# بررسی توسعه‌ی انسانی کشورها
# با استفاده از روش تحلیل ممیزی آمیخته

دکتر رحیم چینی پرداز، قاسم رکابدار و رضا یوسفی حاجی آباد*

تاریخ وصول: 84/9/29 تاریخ پذیرش: 85/4/1

**چکیده:**

***تحلیل ممیزی خطی[1](LDA) ابزار مهمی برای رده بندی دو یا چند گروه است. این روش برای حالتی که کلاس‌ها دارای توزیع نرمال چند متغیره هستند، معادل با روش ماکزیمم درستنمـــایی است. در حالتی که کلاس‌هـــا نرمال نباشنـــد، ممیزی منجر به روش‌های ناپارامتری می‌شود. در این مقاله فرض می‌شود که نرمال بودن کلاس‌ها مشخص نیست، اما هر کلاس شامل زیر کلاس‌هایی است که دارای توزیع نرمال هستند. بر اساس این فرض از تحلیل ممیزی آمیخته (MDA)[2] استفاده شده که در این حالت برآورد پارامترها تنها با استفاده از الگوریتم EM[3] قابل محاسبه است. برای مقایسه‌ی روش‌های ممیزی نرمال و تحلیل ممیزی درجه‌ی دوم[4] (QDA) با تحلیل ممیزی آمیخته از داده‌های توسعه‌ی انسانی سال 2002 بر اساس تقسیم بندی کشورها به سه گروه با توسعه‌ی انسانی بالا، توسعه‌ی انسانی متوسط و توسعه‌ی انسانی ضعیف استفاده شده است. نتایج تحقیق نشان داد که تحلیل ممیزی آمیخته دارای رده بندی مؤثرتری نسبت به تحلیل ممیزی خطی و تحلیل ممیزی درجه‌ی دوم است.***

***طبقه‌بندی JEL: O15***

***واژه‌های کلیدی:*** *رده بندی، شناسایی الگو، خوشه بندی، الگوریتم **EM**، تحلیل ممیزی خطی، شاخص توسعه‌ی انسانی*

پژوهشگاه علوم انسانی و مطالعات فرهنگی

---

* به ترتیب، استاد آمار دانشکده‌ی علوم ریاضی و آمار دانشگاه شهید چمران اهواز، کارشناس ارشد کاربرد آمار در اقتصاد و علوم اجتماعی و کارشناس ارشد اقتصاد (chinipardaz_r@cua.ac.ir)

[1] Linear Discriminant Analysis
[2] Mixture Discriminant Analysis
[3] EM algorithm
[4] Quadratic Discriminant Analysis

## 1- مقدمه

ویژگی مهم جهان امروز، وجود تفاوت‌های شدید در توسعه‌ی اقتصادی، سیاسی، اجتماعی و انسانی است. در سال‌های اخیر تاکید فراوانی بر رویکرد توسعه‌ی انسانی صورت‌گرفته است. توسعه‌ی انسانی را می‌توان به معنای فرایند بسط انتخاب‌های انسانی دانست. هدف اساسی توسعه‌ی انسانی، بهره‌مند ساختن واقعی و بهبود سطح زندگی انسان بوده است، به طوری که افزایش درآمد و گسترش اشتغال، از امکانات و ابزارهای ضروری توسعه‌ی انسانی هستند. توسعه‌ی انسانی استراتژی متحرک کردن توان، مهارت، استعداد و خلاقیت افراد است و دولت‌ها را برای ارایه‌ی خدمات مناسب و با کیفیت پاسخگو ساخته، در نهایت منجر به شکل گیری یک جامعه‌ی مدنی نیرومند می‌شود.

در کنار مفهوم توسعه‌ی انسانی، رویکرد توسعه‌ی منابع انسانی وجود دارد. این رویکرد نیز به ایجاد ظرفیت‌ها و پرورش توانمندی‌ها توجه دارد و رشد بهره‌وری افراد را برای نیل به رشد اقتصادی، در سر لوحه‌ی کار خود قرار می‌دهد. بر این اساس، در مفهوم توسعه‌ی انسانی، تمرکز بر مردم به عنوان سرمایه‌های انسانی و وسایل تولید کالاها و خدمات نیست، بلکه همراه با تاکید بر زندگی بهتر، بر ایجاد ظرفیت‌ها به جای تاکید بر مصرف کالاها و خدمات، تمرکز داشته، پرورش استعدادهای ذهنی در کنار رشد ظرفیت‌های مادی را هدف قرار می‌دهد. مدت‌ها توسعه به معنای رشد تولید ناخالص ملی تلقی می‌شد، در حالی‌که سطح بالای درآمد به خودی خود پیشرفت انسانی را تضمین نمی‌کند. به بیانی دیگر، رشد اقتصادی شرط لازم توسعه‌ی انسانی است، اما شرط کافی نیست و کیفیت رشد اقتصادی به اندازه‌ی کمیت آن در نیل به توسعه اهمیت دارد؛ به همین دلیل، مفهوم توسعه‌ی انسانی در عین تاکید بیشتر بر جنبه‌ی کیفی، به هر دو جنبه‌ی رشد نظر دارد.

توسعه‌ی انسانی هدف نهایی و بهترین وسیله‌ی ممکن برای پیشبرد توسعه‌ی اقتصادی است. از مزایای توسعه‌ی انسانی این است که نقش مستقیمی در رفاه جامعه دارد و بر پایه‌ی فرصت‌های برابر بنا شده است؛ و بین انواع مختلف سرمایه‌گذاری برای توسعه‌ی نیروی انسانی ارتباط ایجاد می‌کند. این مزایا تنها با هدایت و رهبری دولت حاصل می‌شود. به دیگر سخن، تا حصول نتایج مورد انتظار توسعه‌ی انسانی مداخله‌ی دولت لازم است. در این راه کاهش تمرکز و مشارکت مردم به عنوان یک مشخصه‌ی اساسی رویکرد توسعه، اهمیت زیادی دارد. جنبه‌ی

دیگر رویکرد توسعه‌ی انسانی (علاوه بر هدف و وسیله‌ی تحقق رفاه)، حفظ قابلیت افراد یا تامین اقتصادی، یعنی تامین افراد جامعه از لحاظ سطح خدمات بهداشتی، تامین مواد غذایی، شغل و سهمی عادلانه از درآمد و ثروت، در جهت کاهش محرومیت افراد جامعه است.

بررسی وضعیت توسعه‌ی انسانی هر کشور، می‌تواند راهنمای مسولان آن جامعه برای تدوین سیاست‌های مناسب اقتصادی همچون ایجاد اشتغال، توزیع عادلانه‌ی درآمد، حذف انحصارات، تخصیص بهینه‌ی منابع و هزینه‌ها، کاهش فقر و شتاب در دگرگونی‌های نهادی است. هدف تحقیق حاضر، تعیین وضعیت توسعه‌ی انسانی در 115 کشور مورد بررسی، به روش تحلیل ممیزی و با استفاده از داده‌های شاخص توسعه‌ی انسانی است. بررسی روش‌های تحلیل ممیزی نرمال ($LDA$)، تحلیل ممیزی آمیخته ($MDA$) و تحلیل ممیزی درجه‌ی دوم ($QDA$)، از نظر کارایی جهت رتبه‌بندی کشورهای نمونه، از لحاظ وضعیت توسعه‌ی انسانی، در این تحقیق مورد بررسی قرار گرفته است.

در این مقاله تحلیل ممیزی آمیخته، بررسی شده و الگوریتم $EM$ برای برآورد پارامترها و ممیزی پیشنهاد شده است. این الگوریتم به وسیله‌ی هاستی و تیبشیرانی[5] (1996) پیشنهاد و به‌وسیله‌ی نویسندگان مقاله انطباق و سپس بر اساس آن سطح توسعه‌ی انسانی کشورها بررسی شده است. در بخش دوم ممیزی خطی مرور می‌شود. بخش سوم توسعه $LDA$ به $MDA$ بررسی می‌شود. بخش چهارم به انطباق الگوریتم $EM$ برای استفاده در ممیزی $MDA$ می‌پردازد. در بخش پنجم داده‌های توسعه‌ی انسانی برای مقایسه‌ی $MDA$ با $LDA$ و $QDA$ ارائه می‌شود. بخش نهایی به نتیجه گیری در خصوص $MDA$ اختصاص دارد.

## 2- روش تحقیق

رده بندی دو یا چند کلاس یا ممیزی از موضوعات مهم در آمار چند متغیره بوده که دارای کاربرد در حوزه‌های گوناگون است. در حالت کلی، ممیزی اختصاص یک یا چند مشاهده‌ی $x$ با کلاس نامعلوم به جوامع معلوم $w_1, w_2, ..., w_J$ است. بر این اساس، مجموعه‌ی راهنما $\{x_i, g_i\}_{i=1}^{n}$ است که در آن $x_i \in R^P$ موجود و کلاس

[5] Hastie and Tibshirani

آنها معلوم است. با استفاده از مجموعه‌ی راهنما پارامترها در مدل برآورد می‌شود و سپس با استفاده از پارامترهای برآورد شده، یک قاعده برای ممیزی، به صورت یک تابع برای پیش بینی مشاهده‌ی جدید $x$ در یکی از کلاس‌ها به دست می‌آید.

روش سنتی و کلاسیک آماری برای این منظور، ممیزی خطی و رگرسیون لجستیک است که این روش‌ها تا حدود زیادی توسعه یافته است (ماردیا و همکاران،[6] 1979 و آندرسن،[7] 1984). هرگاه جوامع نرمال باشند، تابع ممیزی خطی نیز نرمال است و خطای ممیزی نیز براحتی قابل محاسبه خواهد بود. روش درجه‌ی دوم در ممیزی به دلیل پیچیده بودن توزیع ممیزی به وسیله‌ی آماردانان هنوز در حال توسعه است.

متأسفانه فرض نرمال بودن کلاس‌ها بندرت صادق است. در چنین صورتی، تابع ممیزی، داده‌ها را به صورت مؤثری از هم جدا نمی‌کند و خطای ممیزی افزایش خواهد یافت. یک راه برای حل این مسأله، فرض تعمیم $LDA$ است. بر این اساس فرض می‌شود که کلاس‌ها شامل زیر کلاس‌هایی نرمال‌اند؛ یعنی کلاس‌ها دارای توزیع آمیخته‌اند و بنابراین، ممیزی به کار رفته نیز آمیخته خواهد بود. این موضوع در نوشته‌های مک لاچلان[8] (1992) و تکست و دیگران[9] (1991) پیشنهاد شده است. هاستی و تیبشیرانی (1996) این موضوع را گسترش داده‌اند. آنها فرض کرده‌اند که هر کلاس، از زیر کلاس‌هایی با میانگین متفاوت و ماتریس کواریانس درون و بین کلاس‌های مشترک هستند.

### 2-1- ممیزی خطی $LDA$

فرض کنید در مساله‌ی ممیزی، مجموعه‌ی راهنما $\{x_i, g_i\}_{i=1}^{n}$ و توزیع $x_i$ به صورت $X_i \sim N_p(\mu_j, \Sigma)$ برای $j = 1, 2, \ldots, J$ باشد. در این‌صورت چگالی شرطی $x$ در کلاس $j$ ام به صورت زیر است:

$$\varphi(\mu_j, \Sigma) = P(X = x \mid G = j) = |2\pi\Sigma|^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2} D(x_i, \mu_j)\right\} \quad (1)$$

لگاریتم درستنمایی مشاهدات به صورت زیر است.

---

[6] Mardia
[7] Anderson
[8] Mclachlan
[9] Taxt

$$2\log(\mu_j, \Sigma) = -\sum_{j=1}^{J} \sum_{g_j = j} D(x_i, \mu_j) - n\log|\Sigma| \quad (2)$$

$$D(x; \mu) = (x - \mu)' \Sigma^{-1} (x - \mu)$$

فاصله‌ی ماهالانوبیس[10] بین $x$ و $\mu$ و $\sum_{g_i=j}$ به معنی مجموع مشاهدات در کلاس $j$ ام است. با این فرض که احتمال پیشین کلاس $j$ ام به صورت $P(G=j)=\Pi_j$ باشد که معمولاً از قبل معلوم است یا از مجموعه‌ی راهنما برآورد می‌شود. اگر هیچ دلیل قبلی بر ترجیح کلاس‌ها بر یکدیگر نباشد، $P(G=j)=\frac{1}{J}$ در نظر گرفته می‌شود. اگر احتمال پسین $P(G \mid x)$ باشد و رابطه‌ی زیر برقرار باشد، آنگاه مشاهده‌ی $x_0$ در کلاس $j$ ام رده بندی می‌شود؛ یعنی $C(x_0)=j$ است.

$$P(G=j \mid X=x_0) = \max_{\ell} P(G=\ell \mid X=x_0) \quad (3)$$

هر گاه $P(G=j)=\frac{1}{J}$ باشد، این روش با روش ممیزی ماکزیمم درستنمایی معادل می‌شود. با توجه به رابطه‌ی (1)، چگالی پسین کلاس $j$ ام به صورت زیر خواهد بود:

$$P(G=j \mid X=x) = \frac{\Pi_j \varphi(\mu_j, \Sigma)}{\sum_{\ell=1}^{J} \Pi_\ell \varphi(\mu_\ell, \Sigma)} = -\int \exp\{x' \Sigma^{-1} \mu_j - \tfrac{1}{2} \mu_j' \Sigma^{-1} \mu$$

$$+ \log\Pi_j = \int = \exp\{X' \beta_j + a_j\} \quad (4)$$

در رابطه‌ی فوق ثابت بودن مخرج کسر نسبت به $j$ تأثیری در ماکزیمم کردن کلاس‌ها ندارد. همچنین، کران مرزی بین دو کلاس مانند $i$ و $j$ ام به صورت مجموعه‌ای از نقاط تعریف می‌شود که دارای احتمالات پسین مساوی هستند؛ یعنی:

$$b_{i,j} = \{x \in R^p ; P(G=j \mid X=x) = P(G=i \mid X=x)\} \quad (5)$$

$$b_{i,j} = \{x \in R^p ; D(\mu_j, \Sigma) = D(\mu_i, \Sigma)\} \quad (6)$$

که تابعی خطی از $x$ است. تابع ممیزی برای کلاس $j$ ام به صورت زیر است:

$$\delta_j(x) = -(x'\beta_j + \alpha_j) \quad (7)$$

[10] Mahalanobis

که تعریف و قاعده‌ی رده بندی انتساب مشاهده جدید $x_0$ است. اگر رابطه‌ی (8) برقرار باشد، $\alpha_L$ و $\beta_L$ با استفاده از مجموعه‌ی راهنما برآورد می‌شوند.

$$\delta_j(x_0) = \min_{\ell} (x_0'\beta_\ell + \alpha_\ell) \tag{8}$$

همان‌طورکه گفته شد، برآوردهای $LDA$ معادل برآوردهای ماکزیمم درستنمایی هستند. بنابراین، با ماکزیمم کردن رابطه‌ی (2) خواهیم داشت:

$$\hat{\mu}_j = \frac{1}{n_j} \sum_{g_i = j} x_i$$
$$\hat{\Sigma} = \frac{1}{n} \sum_{g_i = j}^{J} \sum_{g_i = j} (x_i - \hat{\mu}_j)(x_i - \hat{\mu}_j)' \tag{9}$$

لازم به‌ذکر است که اگر کلاس‌ها دارای ماتریس کواریانس مختلف باشند، در آن صورت توزیع $x_i$ به صورت زیر است:

$$X_i \sim N_P(\mu_j, \Sigma_j) \tag{10}$$

و کران تصمیم بین دو کلاس $i$ و $j$ به صورت زیر است:

$$b_{i,j} = \left\{ x \in R^p ; D(x, \mu_i, \Sigma_i) + log|\Sigma_j| = D(x, \mu_j, \Sigma_j) + log|\Sigma_i| \right\} \tag{11}$$

که غیر خطی و به صورت تابع درجه‌ی دوم از $x$ به صورت زیر است.

$$x'(\Sigma_i^{-1} - \Sigma_j^{-1})x \tag{12}$$

**2-2- گسترش $LDA$ به کلاس‌های آمیخته‌ی نرمال**

فرض کنید در بخش قبلی تابع چگالی در هر کلاس $P(X = x \mid G = j)$ چگالی آمیخته نرمال باشد که در آن هر مؤلفه دارای میانگین مربوط به خود و ماتریس کواریانس مشترک با بقیه‌ی مؤلفه‌ها باشد. بنابراین، در کلاس $j$ ام آمیخته‌ی نرمال به صورت رابطه‌ی زیر است.

$$m_j(X) = \sum_{r=1}^{R_j} \pi_{jr} \varphi_{jr}(\mu_{jr}, \Sigma) = |2\pi\Sigma|^{-\frac{1}{2}} \sum_{r=1}^{R_j} \pi_{jr} exp\left\{ -\frac{1}{2} D(\mu_{jr}, \Sigma) \right\} \tag{13}$$

در رابطه‌ی فوق، $R_j$ تعداد زیرکلاس‌های گروه $j$ ام و $\pi_{jr}$ نسبت‌های آمیختگی هر یک از مؤلفه‌های نرمال و $\sum_{r=1}^{R_j}\pi_{jr}=1$ است. در چنین حالتی احتمال پسین کلاس $j$ ام در صورتی‌که $x$ مشاهده شده باشد، به صورت زیر است:

$$P(\quad G = j \mid X = x) = \frac{\sum_{r=1}^{R_j} \pi_{jr}\, \varphi_{jr}(\mu_{ir}, \Sigma)}{\sum_{j=1}^{J}\sum_{r=1}^{R_j} \pi_{jr}\, \varphi_{jr}(\mu_{jr}, \Sigma)} \qquad (14)$$

و لگاریتم درستنمایی شرطی به صورت زیر است.

$$\ell_{mix}(\mu_j, \Sigma, \pi_{jr}) \propto \sum_{j=1}^{J}\sum_{g_i=j} log\Big(\sum_{r=1}^{R_j} \pi_{jr} exp \Big\{-\frac{1}{2} D(x_i, \mu_{jr})\Big\}\Big) - \frac{n}{2} log \mid \Sigma \mid \qquad (15)$$

در اینجا مشاهده‌ی جدید $x_0$ با ماکزیمم کردن رابطه‌ی (13) به یکی از کلاس‌های $j=1,...,J$ نسبت داده می‌شود. یعنی $x_0$ به کلاس $j$ ام نسبت داده می‌شود اگر رابطه‌ی زیر برقرار باشد.

$$P(G=j \mid X=x_0) = \max_{e} P(G=\ell \mid X=x_0)$$

که در آن $P(G=\ell \mid X=x_0)$ از رابطه‌ی (13) به‌دست می‌آید.

پارامترهای مدل، یعنی $\mu_{jr}$، $\pi_{jr}$ و $R_j$ برای $r=1,...,R_j$ و $j=1,...,J$ باید از طریق مجموعه‌ی راهنما برآورد شوند. بدیهی است که ماکزیمم کردن رابطه‌ی (14) از طریق مشتق گیری امکان پذیر نیست و باید با استفاده از روش‌های عددی ماکزیمم شود.

### 2-3- الگوریتم *EM* برای برآورد پارامترها

دمپستر و دیگران[11] (1977) الگوریتمی دو مرحله‌ای *(EM)* را برای برآورد پارامترها مطرح کرده‌اند که در هنگام ناکامل بودن مشاهدات کاربرد دارد. این

[11] Dempster

الگوریتم به وسیله‌ی بعضی از آماردانان توسعه یافته است. شوم وی[12] (1989) در برآورد پارامترهای سری‌های زمانی از این الگوریتم استفاده کرده است. نایت[13] (2000) برتری این روش را نسبت به روش‌های دیگر مانند استفاده از الگوریتم نیوتون - رافسون نشان داده است. الگوریتم $EM$ روشی تکراری تا رسیدن به همگرایی است که در هر تکرار (مرحله) امید ریاضی ($E$) و برآورد ماکزیمم درستنمایی ($M$) وجود دارد. در اینجا با استفاده از این روش برآوردهای ماکزیمم درستنمایی به‌دست می‌آیند.

### 2-3-1- مرحله‌ی $E$

با فرض درستی پارامترها، احتمال مشاهده‌ی $x_i$ متعلق به ریز کلاس $C_{jr}$ برای همه‌ی مشاهدات مجموعه‌ی راهنما در کلاس $g_i = j$ محاسبه می‌شود.

$$P\left(C_{jr}(x_i) \mid x_i, j\right) = \frac{\pi_{jr}\,\varphi(x_i, \mu_{jr}, \Sigma)}{\sum_{\ell=1}^{R_j} \pi_{j\ell}\,\varphi(x_i, \mu_{j\ell}, \Sigma)} \qquad (16)$$

### 2-3-2- مرحله‌ی $M$

با استفاده از وزن‌های برآورد شده در مرحله‌ی $E$، برآوردهای ماکزیمم درستنمایی (*MLE*) برای پارامترهای هر مؤلفه‌ی نرمال و در هر کلاس محاسبه می‌شود.

$$\hat{\pi}_{jr} \propto \sum_{g_i=j} P(C_{jr}(x_i) \mid x_i, j), \qquad \sum_{r=1}^{R_j} \hat{\pi}_{jr} = 1 \qquad (17)$$

$$\hat{\mu}_{jr} = \frac{\sum_{g_i=j} P(C_{jr}(x_i) \mid x_i, j)\, x_i}{\sum_{g_2=i} P(C_{jr}(x_i) \mid x_i, j)} \qquad (18)$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{j=1}^{J} \sum_{g_i=j} \sum_{r=1}^{R_j} P\left(C_{jr}(x_i) \mid x_i, j\right)\left(x_i - \hat{\mu}_{jr}\right)\left(x_i - \hat{\mu}_{jr}\right)' \qquad (19)$$

[12] Shumway
[13] Knight

همان‌طورکه از مقایسه‌ی رابطه‌ی (9) با رابطه‌های (18) و (19) دیده می‌شود، برآوردهای ماکزیمم درستنمایی مانند حالت نرمال کامل هستند؛ با این تفاوت که وزن‌های $P(C_{jr}(x_i)|x_i,j)$ جایگزین تابع نشانگر کلاس‌ها شده‌اند. با توجه به اینکه وزن $P(C_{jr}(x_i)|x_i,j)$ خود تابعی از $\mu_{jr}$ و $\Sigma$ است، رابطه‌های (18) و (19) در رابطه‌ی (16) جایگزین شده تا وزن‌های بعدی محاسبه شوند. این مراحل آنقدر تکرار می‌شود تا همگرایی در لگاریتم درستنمایی شرطی (14) برقرار شود؛ یعنی اگر از مرحله‌ای به مرحله‌ی بعد، رابطه‌ی (14) تغییر کمی داشته باشد، الگوریتم متوقف می‌گردد و برآوردهای نهایی حاصل می‌شوند.

تکرارهای الگوریتم $EM$ به اندازه‌ی خوشه‌های $R_j$ و مقادیر آغازین برای $\mu_{jr}$ و $\Sigma$ و احتمال $P(C_{jr}(x_i)|x_i,j)$ بستگی دارد. به این منظور، می‌توان از الگوریتم خوشه بندی میانگین $K$ ام یک مقدار ثابت برای خوشه‌ها تعیین کرد و سپس از این الگوریتم برای برآورد مراکز زیر کلاس‌ها $\mu_{jr}$ در هر کلاس استفاده شود. اگر $\mu_{jr}$ نزدیکترین مرکز به $x_i$ باشد، آنگاه برای همه‌ی مشاهدات در کلاس $j$ ام احتمال $P(C_{jr}(x_i)|x_i,j)$ برابر با یک است و در غیر این‌صورت صفر است.

پس از برآورد پارامترها، مشاهده‌ی جدید $x_0$ مطابق با رابطه‌ی (13) به خوشه‌ی مورد نظر نسبت داده می‌شود. در این حالت، ناحیه‌ی مرزی بین کلاس $i$ و $j$ ام به‌صورت زیر است.

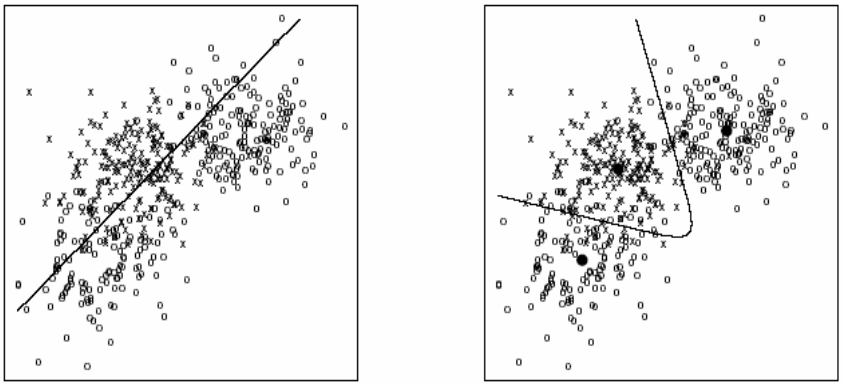
$$\left\{\begin{array}{l} b_{ij}=x\in R^P \\ log\left[\sum_{r=1}^{R_j}\pi_{jr}exp\left\{-\frac{1}{2}D(x,\mu_{jr})\right\}\right]=log\left[\sum_{r=1}^{R_i}\pi_{ir}exp\left\{-\frac{1}{2}D(x,\mu_{ir})\right\}\right] \end{array}\right\} \quad (20)$$

بدیهی است که این رابطه شکل خطی یا درجه‌ی دوم نداشته و بنابراین ناحیه‌های مرزی در $MDA$ پیچیده و غیر خطی است.

نمودار (1) یک مثال ساده شبیه سازی شده برای مقایسه‌ی کران‌های تصمیم خطی و غیرخطی است. در کلاس اول 40 مشاهده دارای توزیع نرمال دو متغیره با میانگین (0 و 0) و 40 مشاهده‌ی دیگر دارای توزیع نرمال دو متغیره با میانگین (4 و 4) است. در کلاس دوم مشاهدات دارای توزیع نرمال با میانگین (2 و 2) هستند و ماتریس کواریانس در هر دو کلاس به صورت یکانی $\Sigma=I_2$ است با توجه به این

نمودار مشخص می‌شود که در *MDA* حتی با فرض مساوی بودن ماتریس کواریـانس کلاس‌ها، کران‌های تصمیم بین کلاس‌ها غیر خطی است.

**نمودار** 1: کران تصمیم بین دو کلاس



ممیزی خطی ممیزی آمیخته

## 3- مروری بر سوابق تحقیق

در گذشته روش تحلیل ممیزی اغلب به عنوان آزمون برابری دو یا چند توزیع بوده، اما امروزه این روش به همراه دیگر روش‌های آماری، دارای کاربرد فراوانی در زمینه‌های اقتصادی، پزشکی، جمعیت شناسی و روانشناسی است.

گرین[14] (1967) با اندازه‌گیری 14 متغیر شامل جمعیت شهری، درآمد و تعداد جراید منتشر شده، تعداد 88 شهر اروپایی را به گروه‌های مختلف، تقسیم نموده و توسعه یافتگی آنها را بررسی کرده است. نتایج تحقیق نشان داد که تعداد 13 شهر دارای توسعه یافتگی بالا، 33 شهر دارای توسعه‌یافتگی متوسط و 42 شهر از توسعه‌یافتگی کمی برخوردار بوده‌اند.

رائو[15] (1950) با در نظر گرفتن صفاتی همانند اندازه‌ی قامت، اندازه‌ی بالاتنه، گودی بینی و بلندی قد و با استفاده از روش نمونه‌گیری جمعیتی، به تفکیک و متمایز نمودن سه جمعیت برهمن، کوارا و آرتیزان در هند اقدام نمود.

[14] Green
[15] Rao

کومانز و دیگران[16](1983) با استفاده از روش تحلیل ممیزی، به رتبه بندی 215 بیمار تیروئیدی، به سه گروه غدد تیروئیدی طبیعی، کم‌کار و پرکار پرداخته‌اند. در این تحقیق متغیرهای اندازه‌گیری شده تعیین مشاهدات موجود در هر کلاس (گروه) شامل آزمایش برداشت عصاره تری یدوتیرونین و سطح کلی هورمون تیروکسن در سرم بوده است. نتیجه تحقیق نشان داد که در کلاس اول 150 مشاهده، در کلاس دوم 35 مشاهده و در کلاس سوم 30 مشاهده موجود است.

پردل نوغابی(1379)، در تحقیقی دیگر، با استفاده از روش آنالیز خوشه‌ای، به بررسی وضعیت 127 کشور از لحاظ وضعیت توسعه‌ی انسانی و تفکیک آنها به سه گروه کشورهایی با توسعه‌ی انسانی بالا، توسعه‌ی انسانی متوسط و توسعه‌ی انسانی پایین، اقدام کرد. در این مورد، متغیرهایی همانند نرخ با سوادی بزرگسالان، انتظار زنده ماندن در بدو تولد، میانگین سال‌های تحصیلی، سرانه‌ی تولید ناخالص داخلی ($GDP$)، نرخ باروری، جمعیت شهری، نیروی کار، درصد نیروی کار شاغل در بخش کشاورزی و صنعت، و همچنین تعداد تلفن به ازای هر 100 نفر و میزان جمعیت به ازای هر پزشک، رتبه‌بندی کشورهای مورد نظر مورد ارزیابی قرارگرفته است.

## 4- رده بندی کشورها بر اساس توسعه‌ی انسانی

از سال 1990 سازمان ملل هر ساله گزارشی تحت عنوان گزارش توسعه‌ی انسانی، منتشر می‌کند، در این گزارش اندازه‌ی متغیرهای امید به زندگی، نرخ باسوادی، نرخ رشد نام نویسی در دوره‌های تحصیلی و سرانه‌ی محصول ناخالص داخلی برای رده بندی کشورها از نظر توسعه‌ی انسانی گزارش می‌شود.

جمع آوری اطلاعات معمولاً با استفاده از اطلاعات رسمی منتشره‌ی کشورها و در صورت غیر قابل اعتماد بودن اطلاعات، به وسیله‌ی برنامه‌ی عمران سازمان ملل صورت می‌گیرد. در این گزارش، با استفاده از داده‌های به دست آمده، شاخص توسعه‌ی انسانی[17] ($HDI$) محاسبه می‌شود و بر اساس این شاخص کشورها به سه رده‌ی کشورهای با توسعه‌ی انسانی بالا، متوسط و ضعیف تقسیم می‌شوند.

پرتال جامع علوم انسانی

[16] Coomans
[17] Human Development Index

**4-1- رده بندی کشورها با استفاده از تحلیل ممیزی**

برای ممیزی بین کشورها 115 کشور به طور تصادفی به عنوان مجموعه‌ی راهنما از گزارش توسعه‌ی انسانی سال 2002 استخراج شده است. نرخ خطای رده بندی نادرست برای تحلیل ممیزی ($LDA$)، تحلیل ممیزی درجه‌ی دوم ($QDA$) و تحلیل ممیزی آمیخته ($MDA$) برای مجموعه‌ی راهنما در جدول شماره‌ی (1) آمده است. ممیزی با روش $LDA$ و $QDA$ با استفاده از نرم افزار $Splus/2000$ صورت گرفته است. برای $MDA$ با توجه به الگوریتم $EM$ داده شده در بخش چهارم نرم افزار $MDA$ به $Splus/2000$ اضافه گردیده تا برای ممیزی $MDA$ و داده‌های توسعه‌ی انسانی منطبق گردد. $MDA$ با سه زیر کلاس در هر کلاس برای مجموعه‌ی راهنما، کمترین خطای رده بندی نادرست را دارد.

**جدول** 1: نرخ خطای رده بندی نادرست برای مجوعه‌ی راهنما

| $LDA$ | $QDA$ | $MDA$ |
|---|---|---|
| 0/103 | 0/096 | 0/026 |

ماخذ: نتایج تحقیق

برای بررسی اعتبار مدل‌های ارائه شده، 58 مشاهده‌ی دیگر به عنوان مجموعه‌ی آزمون در نظر گرفته شده است نرخ خطای رده بندی نادرست برای مشاهدات مجموعه‌ی آزمون با استفاده از توابع ممیزی برآورد شده توسط مجموعه‌ی راهنما در جدول ( 2 ) آمده است. مطابق اطلاعات این جدول، نرخ خطای رده بندی نادرست برای مجموعه‌ی آزمون توسط تحلیل ممیزی آمیخته کمترین مقدار است. بنابراین، با توجه به جداول (1) و (2) برای رده بندی کشورها، تابع ممیزی $MDA$ مناسب است.

**جدول** 2: نرخ خطای رده بندی نادرست برای مجموعه‌ی آزمون

| $LDA$ | $QDA$ | $MDA$ |
|---|---|---|
| 0/103 | 0/121 | 0/052 |

ماخذ: نتایج تحقیق

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

### 4-2- رده بندی کشورها با استفاده از *MDA*

در گزارش توسعه‌ی انسانی 2002 آمار برای 173 کشور گزارش شده است. با استفاده از تابع ممیزی *MDA*، به‌دست آمده از مجموعه‌ی راهنما، 51 کشور دارای توسعه‌ی انسانی بالا، 82 کشور دارای توسعه‌ی انسانی متوسط و 40 کشور دارای توسعه‌ی انسانی ضعیف است.

#### 4-2-1- کشورهای با توسعه‌ی انسانی بالا

بر اساس رده بندی *MDA* کشورهای آرژانتین، آنتی گوا و باربودا، آلمان، اتریش، اروگوئه، اسپانیا، استرالیا، استونی، اسلواکی، اسلوونی، ایتالیا، امارات متحده عربی، انگلستان، ایالات متحده امریکا، ایرلند، ایسلند، باربادوس، باهاماس، بحرین، برونئی، بلژیک، پرتغال، ترنیداد و توباگو، جمهوری چک، دانمارک، ژاپن، سنگاپور، سوئد، سوئیس، سیشل، شیلی، فرانسه، اسرائیل، فنلاند، قبرس، قطر، کاستاریکا، کانادا، کرواسی، کره جنوبی، کویت، لوکزامبورگ، لهستان، ماریتیوس، مالت، مالزی، مجارستان، نروژ، نیوزلند، هلند، هنگ کنگ و یونان کشورهایی با توسعه‌ی انسانی بالا رده بندی می‌شوند.

#### 4-2-2- کشورهای با توسعه‌ی انسانی متوسط

بر اساس رده بندی *MDA*. کشورهای آذربایجان، آفریقای جنوبی، آلبانی، اردن، ارمنستان، ازبکستان، اکراین، اکوادور، اکوتوریال، الجزایر، السالوادور، اندونزی، ایران، برزیل، بلغارستان، بلیز، بولیوی، بوتسوانا، پاپوآ، پاراگوئه، پاناما، پرو، تاجیکستان، تایلند، ترکمنستان، ترکیه، تونس، جامائیکا، جزایر سلیمان، جمهوری دمنیکا، جمهوری دمنیکن، چین، روسیه، روسیه سفید، رومانی، زیمباوه، سآموا، سآوتومو، سری لانکا، سنت اویسا، سنت وینت، سوازیلند، سورینام، سوریه، عربستان، عمان، غنا، فلیپین، فیجی، قرقیزستان، قزاقستان، کاپ ورد، کامبوج، کلمبیا، کنگو، کنیا، کوبا، گرجستان، گرانادا، گواتمالا، گوانا جدید، گویان، مالدیو، مراکش، مصر، مغولستان، مقدونیه، مکزیک، ملداوی، میانمار، لاتویا، لیتوانی، لبنان، لسوتو، لیبی، لیتونی، نامبیا، نیکاراگوئه، ونزوئلا، ویتنام، هند و هندوراس کشورهایی با توسعه‌ی انسانی متوسط رده بندی می‌شوند.

### 4-2-3- کشورهای با توسعه‌ی انسانی ضعیف

بر اساس رده بندی *MDA*، کشورهای آفریقای مرکزی، آنگولا، اتیوپی، اریتره، اوگاندا، بنگلادش، بنین، بوتان و بورکینافاسو، بوروندی، پاکستان، تانزانیا، توگو، چاد، جمهوری مردمی لئو، جیبوتی، روآندا، زامبیا، سنگال، سودان، سیرالئون، کامرون، کنگو، کوته دویری، کوموروس، گابن، گامبیا، گینه، گینه بیسائو، ماداگاسکار، مالاوی، مالی، موریتانی، موزامبیک، نپال، نیجر، نیجریه، وانوتوآ، هائیتی و یمن جزء کشورهای با توسعه‌ی انسانی ضعیف به شمار می‌روند.

## 5- نتیجه گیری

به دلیل غیرنرمال بودن کلاس‌ها، روش‌های ممیزی متداول کلاسیک توان کافی در جدا سازی کلاس‌ها ندارند. در این حالت، می‌توان از یک روش پارامتری ممیزی استفاده کرد که با فرض توزیع نرمال آمیخته در کلاس‌ها، ممیزی کلاس‌ها امکان پذیر است. روش ممیزی آمیخته اگر چه منجر به محاسبات پیچیده و استفاده از روش‌های عددی به جای تحلیلی می‌گردد، ولی با توجه به دسترسی به رایانه‌های با سرعت بالا و نرم افزارهای قوی می‌توان از روش *MDA* استفاده کرد. نتایج به دست آمده در مثال توسعه‌ی انسانی و نتایج جداول (1) و (2) بیانگر کارآیی مفید و بهتر *MDA* بر *LDA* و *QDA* است. بدیهی است که برای رسیدن به یک نتیجه‌ی کلی‌تر لازم است که داده‌های بیشتری استفاده شود.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

## فهرست منابع:

پردل نوغابی، احمد رضا، آنالیز خوشه‌ای و مقیاس بندیهای چند بعـدی و اسـتفاده از آنهـا در تقسیم بندی کشورها، پایانه کارشناسی ارشد، دانشگاه شهید چمران اهواز، 1379.

Anderson, T.W., An Introduction to Multivariate Statistical Analysis, John Wiley, New York, 1984.

Coomans, D., Broeckeart, M.J. and Massart, D.L., "Comparison of Multivarite Discriminant Techniques for Clinical Data-Application to the Thyroid Functional State," Meth. Inform, Med 22, 1983, pp. 93-101.

Dempster, A.P. Laird, N.M. and Rubin, D.B., "Maximum Likelihood form in Complete Data via EM Algorithm (with discussion)," R. Statist. Sac. Vol. 39, No. 1, 1977, pp. 1-38.

Green, P.E., "Cluster Analysis in Test Market Selection," Management Science, Vol. 13, 1967, pp. 387-400.

Hastie, T. and Tibshirani, R., "Discriminant Analysis by Gaussian Mixtures," Journal of the Royal statistical society, series B, Vol. 58, 1996, pp. 155-176.

Knight, K., Mathematical Statistics, Chapman and Hall / CRC, London, 2000.

Mardia, K., Kent, J. and Bibby, J., Multivariate Analysis, Academic Press, London, 1979.

McLachlan, G.J., Discriminant Analysis and Statistical Pattern Recognition, John Wiley, New York, 1992.

Rao, C.R., "Statistical Infrence Applied to Classification Problems," Sankhya, Vol. 10, 1950, pp. 229-256.

Shumway, R.H., Applied Statistical Time Series Analysis, Prentice - Hall Publication, London, 1984.

Taxt, T., Hjot, N. and Eikvil, L., "Statistical Classification Using a Linear Mixtures of Multinormal Probability Densities," Pattern Recognition Letters, No. 12, 1991, pp. 731-737.

UNDP, Human Development Report 2002, New York, 2002.

Webb, A., Statistical Pattern Recognition, Arnold, London, 1999.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

# Human Development of Countries Using Mixtures Discriminant Analysis

Rahim Chinipardaz (Ph.D.), Ghasem Rekabdar (M.Sc.) and Reza Yousefi Hajeabad (M.Sc.) *

## Abstract:

Linear Discrimination Analysis (LDA) is based on normality assumption. When this condition is not the case, nonparametric methods have to be used. In this paper it is assumed that the classes are not normal, but each class includes subclasses with the normal distribution. This leads to mixture discrimination analysis, MDA, and EM algorithm needed for classification between classes. Then, the Human Development data in 2002 has been used and MDA software is adapted to classified countries into three categories. It has been shown that the proposed method has more performance than LDA method.

**JEL** Classification**:** ***O15***

**Keywords:** Classification, pattern recognition, clustering, EM algorithm; Human Development Index, linear discriminant analysis

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

* Professor of statistics, graduate students of applied statistics and economics, respectively - Shahid Chamran University, Ahvaz, Iran