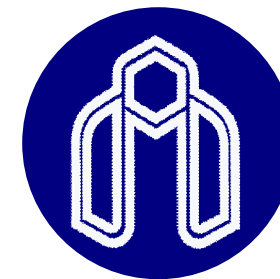


روشها و سیستمهای فازی

جلسه چهاردهم: خوشه بندی فازی

ارائه دهنده: مرتضی زاهدی

zahedi@ganjineh.co.ir

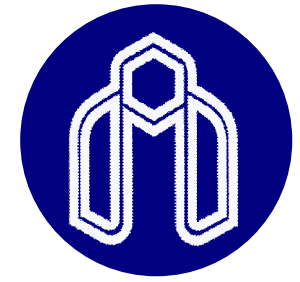


دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

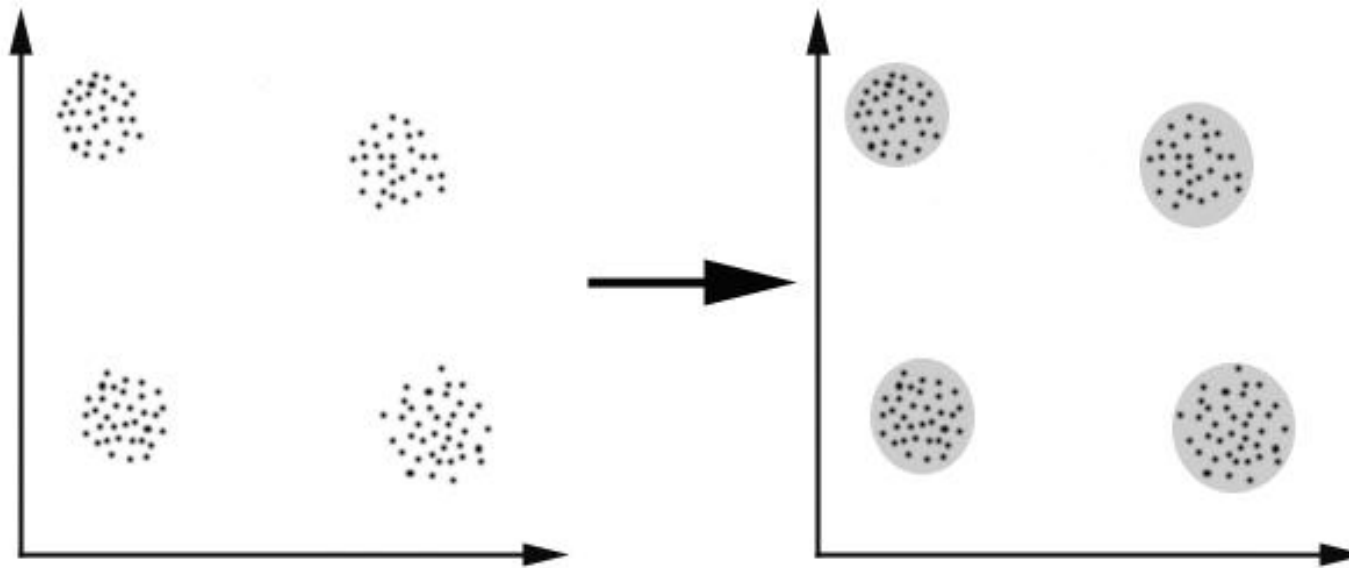
فهرست مطالب

- مفهوم خوشه بندی
- گامهای اساسی در خوشه بندی
- انواع الگوریتمهای خوشه بندی
- خوشه بندی فازی
- خوشه بندی میانگین فازی (FCM)
- خوشه بندی PCM
- خوشه بندی میانگین امکانی فازی (FPCM)
- خوشه بندی فازی ژنتیک (FGC)
- معیارهای کارایی در خوشه بندی
- کاربرد خوشه بندی در قطعه بندی تصاویر

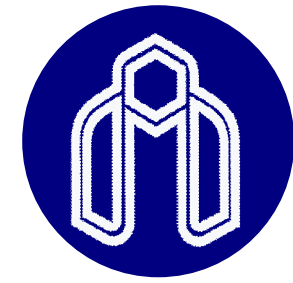
مفهوم خوشه بندی



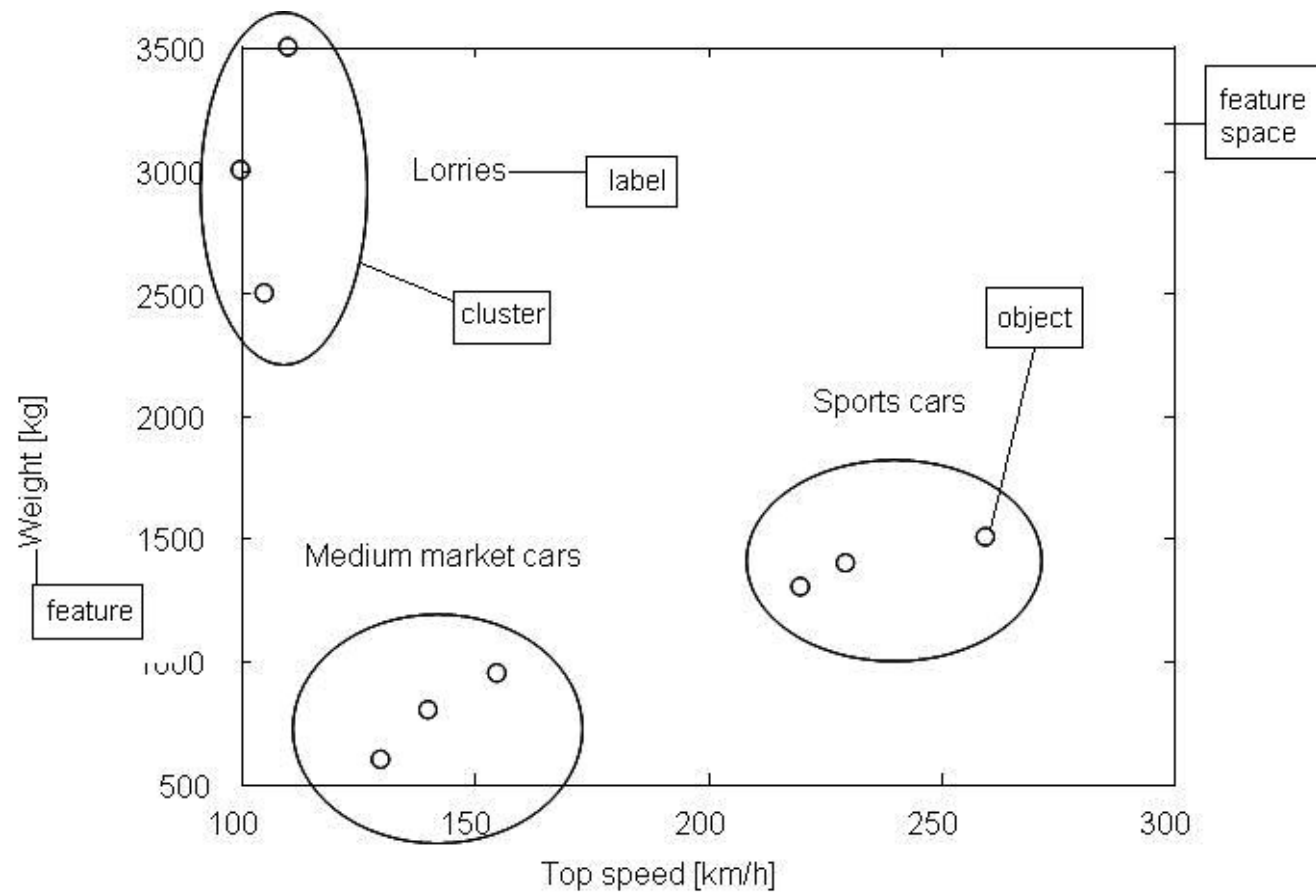
دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

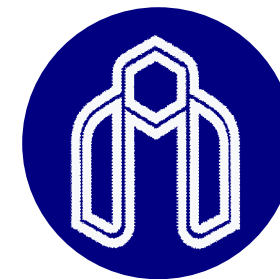


مفهوم خوشه بندی



دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی





دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

مفهوم خوشه بندی

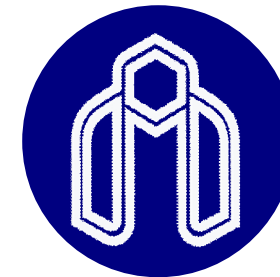
خوشه بندی ← → طبقه بندی

خوشه بندی:

– نمونه های ورودی دارای برچسب اولیه نمی باشند.

طبقه بندی:

– نمونه های ورودی برچسب گذاری شده اند.

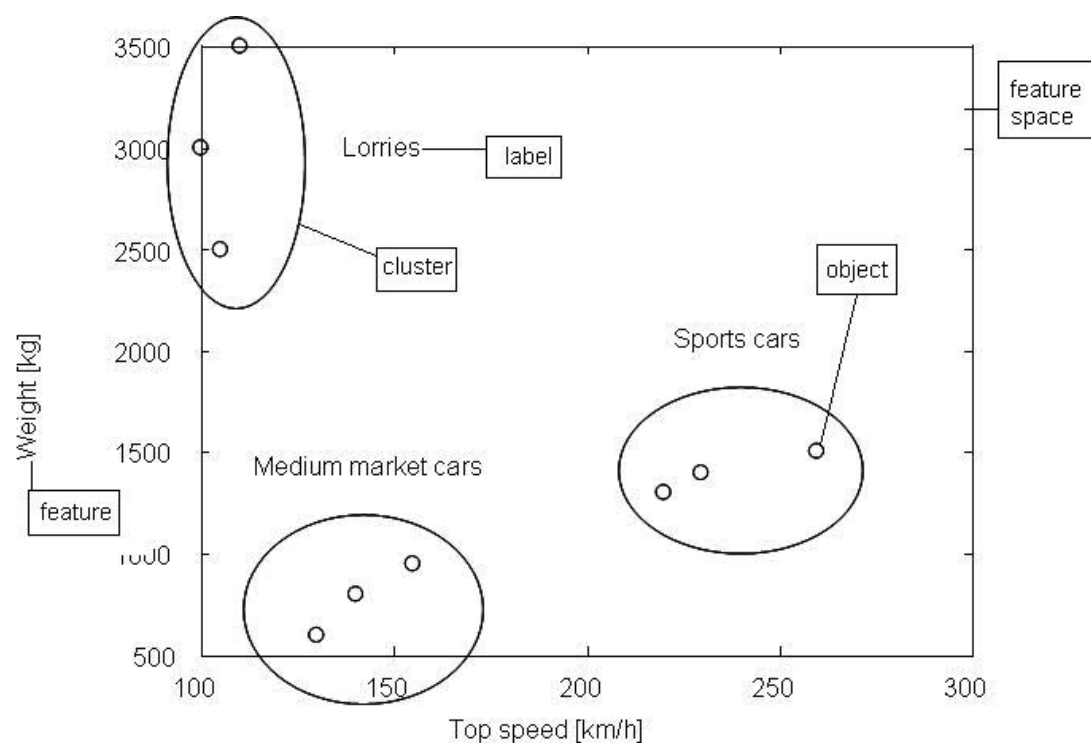


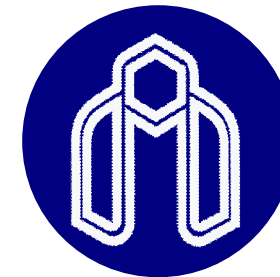
دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

مفهوم خوشه بندی

انواع خوشه ها:

- ۱- خوشه های به خوبی تفکیک شده
- ۲- خوشه های مبتنی بر مرکز
- ۳- خوشه های مبتنی بر مجاورت و نزدیکی

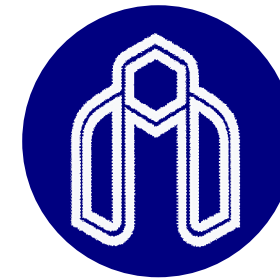




دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

گامهای اساسی در خوشه بندی

- خوشه بندی:
- انتخاب ویژگی
- مقیاس نزدیکی
- معیار دسته بندی
- الگوریتم خوشه بندی
- بررسی اعتبار نتایج

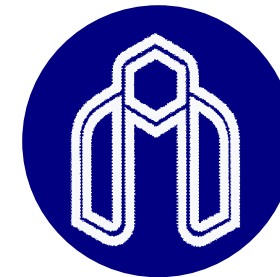


دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

انواع الگوریتمهای خوشه بندی

الگوریتمهای خوشه بندی به سه دسته اصلی زیر تقسیم می شوند:

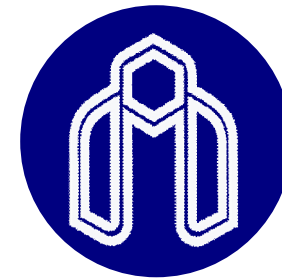
- الگوریتمهای خوشه بندی ترتیبی
- الگوریتمهای خوشه بندی سلسله مراتبی
- الگوریتمهای خوشه بندی مبتنی بر بهینه سازی تابع هزینه (تابع هدف)



دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

انواع الگوریتمهای خوشه بندی

- الگوریتمهای خوشه بندی ترتیبی
 - تعداد خوشه ها از قبل مشخص نیست
 - ایجاد خوشه های جدید در حین اجرای برنامه
 - وابستگی به ترتیب بردار ویژگیهای ورودی



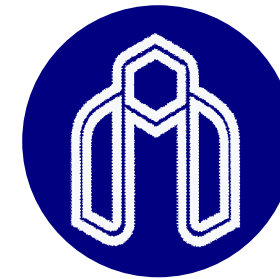
دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

انواع الگوریتمهای خوشه بندی

○ الگوریتمهای خوشه بندی سلسله مراتبی

۱. تخصیص هر نمونه به یک خوشه (N خوشه به ازای N نمونه)
۲. با یک معیار شباهت مثلاً فاصله اقلیدسی، دو خوشه های را که به هم نزدیکتر هستند ادغام میکنیم.
۳. فاصله خوشه جدید را با هر یک از خوشه های قدیمی محاسبه میکنیم.
۴. مراحل ۲ و ۳ را تکرار میکنیم تا در نهایت به تعداد خوشه تعیین شده برسیم.

خوشه بندی سلسله مراتبی یک نوع خوشه بندی متراکم شونده است.



دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

انواع الگوریتمهای خوشه بندی

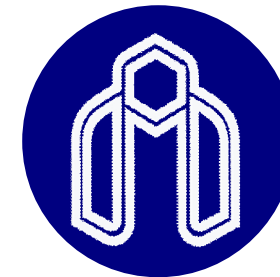
○ الگوریتمهای خوشه بندی مبتنی بر بهینه سازی تابع هزینه (تابع هدف)

○ الگوریتم خوشه بندی C میانگین

این الگوریتم تلاش میکند تا تابع هدف خود را که در اینجا تابع مربع خطا میباشد، مینیمم کند:

$$J = \sum_{j=1}^c \sum_{i=1}^n \|x_i^{(j)} - c_i\|^2$$

$\|x_i^{(j)} - c_i\|^2$ یک معیار فاصله بین نقاط داده $x_i^{(j)}$ و مرکز خوشه C_j میباشد که j مشخص کننده فاصله n نقطه داده از مرکز خوشه مربوطه است.

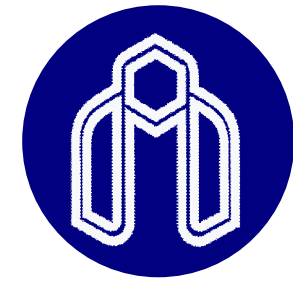


دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

انواع الگوریتمهای خوشه بندی

الگوریتم خوشه بندی C میانگین:

۱. مشخص کردن مراکز اولیه دسته ها
 ۲. تخصیص هر الگو به دسته های که نزدیکترین فاصله را با آن دسته دارد.
 ۳. وقتی که تمام الگوها تخصیص داده شدند، موقعیت C مرکز دوباره محاسبه میشود . (مراکز جدید به صورت میانگین الگوهای موجود در آن دسته محاسبه میشوند).
- تکرار مراحل ۲ و ۳ ادامه می یابد تا مراکز بدون تغییر باقی بمانند.

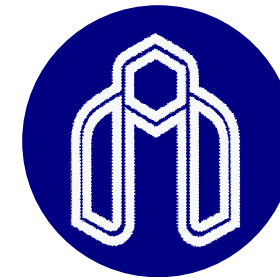


دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

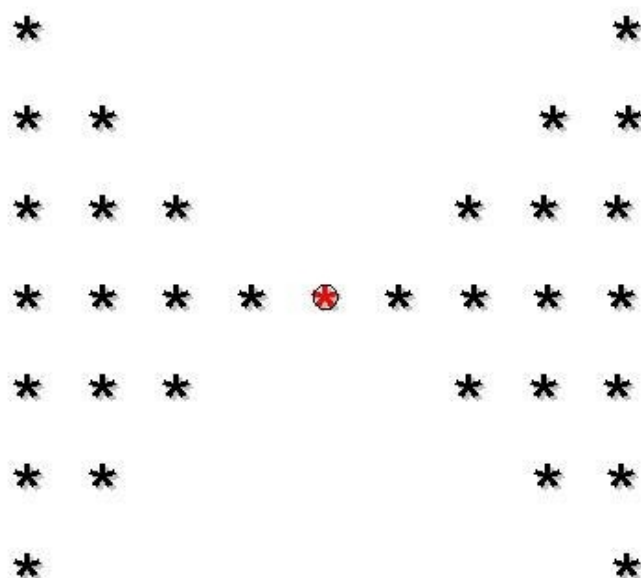
انواع الگوریتمهای خوشه بندی

تابع فاصله	فرمول
Euclidian distance	$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
Haming (city block) distance	$d(x, y) = \sum_{i=1}^n x_i - y_i $
Tchebyshev distance	$d(x, y) = \max_{i=1,2,\dots,n} x_i - y_i $
Minkowski distance	$d(x, y) = \sqrt[p]{\sum_{i=1}^n (x_i - y_i)^p}, p > 0$
Canberra distance	$d(x, y) = \sum_{i=1}^n \frac{ x_i - y_i }{x_i + y_i}, x_i \text{ and } y_i \text{ are positive}$
Angular separation	$d(x, y) = \frac{\sum_{i=1}^n x_i y_i}{\left[\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2 \right]^{1/2}}$

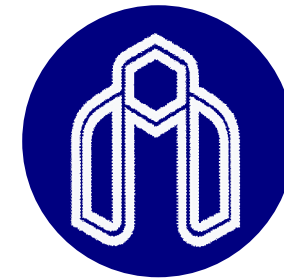
خوشه بندی فازی



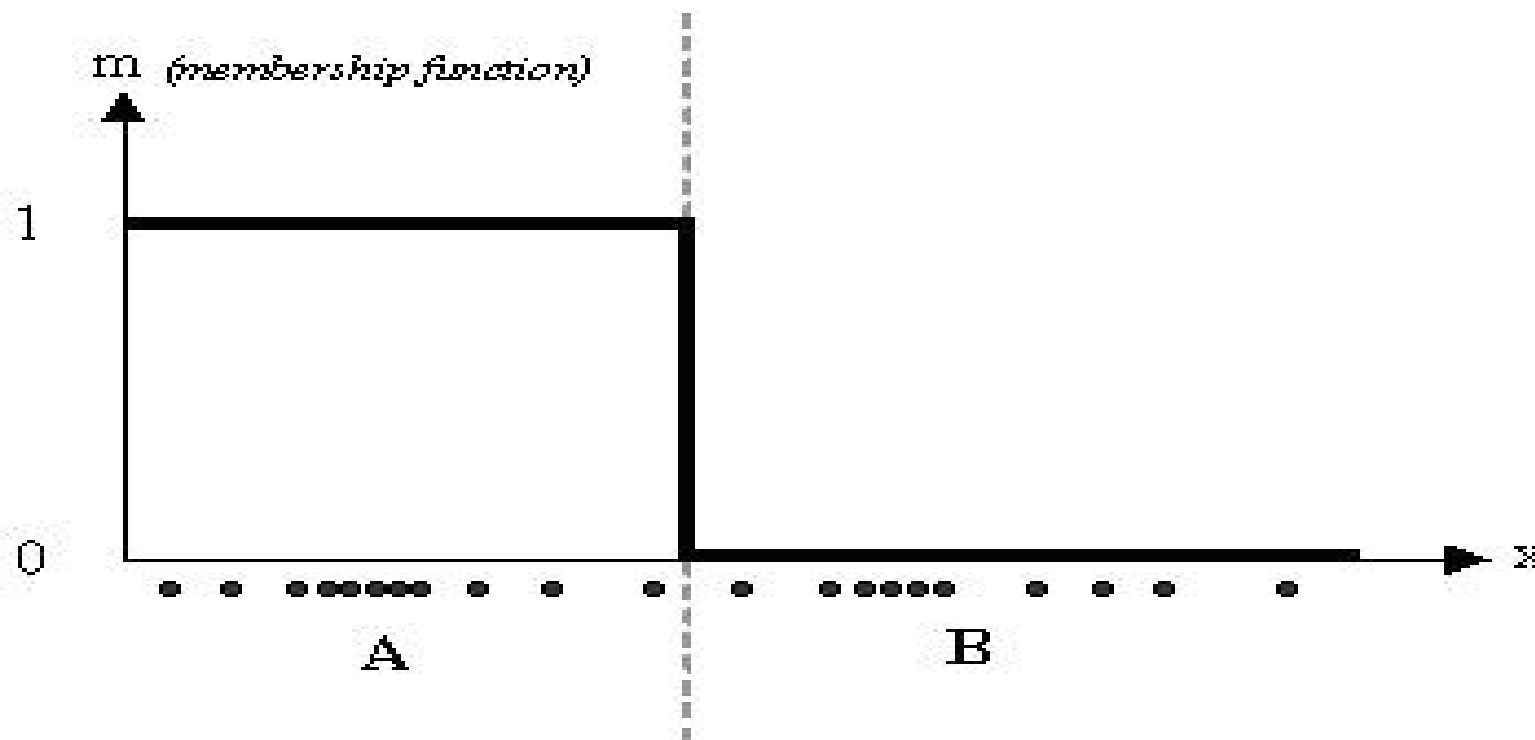
دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی



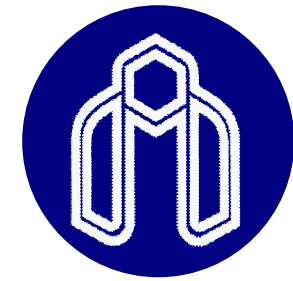
خوشه بندی فازی



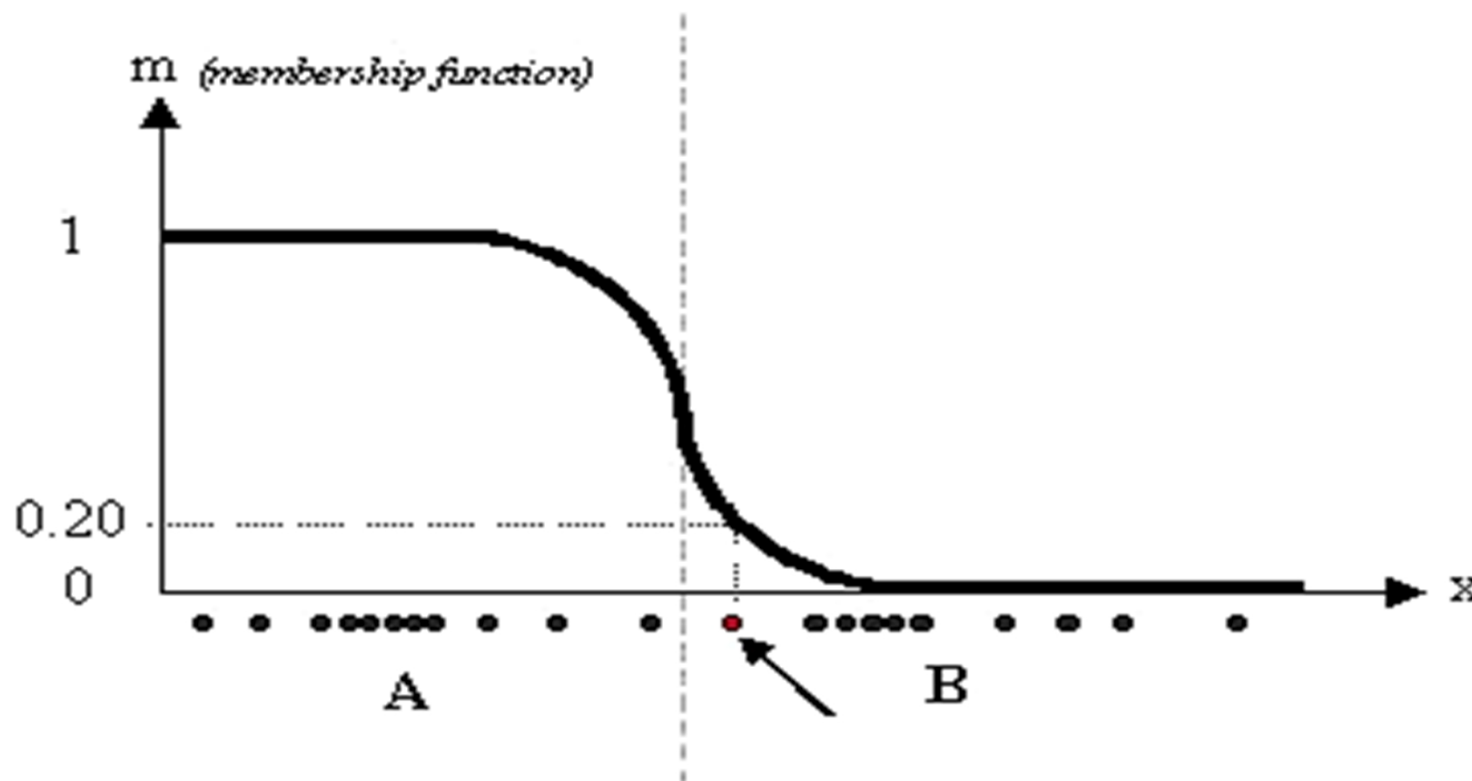
دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

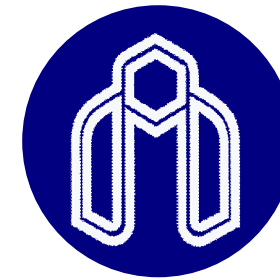


خوشه بندی فازی



دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی





دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

خوشه بندی FCM

$$J = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m d_{ik}^2 = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \|x_k - v_i\|^2$$

محدودیت‌های وضع شده:

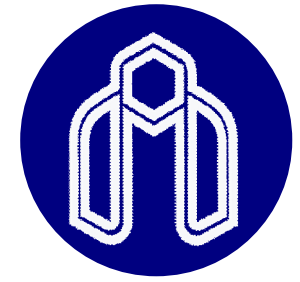
۱- هیچ خوشه‌ای نباید تهی باشد:

$$\left(\sum_{j=1}^n u_{ij} > 0 \quad \forall i \in \{1, \dots, c\} \right)$$

۲- محدودیت نرمالسازی

$$\sum_{i=1}^c u_{ij} = 1, \forall j = 1, \dots, n$$

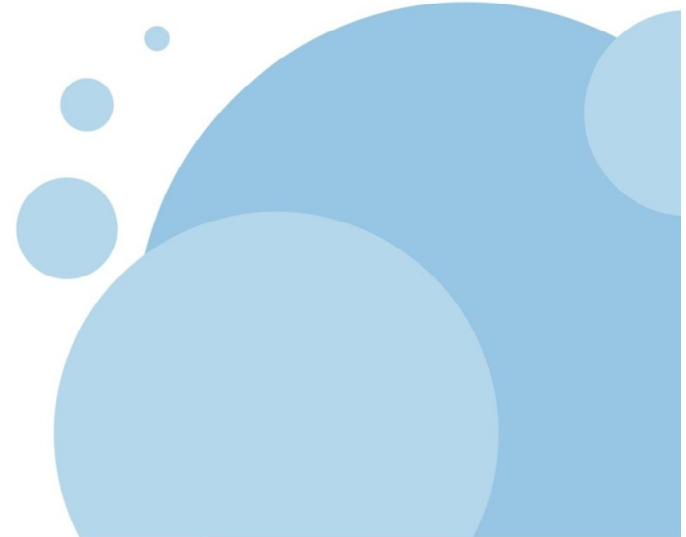
خوشه بندی FCM

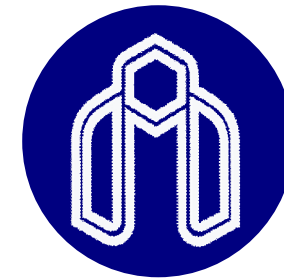


دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

$$J = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m d_{ik}^2 = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \|x_k - v_i\|^2$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}} \quad v_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m}$$





دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

خوشه بندی FCM

الگوریتم خوشه بندی c میانگین فازی:

۱. مقداردهی اولیه برای c (تعداد خوشه ها)، m (میزان فازی بودن) و مراکز خوشه ها

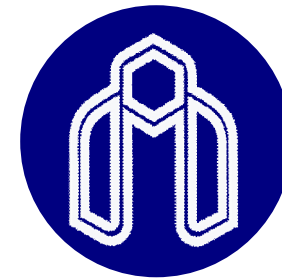
($V_1 \dots V_k$)، انتخاب مراکز خوشه ها میتواند به صورت تصادفی باشد)

۲. در نظر گرفتن $p=0$ ، که p اندیس تکرار است.

۳. افزودن یک مقدار به p ($p=p+1$)

۴. محاسبه تابع وابستگی فازی $u_k(x_i)$ در هر مرحله به صورت:

$$u_k^p(x_i) = \frac{\left(\frac{1}{d^2(x_i, V_k^p)} \right)^{\frac{2}{m-1}}}{\sum_{k=1}^K \left(\frac{1}{d^2(x_i, V_k^p)} \right)^{\frac{2}{m-1}}} \begin{cases} i = 1, 2, \dots, N \\ k = 1, 2, \dots, K \end{cases}$$



دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

خوشه بندی FCM

۵. محاسبه مراکز فازی جدید $(V_k, k = 1, 2, \dots, c)$:

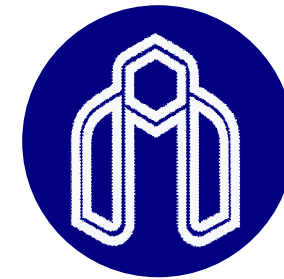
$$V_k^{p+1} = \frac{\sum_{i=1}^N \mu_k^p (x_i)^m x_i}{\sum_{i=1}^N \mu_k^p (x_i)^m}$$

۶. اگر $|u^p - u^{p+1}| < \varepsilon$ الگوریتم خاتمه می یابد. در غیر این صورت، به مرحله ۳ برو.

در الگوریتم فوق x بیانگر داده موجود، N تعداد داده ها، $u_k(x_i)$ تابع وابستگی فازی و

$d(x, y)$ بیانگر معیار فاصله تعریف شده (مثلاً معیار فاصله اقلیدسی) میباشد:

$$d(x, y) = \sum_{i=1}^N \sqrt{(x_i - y_i)^2}$$



دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

خوشه بندی FCM با نمونه های برچسب گذاری شده

فرض کنید که تعداد نمونه n باشد و M نمونه از این تعداد برچسبدار باشند.

$$J_{m,\alpha}(U,V;X) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2 + \alpha \sum_{i=1}^c \sum_{j=1}^n (u_{ij} - l_{ij} b_j)^m d_{ij}^2$$

$$u_{ij} = \frac{1}{1+\alpha} \left\{ \frac{1 + \left(1 - b_j \sum_{i=1}^c l_{ij} \right) + \alpha l_{ij} b_j}{\sum_{i=1}^c \frac{d_{ij}^2}{d_{ij}^2}} \right\}$$

در این نوع خوشه بندیها :

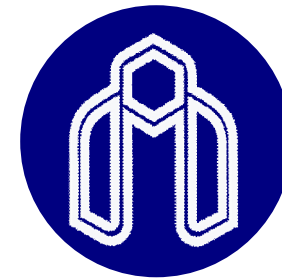
- اگر $M=n$ باشد، الگوریتم خوشه بندی را با سرپرست گویند.
- اگر $M < n$ باشد، الگوریتم خوشه بندی را با سرپرست جزئی گویند.
- اگر $M=0$ باشد، الگوریتم خوشه بندی را بدون سرپرست گویند.

خوشه بندی PCM



این روش، محدودیت نرمالسازی در FCM را به صورت $\max u_{ij} > 0 \quad \forall j$ ساده می کند، یعنی داده باید حداقل به یکی از خوشه ها تعلق داشته باشد. البته یک عبارت پینالیتی نیز به تابع هدف اضافه می شود:

$$J_f(U, V; X) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2 + \sum_{i=1}^c \eta_i \sum_{j=1}^n (1 - u_{ij})^m$$



دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

خوشه بندی PCM

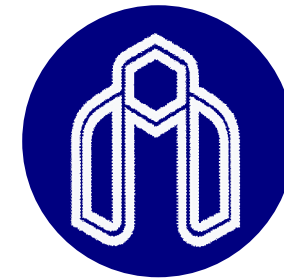
برای مشخص شدن شکل خوشه، تخمین مناسب برای مقدار η_i اهمیت بسیاری دارد:

$$\eta_i = \frac{\sum_{j=1}^n u_{ij}^m d_{ij}^2}{\sum_{j=1}^n u_{ij}^m}$$

رابطه فوق بیان مینماید که هر فاصله (d_{ij}) یک درجه عضویت به خوشه i دارد، پس برای تخمین η_i یک میانگین وزندار روی درجه عضویت داده ها میگیریم.

با توجه به اینکه روش PCM، شرط دوم روش FCM (محدودیت نرمالسازی) را حذف می کند، فرمول به روزرسانی درجه عضویت داده ها به صورت زیر تغییر خواهد کرد:

$$u_{ij} = \frac{1}{1 + \left(\frac{d_{ij}^2}{\eta_i} \right)^{1/(m-1)}}$$



دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

خوشه بندی FPCM

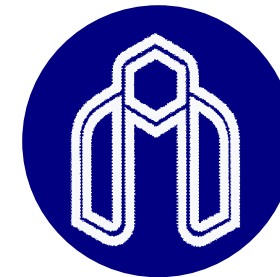
این روش از ترکیب دو روش خوشه بندی FCM و PCM حاصل شده است.
تابع هدف این نوع خوشه بندی توسط رابطه زیر تعریف می شود:

$$J = \sum_{i=1}^c \sum_{j=1}^n (u_{ij}^m + t_{ij}^{\eta}) d_{ij}^2 \quad \sum_{j=1}^n t_{ij} = 1$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}}$$

$$t_{ij} = \left[\sum_{k=1}^n \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{\eta-1}} \right]^{-1}$$

$$v_i = \frac{\sum_{k=1}^n (u_{ik}^m + t_{ik}^{\eta}) x_i}{\sum_{k=1}^n (u_{ik}^m + t_{ik}^{\eta})}$$

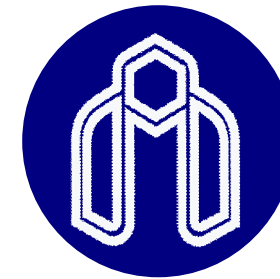


دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

خوشه بندی FPCM

الگوریتم خوشه بندی c میانگین فازی امکانی:

۱. تعیین مقادیر اولیه c ، m و v که $2 < c < n$ و $m > 1$ باشند. همچنین، ماتریسهای تصادفی U و T تولید شوند.
 ۲. برای مقادیر $c, 2, \dots, i=1$ بردار میانگین $\{v_i\}$ را با استفاده از رابطه گفته شده به دست آورید.
 ۳. به روزرسانی ماتریس U .
 ۴. به روزرسانی ماتریس T .
- در صورت عدم ارضای شرط خاتمه (نرسیدن بهبود در عدد J_m به یک مقدار آستانه) مراحل دوم تا چهارم تکرار شود.



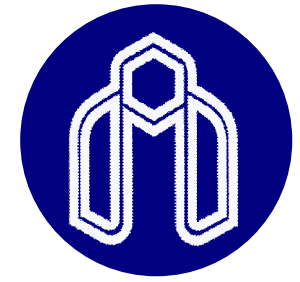
دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

معیارهای کارایی

دو شرط زیر باید در انتخاب تعداد خوشه‌ها مدنظر قرار بگیرد:

۱. تا حد امکان، نمونه‌های موجود در یک خوشه، شبیه به یکدیگر باشند.
۲. تا حد امکان، نمونه‌های متعلق به خوشه‌های متفاوت، نامتشابه با یکدیگر باشند.



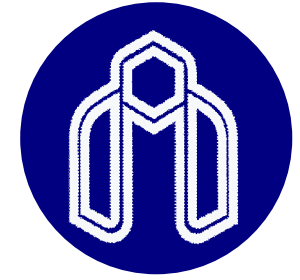


- تابع ارزیابی ضریب افراز:

$$v_{PC}(U) = \frac{1}{n} \left(\sum_{j=1}^n \sum_{i=1}^c u_{ij}^2 \right)$$

- تابع ارزیابی آنتروپی افراز:

$$v_{PE}(U) = -\frac{1}{n} \left(\sum_{j=1}^n \sum_{i=1}^c u_{ij} \log(u_{ij}) \right)$$

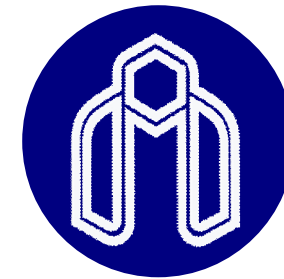


- تابع **Fukuyama and Sugeno**:

$$v(U; V; X) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \left(\|x_j - v_i\|^2 - \|v_i - \bar{v}\|^2 \right)$$

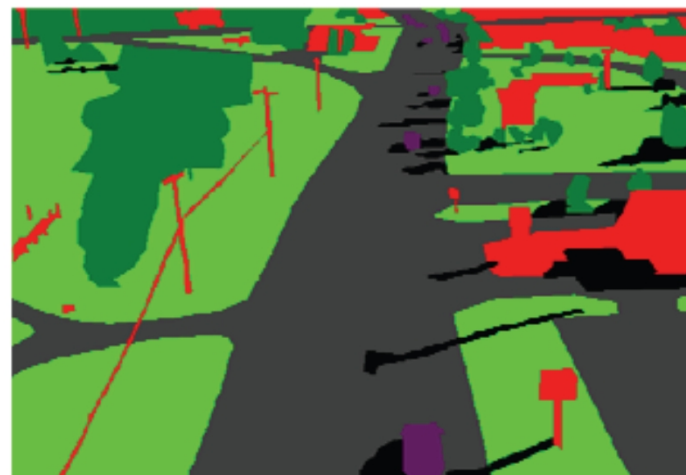
- تابع **Xie and Beni**:

$$v(U; V; X) = \frac{\sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - v_i\|^2}{n(\min\{v_i - v_j\})}$$



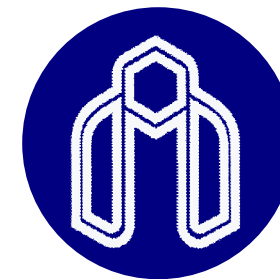
دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

کاربرد خوشه بندی فازی در قطعه بندی تصاویر



قطعه بندی تصویر:

- ۱- روشهای مبتنی بر هیستوگرام (Histogram-based)
- ۲- روشهای مبتنی بر انجام خوشه بندی (Clustering-Based)

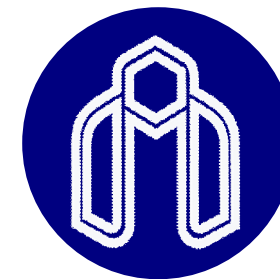


دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

کاربرد خوشه بندی فازی در قطعه بندی تصاویر

بلوک بندی اولیه تصویر جهت تحلیل آماری مقادیر پیکسلها برای قطعه بندی





دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

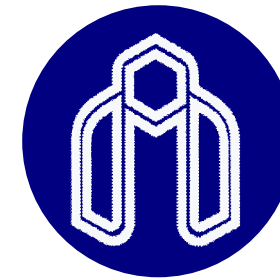
کاربرد خوشه بندی فازی در قطعه بندی تصاویر



(ب) قطعه بندی با استفاده از K-Means
بلوک بندی شده



(الف) قطعه بندی با استفاده از K-Means



دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

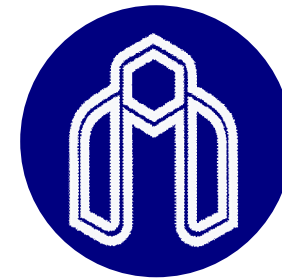
کاربرد خوشه بندی فازی در قطعه بندی تصاویر



(د) قطعه بندی با استفاده از Fuzzy C-Means
بلوک بندی شده



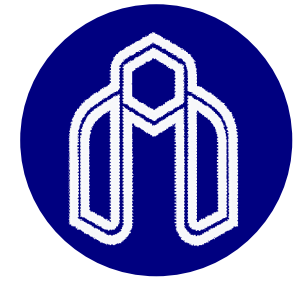
(ج) قطعه بندی با استفاده از Fuzzy C-Means



دانشگاه صنعتی شاهرود
مرکز آموزش های الکترونیکی

خلاصه مطالب

- مفهوم خوشه بندی
- گامهای اساسی در خوشه بندی
- انواع الگوریتمهای خوشه بندی
- خوشه بندی فازی
- خوشه بندی میانگین فازی (FCM)
- خوشه بندی PCM
- خوشه بندی میانگین امکانی فازی (FPCM)
- خوشه بندی فازی ژنتیک (FGC)
- معیارهای کارایی در خوشه بندی
- کاربرد خوشه بندی در قطعه بندی تصاویر



با تشکر از توجه شما

ارائه دهنده: مرتضی زاهدی

zahedi@ganjineh.co.ir