طبقه‌بندی بیزی

روح اللّه محمودی [۱] و منوچهر خردمندنیا [۲]

چکیده:

در این مقاله فرض شده است که k جامعه نرمال p–متغیره $\pi_1, \pi_2, \ldots, \pi_k$ داریم و می‌خواهیم مشاهده p–متغیره جدید y را با روش بیزی در یکی از این k جامعه طبقه‌بندی کنیم. طبقه‌بندی کلاسیک (غیربیزی) در اغلب کتب چند متغیره به تفصیل تشریح شده است. در مقاله حاضر چگونگی طبقه‌بندی بیزی را تشریح می‌کنیم. ایده‌های اساسی مقاله حاضر از مراجع پرس [٦] و جیزر [۲] الهام گرفته شده‌است اما در مقاله حاضر اثبات‌ها ساده‌تر و احتمالاً جذابتر همراه با مثال عددی ارائه شده‌است.

واژه‌های کلیدی: تحلیل ممیزی، طبقه‌بندی ، توزیع نرمال چند متغیره.

۱ مقدمه، تعاریف و علائم اختصاری

در تحلیل ممیزی هدف تعیین قاعده‌ای برای تشخیص جوامع چند متغیره رقیب از یکدیگر می‌باشد. اهمیت دیدگاه بیزی در ممیزی این است که اگر اطلاعاتی در خارج از حیطه داده‌ها وجود داشته باشد امکان مداخله آن‌ها در طبقه‌بندی وجود دارد. حتی اگر هیچ اطلاع دیگری غیر از داده‌ها وجود نداشته باشد، در روش بیزی احتمال تعلق مشاهده p–متغیره به هر یک از k جامعه رقیب مشخص می‌شود و لذا میزان قاطعیت در طبقه‌بندی بهتر خواهد بود. این در حالی است که در دیدگاه غیربیزی وضعیت تعلق یک مشاهده به یک جامعه خاص فقط به صورت " تعلق–عدم تعلق " روشن می‌شود. قبل از ورود به مبحث اصلی در این بخش چند تعریف و چند علامت اختصاری ارائه می‌شوند.

۱.۱ ماتریس داده‌های نرمال

ماتریس X را تصادفی می‌گویند هرگاه هر یک از درایه‌های آن یک متغیر تصادفی باشد. ماتریس تصادفی $X_{n\times p} = (x_1, x_2, \ldots, x_n)^T$ را در نظر بگیرید که در آن x'_r سطر r ام آن است. اگر $x_1, x_2, \ldots, x_n$ یک نمونه تصادفی از $N_p(\mu, \Sigma)$ باشد، آن‌گاه می‌گوییم که X یک ماتریس تصادفی داده‌ها از $N_p(\mu, \Sigma)$ است و یافته‌ای از X را با x نشان می‌دهیم. بسهولت می‌توان نشان داد که تابع چگالی احتمال X که در واقع تابع چگال احتمال توأم سطرهای X است به صورت زیر می‌باشد:

$$P(x|\mu, \Sigma) = (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \exp\{-\frac{1}{2}\mathrm{tr}\Sigma^{-1}V\} \quad (1)$$

که در آن

$$V = nS + n(\bar{x} - \mu)(\bar{x} - \mu)^T$$

[۱] فارغ التحصیل کارشناسی ارشد آمار دانشگاه اصفهان
[۲] عضو هیئت علمی گروه آمار دانشگاه اصفهان

و $\bar{x}$ و S به ترتیب بردار میانگین و ماتریس کوواریانس نمونه هستند.

۲.۱ توزیع t چند متغیره

گوئیم که بردار تصادفی T دارای توزیع t ی p–متغیره با پارامتر مکانی $\mu_{p\times 1}$ و پارامتر قیاس $\phi_{p\times p}$ (یک ماتریس معین مثبت) و درجه آزادی $d > 0$ است و می‌نویسیم $T \sim t_p(\mu, \phi, d)$ هرگاه تابع چگالی احتمال آن به فرم زیر باشد (ماردیا و همکاران [۵])

$$P(t|\mu,\phi,d) = \frac{\Gamma(\frac{d+p}{2})}{(\pi d)^{\frac{p}{2}}}|\phi|^{-\frac{1}{2}}\left[1+\frac{1}{d}(t-\mu)^t\phi^{-1}(t-\mu)\right]^{-\frac{d+p}{2}}$$

که در آن $\Gamma(.)$ تابع گاما است.

۳.۱ توزیع وارون ویشارت

ماتریس تصادفی متقارن $M_{p\times p}$ را دارای توزیع وارون ویشارت با پارامتر مقیاس $\psi_{p\times p}$ و d درجه آزادی گویند و می‌نویسند $M \sim IW_p(\psi, d)$ هرگاه تابع چگالی احتمالی آن به فرم زیر باشد

$$\begin{aligned} P(M|\psi,d) &= \left\{2^{\frac{pd}{2}}\Gamma_p(\frac{d}{2})|\psi|^{\frac{d}{2}}\right\}^{-1}|m|^{-\frac{(d+p+1)}{2}} \\ &\times \exp\left\{-\frac{1}{2}tr\psi^{-1}m^{-1}\right\} \end{aligned}$$

که در آن $\Gamma_p(.)$ تابع گامای p–متغیره است. برای توضیح بیشتر به عنوان مثال [۳] را ملاحظه کنید. می‌توان نشان داد که

$$E(M|\psi,d) = \frac{\psi^{-1}}{(d-p-1)}$$

برای اثبات به [۵] مراجعه کنید.

۲ برآورد بیزی پارامترهای نرمال چند متغیره

فرض کنید $X_{n\times p}$ یک ماتریس داده‌ها از $N_p(\mu,\Sigma)$ است. بنابراین درستنمایی به فرم (۱) می‌باشد. با پیشین آگاهی نابخش جفریز، یعنی پیشین $P(\mu,\Sigma) \propto |\Sigma|^{-\frac{p+1}{2}}$ توزیع پسین پارامترها عبارت است از:

$$\begin{aligned} P(\mu,\Sigma|x) &\propto P(\mu,\Sigma)P(x|\mu,\Sigma) \\ &\propto |\Sigma|^{-\frac{n+p+1}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\Sigma^{-1}V\right\} \end{aligned}$$

پسین کناری μ با انتگرال گیری از پسین توأم فوق روی Σ حاصل می‌شود که عبارت است از:

$$\mu|x \sim t_p(\bar{x}, \frac{S}{n-p}, n-p)$$

یعنی پسین کناری μ یک توزیع t ی p–متغیره با میانگین $\bar{x}$ و ماتریس مقیاس $\frac{S}{n-p}$ و درجه آزادی $n-p$ می‌باشد.

پسین کناری Σ با انتگرال گیری از پسین توأم فوق روی μ حاصل می‌شود که عبارت است از:

$$\Sigma|x \sim IW_p((nS)^{-1}, n-1)$$

یعنی پسین کناری Σ یک توزیع وارون ویشارت با پارامترهای $(nS)^{-1}$ و $n-1$ می‌باشد.

میانگین توزیع پسین یک پارامتر برآورد نقطه‌ای بیزی مناسبی برای آن پارامتر می‌باشد. در اینجا میانگین‌های توزیع‌های پسین عبارت اند از:

$$E(\Sigma|x) = nS/(n-p-2), \quad E(\mu|x) = \bar{x}$$

پیشین آگاهی نابخش خاصی که برای (μ,Σ) به کار برده‌ایم منجر به این برآوردهای نقطه‌ای شده است.

۳ طبقه‌بندی بیزی

فرض کنید X_i یک ماتریس $n_i \times p$ تصادفی داده‌ها از $N_p(\mu_i, \Sigma_i)$ (از جامعه π_i) است. فرض کنید بردار تصادفی $Y_{p\times ۱}$ نشان دهنده یک نمونه جدید با یافته y است. می‌خواهیم با روش بیزی y را به یکی از k جامعه رقیب $\pi_۱, \pi_۲, \ldots, \pi_k$ طبقه‌بندی کنیم. برای اینکار کافی است احتمال پسین متعلق بودن y به π_i را برای هر i محاسبه کنیم. اگر احتمال پیشین (قبل مشاهده y) تعلق Y به π_i را با

$$P(Y \in \pi_i | x_i)$$

نشان دهیم آن‌گاه احتمال پسین تعلق Y به π_i بر اساس قاعده بیز عبارت است از:

$$P(Y \in \pi_i | x_i, Y = y) \propto P(Y \in \pi_i | x_i) p(y|\pi_i) \quad (۲)$$

که در آن $p(y|\pi_i)$ را معمولا چگالی پیش بین می‌نامند. چگالی پیش بین را به طریق زیر می‌توان محاسبه نمود

$$\begin{aligned} p(y|\pi_i) &= \int\int P(y, \mu_i, \Sigma_i | x_i) d\mu_i d\Sigma_i \\ &= \int\int P(y|\mu_i, \Sigma_i, x_i) P(\mu_i, \Sigma_i | x_i) d\mu_i d\Sigma_i \quad (۳) \end{aligned}$$

که در آن $P(\mu_i, \Sigma_i | x_i)$ پسین توأم μ_i, Σ_i است و $P(y|\mu_i, \Sigma_i, x_i)$ درستنمایی در جامعه π_i می‌باشد. توجه کنید که احتمال پیشین را با علامت $P(Y \in \pi_i | x_i)$ و احتمال پسین را با علامت $P(Y \in \pi_i | x_i, Y = y)$ نشان دادیم. شرطی کردن احتمال پیشین $Y \in \pi_i$ روی x_i به این معنی است که قبل از مشاهده $Y = y$، این اطلاع را که X_i یک ماتریس داده ها از جامعه π_i است را داشته ایم . طبیعی است که احتمال پسین $Y \in \pi_i$ نیز باید روی x_i شرطی شود. اکنون که نحوه محاسبه احتمال پسین را با استفاده از چگالی پیش‌بین می‌دانیم با فرضیات متفاوت به محاسبه احتمال پسین متعلق بودن y به π_i می‌پردازیم. از این به بعد احتمال‌های پیشین و پسین را با علائم

$$P_{*i} = P(Y \in \pi_i | x_i, Y = y), P_{oi} = P(Y \in \pi_i)$$

نشان می‌دهیم و لذا عبارت (۲) به صورت زیر نوشته می‌شود

$$P_{*i} \propto P_{oi} p(y|\pi_i) \quad (۴)$$

۱.۳ طبقه‌بندی بیزی با فرض معلوم بودن پارامترها

اگر μ_i ها و Σ_i ها معلوم باشند چگالی پیش بین نیز برای هر i معلوم و عبارت از $Y|\pi_i \sim N_p(\mu_i, \Sigma_i)$. در این حالت با توجه به (۴) احتمال پسین برابر است با:

$$\begin{aligned} & P_{*i} \propto P_{oi} |\Sigma_i|^{-\frac{۱}{۲}} \\ & \times \exp\left\{-\frac{۱}{۲} \operatorname{tr}\Sigma_i^{-۱}(y-\mu_i)(y-\mu_i)^T\right\} \quad (۵) \end{aligned}$$

در بسیاری از متون آماری، وقتی μ_i ها و Σ_i ها نامعلومند با در نظر گرفتن برآوردهای ماکسیمم درستنمایی (ML) این پارامترها به عنوان مقادیر معلوم آنها احتمال‌های پسین را محاسبه می‌کنند. این روش را روش شبه بیزی و یا روش درستنمایی تعمیم یافته نیز می‌نامند [۴]. اگر در (۵) به‌جای μ_i و Σ_i برآوردهای ML آنها که به ترتیب $\bar{x}_i$ و S_i هستند را قرار دهیم احتمال پسین که در واقع

برآورد احتمال پسین است برابر است با

$$P_{*i} \propto P_{oi}|S_i|^{-\frac{1}{2}} \times \exp\left\{-\frac{1}{2}trS_i^{-1}(y-\bar{x}_i)(y-\bar{x}_i)^T\right\} \quad (6)$$

y به جامعه‌ای رده‌بندی می‌شود که ماکسیمم احتمال پسین را نسبت به دیگر جوامع داشته باشد.

در حالت خاص $P_{oi} = \frac{1}{k}$ روش اخیر معادل روش طبقه‌بندی غیربیزی موسوم به روش درجه دوم[3] می‌باشد.

در صورتی که ماتریس‌های کوواریانس جوامع رقیب همگون فرض شوند برآورد ML ماتریس کوواریانس آمیخته برابر

$$S_p = \frac{\sum_{i=1}^{k} n_i S_i}{\sum_{i=1}^{k} n_i} \quad (7)$$

است که در آن n_i حجم نمونه‌ی است که از جامعه π_i گرفته شده است. در این حالت با توجه به اینکه S_p به i بستگی ندارد احتمال پسین برابر است با:

$$P_{*i} \propto P_{oi}\exp\left\{-\frac{1}{2}\mathrm{tr}S_p^{-1}(y-\bar{x}_i)(y-\bar{x}_i)^T\right\} \quad (8)$$

در حالت خاص $P_{oi} = \frac{1}{k}$ روش اخیر معادل روش غیربیزی موسوم به روش خطی[4] (روش فیشر[5]) می‌باشد.

۲.۳ پارامترهای نامعلوم و ماتریس های کوواریانس ناهمگون

با این فرض که پارامترهای μ_i و Σ_i، $i = 1,2,\ldots,k$ هر دو نامعلوم و ماتریس‌های کوواریانس ناهمگون، به محاسبه احتمال پسین تعلق y به π_i می‌پردازیم. همانطور که در بخش ۳ اشاره نمودیم چگالی پیش بین از رابطه (۳)محاسبه می‌گردد و با اتخاذ پیشین جفریز، پسین توأم برابر است با

$$P(\mu_i,\Sigma_i|x_i) \propto |\Sigma_i|^{-\frac{-n_i+p+1}{2}} \times \exp\left\{-\frac{1}{2}tr\Sigma_i^{-1}V_i\right\} \quad (9)$$

که در آن $V_i = n_iS_i + n_i(\bar{x}_i-\mu_i)(\bar{x}_i-\mu_i)^T$ و $\bar{x}_i$ و S_i به ترتیب بردار میانگین و ماتریس کوواریانس نمونه مشاهده شده در جامعه π_i هستند.

بر اساس فرضیات داریم

$$Y|\mu_i,\Sigma_i,x_i \sim N_p(\mu_i,\Sigma_i)$$

بنابراین

$$P(y|\mu_i,\Sigma_i,x_i) \propto |\Sigma_i|^{-\frac{1}{2}} \times \exp\left\{-\frac{1}{2}tr\Sigma_i^{-1}(y-\mu_i)(y-\mu_i)^T\right\} \quad (10)$$

با جایگزینی (۱۰) و (۹) در (۳) می‌توان نوشت:

$$p(y|\pi_i) \propto \int\int |\Sigma_i|^{-\frac{n_i+p+2}{2}} \times \exp\left\{-\frac{1}{2}tr\Sigma_i^{-1}A_i\right\} d\mu_i d\Sigma_i \quad (11)$$

که در آن

$$A_i = n_iS_i + n_i(\bar{x}_i-\mu_i)(\bar{x}_i-\mu_i)^T + (y-\mu_i)(y-\mu_i)^T \quad (12)$$

برای انتگرال‌گیری نسبت به Σ_i توجه کنید که عبارت داخل انتگرال، صرف نظر از ثابت نرمال ساز، به صورت

[3] Quadratic Method
[4] Linear Method
[5] Fisher Method

توزیع وارون ویشارت است. بنابراین با توجه به ثابت نرمال ساز داریم:

$$P(y|\pi_i) \propto \int |A_i|^{-\frac{n_i+1}{2}} d\mu_i \qquad (13)$$

برای حل انتگرال اخیر از تساوی کلیدی زیر استفاده می‌کنیم

$$A_i = (n_i+1)(\mu_i-\alpha_i)(\mu_i-\alpha_i)^T + C_i \qquad (14)$$

که در آن

$$\alpha_i = \frac{(n_i\bar{x}_i - y)}{n_i+1}$$

و

$$C_i = \frac{n_i}{n_i+1}(\bar{x}_i - y)(\bar{x}_i - y)^T + n_i S_i$$

توجه کنید که C_i یک ماتریس $p \times p$ است که به پارامتر مجهولی بستگی ندارد و فقط تابعی از مشاهدات نمونه است.

اکنون ماتریس $(p+1) \times (p+1)$ زیر را در نظر بگیرید

$$B = \begin{bmatrix} C_i & -(\mu_i-\alpha_i) \\ (\mu_i-\alpha_i)^T & (n_i+1)^{-1} \end{bmatrix}$$

بر اساس خواص دترمینان می‌توان نوشت:

$$\begin{aligned} |B| &= |C_i|((n_i+1)^{-1} + (\mu_i-\alpha_i)^T C_i^{-1}(\mu_i-\alpha_i)) \\ &= (n_i+1)^{-1}|C_i + (n_i+1)(\mu_i-\alpha_i)(\mu_i-\alpha_i)^T| \\ &= (n_i+1)^{-1}|A_i| \end{aligned}$$

بنابراین با توجه به (۱۳) می‌توان نوشت:

$$\begin{aligned} P(y|\pi_i) &\propto |C_i|^{-\frac{n_i+1}{2}} \int \Big[(n_i+1)^{-1} \\ &+ (\mu_i-\alpha_i)^T C_i^{-1}(\mu_i-\alpha_i)\Big]^{-\frac{n_i+1}{2}} d\mu_i \end{aligned}$$

عبارت داخل انتگرال صرف نظر از ثابت نرمال ساز به فرم توزیع t ی p–متغیره است لذا با ضرب و تقسیم کردن ثابت نرمال ساز و با توجه به اینکه ثابت نرمال ساز به y بستگی ندارد می‌توان نوشت

$$\begin{aligned} P(y|\pi_i) &\propto |C_i|^{-\frac{n_i+1}{2}}|C_i|^{\frac{1}{2}} = |C_i|^{-\frac{n_i}{2}} \\ &\propto \left|S_i + \frac{1}{n_i+1}(\bar{x}_i - y)(\bar{x}_i - y)^T\right|^{-\frac{n_i}{2}} \end{aligned}$$

با استفاده از خواص دترمینان با عملیاتی نظیر آنچه در مورد ماتریس B انجام دادیم می‌توان نشان داد که

$$\begin{aligned} &\left|S_i + \tfrac{1}{n_i+1}(y-\bar{x}_i)(y-\bar{x}_i)^T\right|^{-\frac{n_i}{2}} \\ &= |S_i|^{\frac{n_i}{2}}\left[1 + \tfrac{1}{n_i+1}(y-\bar{x}_i)^T S_i^{-1}(y-\bar{x}_i)\right]^{-\frac{n_i}{2}} \end{aligned}$$

بنابراین با توجه به اینکه S_i به y بستگی ندارد می‌توان نوشت

$$\begin{aligned} P(y|\pi_i) &\propto \Bigg[1 + \frac{1}{n_i-p}(y-\bar{x}_i)^T \\ &\times \left(\frac{(n_i+1)S_i}{n_i-p}\right)^{-1}(y-\bar{x}_i)\Bigg]^{-\frac{n_1-p+p}{2}} \end{aligned}$$

ملاحظه می‌شود که

$$Y|\pi_i \sim t_p\left(\bar{x}_i, \frac{(n_i+1)S_i}{n_i-p}, n_i-p\right) \qquad (15)$$

اکنون بعد از محاسبه چگالی پیش بین، می‌توانیم بر اساس (۴) احتمال پسین تعلق y به π_i را محاسبه نماییم. احتمال پسین برابر است با:

$$P_{*i} = cP(y|\pi_i)P_{oi}$$

که در آن c یک ثابت نرمال ساز است که با استفاده از تساوی $\sum_{i=1}^{k} P_{*i} = 1$ قابل محاسبه است. در عمل

برای محاسبه $P(y|\pi_i)$ با این مشکل رو به رو هستیم که لازم است عبارتی به فرم $(a)^{-\frac{n_i+۱}{۲}}$ را محاسبه کنیم که a یک اسکالر است. مشاهده جدید y به جامعه‌ای رده‌بندی می‌شود که دارای بیشترین احتمال پسین باشد.

مثال ۳-۱ برای داوطلبینی که در کنکور دانشگاه شرکت می‌کنند یکی از سه نتیجه قبولی $(\pi_۱)$، ردی $(\pi_۲)$ و ذخیره $(\pi_۳)$ امکان پذیر است. فرض کنید می‌خواهیم قاعده‌ای طراحی کنیم که بر اساس اطلاع از معدل دو سال آخر دیپلم داوطلب پیش بینی کنیم که آیا وی در کنکور قبول، رد یا ذخیره می‌شود.

در ادامه سه ماتریس داده‌ها بر اساس افرادی که به طور تصادفی انتخاب شده اند، ارائه شده است. یکی از ماتریس‌ها، معدل دو سال آخر را برای ۵ نفر از قبولی‌ها نشان می‌دهد. ماتریس دیگر معدل دو سال آخر را برای سه نفر از رد شده‌ها و آخرین ماتریس، معدل دو سال آخر را برای چهار نفر از افراد ذخیره ارائه می‌دهد.

$$\begin{array}{ccc} \pi_۱ & \pi_۲ & \pi_۳ \\ \begin{bmatrix} ۱۴/۹ & ۱۷/۸ \\ ۱۵/۱ & ۱۲/۹ \\ ۱۵/۲ & ۱۳/۲ \\ ۱۵/۳ & ۱۵ \\ ۱۵/٦ & ۱۴/۲ \end{bmatrix} & \begin{bmatrix} ۱۴/۵ & ۱۱/۸ \\ ۱۴/۴ & ۱۱ \\ ۱۴/۲ & ۱۲/۹ \end{bmatrix} & \begin{bmatrix} ۱۴/۸ & ۱۳/۷ \\ ۱۴/۹ & ۱۳/۱ \\ ۱۵/۱ & ۱۰/۷ \\ ۱۵/۲ & ۸/۸ \end{bmatrix} \end{array}$$

به منظور ارزیابی قاعده رده‌بندی بیزی فوق هر یک از ۱۲ مشاهده دو متغیره فوق را به عنوان یک مشاهده مستقل جدید تلقی نموده آنرا بر اساس روش بیزی و با فرض اینکه $P_{oi} = \frac{۱}{k} = \frac{۱}{۳}$ طبقه‌بندی نمودیم. نتیجه رده‌بندی را می‌توان در جدول (۱) ملاحظه نمود. در این مثال با اتخاذ پیشین $P_{oi} = \frac{۱}{۳}$ در واقع فرض کرده‌ایم که هیچ اطلاع پیشینی راجع به احتمال تعلق هر یک از مشاهدات به جوامع رقیب وجود ندارد و لذا طبقه‌بندی صرفاً بر اساس اطلاعاتی که در داده‌ها وجود دارد صورت می‌گیرد. با اتخاذ این پیشین نتایج طبقه‌بندی دقیقاً با نتایج طبقه‌بندی غیربیزی یکسان است ولی این تفاوت مهم وجود دارد که در روش بیزی احتمال‌های پسین تعلق هر مشاهده به هر یک از سه جامعه رقیب نیز قابل محاسبه می‌باشند. در این مثال محاسبات احتمال‌های پسین بر اساس عبارت (٦) صورت گرفته است.

جدول (۱) احتمال های پسین تعلق هر یک از ۱۲ مشاهده به π_i، $i = ۱, ۲, ۳$

نتیجه رده‌بندی	احتمال متعلق بودن به جامعه $\pi_۳$	احتمال متعلق بودن به جامعه $\pi_۲$	احتمال متعلق بودن به جامعه $\pi_۱$	y^T	جامعه
درست	۰/۰۱۰۴	۰/۰۲۰۷	۰/۹۶۸۹	$y_۱^T = (۱۴/۹, ۱۷/۸)$	$\pi_۱$
درست	۰/۰۸۱۵	۰/۰۵۰۳	۰/۸۶۸۲	$y_۲^T = (۱۵/۱, ۱۲/۹)$	
درست	۰/۰۱۱۲	۰/۰۲۶۱	۰/۹۶۲۸	$y_۳^T = (۱۵/۲, ۱۳/۲)$	
درست	۰/۰۰۱۱	۰/۰۰۹۴	۰/۹۸۹۵	$y_۴^T = (۱۵/۳, ۱۵)$	
درست	$۵/۹۰۱۸e - ۰۰۴$	۰/۰۱۳۵	۰/۹۸۵۹	$y_۵^T = (۱۵/۶, ۱۴/۲)$	
درست	۰/۰۰۱۳	۰/۹۷۴۸	۰/۰۲۳۸	$y_۱^T = (۱۴/۵, ۱۱/۸)$	$\pi_۲$
درست	$۴/۲۷۹۶e - ۰۰۴$	۰/۹۸۶۸	۰/۰۱۲۷	$y_۲^T = (۱۴/۴, ۱۱)$	
درست	$۳/۲۱۱۶e - ۰۰۴$	۰/۹۸۷۹	۰/۰۱۱۸	$y_۳^T = (۱۴/۲, ۱۲/۹)$	
درست	۰/۸۹۸۹	۰/۰۱۹۸	۰/۰۸۱۳	$y_۱^T = (۱۴/۸, ۱۳/۷)$	$\pi_۳$
درست	۰/۹۱۲۹	۰/۰۱۴۲	۰/۰۷۲۹	$y_۲^T = (۱۴/۹, ۱۳/۱)$	
درست	۰/۹۳۴۶	۰/۰۲۱۲	۰/۰۴۴۲	$y_۳^T = (۱۵/۱, ۱۰/۷)$	
درست	۰/۹۴۴۵	۰/۰۳/۷	۰/۰۲۴۸	$y_۴^T = (۱۵/۲, ۸/۸)$	

۴ مباحث تکمیلی

در این بخش به اختصار مطالب تکمیلی ارائه می‌شوند.

۱.۴ پارامترها نامعلوم و ماتریس‌های کوواریانس همگون

فرض کنید X_i یک ماتریس $n_i \times p$ تصادفی داده‌ها از $N_p(\mu_i, \Sigma)$ است $(i = ۱, ۲, \ldots, n)$. با پیشینی به فرم $P(\mu_i, \Sigma) \propto |\Sigma|^{-\frac{p+۱}{۲}}$ و با روشی مشابه بخش ۲.۳ می‌توان نشان داد، برای حالتی که ماتریس‌های کوواریانس همگون هستند، چگالی پیش بین عبارت است از

$$Y|_{\pi_i} \sim t_p\left(\bar{x}_i, \frac{(n_i + ۱)S_p}{(n_i - p)}, n_i - p\right)$$

که در آن

$$S_p = \frac{\sum_{i=۱}^{k} n_i S_i}{\sum_{i=۱}^{k} n_i}$$

اثبات و توضیحات مفصل در [۱] داده شده است. مراجعه کنید.

۲.۴ طبقه‌بندی نیرومند

در حالتی که پارامترها مجهول هستند، برآورد آنها می‌تواند به شدت تحت تأثیر داده‌های آلوده قرار گیرد. به این جهت معقول است که قبل از اعمال هر روش ابتدا با استفاده از یک روش نیرومند بردار میانگین و ماتریس کوواریانس برآورد شوند. به عنوان مثال هوبرت و

همکاران [۴] از روش نیرومند مینیمم دترمینان کوواریانس (MCD) برای برآورد پارامترها استفاده کرده‌اند. برای توضیحات مفصل راجع به طبقه‌بندی بیزی سریع و نیرومند به [۱] مراجعه کنید.

مراجع

[۱] محمودی، ر. (۱۳۸٦)، تحلیل ممیزی بیزی سریع و نیرومند، پایان‌نامه کارشناسی ارشد آمار، گروه آمار دانشگاه اصفهان.

[2] Geisser, S. (1982). *Bayesian Discrimination*, in Krishnaiah, P. R. and Kanal, L. N. (ed.s.), *Handbook of Statistics*, 2, 101-120, Amsterdam, North Holland.

[3] Gupta, A.K. and Nagar, D.K. (2000). *Matrix Variate Distributions*, Chapman and Hall, New York.

[4] Hubert, M. and Driessin, K. V. (2004). Fast and robust discriminant analysis, *Computational Statistics and Data Analysis*, 45,301-320.

[5] Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). *Multivariate Analysis*, Academic Press, Sandigo

[6] Press, S.J. (2003). *Subjective and Objective Bayesian Statistics*, 2nd Ed., Wiley-Interscience, New York.

[7] Srivastava, M. S. (2002). *Methods of Multivariate Statistics*, Wiley-International, New York.