x_1 + 16x_2 - 59 - 4ln2 = 0$$
$$\Rightarrow x_2 = \frac{3}{16}x_1^2 - \frac{18}{16}x_1 + \frac{59 - 4ln2}{16}$$
$$\Rightarrow x_2 = 0.1875x_1^2 - 1.125x_1 + 3.5142$$



تخمین توابع چگالی احتمال نامعلوم

در بیشتر موارد، توابع چگالی احتمال کلاسها ناشناخته بوده و مجبور به تخمین‌زدن آن از روی داده موجود می‌باشیم.
گاهی اوقات شکل توزیع (گوسی یا ریلی) معین و پارامترها (میانگین، واریانس) نامعین؛ و در برخی موارد توزیع نامعین و پارامترها (میانگین، واریانس) معین می باشند.

۲,۵تخمین پارامتر با روش حداکثر شباهت (درستنمایی)

در یک مسئله M کلاسه بردارهای ویژگی بصورت توابع شباهت $p(x|\omega_i)$ در شکل پارامتری به بردارهای ناشناخته θ_i وابسته است. هدف تخمین پارامترهای تابع بصورت $p(x|\omega_i;\theta_i)$ از روی یک مجموعه بردار ویژگی معین (مجموعه داده آموزش) برای هر کلاس می باشد.
فرض کنید داده هر کلاس مستقل از کلاسهای دیگر می‌باشد (جهت تخمین پارامترها)

$x_1, x_2, \ldots, x_N$ نمونه‌های تصادفی از تابع $p\ (x\ ;\theta)$ می‌باشند و تابع چگالی احتمال توام $p(X\ ;\ \theta)$ را از روی مجموعه $X = \{x_1, \ldots, x_N\}$ تشکیل می‌دهیم

با فرض استقلال آماری بین نمونه‌ها، داریم

$$p(X;\boldsymbol{\theta}) \equiv p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N;\boldsymbol{\theta}) = \prod_{k=1}^{N} p(\boldsymbol{x}_k;\boldsymbol{\theta})$$

تابع بالا را تابع درستنمایی θ برحسب X نامیده و روش **حداکثر درستنمایی** (ML) مقدار θ را برای ماکزیمم کردن این تابع تخمین می‌زند. یعنی:

$$\hat{\boldsymbol{\theta}}_{ML} = arg\max_{\boldsymbol{\theta}} \prod_{k=1}^{N} p(\boldsymbol{x}_k;\boldsymbol{\theta})$$

شرط لازم برای ماکزیمم شدن تابع بالا آن است که گرادیان تابع درستنمایی برحسب θ صفر باشد. یعنی:

$$\frac{\partial \prod_{k=1}^{N} p(\boldsymbol{x}_k;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}$$

با توجه به یکنوا بودن تابع لگاریتم، تابع درستنمایی لگاریتمی را به صورت زیر تعریف می کنیم:

$$L(\boldsymbol{\theta}) \equiv \ln \prod_{k=1}^{N} p(\boldsymbol{x}_k;\boldsymbol{\theta})$$

با جایگزینی تابع لگاریتمی به جای تابع چگالی، داریم:

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{k=1}^{N} \frac{\partial \ln p(\boldsymbol{x}_k;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{k=1}^{N} \frac{1}{p(\boldsymbol{x}_k;\boldsymbol{\theta})} \frac{\partial p(\boldsymbol{x}_k;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0} \qquad (2.59)$$

شکل۲.۱۴ این روش را در حالتی که پارامتر نامعلوم منحصر بفرد است شرح می دهد. برآوردگر ML متناظر با نقطه پیک یا ماکزیمم تابع لگاریتم درستنمایی می باشد.

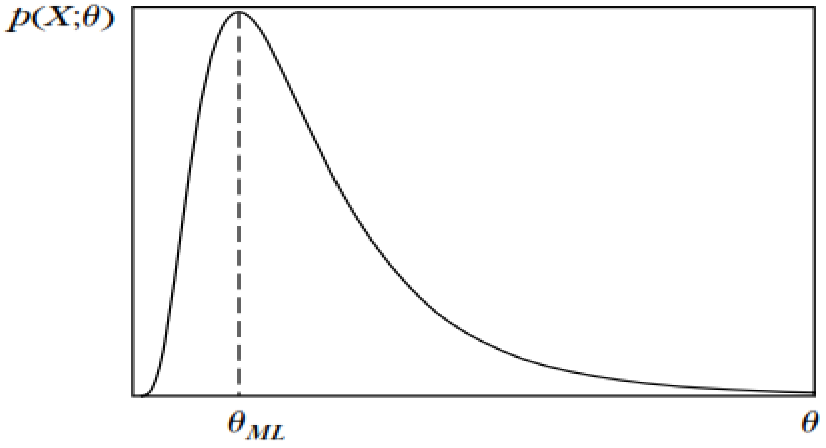


FIGURE 2.14

The maximum likelihood estimator θ_{ML} corresponds to the peak of $p(X;\theta)$.

برآوردگر ML دارای خواص خیلی مطلوبی می باشد که در اینجا به آن ها اشاره می کنیم:

۱. برآوردگر ML مجانباً نااریب می باشد. یعنی:

$$\lim_{N\to\infty} E[\hat{\boldsymbol{\theta}}_{ML}] = \boldsymbol{\theta}_0$$

آمارهٔ $\hat{\Theta}$ یک برآوردگر **سازگار** پارامتر θ است اگر و تنها اگر به‌ازای هر ثابت مثبت c

$$\lim_{n\to\infty} P(|\hat{\Theta} - \theta| < c) = ۱$$

۲. برآوردگر ML مجانباً سازگار می باشد. یعنی برای مقادیر بزرگ از N ،واریانس تخمین به صفر میل می کند:

$$\lim_{N\to\infty} E[||\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0||^2] = 0$$

توجه کنید که سازگاری یک **خاصیت مجانبی**، یعنی، خاصیت حدی یک برآوردگر است؛

قضیه: اگر $\hat{\Theta}$ برآوردگری ناریب برای پارامتر θ باشد و وقتی $n \to \infty$، $\operatorname{var}(\hat{\Theta}) \to 0$،
آنگاه $\hat{\Theta}$ برآوردگری سازگار برای θ است.

قضیه: اگر $\hat{\Theta}$ یک برآوردگر ناریب θ باشد و

$$\operatorname{var}(\hat{\Theta}) = \frac{1}{n \cdot E\left[\left(\frac{\partial \ln f(X)}{\partial \theta}\right)^2\right]}$$

آنگاه، $\hat{\Theta}$ یک برآوردگر ناریب با کمترین واریانس θ است.

در اینجا، کمیت واقع در مخرج کسر را اطلاع دربارهٔ θ نامند که به‌وسیلهٔ نمونه تأمین می‌شود
بنابراین، هرچه واریانس کمتر باشد، اطلاع بیشتر است.

برآوردگرهای نااریب معمولاً برحسب واریانسهایشان با می‌شوند. اگر $\hat{\Theta}_1$ و $\hat{\Theta}_2$ دو برآوردگر نااریب پارامتر θ باشند و واریانس $\hat{\Theta}_1$ کوچکتر $\hat{\Theta}_2$ باشد، گوییم که $\hat{\Theta}_1$ **به‌طور نسبی کاراتر** از $\hat{\Theta}_2$ است.

۳.برآوردگر ML مجانباً کارا می باشد. یعنی کران پایین کرامر-رائو(کوچکترین واریانس ممکن)رابرآورد می کند.

۴.برای مقادیر بزرگ N این برآوردگر دارای توزیع نرمال می‌باشد.

مثال۱:

اگر $X_۱$، $X_۲$، ...، X_n نمونه‌ای تصادفی از جامعه‌ای با چگالی

$$f(x) = \begin{cases} e^{-(x-\delta)} & x > \delta \\ ۰ & \text{سایر جاها} \end{cases}$$

باشد، نشان دهید که $\overline{X}$ برآوردگری اریب برای δ است.

چون میانگین جامعه عبارت است از

$$\mu = \int_{\delta}^{\infty} x \cdot e^{-(x-\delta)} dx = ۱ + \delta$$

نتیجه می‌شود که $E(\overline{X}) = ۱ + \delta \neq \delta$ و بنابراین $\overline{X}$ یک برآوردگر اریب δ است.

مثال ۲:

نشان دهید که $\overline{X}$ یک برآوردگر نااریب با کمترین واریانس، برای میانگین μی جامعه‌ای نرمال است.

حل. چون

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad , \quad -\infty < x < \infty$$

نتیجه می‌شود که

$$\ln f(x) = -\ln \sigma\sqrt{2\pi} - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$$

به‌طوری که

$$\frac{\partial \ln f(x)}{\partial \mu} = \frac{1}{\sigma}\left(\frac{x-\mu}{\sigma}\right)$$

و بنابراین

$$E\left[\left(\frac{\partial \ln f(X)}{\partial \mu}\right)^{۲}\right] = \frac{۱}{\sigma^{۲}} \cdot E\left[\left(\frac{X-\mu}{\sigma}\right)^{۲}\right] = \frac{۱}{\sigma^{۲}} \cdot ۱ = \frac{۱}{\sigma^{۲}}$$

در نتیجه

$$\frac{۱}{n \cdot E\left[\left(\frac{\partial \ln f(X)}{\partial \mu}\right)^{۲}\right]} = \frac{۱}{n \cdot \frac{۱}{\sigma^{۲}}} = \frac{\sigma^{۲}}{n}$$

و چون $\overline{X}$ ناریب است و طبق قضیهٔ $\operatorname{var}(\overline{X}) = \frac{\sigma^{۲}}{n}$، نتیجه می‌شود که $\overline{X}$ یک برآوردگر ناریب با کمترین واریانس برای μ است.

مثال ۳:

برآورد ML از یک داده N نقطه‌ای با میانگین معلوم μ و واریانس نامشخص را بیابید. این نقاط توسط یک تابع چگالی احتمال گوسی یک بعدی تولید شده‌اند.

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

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

$$-\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{k=1}^{N}(x_k-\mu)^2 = 0$$

$$\hat{\sigma}^2_{ML} = \frac{1}{N}\sum_{k=1}^{N}(x_k-\mu)^2$$

$$E[\hat{\sigma}^2_{ML}] = \frac{1}{N}\sum_{k=1}^{N}E[(x_k-\mu)^2] = \frac{N-1}{N}\sigma^2$$

تخمین بالا دارای بایاس می باشد. حال اگر N بسمت بی نهایت میل کند خواهیم داشت:

$$E[\hat{\sigma}^2_{ML}] = (1 - \frac{1}{N})\sigma^2 \approx \sigma^2$$

مثال ٤:

فرض کنید $x_1, x_2, \ldots, x_N$ بردارهای ناشی از یک توزیع نرمال با ماتریس کوواریانس معلوم و میانگین نامعلوم باشند. یعنی:

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

برآورد ML از بردار میانگین نامعلوم را بدست می آوریم:
به ازای N نمونه موجود خواهیم داشت:

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

با استفاده از گرادیان تابع درستنمایی نسبت به μ داریم:

$$\hat{\boldsymbol{\mu}}_{ML} = \frac{1}{N} \sum_{k=1}^{N} \boldsymbol{x}_k$$

مثال۵:

اگر $X_1, X_2, \ldots, X_n$ نمونه‌ای تصادفی به اندازه‌ی n از جامعه‌ی نرمال با میانگین μ و واریانس σ^2 باشد، برآوردهای درستنمایی ماکسیمم توأم این دو پارامتر را پیدا کنید.

حل. چون تابع درستنمایی به صورت

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

است، با گرفتن مشتق جزئی $L(\mu,\sigma^2)$ نسبت به μ و σ^2 نتیجه می‌شود که

$$\frac{\partial[\ L(\mu,\sigma^2)]}{\partial\mu} = \frac{1}{\sigma^2}\cdot\sum_{k=1}^{N}(x_k-\mu)^2$$

$$\frac{\partial[\ L(\mu,\sigma^2)]}{\partial\sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{k=1}^{N}(x_k-\mu)^2$$

از برابر گرفتن اوّلین مشتق جزئی با صفر و حل آن برحسب μ به‌دست می‌آوریم.

$$\hat{\mu}_{ML} = \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{x}_k$$

و با برابر گرفتن دوّمین مشتق جزئی با صفر و حل آن برحسب σ^2

$$\hat{\sigma}^2_{ML} = \frac{1}{N}\sum_{k=1}^{N}(x_k - \hat{\mu}_{ML})^2$$

لازم است توجه شود که ما ثابت نکرده‌ایم که $\hat{\sigma}$ یک برآورد درستنمایی ماکسیمم σ است، بلکه تنها ثابت کرده‌ایم که $\hat{\sigma}^2$ یک برآورد درستنمایی ماکسیمم σ^2 است. با این حال، می‌توان نشان داد که برآوردکننده‌های درستنمایی ماکسیمم دارای این **خاصیت ناوردایی** هستند که اگر $\hat{\Theta}$ یک برآوردکننده‌ی درستنمایی ماکسیمم θ و تابع مفروض $g(\theta)$ پیوسته باشد، آنگاه $g(\hat{\Theta})$ نیز یک برآوردکننده‌ی درستنمایی ماکسیمم $g(\theta)$ است. از اینجا نتیجه می‌گیریم که

$$\hat{\sigma} = \sqrt{\frac{1}{n}.\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

نیز یک برآورد درستنمایی ماکسیمم σ است، که از s متفاوت است زیرا در آن، Σ را به‌جای $n-1$ بر n تقسیم می‌کنیم.

مثال ۶:

اگر $x_1, x_2, \ldots, x_n$ مقادیر یک نمونه‌ی تصادفی از جامعه‌ای نمایی باشند، برآورد درستنمایی ماکسیمم پارامتری θی جامعه را پیدا کنید.

حل. چون تابع درستنمایی به صورت

$$L(\theta) = f(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta) = \left(\frac{1}{\theta}\right)^n . e^{-\frac{1}{\theta}(\sum_{i=1}^{n} x_i)}$$

است، با مشتق‌گیری از $\ln L(\theta)$ نسبت به θ

$$\frac{d[\ln L(\theta)]}{d\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} . \sum_{i=1}^{n} x_i$$

به‌دست می‌آید. از برابر گرفتن این مشتق با صفر و حل آن نسبت به θ، برآورد درستنمایی ماکسیمم زیر را به‌دست می‌آوریم

$$\hat{\theta} = \frac{1}{n}.\sum_{i=1}^{n} x_i = \bar{x}.$$

بنابراین، برآوردکننده‌ی درستنمایی ماکسیمم عبارت است از $\hat{\Theta} = \bar{X}$.

مثال۷:

با مفروض بودن x «موفقیت» در n آزمایش، برآورد درستنمایی ماکسیمم پارامتر θ را در توزیع دوجمله‌ای نظیر پیدا کنید.

حل. برای پیدا کردن مقداری از θ که

$$L(\theta)=\binom{n}{x}\theta^{x}(1-\theta)^{n-x}$$

را ماکسیمم می‌کند، بهتر است از این واقعیت استفاده کنیم که مقدار θ که $L(\theta)$ را ماکسیمم می‌کند، تابع

$$\ln L(\theta)=\ln\binom{n}{x}+x.\ln\theta+(n-x).\ln(1-\theta)$$

را نیز ماکسیمم می‌کند. بنابراین، نتیجه می‌گیریم که

$$\frac{d[\ln L(\theta)]}{d\theta}=\frac{x}{\theta}-\frac{n-x}{1-\theta}$$

و با برابرگرفتن این مشتق با ۰ و حل آن نسبت به θ، درمی‌یابیم که تابع درستنمایی دارای ماکسیممی در $\theta = \frac{x}{n}$ است. این مقدار، برآورد درستنمایی ماکسیمم پارامتر θ در توزیع دوجمله‌ای است و $\hat{\Theta} = \frac{X}{n}$ را برآوردکننده‌ی درستنمایی ماکسیمم نظیر می‌نامیم.

تخمین ماکزیمم احتمال پسین (Maximum A Posteriori Probability)

برای رسیدن به برآورد ML، θ را پارامتر نامعلوم در نظر می گیریم. در این تخمین θ را بردار تصادفی فرض می‌کنیم و مقدارش را بشرط مشاهده نمونه‌های داده تخمین می‌زنیم. نقطه شروع ما $p(\theta|X)$ می باشد. با توجه به قضیه بیز داریم:

$$p(\boldsymbol{\theta})p(X|\boldsymbol{\theta}) = p(X)p(\boldsymbol{\theta}|X)$$

$$p(\boldsymbol{\theta}|X) = \frac{p(\boldsymbol{\theta})p(X|\boldsymbol{\theta})}{p(X)}$$

درستنمایی

تخمین ماکزیمم احتمال پسین (MAP) در نقطه ای تعریف می شود که احتمال $p(\theta|X)$ را ماکزیمم کند:

$$\hat{\boldsymbol{\theta}}_{MAP} : \frac{\partial}{\partial \boldsymbol{\theta}} p(\boldsymbol{\theta}|X) = 0 \quad \text{or} \quad \frac{\partial}{\partial \boldsymbol{\theta}} (p(\boldsymbol{\theta})p(X|\boldsymbol{\theta})) = 0$$

لازم به ذکر است که اختلاف تخمین‌های ML و MAP در وجود $p(\theta)$ می‌باشد.

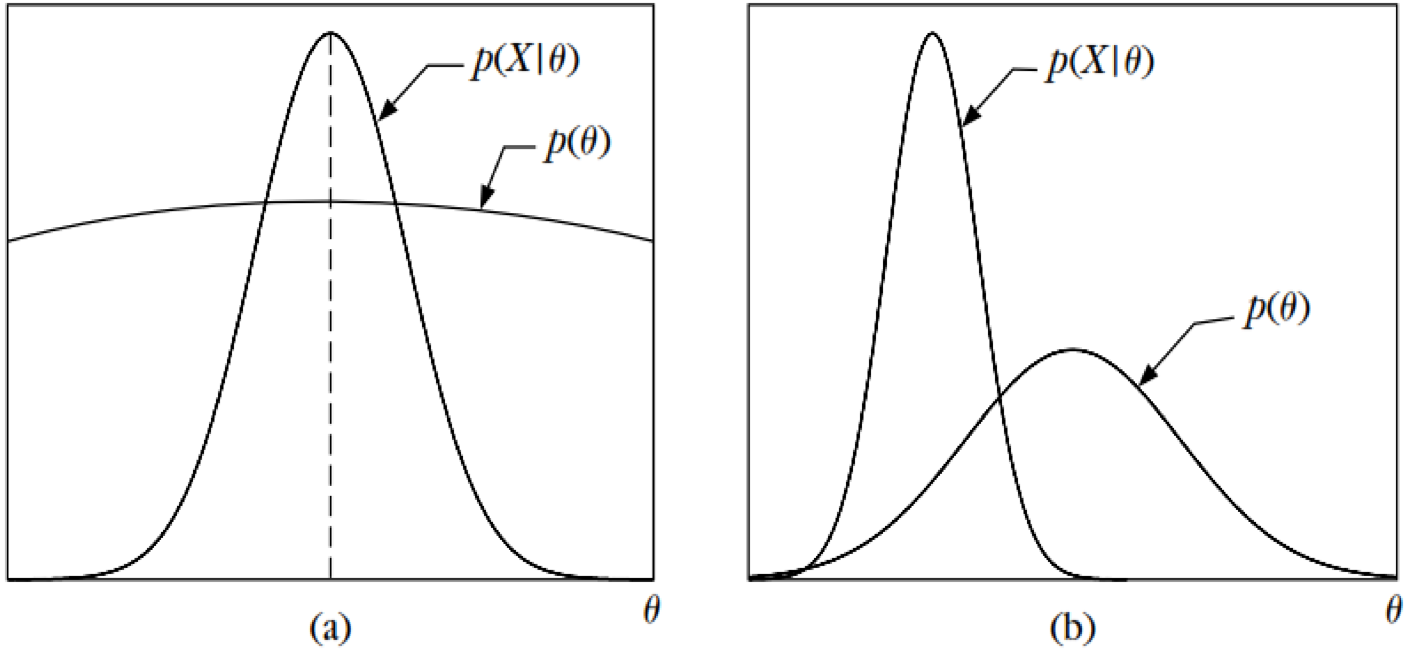
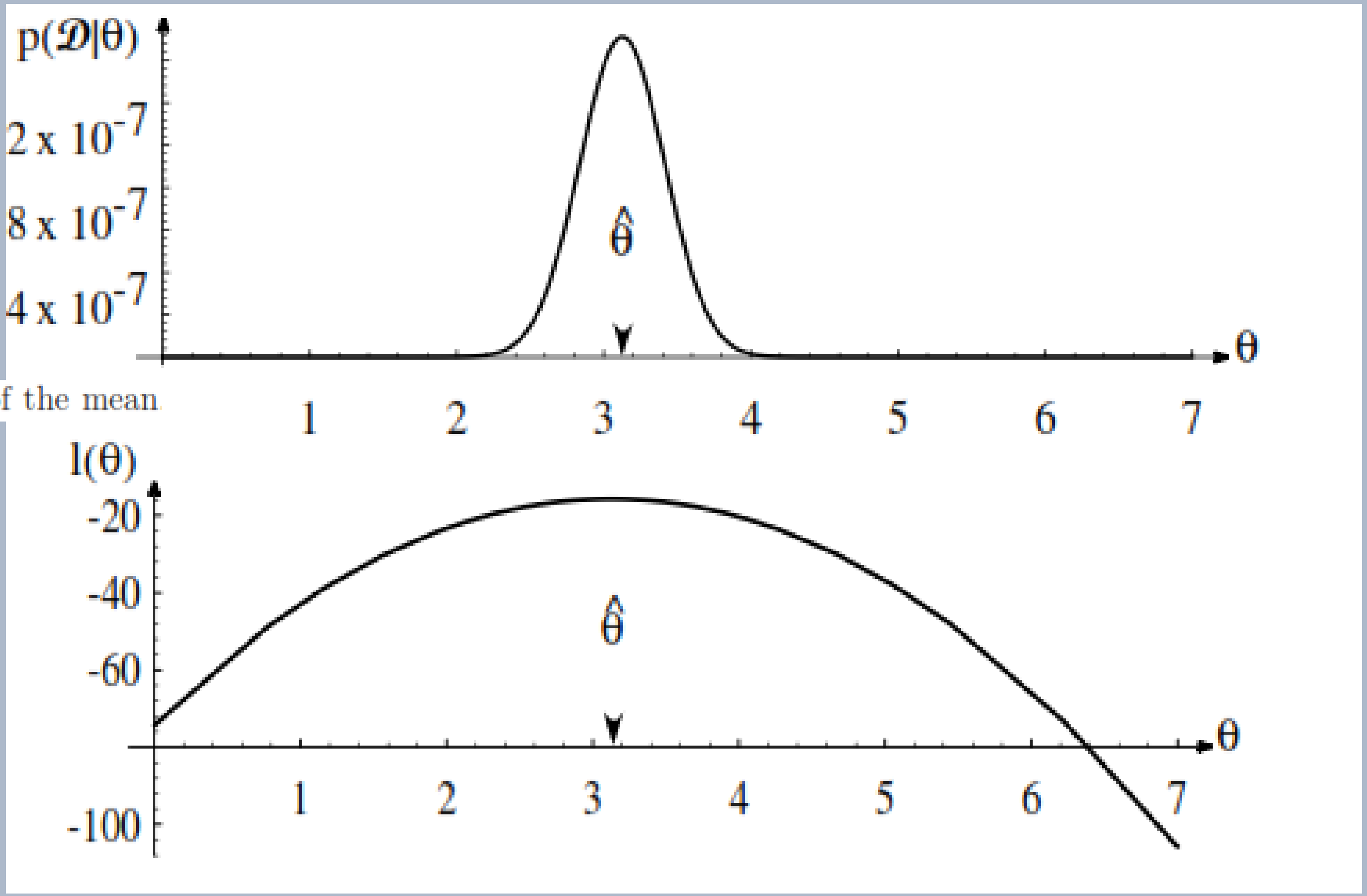


FIGURE 2.15

ML and MAP estimates of θ will be approximately the same in (a) and different in (b).



likelihood $p(\mathcal{D}|\theta)$ as a function of the mean

فرض کنید: $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)^t$

$$l(\boldsymbol{\theta}) \equiv \ln p(\mathcal{D}|\boldsymbol{\theta})$$

$$p(\mathcal{D}|\boldsymbol{\theta}) = \prod_{k=1}^{n} p(\mathbf{x}_k|\boldsymbol{\theta}).$$

$$\nabla_{\boldsymbol{\theta}} l = \sum_{k=1}^{n} \nabla_{\boldsymbol{\theta}} \ln p(\mathbf{x}_k|\boldsymbol{\theta}).$$

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} l(\boldsymbol{\theta}),$$

$$\boxed{\nabla_{\boldsymbol{\theta}} l = \mathbf{0}.}$$

مثال۸

فرض کنید تابع مورد نظر دارای توزیع نرمال با μ نامعلوم باشد:

$$\ln p(\mathbf{x}_k|\boldsymbol{\mu}) = -\frac{1}{2}\ln\left[(2\pi)^d|\boldsymbol{\Sigma}|\right] - \frac{1}{2}(\mathbf{x}_k - \boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}_k - \boldsymbol{\mu})$$

$$\nabla_{\boldsymbol{\theta}} \ln p(\mathbf{x}_k|\boldsymbol{\mu}) = \boldsymbol{\Sigma}^{-1}(\mathbf{x}_k - \boldsymbol{\mu}).$$

$$\sum_{k=1}^{n} \boldsymbol{\Sigma}^{-1}(\mathbf{x}_k - \hat{\boldsymbol{\mu}}) = \mathbf{0},$$

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}_k.$$

مثال۹

فرض کنید تابع مورد نظر دارای توزیع نرمال با μ و ماتریس کوواریانس نامعلوم باشد:

با فرض $\theta_1 = \mu$ و $\theta_2 = \sigma^2$

$$\ln p(x_k|\boldsymbol{\theta}) = -\frac{1}{2}\ln 2\pi\theta_2 - \frac{1}{2\theta_2}(x_k - \theta_1)^2$$

$$\nabla_{\boldsymbol{\theta}} l = \nabla_{\boldsymbol{\theta}} \ln p(x_k|\boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{\theta_2}(x_k - \theta_1) \\ -\frac{1}{2\theta_2} + \frac{(x_k-\theta_1)^2}{2\theta_2^2} \end{bmatrix}.$$

$$\sum_{k=1}^{n} \frac{1}{\hat{\theta}_2}(x_k - \hat{\theta}_1) = 0$$

$$-\sum_{k=1}^{n} \frac{1}{\hat{\theta}_2} + \sum_{k=1}^{n} \frac{(x_k - \hat{\theta}_1)^2}{\hat{\theta}_2^2} = 0,$$

که در آن $\hat{\theta}_1$ و$\hat{\theta}_2$ برآورد های ML برایθ_1 و θ_2 می باشند. با جایگزینی $\hat{\mu} = \hat{\theta}_1,\ \hat{\sigma}^2 = \hat{\theta}_2$ خواهیم داشت:

$$\hat{\mu} = \frac{1}{n}\sum_{k=1}^{n} x_k$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \hat{\mu})^2.$$

آنتروپی

جدول ۱ـ۱. ماتریس تصمیم‌گیری چند شاخصه

گزینه‌ها \ شاخص‌ها	C_1	$C_2 \cdots$	$C_j \cdots$	C_n
A_1	a_{11}	$a_{12} \cdots$	$a_{1j} \cdots$	a_{1n}
A_2	a_{21}	$a_{22} \cdots$	$a_{2j} \cdots$	a_{2n}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_i	a_{i1}	$a_{i2} \cdots$	$a_{ij} \cdots$	a_{in}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_m	a_{m1}	$a_{m2} \cdots$	$a_{mj} \cdots$	a_{mn}

می‌توان گفت A_i نشان‌دهنده‌ی گزینه i ام، c_j نشان‌دهنده‌ی شاخص j ام و a_{ij} نشان‌دهنده ارزش گزینه i ام، از نظر شاخص j ام می‌باشد.

مثال ۱ـ۱

فرض کنید یک دانش آموخته‌ی دانشگاهی، قصد دارد از بین ٤ شغل، با توجه به ٥ شاخص، یکی را انتخاب کند. «شاخص‌ها» عبارتند از درآمد، وجهه‌ی اجتماعی، سختی کار، مسافت و امنیت اجتماعی (درآمد، برحسب ده‌هزار تومان و مسافت، برحسب کیلومتر است). ارزش هر شغل از نظر هر شاخص، در جدول ۲ـ۱ آمده است.

جدول ۲ـ۱. ماتریس تصمیم برای مثال ۱ـ۱

A_i \ c_j	درآمد c_1	وجهه‌ی اجتماعی c_2	سختی c_3	مسافت c_4	امنیت c_5
A_1	۱۵	زیاد	نسبتاً زیاد	۱۰	زیاد
A_2	۱۲	متوسط	متوسط	۳	خیلی زیاد
A_3	۲۰	خیلی زیاد	زیاد	۳۰	متوسط
A_4	۳۰	کم	خیلی زیاد	۱	کم

در ماتریس تصمیم (جدول ۲ـ۱)، ملاحظه می‌شود که از پنج شاخص موجود (c_j)، دو شاخص (c_4 و c_1) کمّی بوده و بقیه‌ی آن‌ها، کیفی می‌باشد.

اغلب، شاخص‌ها در مدل‌های MADM از مقیاس‌های مختلف بوده و غالباً در تعارض با یکدیگرند. در اغلب مواقع، گزینه‌ای که بهینه باشد (ایده‌آل از هر شاخص را تامین کند)، وجود ندارد. علاوه بر این برخی شاخص‌ها جنبه‌ی مثبت (c_j^+) و برخی جنبه‌ی منفی (c_j^-)دارند. بنابراین «گزینه‌ی بهینه» در یک مدل MADM، یک گزینه‌ی ذهنی A خواهد بود که بهترین ارزش از هر شاخص را تامین کند. در اکثر مواقع، دسترسی به A غیرممکن است؛ اما انتخاب مناسب‌ترین گزینه به‌طور نسبی در هر صورت، امکان‌پذیر خواهد بود.

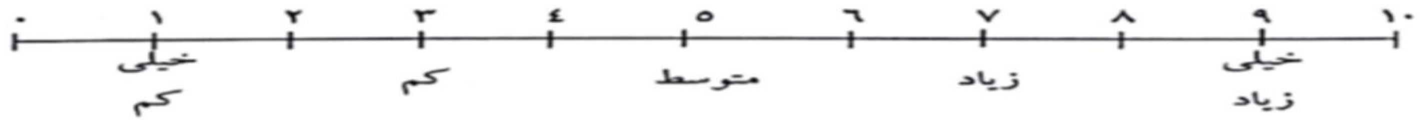
۵ـ۱. تبدیل شاخص‌های کیفی به کمّی

همان‌طور که گفتیم، می‌توان راهکارهای انتخابی را توسط دو نوع شاخص توصیف کرد:

۱ـ شاخص‌های کمّی (قیمت، درآمد، مسافت، و....)،

۲ـ شاخص‌های کیفی (وجهه‌ی اجتماعی، سختی، امنیت، زیبایی، و....).

می‌توان با استفاده از روش‌های مختلفی، شاخص‌های کیفی را به شاخص‌های کمّی تبدیل کرد؛ ولی بهترین روش، روش‌هایی هستند که از مقیاس‌های فاصله‌ای و رتبه‌ای یا مقیاس دوقطبی استفاده می‌نمایند. یک روش عمومی در اندازه‌گیری یک شاخص کیفی با مقیاس فاصله‌ای، استفاده از «مقیاس دوقطبی فاصله‌ای» است که به قرار زیر می‌باشد:



این اندازه‌گیری، براساس یک مقیاس یازده نقطه‌ای می‌باشد که صفر، کمترین ارزش و ۱۰، بیشترین ارزش را به خود اختصاص می‌دهد. این اندازه‌گیری، برای

۷ـ۱. ارزیابی اوزان شاخص‌ها

همان‌طور که در قسمت‌های قبل گفته شد، هر مساله‌ای که فرد تصمیم‌گیرنده با آن مواجه است، ممکن است دارای چندین شاخص باشد. بنابراین، «دانستن اهمیت نسبی شاخص‌ها»، ضرورت دارد. از این رو، به هر شاخص، یک وزن داده می‌شود، به صورتی که مجموع اوزان شاخص‌ها، برابر با یک باشد. این وزن‌ها، اهمیت نسبی (درجه‌ی ارجحیت) هر شاخص را نسبت به بقیه، برای تصمیم‌گیری مورد نظر نشان می‌دهد. برای ارزیابی اوزان شاخص‌ها، روش‌های مختلفی وجود دارد که برخی از آن‌ها عبارتند از :

الف) روش آنتروپی

ب) روش لینمپ[1]

ج) روش کمترین مجذورات موزون

د) روش بردار ویژه.

روش آنتروپی و لینمپ ، براساس «ماتریس تصمیم‌گیری» است در حالی که روش کمترین مجذورات موزون و روش بردار ویژه، نیاز به ماتریس تصمیم‌گیری ندارد. اکنون روش آنتروپی توضیح داده می‌شود، روش بردار ویژه (که گاهی روش AHP گفته می‌شود) در قسمت ٤ـ٨ـ١ توضیح داده می‌شود.

روش آنتروپی

«آنتروپی[2]» یک، مفهوم بسیار با اهمیت در علوم اجتماعی، فیزیک و تئوری اطلاعات می‌باشد. وقتی که داده‌های یک ماتریس تصمیم‌گیری، به طور کامل مشخص شده باشد، می‌توان از روش آنتروپی، برای ارزیابی وزن‌ها استفاده کرد.

---

1. linmap
2. entropy

ایده‌ی روش فوق، این است که هر چه پراکندگی در مقادیر یک شاخص، بیشتر باشد، آن شاخص از اهمیت بیشتری برخوردار است.

آنتروپی در نظریه‌ی اطلاعات، یک معیار عدم اطمینان است که با توزیع احتمال مشخص P_i بیان می‌شود.

اندازه‌گیری این عدم اطمینان (E_i)، توسط شانون[1]، به صورت زیر بیان شده است:

$$E_i = S(p_1, p_2, \ldots, p_n) = -k\sum_{i=1}^{n}[p_i \times I_n p_i]$$

k مقداری ثابت است و به منظور این‌که E_i بین صفر و یک باشد، اعمال می‌شود.

E از توزیع احتمال p_i براساس مکانیزم آماری، محاسبه شده و مقدار آن در صورت تساوی p_i ها با یکدیگر (یعنی $p_i = \frac{1}{n}$)، ماکزیمم مقدار ممکن خواهد بود که بدین صورت محاسبه می‌شود:

$$-k\sum_{i=1}^{n} p_i - \ln p_i = -k\left\{\frac{1}{n}\ln\frac{1}{n} + \frac{1}{n}\ln\frac{1}{n} + \ldots + \frac{1}{n}\ln\frac{1}{n}\right\} = -k\left\{\ln\frac{1}{n}\left(\frac{n}{n}\right)\right\}$$
$$= -k \times \ln\frac{1}{n}$$

k به عنوان مقدار ثابت، به صورت زیر محاسبه می‌شود :

$$k = \frac{1}{ln(m)}$$

«ماتریس تصمیم‌گیری»، حاوی اطلاعاتی است که آنتروپی می‌تواند به عنوان معیاری برای ارزیابی آن به کار رود. فرض کنید که ماتریس تصمیم‌گیری، به این صورت باشد:

1. Shanon

گزینه‌ها \ شاخص‌ها	C_1	C_2	$\cdots$	C_n
A_1	a_{11}	a_{12}	$\cdots$	a_{1n}
A_2	a_{21}	a_{22}	$\cdots$	a_{2n}
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
A_m	a_{m1}	a_{m2}	$\cdots$	a_{mn}

با استفاده از این ماتریس، P_{ij} به صورت زیر محاسبه می‌شود :

$$P_{ij} = \frac{a_{ij}}{\sum_{i=1}^{m} a_{ij}} \quad ; \quad \forall_{i,j}$$

و آنتروپی شاخص j ام (E_j) به صورت زیر محاسبه می‌شود :

$$E_j = -k\sum_{i=1}^{m}\left[p_{ij} \ln p_{ij}\right] \quad ; \quad \forall_j$$

عدم اطمینان یا درجه‌ی انحراف (d_j) از اطلاعات به‌دست آمده برای شاخص j، بیان می‌کند که شاخص مربوطه (j)، چه میزان اطلاعات مفید برای تصمیم‌گیری، در اختیار تصمیم‌گیرنده قرار می‌دهد. مقدار (d_j) به صورت زیر به‌دست می‌آید:

$$d_j = 1 - E_j \quad ; \quad \forall_j$$

سپس مقدار وزن W_j به صورت زیر به‌دست می‌آید :

$$w_j = \frac{d_j}{\sum_{j=1}^{n} d_j} \quad ; \quad \forall_j$$

اگر تصمیم‌گیرنده از قبل، وزن ذهنی مشخصی مثل λ_j را برای شاخص j در نظر گرفته باشد، در این صورت وزن تعدیل شده (W_j) ،به شرح زیر محاسبه می‌شود:

$$w'_j = \frac{\lambda_j w_j}{\sum_{j=1}^{n} \lambda_j w_j} \quad ; \quad \forall_j$$

به طور خلاصه، می‌توان برای به‌دست آوردن اوزان شاخص‌ها، گام‌های زیر را طی کرد:

گام ۱ـ محاسبه‌ی P_{ij}

$$P_{ij} = \frac{a_{ij}}{\sum_{i=1}^{m} a_{ij}} \quad ; \quad \forall_j$$

گام ۲ـ محاسبه‌ی مقدار آنتروپی E_j

$$E_j = -k\sum_{i=1}^{m}\left[p_{ij} \ln p_{ij}\right] \quad ; \quad \forall_j$$

گام ۳ـ محاسبه‌ی مقدار عدم اطمینان d_j

$$d_j = 1 - E_j \quad ; \quad \forall_j$$

گام ٤ـ محاسبه‌ی اوزان w_j

$$w_j = \frac{d_j}{\sum_{j=1}^{n} d_j} \quad ; \quad \forall_j$$

گام ٥ـ محاسبه‌ی اوزان تعدیل شده w'_j

$$w'_j = \frac{\lambda_j w_j}{\sum_{j=1}^{n} \lambda_j w_j} \quad ; \quad \forall_j$$

λ_j اوزان ذهنی (قضاوتی) می‌باشند(توجه: در صورتی که اوزان ذهنی (قضاوتی) موجود نباشد، مرحله‌ی ٥ منتفی است.)

دقت کنید که در روش آنتروپی، مثبت یا منفی بودن شاخص‌ها، تاثیری در روش محاسبه‌ی وزن‌ها نخواهد داشت.

مثال ٤ـ١

دوباره مساله‌ی فرد دانش آموخته را در نظر بگیرید و با روش آنتروپی، وزن شاخص‌ها را حساب کنید.

جدول ۶ـ۱. ماتریس تصمیم فرد دانش آموخته

A_i \ C_j	C_1	C_2	C_3	C_4	C_5
A_1	۱۵	زیاد	نسبتاً زیاد	۱۰	زیاد
A_2	۱۲	متوسط	متوسط	۳	خیلی‌زیاد
A_3	۲۰	خیلی‌زیاد	زیاد	۳۰	متوسط
A_4	۳۰	کم	خیلی‌زیاد	۱	کم

ابتدا مقادیر شاخص‌های کیفی را طبق روش‌های گفته شده، به کمّی تبدیل می‌کنیم. نتایج آن، در جدول ۷ـ۱ آورده شده است.

جدول ۷ـ۱. ماتریس کمّی

A_i \ C_j	C_1	C_2	C_3	C_4	C_5
A_1	۱۵	۷	٤	۱۰	۷
A_2	۱۲	٥	٥	۳	۹
A_3	۲۰	۹	۳	۳۰	٥
A_4	۳۰	۳	۱	۱	۳
Σ	۷۷	۲٤	۱۳	٤٤	۲٤

گام اول : در این مرحله، با استفاده از فرمول زیر، مقدار P_{ij} را برای تمامی شاخص‌ها و راهکارها به‌دست می‌آوریم.

$$p_{ij} = \frac{a_{ij}}{\sum_{i=1}^{n} a_{ij}} \Rightarrow \quad p_{11} = \frac{۱۵}{۷۷} = ۰/۱۹۵, \quad p_{21} = \frac{۱۲}{۷۷} = ۰/۱۵۶$$

$$p_{31} = \frac{۲۰}{۷۷} = ۰/۲۶۰ \quad , \quad p_{41} = \frac{۳۰}{۷۷} = ۰/۳۸۹$$

نتایج گام اول، در جدول ۸ـ۱ آورده شده است.

جدول ۸ـ۱. نتایج گام اول

A_i \ C_j	C_1	C_2	C_3	C_4	C_5
A_1	۰/۱۹۵	۰/۲۹۲	۰/۳۰۸	۰/۲۲۷	۰/۲۹۲
A_2	۰/۱۵۶	۰/۲۰۸	۰/۳۸۵	۰/۰۶۸	۰/۳۷۵
A_3	۰/۲۶۰	۰/۳۷۵	۰/۲۳۱	۰/۶۸۲	۰/۲۰۸
A_4	۰/۳۸۹	۰/۱۲۵	۰/۰۷۷	۰/۰۲۳	۰/۱۲۵

گام دوم: در این گام، با استفاده از فرمول $E_j = -k\sum[p_{ij} \ln p_{ij}]$ ، مقدار اطمینان را به‌دست می‌آوریم:

$$E_1 = -۰/۷۲۱[۰/۱۹۵ \ln(۰/۱۹۵) + ۰/۱۵۶ \ln(۰/۱۵۶) + ۰/۲۶۰ \ln(۰/۲۶۰) + ۰/۳۸۹ \ln(۰/۳۸۹)] = ۰/۹۵۶$$

$$E_2 = -۰/۷۲۱[۰/۲۹۲ \ln(۰/۲۹۲) + ۰/۲۰۸ \ln(۰/۲۰۸) + ۰/۳۷۵ \ln(۰/۳۷۵) + ۰/۱۲۵ \ln(۰/۱۲۵)] = ۰/۹۴۷$$

$$E_3 = -۰/۷۲۱[۰/۳۰۸ \ln(۰/۳۰۸) + ۰/۳۸۵ \ln(۰/۳۸۵) + ۰/۲۳۱ \ln(۰/۲۳۱) + ۰/۰۷۷ \ln(۰/۰۷۷)] = ۰/۹۱۳$$

$$E_4 = -۰/۷۲۱[۰/۲۲۷ \ln(۰/۲۲۷) + ۰/۰۶۸ \ln(۰/۰۶۸) + ۰/۶۸۲ \ln(۰/۶۸۲) + ۰/۰۲۳ \ln(۰/۰۲۳)] = ۰/۶۲۵$$

$$E_5 = -۰/۷۲۱[۰/۲۹۲ \ln(۰/۲۹۲) + ۰/۳۷۵ \ln(۰/۳۷۵) + ۰/۲۰۸ \ln(۰/۲۰۸) + ۰/۱۲۵ \ln(۰/۱۲۵)] = ۰/۹۴۷$$

همان‌طور که گفتیم، K به این ترتیب به‌دست می‌آید: $k = \frac{۱}{\ln(m)} = \frac{۱}{\ln(۴)} = ۰/۷۲۱$

جدول ۹ـ۱. نتایج گام دوم

E_1	E_2	E_3	E_4	E_5
۰/۹۵۶	۰/۹۴۷	۰/۹۱۳	۰/۶۲۵	۰/۹۴۷

گام سوم: در این گام، مقدار عدم اطمینان d_j را به‌دست می‌آوریم (جدول ۱۰ـ۱).

جدول ۱۰ـ۱. نتایج گام سوم

d_j	d_1	d_2	d_3	d_4	d_5	$\sum d_j$
$1-E_j$	۰/۰۴۴	۰/۰۵۳	۰/۰۸۷	۰/۳۷۵	۰/۰۵۳	۰/۶۱۲

گام چهارم: اکنون اوزان شاخص‌ها را به‌دست می‌آوریم:

$$w_j = \frac{d_j}{\sum d_j} \Rightarrow \quad w_1 = \frac{0/044}{0/612} = 0/072 \quad w_2 = \frac{0/053}{0/612} = 0/087$$

$$w_3 = \frac{0/087}{0/612} = 0/142 \quad w_4 = \frac{0/375}{0/612} = 0/613 \quad w_5 = \frac{0/053}{0/612} = 0/087$$

جدول ۱۱ـ۱. نتایج گام چهارم

w_1	w_2	w_3	w_4	w_5
۰/۰۷۲	۰/۰۸۷	۰/۱۴۲	۰/۶۱۳	۰/۰۸۷

توجه داشته باشید که $\sum_{j=1}^{n} w_j = 1$ می‌باشد. اگر تصمیم‌گیرنده هیچ اوزان قضاوتی (ذهنی) ارایه نکرده باشد همین اوزان را به عنوان اوزان شاخص‌ها در نظر می‌گیریم؛ اما اگر از سوی فرد تصمیم‌گیرنده اوزانی پیشنهاد شده باشد، به گام بعدی می‌رویم.

گام پنجم: در این قسمت، اوزان تعدیل شده را محاسبه می‌کنیم. فرض کنید که تصمیم‌گیرنده به شاخص‌ها، وزن‌های ۰/۱ ، ۰/۲ ، ۰/۱۵ ، ۰/۳، و ۰/۲۵ را به‌ترتیب برای شاخص اول تا پنجم نسبت داده باشد. در این حالت، ابتدا باید $\sum \lambda_j w_j$ را به‌دست آوریم و سپس اوزان تعدیل شده را محاسبه کنیم.

جدول ۱۲ـ۱. نتایج گام پنجم

	۵	۴	۳	۲	۱	جمع
λ_j	۰/۱	۰/۲	۰/۱۵	۰/۳۰	۰/۲۵	۱
w_j	۰/۰۷۲	۰/۰۸۷	۰/۱۴۲	۰/۶۱۳	۰/۰۸۷	۱
$\lambda_j w_j$	۰/۰۰۷	۰/۰۱۷	۰/۰۲۱	۰/۱۸۴	۰/۰۲۲	۰/۲۵۱
w'_j	۰/۰۲۸	۰/۰۶۸	۰/۰۸۴	۰/۷۳۳	۰/۰۸۷	۱

برای مثال، w'_1 و w'_2 به این صورت محاسبه شده‌اند:

$$w'_1 = \frac{\lambda_1 w_1}{\sum \lambda_j w_j} = \frac{0/007}{0/251} = 0/028 \quad \text{و} \quad w'_2 = \frac{\lambda_2 w_2}{\sum \lambda_j w_j} = \frac{0/017}{0/251} = 0/068$$

بنا به تجربه‌ی نگارنده، روش آنتروپی به شاخص‌ها وزن‌های دور از انتظاری می‌دهد. روش AHP (که بعداً گفته می‌شود) وزن‌های معقول‌تری برای شاخص‌ها نسبت به آنتروپی محاسبه می‌کند؛ زیرا اساس کارش، نظرات تصمیم‌گیرنده (و نه ماتریس تصمیم‌گیری) است.

با تشکر