

# الگوریتم جدیدی برای خوشه بندی داده ها با استفاده از تحلیل پوششی داده ها

حامد ژبانی رضایی، فهیمه شبانی\*

گروه ریاضی، دانشکده علوم، دانشگاه آزاد اسلامی، واحد مشهد

رسید مقاله: ۸۸/۶/۲۰

پذیرش مقاله: ۸۸/۱۱/۱۰

## چکیده

اخیراً روش جدیدی برای خوشه بندی داده ها، توسط گو و همکارانش در [13] ارائه شده است، که از تحلیل پوششی داده ها (DEA) برای خوشه بندی استفاده می کند. این روش، برای خوشه بندی داده هایی با مؤلفه های ورودی و خروجی، از توابع تولید قطعه قطعه خطی به دست آمده از مدل های DEA استفاده می کند. اما الگوریتمی که به این منظور پیشنهاد شده است، به دلیل وجود جواب های چندگانه برای مدل های DEA، یک مشکل اجرایی دارد. لذا در مقاله حاضر، مشکل الگوریتم مذکور تشریح و الگوریتم جدیدی جهت خوشه بندی صحیح یک گروه از داده ها با استفاده از DEA، ارائه خواهد شد.

**کلمات کلیدی:** تحلیل پوششی داده ها، خوشه بندی با استفاده از DEA، مدل های CCR و BCC، تابع تولید قطعه قطعه خطی.

## ۱ مقدمه

تحلیل خوشه ای، شاخه ای در تحلیل چند متغیره می باشد که به دسته بندی یک مجموعه داده در گروه هایی با ویژگی های مشابه می پردازد. وجه تمایز این روش با سایر روش های تجزیه چندمتغیره در این است که اولاً، تعداد خوشه ها از ابتدا معلوم نیست، ثانیاً، ارتباط بین خوشه ها به کمک روابط ریاضی توجیه و اندازه گیری می شود. در بسیاری از روش های خوشه بندی، فاصله (میزان شباهت) داده های متعلق به یک خوشه نسبت به فاصله آن ها با داده های دیگر خوشه ها، معیار تعیین خوشه در نظر گرفته می شود. در ساده ترین روش، هر داده می تواند با در نظر گرفتن میزان شباهت یا کمترین فاصله اش تا خوشه ها، دقیقاً در یکی از خوشه ها عضویت داشته باشد. در نهایت داده ها طوری بین خوشه ها تقسیم می شوند که اشتراک دودویی خوشه ها تهی بوده و اجتماع آن ها کل داده ها را در بر گیرد. به این نوع خوشه بندی تجزیه سخت نیز گفته می شود [12,6].

\*عهده دار مکاتبات

آدرس الکترونیکی: [sheybanif@yahoo.com](mailto:sheybanif@yahoo.com)

ثیانی و بکار، الگوریتم جدیدی برای خوش‌بندی داده‌ها استفاده از تحلیل پوششی داده‌ها

اما در نوعی دیگر از روش‌های خوشه‌بندی که به تجزیه نرم نیز معروفند، هر داده می‌تواند هم زمان عضو بیش از یک خوشه باشد بطوری که با توجه به میزان شباهتی که به هر کدام از آن‌ها دارد، درجه عضویتی به آن تعلق می‌گیرد. این درجه عضویت در روش‌های تجزیه‌ای سخت تنها شامل  $\{0,1\}$  می‌باشد، اما با استفاده از مفهوم مجموعه‌های فازی [15]، الگوریتم‌های خوشه‌بندی فازی می‌توانند داده‌ها را با درجات عضویتی در بازه  $[0,1]$  در بیش از یک خوشه دسته‌بندی کنند [16]. الگوریتم‌های خوشه‌بندی فازی موجود، روش‌های خوشه‌بندی با روابط فازی [8]، تابع هدف فازی [2] و یا شبکه‌های عصبی [11] را در بر می‌گیرند. نوع سوم، خوشه‌بندی احتمالی است که در آن امکان حضور هر داده در خوشه متناظر آن، می‌تواند به عنوان درجه عضویتش در آن خوشه باشد [5,1]. روش‌های مذکور همگی الگوریتم‌های تجزیه‌ای هستند، اما در نوع دیگری معروف به خوشه‌بندی سلسله‌مراتبی، الگوریتم بر اساس اجتماع خوشه‌هایی که دوبه دو نزدیکترین فاصله را به هم دارند عمل می‌کند و در نهایت، این خوشه‌ها را در یک نمودار درختی نمایش می‌دهند [7].

معیار خوشه‌بندی در روش‌های فوق، فاصله بین داده‌هاست که با تعریف متری به عنوان فاصله، میزان شباهت آن‌ها اندازه‌گیری می‌شود [14]. اما اگر داده‌ها به شکلی باشند که قسمتی از هر داده، مقادیر ورودی و قسمتی از آن، مقادیر خروجی باشد، به طوری که خروجی‌ها با بکارگیری ورودی‌ها تولید شوند، آن‌گاه می‌توان بوسیله رابطه‌ای که بین ورودی‌ها و خروجی‌ها وجود دارد، تابع تولیدی (گاهی فقط بطور ضمنی) تعریف کرد و آن را معیار خوشه‌بندی قرار داد. روش پیشنهاد شده در مرجع [13]، برای خوشه‌بندی چنین داده‌هایی از مفاهیم DEA استفاده می‌کند. تحلیل پوششی داده‌ها (DEA) یک روش غیر پارامتری برای ارزیابی کارایی نسبی واحدهای تصمیم‌گیرنده (DMUها) می‌باشد. به هر DMU مقادیری به عنوان ورودی و مقادیری به عنوان خروجی نسبت داده می‌شود و برای آن‌ها ناحیه‌ای که همه داده‌ها در آن امکان تولید داشته باشند، تعریف می‌شود. مرز این ناحیه رویه‌ای قطعه‌قطعه خطی است که معیار خوشه‌بندی داده‌ها در [13] می‌باشد. در این روش بعضی از داده‌ها تنها به یک خوشه و برخی به بیش از یک خوشه تعلق می‌گیرند اما این عضویت قطعی است (احتمالی یا فازی نیست) یعنی یک داده یا به خوشه‌ای تعلق دارد یا ندارد. در عین حال اشتراک خوشه‌ها نیز لزوماً تهی نیست، لذا از روش‌های تجزیه‌ای سخت نیز متمایز خواهد بود.

آنچه در این روش معیار عضویت داده‌ها قرار می‌گیرد ابرصفحه‌های سازنده رویه قطعه‌قطعه خطی مذکور می‌باشند که ضرایب معادله‌های آن‌ها معیار ارزیابی‌های بعدی قرار خواهد گرفت. لذا به دست آوردن معادله دقیق این ابرصفحه‌ها یا توابع تولید و سپس تعیین درست‌اعضای خوشه داده‌ها، در تحلیل‌هایی که از خوشه‌بندی به دست می‌آید بسیار مؤثر خواهد بود. در الگوریتمی که در مرجع [13] برای این روش ارائه شده است، از مدلی با جواب‌های چندگانه برای تعیین خوشه هر داده استفاده شده است. این مطلب، اجرای الگوریتم را دچار مشکل می‌کند و ممکن است نهایتاً نتیجه درستی حاصل نشود. لذا الگوریتم جدیدی در این مقاله ارائه می‌شود تا با روشی ساده تر و بدون مشکل مذکور، نتایج صحیحی برای خوشه‌بندی داده‌ها با استفاده از DEA به دست آید.

## ۲ تحلیل پوششی داده ها

تحلیل پوششی داده ها (DEA) با تکنیکی غیر پارامتری کارایی نسبی سیستم های با چند ورودی و چند خروجی را ارزیابی می کند. فرض کنید  $DMU_j$  ها،  $(j=1, \dots, n)$ ، مجموعه داده ها را تشکیل دهند، به طوری که برای هر  $j$  داشته باشیم:

$$DMU_j : (x_{1j}, \dots, x_{mj}, y_{1j}, \dots, y_{sj})$$

که در آن  $x_j = (x_{1j}, \dots, x_{mj})$  مقادیر ورودی و  $y_j = (y_{1j}, \dots, y_{sj})$  بردار مقادیر خروجی برای واحد  $j$ ام می باشند،  $X \geq 0, Y \geq 0$  و  $X \neq 0, Y \neq 0$  می باشد. اولین مدلی که برای ارزیابی کارایی نسبی ارایه شده است، مدل CCR می باشد که شکل خطی شده آن را می توان به صورت زیر نشان داد:

$$\begin{aligned} M1. \quad & \text{Max} \quad \text{Eff}_k = u^t y_k \\ & \text{s.t.} \quad u^t y_j - v^t x_j \leq 0, \quad j = 1, \dots, n, \\ & \quad \quad v^t x_k = 1, \\ & \quad \quad u \geq \varepsilon, \quad v \geq \varepsilon. \end{aligned}$$

که در آن  $u^t = (u_1, \dots, u_s)$  و  $v^t = (v_1, \dots, v_m)$ ،  $y_j = (y_{1j}, \dots, y_{sj})^t$ ،  $x_j = (x_{1j}, \dots, x_{mj})^t$  این مدل، به مدل مضربی CCR معروف است. معمولاً برای محاسبه کارایی نسبی  $DMU_k$ ، از مدل زیر که با استفاده از دوآل مدل  $M1$  به دست آمده است، استفاده می شود.

$$\begin{aligned} M2. \quad & \text{Min} \quad \text{Eff}_k = \theta_{k-\varepsilon} (\sum_{i=1}^m s_i^- + \sum_{r=1}^s s_r^+) \\ & \text{s.t.} \quad \sum_{j=1}^n x_{ij} \lambda_j = \theta_k x_{ik} - s_i^-, \quad i = 1, \dots, m, \\ & \quad \quad \sum_{j=1}^n y_{rj} \lambda_j = y_{rk} + s_r^+, \quad r = 1, \dots, s, \\ & \quad \quad \lambda_j \geq 0, \quad s_r^+ \geq 0, \quad s_i^- \geq 0, \quad j = 1, \dots, n. \end{aligned}$$

مدل  $M2$ ، کارایی نسبی  $DMU_k$  را در حالتی که خروجی ها بدتر نشوند و ورودی ها تا حد ممکن کاهش داده می شوند، به دست می آورد (یعنی با مدل ماهیت ورودی).<sup>۱</sup> این مدل را، مدل پوششی (E-دار) می نامند.<sup>۲</sup> اگر مقدار بهینه مدل  $M2$  یعنی  $\text{Eff}_k^*$  برابر با ۱ باشد،  $DMU_k$  پارتو-کارا گفته می شود و در غیراین صورت ناکارایی پارتو خواهد داشت. تعریف ساده ای برای ناحیه ای که  $DMU$ ها در آن مورد ارزیابی قرار می گیرند به صورت زیر ارایه می شود:

$$T = \{ (x, y) \mid y \text{ را تولید کند } x \}$$

که در آن  $x = (x_1, \dots, x_m)$ ،  $y = (y_1, \dots, y_s)$  و فرض می‌شود  $x \geq 0$ ،  $y \geq 0$  باشند، به طوری که لااقل یک مؤلفه  $x$  و لااقل یک مؤلفه  $y$  مثبت اکید باشد. به ناحیه فوق ناحیه امکان تولید (PPS) گفته می‌شود. این ناحیه برای مدل CCR به صورت زیر است:

$$T_c = \{(x, y) \mid x \geq \sum_{j=1}^n x_j \lambda_j, \quad y \leq \sum_{j=1}^n y_j \lambda_j, \quad \lambda_j \geq 0 \quad j = 1, \dots, n\}$$

با اضافه کردن قید  $\sum_{j=1}^n \lambda_j = 1$  به مجموعه  $T_c$ ، مجموعه  $T_v$  به صورت زیر به دست خواهد آمد:

$$T_v = \{(x, y) \mid x \geq \sum_{j=1}^n x_j \lambda_j, \quad y \leq \sum_{j=1}^n y_j \lambda_j, \quad \sum_{j=1}^n \lambda_j = 1, \quad \lambda_j \geq 0 \quad j = 1, \dots, n\}$$

اگر مدل CCR را با ناحیه  $T_v$  فرمول بندی کنیم مدل BCC به دست خواهد آمد. مدل مضربی BCC به صورت زیر می‌باشد:

$$\begin{aligned} M3. \quad & \text{Max} \quad \text{Eff}_k = u^t y_k - u_0 \\ & \text{s.t.} \quad u^t y_j - v^t x_j - u_0 \leq 0, \quad j = 1, \dots, n, \\ & \quad \quad v^t x_k = 1, \\ & \quad \quad u \geq \varepsilon, \quad v \geq \varepsilon, \quad u_0: \text{free}. \end{aligned}$$

فرض کنید  $DMU_k$  را با مدل M2 ارزیابی کرده باشیم. در این صورت مختصات تصویر آن به صورت زیر تعیین می‌شود:

$$\begin{aligned} \hat{x}_k &= \sum_{j=1}^n x_{ij} \lambda_j^* = \theta_k^* x_{ik} - s_i^*, \quad i = 1, \dots, m, \\ \hat{y}_k &= \sum_{j=1}^n y_{rj} \lambda_j^* = y_{rk} + s_r^{+*}, \quad r = 1, \dots, s. \end{aligned}$$

که  $\theta_k^*$ ،  $s_i^*$ ،  $s_r^{+*}$  و  $\lambda_j^*$  مقادیر بهینه به دست آمده از مدل M2 می‌باشند. برای به دست آوردن جواب‌های بهینه مدل پوششی (BCC) مدل‌های زیر را حل می‌کنند:

$$\begin{aligned} M4. \quad & \text{Min} \quad \theta_k \\ & \text{s.t.} \quad \sum_{j=1}^n x_{ij} \lambda_j \leq \theta_k x_{ik}, \quad i = 1, \dots, m, \\ & \quad \quad \sum_{j=1}^n y_{rj} \lambda_j \geq y_{rk}, \quad r = 1, \dots, s, \\ & \quad \quad \sum_{j=1}^n \lambda_j = 1, \\ & \quad \quad \lambda_j \geq 0, \quad j = 1, \dots, n. \end{aligned}$$

$$M5. \quad \text{Max } S = \left( \sum_{i=1}^m s_i^- + \sum_{r=1}^s s_r^+ \right)$$

$$s.t. \quad \sum_{j=1}^n x_{ij} \lambda_j = \theta_k^* x_{ik} - s_i^-, \quad i = 1, \dots, m,$$

$$\sum_{j=1}^n y_{rj} \lambda_j = y_{rk} + s_r^+, \quad r = 1, \dots, s,$$

$$\sum_{j=1}^n \lambda_j = 1,$$

$$s_r^+ \geq 0, \quad s_i^- \geq 0, \quad \lambda_j \geq 0, \quad j = 1, \dots, n.$$

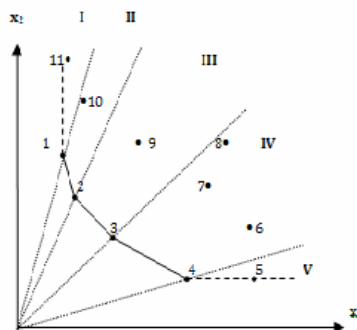
اگر در مدل M4 مقدار بهینه برابر با ۱ باشد،  $DMU_k$  کارا می باشد. اگر علاوه بر آن در مدل M5 مقدار بهینه برابر با صفر باشد،  $DMU_k$  کارای قوی خواهد بود. طبق قضیه ای ثابت می شود که تصویر یک  $DMU$  همیشه کارای قوی می باشد [3].

مرز کارایی مجموعه نقاطی است که در آن همه نقاط (اعم از حقیقی و مجازی) کارایی ۱ داشته باشند (کارایی ضعیف یا قوی). این مرز رویه ای قطعه قطعه خطی و محدب می باشد.

ثابت می شود که اگر  $DMU_k: (x_k, y_k)$  -BCC کارا باشد، ضرایب نرمال شده  $(v, u, u_0)$  از ابرصفحه تکیه کننده ی  $P: uy - vx - u_0 = 0$  بر PPS در  $(x_k, y_k)$ ، یک جواب بهینه مدل BCC می باشد و بالعکس.<sup>۲</sup>

### ۳ خوشه بندی با استفاده از DEA

در این قسمت به خلاصه ای از روش ارایه شده در [13] اشاره می شود. همان طور که بیان شد، ایده اصلی این روش استفاده از توابع تولید قطعه خطی به دست آمده از مدل های DEA برای خوشه بندی داده ها می باشد. شکل ۱ را برای داده های با دو ورودی و یک خروجی در نظر بگیرید. PPS در این حالت یک چندوجهی سه بعدی است که در شکل زیر آنرا با صفحه  $y=1$  قطع داده ایم.



شکل ۱: نمایش خوشه بندی روی ناحیه تولید شده با دو ورودی  $x_1, x_2$  و خروجی برابر با یک

طبق این روش خوشه بندی، همه  $DMU$  هایی که روی قطعه یکسانی تصویر می شوند، در یک خوشه قرار می گیرند. این قطعات باید قسمتی از مرز قوی باشند، لذا  $DMU$  هایی که بر مرز ضعیف تصویر می شوند، مجدداً دسته بندی شده و در نزدیکترین خوشه قرار می گیرند. در نتیجه برای داده های شکل فوق داریم:

خوشه II:  $\{11, 10, 2, 1\}$  خوشه III:  $\{9, 3, 2\}$  خوشه IV:  $\{8, 7, 6, 5, 4, 3\}$

به این ترتیب، در حالت کلی اگر معادله ابر صفحه های تعریف کننده PPS به صورت

$$u_{p1}y_1 + \dots + u_{ps}y_s + v_{p1}x_1 + \dots + v_{pm}x_m = 0$$

باشد، می توان وزن تک تک ورودی-خروجی ها را مورد ارزیابی قرار داد. (که در آن  $p$  تعداد ابر صفحه ها یا توابع تولید می باشد.)

$DMU$  هایی که روی مرز ضعیف تصویر می شوند، در بردار وزن های ورودی-خروجی شان مؤلفه صفر دارند. لذا در خوشه های تهگن قرار می گیرند. برای اینکه بتوانیم تحلیل های فوق را روی این واحدها هم انجام دهیم، آن ها را در نزدیک ترین خوشه با مرز قوی، ادغام می کنیم. ضمناً  $DMU$  هایی که روی مرز مشترک قطعات هستند (یا تصویر می شوند)، هم زمان در خوشه همه قطعات سازنده این مرز مشترک، دسته بندی می شوند.

الگوریتمی که برای اجرای این خوشه بندی ارایه شده است به صورت زیر می باشد:

### الگوریتم ۱

**مرحله اول** کارایی نسبی را برای هر  $DMU$  محاسبه کن، همه توابع تولید را پیدا کرده و سپس  $DMU$  هایی که کارایی نسبی آن ها نیاز به ارزیابی مجدد دارد مشخص گردد. به ترتیب زیر:

$$\text{فرض کن } p = 0, PF(p) = \phi, C(p) = \phi$$

$$\text{فرض کن } q = 0, R(q) = \phi$$

**حلقه:** برای  $k=1$  تا  $n$

با استفاده از مدل مضربی، کارایی نسبی  $Eff_k$  و جواب های بهینه  $v_i^*$ ،  $u_r^*$  و  $(i=1, \dots, m)$ ،  $(r=1, \dots, s)$  را برای  $k$  امین  $DMU$  به دست آور.

$u_r^*$  و  $v_i^*$  به دست آمده یکی از دو حالت زیر را خواهند داشت

**حالت اول**  $u_r^*$  و  $v_i^*$  هر دو غیر صفرند.

مرزهای تولید را بصورت زیر به دست آور:

$$f(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_s) = \sum_{r=1}^s u_r^* y_r - \sum_{i=1}^m v_i^* x_i = 0$$

اگر معادله فوق در یکی از  $PF(1), PF(2), \dots, PF(n)$  وجود دارد، مثلاً  $PF(h)$ ، آن گاه  $K$  امین  $DMU$  را در خوشه  $C(h)$  دسته بندی کن.

در غیراینصورت قرار بده  $p = p + 1$ .

$f(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_s)$  را به عنوان یک تابع تولید در  $PF(p)$  قرار بده و  $DMU_k$  را در  $C(p)$  دسته بندی کن.

**حالت دوم**  $u_r^*$  و  $v_i^*$  هر دو غیر صفر نیستند.

این یعنی  $DMU(x_{k1}, x_{k2}, \dots, x_{km}, y_{k1}, y_{k2}, \dots, y_{ks})$  با یکی از مرزهای ضعیف محصور شده است. بنابراین، کارایی نسبی آن باید مجدداً ارزیابی شود.

قرار بده  $q = q + 1$ .

$k$  امین  $DMU$  را در  $R(q)$  قرار بده.

### پایان حلقه

(هم اکنون توابع تولید در  $PF(1), PF(2), \dots, PF(p)$  با خوشه های  $C(1), C(2), \dots, C(p)$  قرار دارند و به تعداد  $q$  واحد از  $DMU$  ها که با مرزهای ضعیف محصور شده اند در  $R(1), R(2), \dots, R(q)$  می باشند.)  
**مرحله دوم** کارایی نسبی  $DMU$  هایی که با مرزهای ضعیف محصور شده اند را با حلقه زیر مجدداً ارزیابی کن.

**حلقه** برای  $j = 1$  تا  $q$

ورودی های  $R(j)$  امین  $DMU$  را در  $t$  ضرب کن و سپس داده  $(tx_{j1}, tx_{j2}, \dots, tx_{jm}, y_{j1}, y_{j2}, \dots, y_{js})$  را در  $R(j)$  جایگزین کن.

**حلقه:** برای  $w = 1$  تا  $p$

داده  $R(j)$  ام یعنی  $(tx_{R(j)1}, tx_{R(j)2}, \dots, tx_{R(j)m}, y_{R(j)1}, y_{R(j)2}, \dots, y_{R(j)s})$  را در  $PF(w)$  قرار بده. مقدار  $t$  را به دست آور به طوری که،

$$f(tx_{R(j)1}, tx_{R(j)2}, \dots, tx_{R(j)m}, y_{R(j)1}, y_{R(j)2}, \dots, y_{R(j)s}) = 0$$

قرار بده  $t(w) = t$ .

### پایان حلقه

اندیس  $k^*$  را به دست آور به طوری که:

$$t(k^*) = \max\{t(1), t(2), \dots, t(p)\}$$

کارایی نسبی  $DMU_{R(j)}$  را مجدداً ارزیابی کن تا برابر با  $t(k^*)$  شود.

$R(j)$  امین  $DMU$  را در خوشه  $C(k^*)$  دسته بندی کن.

### پایان حلقه

**مرحله سوم** خوشه های نهایی  $C(1), C(2), \dots, C(p)$  را به دست آور. علاوه بر این، کارایی نسبی  $Eff_1, Eff_2, \dots, Eff_n$  نیز به دست می آید.

الگوریتم فوق خوشه های کارآمد (با ضرایب غیر صفر) را انتخاب و خوشه های تبهگن را بطور هم زمان حذف می کند. همچنین  $DMU$  هایی که در خوشه های تبهگن قرار می گیرند را، در نزدیکترین خوشه غیر تبهگن مجدداً دسته بندی و کارایی نسبی آن ها را با مرز جدید محاسبه می کند.

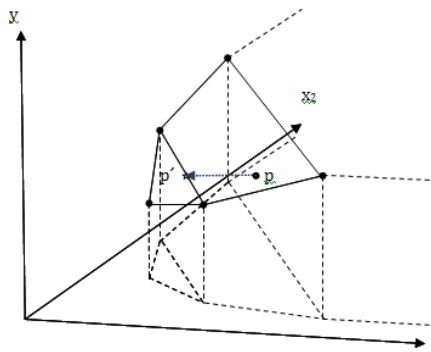
#### ۴ الگوریتم پیشنهادی

در این بخش ابتدا به بیان مشکل الگوریتم ۱ پرداخته و سپس الگوریتم دیگری بجای آن برای خوشه بندی داده ها با استفاده از DEA ارائه می شود. الگوریتم ۱ برای تشخیص خوشه یک DMU با  $u^*, v^*$  غیر صفر، مراحل زیر را دنبال می کند:

۱. DMU را با مدل مضربی ارزیابی کرده و مقادیر بهینه  $Eff^*, v^*, u^*$  را به دست می آورد.
  ۲. بوسیله این وزن های بهینه، تابع تولید این DMU را به صورت  $x^* - v^* y^* = 0$  تشکیل می دهد.
  ۳. اگر این تابع تولید در تکرارهای قبل به دست آمده بود این DMU را در خوشه متناظر آن تابع قرار می دهد. در غیراینصورت این تابع تولید، به توابع تولید به دست آمده از تکرارهای قبل اضافه می شود و خود، متناظر با خوشه جدیدی در نظر گرفته می شود. DMUی مورد ارزیابی هم، در این خوشه جدید قرار می گیرد.
  ۴. مراحل فوق تا زمانی که همه DMUها در خوشه های مناسب جای گیرند تکرار می شوند. لذا بطور هم زمان طی این تکرارها، DMUها خوشه بندی و توابع تولید قطعه قطعه خطی ساخته می شوند.
- همان طور که می دانیم مدل های پوششی و مضربی DEA به دلیل داشتن قيود همگن، تبهگنی شدید دارند، لذا ممکن است جواب های بهینه چند گانه داشته باشند. از آنجا که وزن های بهینه مدل مضربی، متناظر با ضرایب ابرصفحه تکیه کننده بر PPS در نقطه مورد ارزیابی می باشد [3]، جواب های چند گانه، متناظر با ابرصفحه های متعدد در آن نقطه خواهند بود. در نتیجه ممکن است بی نهایت ابرصفحه در ارزیابی یک DMU به دست آید که در صورت استفاده از این الگوریتم، هر کدام از آن ها می توانند خوشه این DMU در نظر گرفته شوند. در حالی که ما فقط ابرصفحه هایی را لازم داریم که سازنده PPS می باشند.
- اگر بخواهیم این الگوریتم را اجرا کنیم، در اولین مرحله این سؤال ایجاد می شود که چگونه می توان برای DMUای با این شرایط، همه جواب های چند گانه را به دست آورد؟! و با فرض به دست آوردن آن ها کدام وزن ها متناظر با ابرصفحه ای هستند که سازنده PPS باشد و بتواند قطعه ای از مرز را تولید کند؟
- این مطلب را روی  $T_v$  توضیح می دهیم. جواب های چند گانه برای مدل CCR هم مشکل ساز خواهد بود اما چون تشریح موضوع روی PPS مدل BCC،  $(T_v)$ ، روشن تر می باشد از آن استفاده شده است. شکل ۲ را در نظر بگیرید.

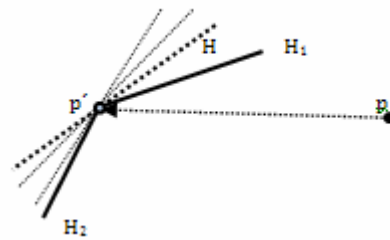
۱. در [13] از مدل CCR استفاده شده است.





شکل ۲: نمایش موقعیت  $DMU_p$  که در ارزیابی با مدل های مضربی جواب های چندگانه خواهد داشت.

در این شکل، نقطه  $p'$ ، تصویر  $DMU_p$  (یا نقطه  $p$ ) روی مرز قطعه قطعه خطی PPS می باشد. فرض کنید  $DMU_p$  را با مدل مضربی BCC ارزیابی کنیم و  $(u^*, v^*, u_0^*)$  جواب بهینه آن باشد. لذا، معادله ابرصفحه تکیه کننده بر PPS در  $p'$  به صورت  $u^*y - v^*x - u_0^* = 0$  خواهد بود. این ابرصفحه می تواند مثلاً ابرصفحه H در شکل ۳ باشد. در این شکل، هر یک از خطوط مماس بر  $p'$  نشان دهنده ابرصفحه ای تکیه کننده بر  $T_v$  در این نقطه می باشد.



شکل ۳: نمایش نقطه  $p'$  وقتی  $T_v$  را بر صفحه گذرنده از مبدأ و نقاط  $p, p'$  تصویر کنیم. این شکل مقطع شکل ۲ می باشد.

قطعاتی که  $DMU_p$  باید در خوشه متناظر آن ها دسته بندی شود فقط  $H_1, H_2$  هستند، در حالیکه جواب های چندگانه مدل مضربی می توانند متناظر با هر یک از خطوط مماس در  $p'$  باشند و این یعنی بی نهایت خوشه برای  $DMU_p$ !

روش هایی برای انتخاب  $(u^*, v^*, u_0^*)$  هایی که متناظر با ابرصفحه های سازنده ناحیه باشند (از بین جواب های چندگانه)، در مراجع [4] و [10] ارایه شده است. اما این دو روش از مدل های صفر و یک برای یافتن این ابرصفحه ها استفاده می کنند که غیرخطی هستند و ممکن است جواب های تقریبی داشته باشند و اصلاً ابرصفحه های سازنده را به دست ندهند. برای رهایی از مشکل جواب های چندگانه مدل مضربی، روش دیگری برای خوشه بندی داده ها در پیش می گیریم.

## الگوریتم ۲

۱. معادله همه توابع تولید را به دست آورده و هریک را متناظر با یک خوشه در نظر می‌گیریم. این توابع تولید (ابرفصله های کارای قوی سازنده PPS) را توسط روش ارایه شده در [9] به دست می‌آوریم.
  ۲. هر یک از  $DMU$  ها را با مدل  $M4$ ،  $M5$  ارزیابی می‌کنیم و سپس با توجه به جواب بهینه به دست آمده،  $DMU$  ها را به صورتی دسته بندی می‌کنیم که برای  $DMU_k$ ،  $k=1, \dots, n$  اگر  $S_k^* = (s_k^{-*}, s_k^{+*}) \neq 0$  آن گاه  $DMU_k$  را در مجموعه  $R$  و در غیر اینصورت آن را در مجموعه  $P$  قرار می‌دهیم.
  ۳. مختصات تصویر هر عضو مجموعه  $P$  را به دست آورده و به جای مقدار اولیه جایگزین می‌کنیم. لذا هم اکنون مجموعه  $P$ ، شامل مختصات تصاویر  $DMU$  هایی است که بر مرز قوی PPS تصویر شده اند.
  ۴. هر عضو  $P$  را در معادله هر یک از توابع تولید قرار می‌دهیم. در صورتی که در آن صدق کرد عضو خوشه متناظر با آن قطعه خواهد شد.
  ۵. مرحله دوم از الگوریتم ۱ را روی اعضای  $R$  اجرا می‌کنیم.
  ۶. خوشه های نهایی را مشخص می‌کنیم.
- با استفاده از این الگوریتم نقطه ای مانند  $p'$  در شکل ۲، چون فقط در معادله دو ابرفصله، که هر دو سازنده ناحیه هم هستند، صدق می‌کند، دقیقاً در خوشه همین دو قطعه قرار می‌گیرد.
- مختصات تصویر  $DMU_k$  با استفاده از مدل پوششی، به صورت زیر به دست می‌آید:

$$\hat{x}_k = \sum_{j=1}^n x_{ij} \lambda_j^* = \theta_k^* x_{ik} - s_i^{-*}, \quad i=1, \dots, m,$$

$$\hat{y}_k = \sum_{j=1}^n y_{rj} \lambda_j^* = y_{rk} + s_r^{+*}, \quad r=1, \dots, s.$$

می‌دانیم که روی ابرفصله های قوی،  $s_i^{-*} = s_r^{+*} = 0$  می‌باشد، لذا با وجود اینکه به دلیل وجود جواب های چندگانه برای مدل پوششی، این تصویر ممکن است از ترکیب  $(\lambda_1^*, \dots, \lambda_n^*)$  های متفاوت به دست آید، اما چون  $\theta_k^*$  مقدار ثابتی می‌باشد، سمت راست تساوی های فوق ثابتند. لذا، تصویر یک  $DMU$  بر مرز قوی، مختصات منحصر بفردی دارد. در نتیجه خوشه بندی آن ها بطور منحصر بفرد انجام خواهد شد.

اما اگر  $DMU_k$  بر مرز ضعیف تصویر شود، چون  $S_k^* = (s_k^{-*}, s_k^{+*}) \neq 0$  می‌باشد، لذا جواب های چندگانه مدل  $M5$ ، ممکن است تصاویر چندگانه به دست دهند. یعنی به ازای هر جواب بهینه،  $DMU_k$  بر یک قطعه قوی جدید از مرز تصویر شود، که این انتخاب خوشه مناسب را دشوار می‌کند. به این جهت، برای تعیین خوشه این  $DMU$  ها، از روشی که الگوریتم ۱ برای آن ها بکار برده است، استفاده شده است. اگر چه در این روش هم ممکن است  $t(k^*)$  منحصر بفرد نباشد، اما چون  $t(1), t(2), \dots, t(p)$  هر کدام متناظر با یک ابرفصله هستند و تعداد ابرفصله ها متناهی است، لذا تعداد خوشه های مناسب برای چنین  $DMU$  ای متناهی خواهد بود.

در حالی که، تعداد تصاویر چندگانه به دست آمده از جواب های چندگانه مدل ۱، ممکن است نامتناهی باشد و نتیجه درستی به دست ندهد.

برای به دست آوردن همه ابرصفحه های قوی سازنده PPS، از روش ارائه شده در [9] استفاده می کنیم. در الگوریتم ارائه شده در این مقاله، بدون استفاده از مدل مضربی و محاسبه مستقیم وزنهای ورودی-خروجی، معادله همه ابرصفحه های قوی سازنده PPS به دست می آید. از آنجا که خوشه های تبهگن در روش خوشه بندی ارائه شده حذف می شوند، نیازی به محاسبه ابرصفحه های ضعیف نداریم. الگوریتم یافتن معادله ابرصفحه های قوی سازنده PPS به صورت زیر خلاصه می شود:

### الگوریتم ۳

مرحله ۱ DMU های مشاهده شده را با شکل مناسب مدل های  $M4, M5$  ارزیابی کنید. اندیس DMU های کارای قوی را در مجموعه  $F$  قرار دهید. فرض کنید،  $|F|=1$ .

مرحله ۲ برای هر  $p, q \in F$  به طوریکه  $p \neq q$ ،  $DMU_k$  را به صورت زیر محاسبه کنید:

$$DMU_k = \frac{1}{2}DMU_p + \frac{1}{2}DMU_q$$

اگر  $DMU_k$  کاراست آن گاه  $p \in Fq, q \in F_p$ .

مرحله ۳ برای  $j = 1, \dots, n$  قرار دهید،  $\bar{F}_j = F - F_j$ .

مرحله ۴ یک زیر مجموعه  $m+s$  تایی از اعضای  $F$  در نظر بگیرید، به طوریکه هیچ یک از اعضای این مجموعه، متعلق به مجموعه  $\bar{F}$  بقیه اعضا نباشد (اگر ناحیه PPS مجموعه  $T_c$  باشد، یکی از این  $m+s$  عضو، مبدأ می باشد). این مجموعه را  $D = \{j_1, \dots, j_{m+s}\}$  بنامید.

برای به دست آوردن معادله یک ابرصفحه، دترمینان زیر را برابر با صفر قرار دهید.

$$\begin{vmatrix} x_1 - x_{1j_1} & \cdots & x_m - x_{mj_1} & y_1 - y_{1j_1} & \cdots & y_s - y_{sj_1} \\ x_{1j_2} - x_{1j_1} & \cdots & x_{mj_2} - x_{mj_1} & y_{1j_2} - y_{1j_1} & \cdots & y_{sj_2} - y_{sj_1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{1j_{m+s}} - x_{1j_1} & \cdots & x_{mj_{m+s}} - x_{mj_1} & y_{1j_{m+s}} - y_{1j_1} & \cdots & y_{sj_{m+s}} - y_{sj_1} \end{vmatrix}$$

که در آن  $x_1, \dots, x_m, y_1, \dots, y_s$  متغیرند و برای  $p=1, \dots, m, q=1, \dots, s, t=1, \dots, m+s$  ،  $x_{p_{jt}}$  ،  $p$  امین ورودی از  $DMU_j$  و  $y_{q_{jt}}$  ،  $q$  امین خروجی آن می باشد.  
فرض کنید معادله ابرصفحه به صورت  $P^t z + \alpha = 0$  به دست آید، که در آن  $z = (x_1, \dots, x_m, y_1, \dots, y_s)$  و  $\alpha$  یک اسکالر می باشد.

**مرحله ۵** اگر  $P$  مؤلفه ای کمتر یا مساوی صفر دارد برو به مرحله ۶، در غیراینصورت فرض کنید،  $W = (x_1^w, \dots, x_m^w, y_1^w, \dots, y_s^w)$  به صورت زیر تعریف شود:

$$x_j^w = \max_{j=1, \dots, m} \{x_{ij}\}, \quad i=1, \dots, m,$$

$$y_j^w = \min_{j=1, \dots, s} \{y_{ij}\}, \quad r=1, \dots, s.$$

$W$  را ایده آل منفی (NI) می نامیم. اگر

$$P^t z_j + \alpha = 0 \quad j \in D$$

$$P^t z_j + \alpha \leq 0 \quad j \in F-D$$

$$P^t w + \alpha < 0$$

آن گاه  $P^t z + \alpha = 0$  یک ابرصفحه قوی سازنده  $PPS$  می باشد.

**مرحله ۶** اگر زیرمجموعه  $m+s$  تایی دیگری از  $F$  می توان یافت برو به مرحله ۴، در غیراینصورت توقف کن.

برای نشان دادن نحوه استفاده از این الگوریتم یک مثال عددی ارائه می کنیم. روی این مثال مطالب زیر را شرح می دهیم:

الف. خوشه بندی با الگوریتم ۱

ب. خوشه بندی با الگوریتم ۲

ج. جواب های چندگانه و تصاویر چندگانه برای  $DMU$  هایی که بر مرزهای ضعیف تصویر می شوند.

## ۵ مثال عددی

**مثال ۱** جدول زیر مقادیر ورودی و خروجی را برای دوازده  $DMU$  نشان می دهد. موقعیت این داده ها در شکل ۴ نشان داده شده است.

DMU	A	B	C	D	E	F	G	H	I	J	K	P	Q
$x_1$	3	2	5	2	7	2	10	10	5	5	7	8	14
$x_2$	3	5	2	7	5	10	2	9	5	4	2	10	4
$y$	2	5	5	7	7	8	8	8	3	6	6	7	5

جدول ۱: مقادیر ورودی و خروجی برای دوازده DMU

### الف اجرای الگوریتم ۱

در این بخش، تنها به مشکل این الگوریتم می پردازیم. حل کامل مثال ۱ و خوشه بندی داده ها، در قسمت

(ب) توسط الگوریتم ۲، ارایه خواهد شد.

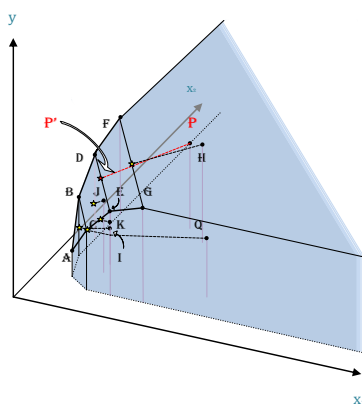
**مرحله اول** طبق الگوریتم ۱، در این مرحله باید کارایی نسبی را با مدل مضربی برای هر DMU محاسبه کرده و

توسط جواب های بهینه به دست آمده، همه توابع تولید را به صورت

$$f(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_s) = \sum_{r=1}^s u_r^* y_r - \sum_{i=1}^m v_i^* x_i - u_0^* = 0$$

به دست آوریم. برای DMU های مثال ۱ ضرایب و توابع تولید به دست آمده از ارزیابی آن ها توسط مدل

مضربی BCC، در جدول ۲ آورده شده است.



شکل ۴: نمایش DMU های مثال ۱ و تصویر آن ها بر مرز کارایی (ضعیف یا قوی).

DMU<sub>K</sub> و DMU<sub>Q</sub> بطور غیر شعاعی نیز کاهش می یابند.

با توجه به شکل ۴ و مقادیر جدول ۲، معادله ابر صفحه تکیه کننده بر:

$$y = 3.5x_1 + 3.5x_2 - 19 \quad : \text{قطعه ABC}$$

$$y = x_1 + x_2 - 2 \quad : \text{قطعه BCED}$$

$$y = 0.32x_1 + 0.32x_2 + 4 \quad : \text{قطعه DEGF}$$

می باشد، که متناظر با ابرصفحه های قوی تکیه کننده بر PPS می باشند و می توان آن ها را به ترتیب متناظر با خوشه های تبهگن I، II و III در نظر گرفت. اما - بدون در نظر گرفتن DMU هایی که در خوشه های تبهگن قرار می گیرند (یعنی K,Q) - با استفاده مرحله اول از الگوریتم ۱ خوشه های زیر تشکیل می شوند:

- I. {B,C}                                  IV. {I}                                  V. {P}
- II. {D,E,J}                                III. {F,G,H}

مشاهده می شود که با سه قطعه، پنج خوشه به دست آمده است! (جواب صحیح در خوشه بندی با الگوریتم ۱ ارایه می شود).

DMU	$v_1^*$	$v_2^*$	$u^*$	$u_0^*$	Eff*	تابع تولید
A	0.2	0.1	0	-1	1	$0.2x_1 + 0.1x_2 = 1$
B	0.14	0.14	0.04	-0.76	1	$y = 3.5x_1 + 3.5x_2 - 19$
C	0.14	0.14	0.04	-0.76	1	$y = 3.5x_1 + 3.5x_2 - 19$
D	0.11	0.11	0.11	-0.22	1	$y = x_1 + x_2 - 2$
E	0.11	0.11	0.11	-0.22	1	$y = x_1 + x_2 - 2$
F	0.08	0.08	0.25	1	1	$y = 0.32x_1 + 0.32x_2 + 4$
G	0.08	0.08	0.25	1	1	$y = 0.32x_1 + 0.32x_2 + 4$
H	0.05	0.05	0.16	0.63	0.63	$y = 0.31x_1 + 0.31x_2 + 3.93$
I	0.1	0.1	0.03	-0.53	0.63	$y = 3.33x_1 + 3.33x_2 - 17.66$
J	0.11	0.11	0.11	-0.22	0.88	$y = x_1 + x_2 - 2$
K	0.5	0	0	-1	1	$x_1 = 2$
P	0.56	0.56	0.05	-0.11	0.5	$y = 11.2x_1 + 11.2x_2 - 2.2$
Q	0.25	0	0	-0.5	0.5	$x_1 = 2$

**جدول ۲:** نتایج به دست آمده از ارزیابی DMU های مثال ۱ با فرم مضربی مدل BCC

## نکات جواب های به دست آمده در جدول ۲

۱. وزن های ارزیابی شده در جدول، از حل مدل مضربی با نرم افزار Lindo به دست آمده اند. این نرم افزار از بین جواب های چندگانه، همیشه (حتی با چندین بار ارزیابی یک  $DMU$ ) همین وزن ها را به دست می دهد. لذا هر  $DMU$  فقط به یک خوشه تعلق خواهد گرفت، در حالی که طبق این روش خوشه بندی،  $DMU$  هایی که روی مرز مشترک چند قطعه تصویر می شوند، به خوشه های متناظر با همه این قطعات تعلق می گیرد. از طرف دیگر، ممکن است نرم افزار دیگری وزن های متفاوتی به دست دهد که در اینصورت خوشه بندی منحصر بفرد نخواهد بود.

۲. با توجه به ضرابی که برای  $DMU_A$  به دست آمده است، با اینکه این  $DMU$  بر مرز کارای قوی هم قرار دارد، اما همه وزن های آن غیر صفر نیستند. یعنی آنرا در خوشه تبهگن دسته بندی می کند و باید مجدداً در مرحله دوم الگوریتم تعیین خوشه شود.

۳. معادلاتی که برای  $DMU_1$  و  $DMU_p$  به دست آمده است، سازنده PPS نمی باشند. اگر فرض کنیم سه ابرصفحه سازنده غیر تبهگن که در این مثال باید متناظر با خوشه ها در نظر گرفته شوند نیز معلوم باشند، با جوابی که برای این دو واحد به دست آمده است، نمی توان تعیین کرد این  $DMU$  ها به کدام یک از این سه قطعه و خوشه تعلق می گیرند. ابرصفحه هایی که برای این واحدها به دست آمده است، یکی از بینهایت ابرصفحه ایست که در این نقاط بر PPS تکیه می کند. اگر روند الگوریتم ۱ را ادامه دهیم باید هر کدام از این ابرصفحه ها را یک خوشه مجزا در نظر بگیریم (خوشه های IV, V). در اینصورت اگر فرض کنیم این بینهایت جواب قابل دست یابی باشند، به بینهایت شکل می توان خوشه بندی را انجام داد!

همه این مشکلات به دلیل چندگانه بودن جواب های مدل مضربی می باشد. اگرچه طبق الگوریتم ۱ باید توابع تولیدی را که سازنده PPS هستند به دست آوریم، اما اینکار با مدل های مضربی (آنطور که در الگوریتم ۱ گفته شده است) عملاً و به سادگی ممکن نیست.

## ب اجرای الگوریتم ۲

### مرحله ۱ یافتن همه ابرصفحه های قوی سازنده PPS

برای یافتن همه توابع تولیدی، طبق الگوریتم ۳، ابتدا باید  $DMU$  های کارای قوی را پیدا کنیم. نتایج حاصل از حل مدل های I و II برای  $DMU$  های داده شده در مثال ۱ در جدول زیر خلاصه می شود:

	$\theta^*$	$s_1^-$	$s_2^-$	$s^*$	$\hat{x}_1$	$\hat{x}_2$	$\hat{y}$
A	1	0	0	0	3	3	2
B	1	0	0	0	2	5	5
C	1	0	0	0	5	2	5
D	1	0	0	0	2	7	7
E	1	0	0	0	7	2	7
F	1	0	0	0	2	10	8
G	1	0	0	0	10	2	8
H	0.63	0	0	0	6.31	5.67	8
I	0.63	0	0	0	3.17	3.17	3
J	0.89	0	0	0	4.45	3.56	6
K	1	1	0	0	7	2	6
P	0.5	0	0	0	4	5	7
Q	0.5	2	0	0	5	2	5

جدول ۳: نتایج حاصل از ارزیابی DMUهای مثال ۱ توسط مدل پوششی دو مرحله‌ای BCC

با توجه به مقادیر جدول ۳، مجموعه  $F = \{A, B, C, D, E, F, G\}$  مجموعه DMUهای کارای قوی خواهد بود. در ادامه اجرای الگوریتم ۲، برای هر یک از اعضای  $F$ ، مجموعه  $F_j$  و  $\bar{F}_j$  را تعیین می‌کنیم. سپس، باید زیر مجموعه‌های  $m+s=3$  از  $F$  را با شرایط بیان شده در مرحله ۳ از الگوریتم ۲، انتخاب و ابرصفحه تشکیل دهنده آن‌ها را به دست آوریم. پس از محاسبات لازم، با انتخاب مجموعه‌های:

مجموعه اول:  $\{A, B, C\}$       مجموعه دوم:  $\{B, C, D\}$       مجموعه سوم:  $\{D, E, F\}$

ابرفصله‌های کارای قوی سازنده PPS، به ترتیب به صورت زیر به دست می‌آیند:

$$y=0.33x_1+0.33x_2+4 \quad \text{III} \quad y=x_1+x_2-2 \quad \text{II} \quad y=3x_1+3x_2-16 \quad \text{I}$$

مرحله ۲ با توجه به مقادیر جدول ۳، مجموعه‌های  $P$  و  $R$  به صورت زیر خواهند بود:

$$P=\{A, B, C, D, E, F, G, H, I, J, P\} \quad R=\{K, Q\}$$

مرحله ۳ مختصات تصاویر اعضای  $P$  را به دست آورده و بجای مقادیر اصلی در مجموعه  $P$  قرار می‌دهیم. این تصاویر در جدول ۳ به صورت  $(\hat{x}_1, \hat{x}_2, \hat{y})$  برای  $\hat{x}_1$  هر DMU نشان داده شده‌اند.

مرحله ۴ مختصات تصاویر به دست آمده را در معادله ابرصفحه‌های قوی یافت شده توسط الگوریتم ۲ قرار می‌دهیم. هر عضو  $P$  در معادله هر کدام از این توابع صدق کند، به خوشه متناظر با آن معادله ملحق می‌شود. پس از این جای‌گذاری، نتیجه زیر حاصل خواهد شد:

$$\begin{aligned} \text{خوشه I با تابع تولید} & \quad \{A, B, C, I\} : y=3x_1+3x_2-16 \\ \text{خوشه II با تابع تولید} & \quad \{B, C, D, E, J\} : y=x_1+x_2-2 \\ \text{خوشه III با تابع تولید} & \quad \{D, E, F, G, H, P\} : y=0.33x_1+0.33x_2+4 \end{aligned}$$



**مرحله ۵** خوشه های اعضای  $R$  را توسط مرحله دوم از الگوریتم ۱ تعیین می کنیم. به این منظور ورودی های  $DMU_Q$  و  $DMU_K$  را در پارامتر  $t$  ضرب می کنیم. سپس برای هر کدام بطور جداگانه، مقدار  $t$  را با قرار دادن مقادیر جدید با پارامتر  $t$ ، در معادله ابرصفحه ها، به دست می آوریم. ماکزیمم مقدار  $t$ ، تعیین کننده خوشه خواهد بود. به این ترتیب، خوشه های نهایی بصورت زیر خواهند بود:

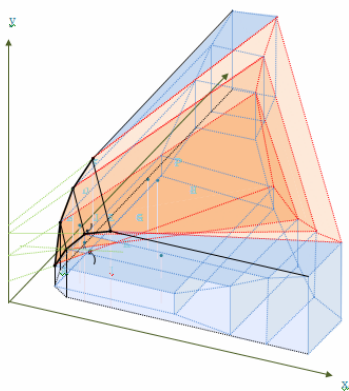
خوشه I :  $\{A,B,C,I,Q\}$

خوشه II :  $\{B,C,D,E,J,K,Q\}$

خوشه III :  $\{D,E,F,G,H,P\}$

که تابع تولید هر خوشه نیز مشخص می باشد.

نمایش هندسی این خوشه ها را می توان در شکل ۵ مشاهده کرد. در این شکل، چندوجهی های نارنجی رنگ خوشه های غیرتبهگن نهایی به دست آمده از روش خوشه بندی با DEA می باشند. نواحی چندوجهی، در طرفین خوشه های اصلی (غیرتبهگن) که با رنگ آبی مشخص شده اند، خوشه های تبهگن هستند که طبق این روش خوشه بندی حذف می شوند.



شکل ۵: نمایش سه بعدی خوشه های به دست آمده در ناحیه PPS

### ج مثالی برای تصاویر چندگانه

مجدداً مثال ۱ را در نظر بگیرید. فرض کنید  $DMU$ ها را با مدل  $M4$  ارزیابی و مقدار  $\theta^*$  را برای هر یک به دست آورده باشیم. همانطور که مشاهده شد،  $DMU_Q$  در این مرحله بر مرز ضعیف تصویر می شود (شکل ۴). بنابراین برای به دست آوردن مختصات تصویر آن بر مرز کارای قوی، لازم است مدل  $M5$  را نیز حل کنیم. مدل  $M5$ ، برای کاهش غیر شعاعی مقادیر به دست آمده از مدل  $M4$ ، به صورت زیر بیان شد:

مدل M5.

$$\begin{aligned} \max \quad & \left( \sum_{i=1}^m s_i^- + \sum_{r=1}^s s_r^+ \right) \\ \text{s.t.} \quad & \sum_{j=1}^n x_{ij} \lambda_j = \theta_k^* x_{ik} = s_i^-, \quad i=1, \dots, m, \\ & \sum_{j=1}^n y_{rj} \lambda_j = y_{rk} + s_r^+, \quad r=1, \dots, s, \\ & \sum_{j=1}^n \lambda_j = 1, \\ & s_r^+ \geq 0 \quad s_i^- \geq 0 \quad \lambda_j \geq 0, \quad j=1, \dots, n. \end{aligned}$$

که در آن  $\theta_k^*$  جواب بهینه به دست آمده از مدل M4 می‌باشد.

جوابی که در مثال ۱ برای DMU<sub>Q</sub>، با حل مدل فوق به دست آمده است عبارت است از:

$$(s_1^-, s_2^-, s^+) = (2, 0, 0)$$

جواب های شدنی زیر را در نظر بگیرید:

$$i. \quad (s_1^-, s_2^-, s^+) = (0, 0, 2)$$

$$ii. \quad (s_1^-, s_2^-, s^+) = (1, 0, 1)$$

$$iii. \quad (s_1^-, s_2^-, s^+) = (0.5, 0, 1.5)$$

هر یک از این جواب ها یک جواب بهینه برای مدل M5 هستند و با مقدار بهینه ۲ برای تابع هدف این مدل، DMU<sub>Q</sub> را بر مرز قوی تصویر می‌کنند.

در مورد اول، (i)، به جای کاهش متغیر کمکی ورودی، متغیر کمکی خروجی تا رسیدن به مرز قوی افزایش می‌یابد. همان طور که در شکل ۴ مشاهده می‌شود، تصویر به دست آمده از جواب مثال ۱ بر DMU<sub>C</sub> و تصویر حاصل از جواب (i) بر DMU<sub>E</sub> واقع می‌شود. به همین ترتیب هر نقطه ای روی پاره خط CE می‌تواند با ترکیبی از  $s_1^-$ ،  $s_2^-$ ،  $s^+$  به عنوان تصویر DMU<sub>Q</sub> در نظر گرفته شود. لذا مدل M5 می‌تواند بینهایت جواب و در نتیجه بینهایت مختصات تصویر برای یک DMU (که بر مرز ضعیف واقع شده است) به دست دهد.

آنچه مسلم است، یافتن همه این نقاط (یا همه جواب های چندگانه مدل M5) ممکن نیست. لذا، تضمینی وجود ندارد که حتماً نقاط C و E جزء جواب های به دست آمده باشند. در حالی که در این مثال فاصله تصویر Q بر مرز ضعیف از DMU<sub>E</sub> و DMU<sub>C</sub> به یک اندازه می‌باشد و Q می‌تواند در خوشه هر دو قرار گیرد.

بنابراین ممکن است خوشه بندی DMU هایی که در خوشه های تبهگن قرار می‌گیرند، به روش یافتن تصاویر آن ها، به درستی انجام نشود. لذا از روش پیشنهاد شده در مرحله دوم از الگوریتم ۱، که مربوط به این نوع DMU ها می‌باشد، استفاده می‌کنیم.

## ۵ نتیجه گیری

وجود جواب های چندگانه مدل های DEA همواره مسئله ساز و مورد توجه بوده است. یکی از مشکلاتی که جواب های چندگانه ایجاد می کنند منحصر بفرد نبودن جواب بهینه و نتایج وابسته به آن است. همانطور که ذکر شد، از آنجا که وزن های بهینه مدل مضربی، متناظر با ضرایب ابرصفحه تکیه کننده بر PPS در نقطه مورد ارزیابی می باشد، جواب های چندگانه، متناظر با ابرصفحه های متعدد در آن نقطه خواهند بود. در نتیجه ممکن است بی نهایت ابرصفحه در ارزیابی یک DMU به دست آید. در حالی که در روش خوشه بندی با DEA، چون اساس کار بر استفاده از توابع تولید یا همان ابرصفحه های سازنده PPS، برای خوشه بندی داده ها می باشد، به دست آوردن این توابع به طور دقیق و منحصر بفرد از اهمیت ویژه ای برخوردار می باشد. پس از به دست آوردن ابرصفحه های مورد نیاز، باید هر DMU طوری در این ابرصفحه ها دسته بندی شود که تا حد ممکن خوشه بندی یکتایی به دست آید. در الگوریتم پیشنهاد شده در این مقاله، سعی شده است که این شرایط تا حد ممکن لحاظ شود، به این ترتیب که اولاً معادله ابرصفحه های قوی سازنده PPS، با استفاده از روش پیشنهاد شده در مرجع [9] بطور دقیق (غیر تقریبی) و منحصر بفرد به دست می آیند. ثانیاً با توجه به اینکه برای تعیین تابع تولید متناظر با هر DMU، از تصویر آن واحد استفاده شده است، DMUهایی که بر مرز قوی PPS تصویر می شوند بطور کاملاً منحصر بفرد خوشه بندی می شوند. از طرفی چون برای خوشه بندی DMUهای تصویر شده بر مرزهای ضعیف، از روش پیشنهاد شده در الگوریتم ۱ (برای این نوع DMUها) استفاده شده است، تعداد حالت های ممکن برای خوشه بندی آن ها متناهی خواهد بود و کاربر می تواند با توجه به سایر شرایط مورد نظر خود، بهترین حالت را انتخاب کند. لذا با استفاده از این الگوریتم، خوشه بندی نهایی، با آنچه برای روش خوشه بندی با DEA تعریف شده است مطابقت خواهد داشت.

## منابع

- [1] Barni, M., Cappellini, V., Mecocci, A. (1996), *Comments on a possibilistic approach to clustering*. IEEE Transactions on Fuzzy Systems 4, 393-396.
- [2] Bezdek, J. C. (1981), *Pattern recognition with fuzzy objective function algorithms*. Plenum Press, New York.
- [3] Cooper, W.W., Li, S., Seiford, L.M., Tone, K. (2007), *Data Envelopment Analysis*. Springer.
- [4] Cooper, W.W., Ruiz, J.L., Sirvent, I. (2007), *Choosing weights from alternative optimal solution of dual multiplier models in DEA*. European Journal of Operational Research 180, 443-458.
- [5] Dubois, D., Prade, H., (1988), *Possibility Theory*. Plenum Press, New York, NY, USA.
- [6] Duda, R., Hart, P., (1973), *Pattern Classification and Scene Analysis*. J. Wiley & Sons, Inc. New York, NY, USA.
- [7] Hartigan, J. A., (1975), *Clustering Algorithms*. Wiley, New York.
- [8] Hu, B., (2004), *The fuzzy theoretical basis*. Wuhan University Press.
- [9] Jahanshahloo, G.R., Lotfi, F.H., Rezaei, H.Z., Balf, F.R., (2007), *Finding strong defining hyperplanes of production possibility set*. European Journal of Operational Research 177, 42-54.

- [10] Jahanshahloo, G.R., Lotfi, F.H., Zohrehbandian, M., (2005), *Finding the piecewise linear frontier production function in data envelopment analysis*. Applied Mathematics and Computation 163,483-488.
- [11] Kohonen, T., (1984), *Self-organization and associative memory*. Springer-Verlag, Berlin.
- [12] Krishnapuram, R. and Keller, J., (1996), *The possibilistic c-means algorithm: insights and recommendations*. IEEE Trans. Fuzzy Systems 4,385-393.
- [13] Po, R.W., Guh, Y.Y., Yang, M.S., (2009), *A new clustering approach using data envelopment analysis*. European Journal of Operational Research .
- [14] Wu, K.L., Yang, M.S., (2002), *Alternative c-means clustering algorithms*. Pattern Recognition 35, 2267-2278.
- [15] Zadeh, I. A., (1965), *Fuzzy sets*. Information Control, 8, 338-353.
- [16] Zhu, J., (1992), *Some key problems and processing methods of application fuzzy mathematic*. Fuzzy System and Mathematics, 11(2), 57-63.