



دانشگاه تهران

دانشکده فنی-مهندسی کامپیوتر

دپارتمان الگوریتم ها و محاسبات

گزارش تمرین شماره ی پنج ، شش و هفت

طراحی الگوریتم های خوشه بندی

نیلوفر آقایی ابیانه

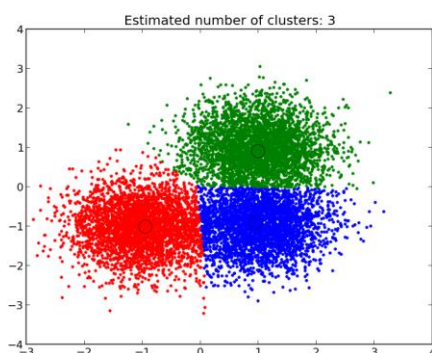
۸۱۰۸۹۰۰۰۱

چکیده

در این پروژه ، از روش های خوشه بندی^۱ برای طبقه بندی مجموعه ای از داده ها استفاده شده است ؛ و نتایج حاصل تحلیل می شوند. در اینجا سه الگوریتم خوشه بندی *کا تا میانگین*^۲، *سلسله مراتبی پایین با بالا*^۳ و *نگاشت خود سازمان دهی شده*^۴ شبیه سازی شده اند. در الگوریتم خوشه بندی نگاشت خود سازمان دهی شده از روش یادگیری البته به صورت کورکورانه^۵ است.

۱. مقدمه

در علم کامپیوتر، الگوریتم های مختلفی برای خوشه بندی وجود دارد. زمانی از خوشه بندی استفاده می شود که اطلاعاتی از داده ی اولیه وجود ندارد. با استفاده از الگوریتم های خوشه بندی توده ی داده ها مشخص می شود سپس با توجه به آنها اطلاعاتی استخراج می شود؛ بنابراین در خوشه بندی هدفی وجود ندارد. شکل ۱-۱ یک دسته از داده ها را در یک فضای دوبعدی نشان می دهد. در این شکل سه خوشه وجود دارد.



شکل ۱-۱ سه خوشه در فضای دوبعدی؛ هر رنگ بیانگر یک خوشه است. هر دایره در خوشه نماینده ی آن خوشه می باشد

الگوریتم های خوشه بندی متعدد از روش های مختلفی برای پیدا کردن خوشه ها^۶ استفاده می کنند. همه ی روش ها از معیار شباهت برای تشخیص نمونه هایی که در یک خوشه استفاده می کنند. در واقع این

^۱ clustering

^۲ kmeans

^۳ Hierarchical bottom-up

^۴ Self-Organizing Maps (SOM)

^۵ unsupervised

^۶ clusters

روش ها سعی بر این دارند نمونه هایی که شباهت بیشتری دارند یا در یک خوشه قرار دهند. خوشه بندی خوب است که بیشترین شباهت درون خوشه ای و کمترین شباهت برون خوشه ای داشته باشد.

یکی از معیار های رایج برای سنجش شباهت، فاصله نمونه ها است. فرمول های مختلفی برای اندازه گیری فاصله دو نمونه وجود دارد؛ به عنوان مثال می توان به فرمول های فاصله ی مینکاسکی^۱ و منهتن^۲ اشاره کرد که به ترتیب در رابطه ها ی ۱-۱ و ۱-۲ آمده اند.

$$d(i, j) = \sqrt[q]{|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \dots |x_{ip} - x_{jp}|^q} \quad (1-1)$$

$$d(i, j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{ip} - x_{jp}| \quad (1-2)$$

که $i = (x_{i1}, x_{i2}, \dots, x_{ip})$ و $j = (x_{j1}, x_{j2}, \dots, x_{jp})$ دو نمونه ی p -بعدی و q یک عدد صحیح است.

معیار فاصله که در سه روش خوشه بندی جلوتر ارائه می شود، فاصله اقلیدسی است. رابطه ی ۱-۳ فرمول این فاصله را نشان می دهد.

$$d(i, j) = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots |x_{ip} - x_{jp}|^2} \quad (1-3)$$

همانطور که اشاره شد، در خوشه بندی هدف وجود ندارد. لذا باید یک الگوریتم خوشه بندی چندین بار اجرا شود تا از بین آنها بهترین خوشه ها انتخاب شود. از طرفی یکی از مشکلات خوشه بندی این است که تعداد خوشه ها مشخص نیست. بنابراین الگوریتم های خوشه بندی باید برای تعداد خوشه های متعدد و به ازای هر کدام چندین بار اجرا شود تا از بین همه ی آنها بهترین انتخاب شود.

در ادامه سه الگوریتم خوشه بندی کا تا میانگین، سلسله مراتبی پایین با بالا و نگاشت خود سازمان دهی شده شبیه سازی شده اند. هر کدام از این الگوریتم ها را روی داده های ایریس^۳ و ست ایمیج^۴ اجرا می شوند.

^۱ Minkowski

^۲ Manhattan

^۳ iris

^۴ satimage

برای این کار الگوریتم ها را برای تعداد خوشه های ۳ تا ۸ و به ازای هر کدام از این خوشه ها ۵ با الگوریتم اجرا می شود.

۲. الگوریتم ها

I خوشه بندی سلسله مراتبی پایین با بالا

در این روش ابتدا هر نمونه به عنوان یک خوشه در نظر گرفته می شود. در مرحله ی بعدی فاصله اقلیدسی دوه دو ی خوشه ها حساب می شود. دو خوشه ای که فاصله ی کمتری نسبت به باقی فاصله ها دارند با هم ادغام می شوند و در یک خوشه نظر گرفته می شوند. سپس میانگین نمونه ها در این خوشه با عنوان نماینده ی آن خوشه در نظر گرفته می شود؛ و این روند مجدداً تکرار می شود. این فرایند تا زمانیکه به تعداد خوشه های مورد نظر برسد تکرار می شود.

II خوشه بندی کا تا میانگین

در این روش ابتدا به تعداد خوشه ها از بین نمونه به صورت تصادفی مرکز^۱ انتخاب می شود (هر یک از این مرکز ها به عنوان یک خوشه در نظر گرفته می شود). سپس فاصله ی هر نمونه تا مرکز ها محاسبه می شود؛ و هر نمونه به خوشه ی مرکزی که فاصله ی کمتری دارد تعلق پیدا می کند. زمانیکه این کار برای همه ی نمونه ها انجام شد، برای هر خوشه میانگین نمونه های قرار گرفته در آن خوشه به عنوان مرکز جدید آن خوشه محاسبه می شود. اگر مرکز ها تغییر نکنند آنگاه الگوریتم متوقف می شود و خوشه ها را تعیین کرده است؛ در غیر این صورت این فرایند برای مرکزهای جدید بدست می آید. در این الگوریتم خلاف الگوریتم خوشه بندی سلسله مراتبی ممکن است شرایطی ایجاد شود که الگوریتم هیچگاه متوقف نشود (در هر بار تکرار مرکز خت تغییر کنند). لذا در اینجا در شبیه سازی ای الگوریتم، برای اجتناب از این مشکل یک شرط تعداد تکرار می گذاریم که اگر پس از ۱۰۰۰ باز تکرار مرکز ها هم چنان تغییر می کنند متوقف شود. از معایب این روش می توان به موارد زیر اشاره کرد:

۱. به انتخاب اولیه مرکز ها بشدت وابسته است

۲. سر بار محاسباتی بالا دارد

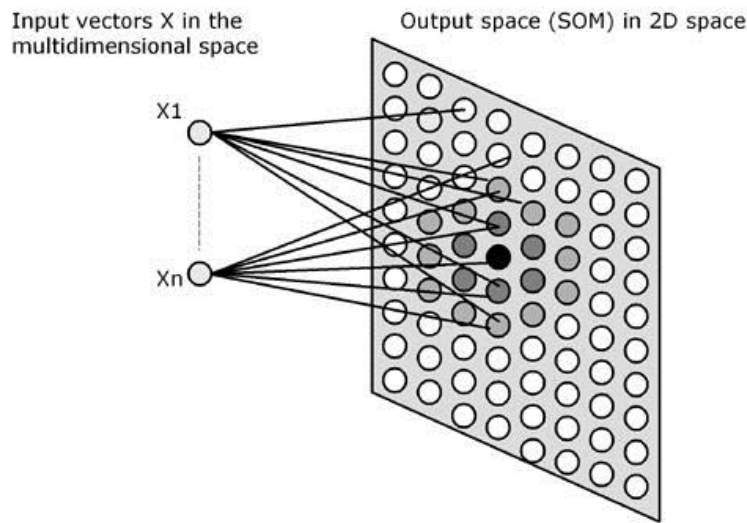
^۱ center

۳. کا بهینه باید پیدا شود

۴. در خیلی مواقع الگوریتم پایان پذیر نیست

III خوشه بندی نگاشت خود سازمان دهی شده

این روش اولین بار توسط مالسبورگ^۱ در سال ۱۹۷۳ ارائه شده و توسط کوهونن^۲ در سال ۱۹۸۲ تکمیل شد. این روش یک شبکه عصبی است که از آموزش رقابتی کورکورانه استفاده می کند و بر اساس مکانیزم کاری مغز انسان است. این شبکه عصبی از دو لایه کاملاً متصل تشکیل شده است: لایه ورودی، لایه خروجی. شکل ۱-۲ ساختار این شبکه عصبی را نشان می دهد.



شکل ۱-۲ ساختار شبکه عصبی نگاشت خود سازمان دهی شده

تعداد نرون های لایه ورودی برابر با ابعاد فضای مسئله (تعداد ویژگی ها) و تعداد نرون های لایه خروجی برابر با تعداد خوشه ها است. هر نرون لایه ورودی به همه ی نرون های لایه خروجی متصل است و این اتصالات دارای وزن می باشند که در طی فرایند آموزش به روز می شوند.

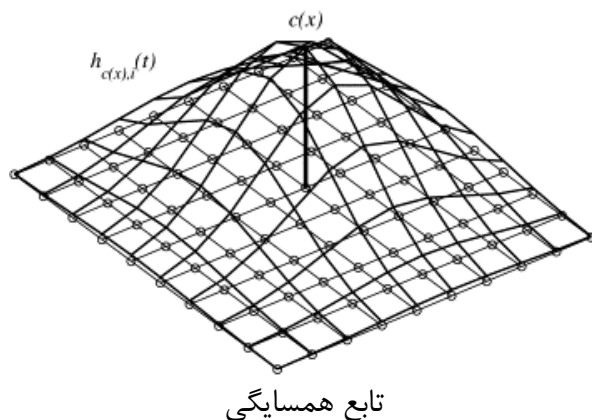
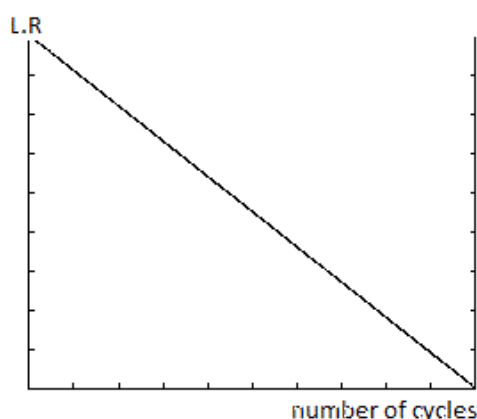
روش کار به این صورت است که یک نمونه به لایه ورودی اعمال می شود. سپس فاصله این نمونه با همه مرکز ها محاسبه می شود؛ در نهایت آن نمونه متعلق به خوشه ای می شود که فاصله ی کمتری دارد پس نرون متناظر در لایه خروجی برنده می شود. فقط وزن یال های متصل به آن نرون طبق قاعده ی یادگیری ۱-۴ به روز می شود،

^۱ C.von der Malsburg

^۲ Kohonen

$$w(n + 1) = w(n) + \eta(X - W) \quad (1-4)$$

که η نرخ آموزش (به صورت پویا)، W بردار ورودی و W بردار وزن های متصل به نرون برنده است. پس از اینکه وزن های نرون برنده به روز شد، وزن های نرون های همسایگی نرون برنده نیز به روز می شوند. در واقع یک تابع همسایگی وجود دارد که این تابع در گذر زمان به سمت صفر میل میکند. همچنین η نیز به سمت صفر میل کی کند. این فرایند تا زمانی که همه نمونه ها خوشه بندی شوند ادامه پیدا می کند (شکل ۱-۳).



شکل ۱-۳ نمودار نرخ آموزش و تابع همسایگی؛ شعاع تابع همسایگی در هر تکرار کم می شود.

۳. الگوریتم های ارزیابی خوشه بندی

همانطور که اشاره شد خوشه بندی خوب است که بیشترین شباهت درون خوشه ای و کمترین شباهت برون خوشه ای داشته باشد، کیفیت خوشه بندی به معیار شباهت استفاده شده در هر روش نیز بستگی دارد.

در اینجا دو روش برای ارزیابی الگوریتم های خوشه بندی ارائه می شود؛ ولی قبل از آن ابتدا چند علامت را توضیح می دهیم:

- n_c : تعداد خوشه ها
- d : تعداد بعد ها
- $d(x, y)$: فاصله ی بین دو نمونه
- \bar{X}_j : میانگین j امین بعد
- $\|X\|$: $\sqrt{X^T X}$ که یک بردار ستونی است.

- n_{ij} : تعداد عناصر در i امین خوشه و j بعد
- n_j : تعداد عناصر در j امین بعد در کل داده
- v_j : مرکز i امین خوشه
- c_i : i امین خوشه
- $\|c_i\|$: تعداد عناصر در خوشه i ام

I شاخص دویس - بولدین^۱

این شاخص براساس معیارهای شباهت خوشه ها (R_{ij}) که از روی معیار پراکندگی یک خوشه (S_i) محاسبه می شود و شبیه نبودن خوشه ها (d_{ij}) است. معیار شباهت خوشه ها می تواند به صورت دلخواه تعریف شود اما باید در شروط زیر صدق کند.

- $R_{ij} \geq 0$
- $R_{ij} = R_{ji}$
- *if $s_i = 0$ and $s_j = 0$ then $R_{ij} = 0$*
- *if $s_j > s_k$ and $d_{ij} = d_{ik}$ then $R_{ij} > R_{ik}$*
- *if $s_j = s_k$ and $d_{ij} < d_{ik}$ then $R_{ij} > R_{ik}$*

معمولا R_{ij} به صورت زیر تعریف می شود:

$$R_{ij} = \frac{s_i + s_j}{d_{ij}}$$

$$d_{ij} = d(v_i, v_j), \quad s_i = \frac{1}{\|c_i\|} \sum_{x \in c_i} d(x, v_i)$$

و بنابراین شاخص دویس - بولدینگ به صورت زیر تعریف می شود:

$$DB = \frac{1}{n_c} \sum_{i=1}^{n_c} R_i \text{ where } R_i = \max_{j=1 \dots n_c, i \neq j} (R_{ij}), \quad i = 1 \dots n_c$$

^۱ Davis Bouldin Index

این شاخص میانگین شباهت بین هر خوشه و خوشه ای که بیشترین شباهت را با آن دارد، اندازه گیری می کند. از آن جایی که خوشه ها باید فشرده و جدا از هم باشند، مقدار کمتر این شاخص نشان دهنده ی خوشه بندی بهتر است.

II شاخص های RS و RMSSTD

شاخص RMSSTD (ریشه-میانگین-مجذور انحراف معیار) معادل واریانس خوشه ها می باشد. در واقع این شاخص شباهت خوشه ها را اندازه گیری می کند. از آن جایی که هدف خوشه بندی شناسایی گروه های مشابه است، بنابراین مقدار کم این شاخص نشان دهنده ی روش خوشه بندی بهتر است.

$$RMSSTD = \sqrt{\frac{\sum_{j=1}^d \sum_{k=1}^{n_j} (x_k - \bar{x}_j)^2}{\sum_{j=1}^d (n_j - 1)}}$$

شاخص RS (مجذور R) میزان عدم شباهت خوشه ها را اندازه گیری می کند. در واقع درجه شباهت بین گروه های مختلف را اندازه گیری می کند. مقدار این شاخص بین صفر و یک است. صفر به معنای داشتن خوشه هایی است که فرقی با هم ندارند و یک نشان دهنده ی تفاوت زیاد بین خوشه ها می باشد.

$$RS = \frac{SS_t - SS_w}{SS_t}$$

$$\text{where } ss_t = \sum_{j=1}^d \sum_{k=1}^{n_j} (x_k - \bar{x}_j)^2, \quad ss_w = \sum_{j=1}^d \sum_{k=1}^{n_j} (x_k - \bar{x}_j)^2$$

۴. توابع در محیط متلب

MainHierarchicalClustering: تابع اصلی برای فراخوانی HierarchicalClustering

HierarchicalClustering: تابع شبیه سازی الگوریتم سلسله مراتبی پایین به بالا

MainKmeans: تابع اصلی برای فراخوانی kmeans

Kmeans: تابع شبیه سازی الگوریتم کا تا میانگین

MainSOM: تابع اصلی برای فراخوانی SOM

SOM: نابع شبیه سازی نگاشت خود سازمان دهی شده
Davies_Bouldin_Index: شاخص دویس - بولدینگ

RMSSTD_and_RS: شاخص های RS و RMSSTD

۵. آزمایش ها

در اجرای این الگوریتم ها از آنجا که هدفی وجود ندارد برای این که به خوشه بندی مناسبی رسید باید هر الگوریتم را چند بار اجرا کرد. از طرفی تعداد بهینه خوشه ها مشخص نیست و از این رو برای به دست آوردن تعداد خوشه مناسب باید الگوریتم را به ازای تعداد های متعددی از خوشه اجرا کرد.

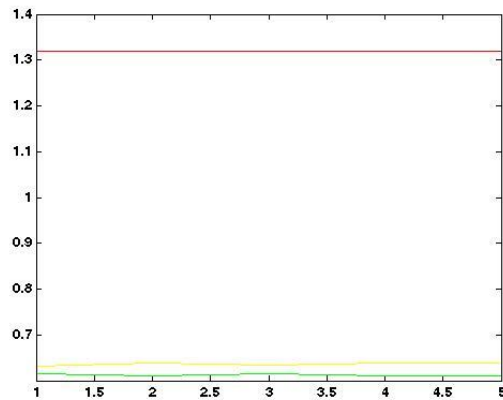
در اینجا هر کدام از الگوریتم های خوشه بندی روی مجموعه داده ای ایریس و ست ایمیج اجرا می شوند. برای مجموعه داده ای ایریس هر الگوریتم برای تعداد خوشه های ۲ تا ۵ اجرا می شود. از طرفی به ازای هر تعداد خوشه مشخص هر الگوریتم ۵ بار اجرا می شود. در نهایت با استفاده از معیار های ارزیابی که گفته شد در هر الگوریتم از بین همه نتایج بهترین انتخاب می شوند. برای مجموعه داده ای ست ایمیج هر الگوریتم برای تعداد خوشه های ۴ تا ۸ اجرا می شود. از طرفی به ازای هر تعداد خوشه مشخص هر الگوریتم ۵ بار اجرا می شود.

برای الگوریتم کاتا میانگین از آنجا که هم اشاره شد ممکن الگوریتم متوقف نشود به صورت دستی شرط پایان در صورتی که خودش متوقف نشود، پس از ۱۰۰۰ بار تکرار گذاشته شد. شرط توقف الگوریتم نگاشت خود سازمان دهی شده زمانی است که وزن ها پس از یک نسل تغییر نکنند.

با اجرای الگوریتم کاتا میانگین، روی مجموعه داده ای ایریس، نتایج جدول های ۱-۱ تا ۱-۴ و شکل های ۴-۱ تا ۷-۱، زیر حاصل شده است. با اجرای الگوریتم سلسله مراتبی پایین به بالا روی مجموعه داده ای ایریس نتایج جدول های ۵-۱ تا ۸-۱ و شکل های ۸-۱ تا ۱۱-۱، زیر حاصل شده است. با اجرای الگوریتم نگاشت خود سازمان دهی روی مجموعه داده ای ایریس نتایج جدول های ۹-۱ تا ۱۱-۱ و شکل های ۱۲-۱ تا ۱۵-۱، زیر حاصل شده است.

جدول ۱-۱
اجرای الگوریتم kmeans برای ایجاد ۲ خوشه
روی مجموعه داده ای iris

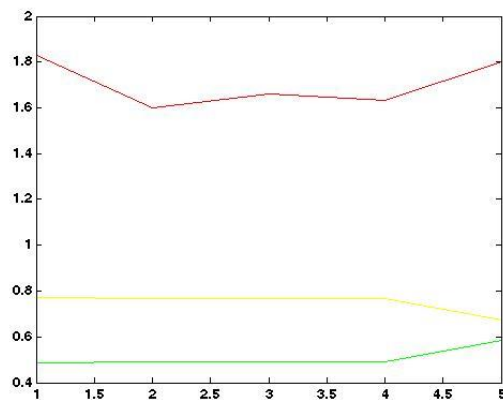
نسل	تعداد عناصر در هر خوشه	تعداد اجرا
۱۰۰۰	۵۰ ۱۰۰	۱
۱۰۰۰	۵۰ ۱۰۰	۲
۱۰۰۰	۵۰ ۱۰۰	۳
۱۰۰۰	۵۰ ۱۰۰	۴
۱۰۰۰	۵۰ ۱۰۰	۵



شکل ۱-۴
نمودار حاصل از توابع ارزیابی برای جدول ۱-۱
شاخص دیوس بلدین با رنگ قرمز
شاخص RMSSDT با رنگ سبز
شاخص RS با رنگ زرد

جدول ۲-۱
اجرای الگوریتم kmeans برای ایجاد ۳ خوشه
روی مجموعه داده ای iris

نسل	تعداد عناصر در هر خوشه	تعداد اجرا
۱۰۰۰	۴۶ ۵۵ ۴۹	۱
۱۰۰۰	۵۹ ۵۰ ۴۱	۲
۱۰۰۰	۴۷ ۵۰ ۵۳	۳
۱۰۰۰	۵۶ ۴۴ ۵۰	۴
۱۰۰۰	۳۷ ۱۰۰ ۱۳	۵



شکل ۱-۵
نمودار حاصل از توابع ارزیابی برای جدول ۲-۱
شاخص دیوس بلدین با رنگ قرمز
شاخص RMSSDT با رنگ سبز
شاخص RS با رنگ زرد

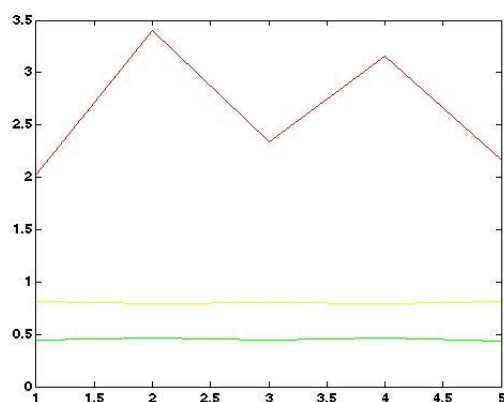
جدول ۳-۱

اجرای الگوریتم kmeans برای ایجاد ۴ خوشه
روی مجموعه داده ای iris

نسل	تعداد عناصر در هر خوشه	تعداد اجرا
۱۰۰۰	۴۱ ۳۵ ۱۵ ۵۹	۱
۲	۳۷ ۳۶ ۵۰ ۲۷	۲
۵	۲۸ ۲۰ ۴۹ ۵۳	۳
۲	۱۱ ۴۵ ۵۰ ۴۴	۴
۳	۲۷ ۳۰ ۴۴ ۴۹	۵

شکل ۶-۱

نمودار حاصل از توابع ارزیابی برای جدول ۳-۱
شاخص دیوس بلدین با رنگ قرمز
شاخص RMSSDT با رنگ سبز
شاخص RS با رنگ زرد



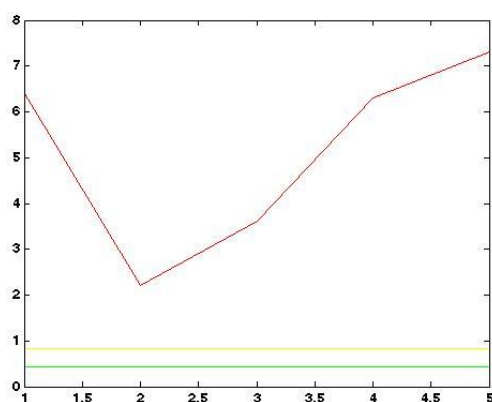
جدول ۴-۱

اجرای الگوریتم kmeans برای ایجاد ۵ خوشه
روی مجموعه داده ای iris

نسل	تعداد عناصر در هر خوشه	تعداد اجرا
۱۰۰۰	۳۸ ۱۸ ۴۹ ۲۳ ۲۲	۱
۱۰۰۰	۴۹ ۱۱ ۴۹ ۱۷ ۲۴	۲
۱۰۰۰	۴۹ ۳۱ ۲۶ ۹ ۳۵	۳
۱۰۰۰	۳۵ ۴۹ ۲۷ ۱۸ ۲۱	۴
۱۰۰۰	۱۶ ۱۹ ۴۹ ۳۳ ۳۳	۵

شکل ۷-۱

نمودار حاصل از توابع ارزیابی برای جدول ۴-۱
شاخص دیوس بلدین با رنگ قرمز
شاخص RMSSDT با رنگ سبز
شاخص RS با رنگ زرد



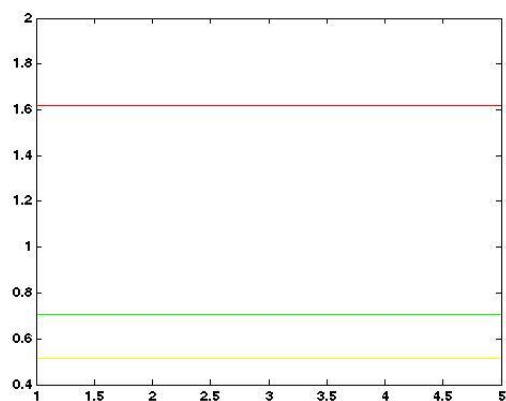
جدول ۵-۱

اجرای الگوریتم Hierarchical برای
ایجاد ۲ خوشه روی مجموعه داده ای iris

تعداد اجزا	تعداد عناصر در هر خوشه
۱	۷۹ ۷۱
۲	۷۹ ۷۱
۳	۷۹ ۷۱
۴	۷۹ ۷۱
۵	۷۹ ۷۱

شکل ۸-۱

نمودار حاصل از توابع ارزیابی برای جدول ۵-۱
شاخص دیوس بلدین با رنگ قرمز
شاخص RMSSDT با رنگ سبز
شاخص RS با رنگ زرد



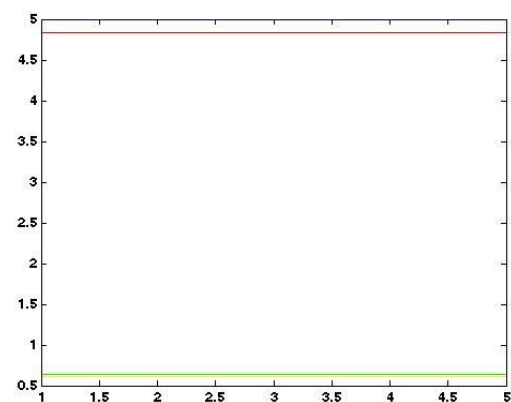
جدول ۶-۱

اجرای الگوریتم Hierarchical برای
ایجاد ۳ خوشه روی مجموعه داده ای iris

تعداد اجزا	تعداد عناصر در هر خوشه
۱	۱۳ ۷۹ ۵۸
۲	۱۳ ۷۹ ۵۸
۳	۱۳ ۷۹ ۵۸
۴	۱۳ ۷۹ ۵۸
۵	۱۳ ۷۹ ۵۸

شکل ۹-۱

نمودار حاصل از توابع ارزیابی برای جدول ۶-۱
شاخص دیوس بلدین با رنگ قرمز
شاخص RMSSDT با رنگ سبز
شاخص RS با رنگ زرد



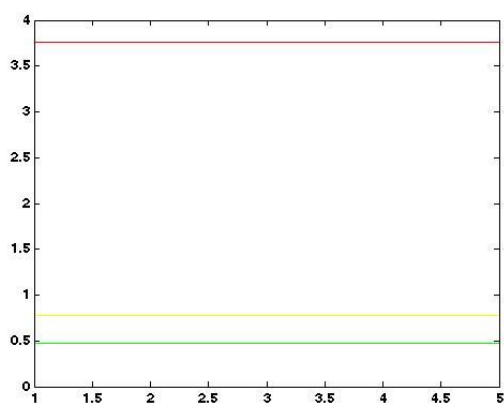
جدول ۷-۱

اجرای الگوریتم Hierarchical برای
ایجاد ۴ خوشه روی مجموعه داده ای iris

تعداد اجزا	تعداد عناصر در هر خوشه			
۱	۱۳	۳۷	۲۱	۷۹
۲	۱۳	۳۷	۲۱	۷۹
۳	۱۳	۳۷	۲۱	۷۹
۴	۱۳	۳۷	۲۱	۷۹
۵	۱۳	۳۷	۲۱	۷۹

شکل ۱۰-۱

نمودار حاصل از توابع ارزیابی برای جدول ۷-۱
شاخص دیوس بلدین با رنگ قرمز
شاخص RMSSDT با رنگ سبز
شاخص RS با رنگ زرد



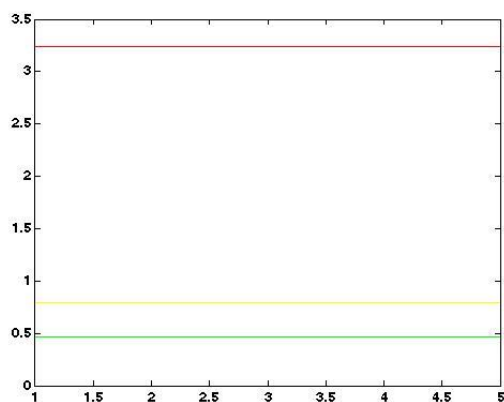
جدول ۸-۱

اجرای الگوریتم Hierarchical برای ایجاد ۵
خوشه روی مجموعه داده ای iris

تعداد اجزا	تعداد عناصر در هر خوشه				
۱	۱	۱۳	۳۶	۲۱	۷۹
۲	۱	۱۳	۳۶	۲۱	۷۹
۳	۱	۱۳	۳۶	۲۱	۷۹
۴	۱	۱۳	۳۶	۲۱	۷۹
۵	۱	۱۳	۳۶	۲۱	۷۹

شکل ۱۱-۱

نمودار حاصل از توابع ارزیابی برای جدول ۸-۱
شاخص دیوس بلدین با رنگ قرمز
شاخص RMSSDT با رنگ سبز
شاخص RS با رنگ زرد



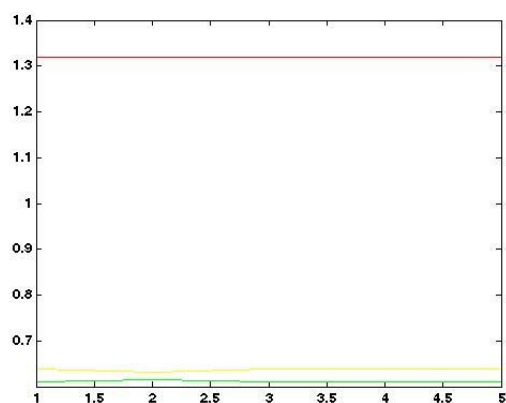
جدول ۹-۱

اجرای الگوریتم SOM برای ایجاد ۲ خوشه روی
مجموعه داده ای iris

تعداد اجرا	تعداد عناصر در هر خوشه	نسل
۱	۵۰ ۱۰۰	۲۲
۲	۱۰۰ ۵۰	۲۲
۳	۵۰ ۱۰۰	۲۲
۴	۵۰ ۱۰۰	۲۲
۵	۵۰ ۱۰۰	۲۲

شکل ۱۲-۱

نمودار حاصل از توابع ارزیابی برای جدول ۹-۱
شاخص دیوس بلدین با رنگ قرمز
شاخص RMSSDT با رنگ سبز
شاخص RS با رنگ زرد



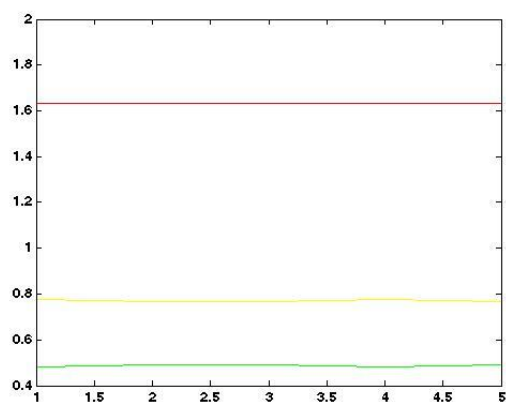
جدول ۱۰-۱

اجرای الگوریتم SOM برای ایجاد ۳ خوشه روی
مجموعه داده ای iris

تعداد اجرا	تعداد عناصر در هر خوشه	نسل
۱	۴۴ ۵۶ ۵۰	۲۴
۲	۴۴ ۵۰ ۵۶	۲۴
۳	۵۶ ۵۰ ۴۴	۲۴
۴	۴۴ ۵۶ ۵۰	۲۴
۵	۵۶ ۵۰ ۴۴	۲۴

شکل ۱۳-۱

نمودار حاصل از توابع ارزیابی برای جدول ۱۰-۱
شاخص دیوس بلدین با رنگ قرمز
شاخص RMSSDT با رنگ سبز
شاخص RS با رنگ زرد



جدول ۱۱-۱

اجرای الگوریتم SOM برای ایجاد ۴ خوشه روی مجموعه داده ای iris

تعداد اجرا	تعداد عناصر در هر خوشه				نسل
۱	۴۹	۴۹	۲۸	۲۴	۳۸
۲	۴۹	۴۹	۲۴	۲۸	۳۸
۳	۴۹	۲۸	۲۴	۴۹	۳۸
۴	۴۹	۲۴	۲۸	۴۹	۳۸
۵	۲۸	۲۴	۴۹	۴۹	۳۸

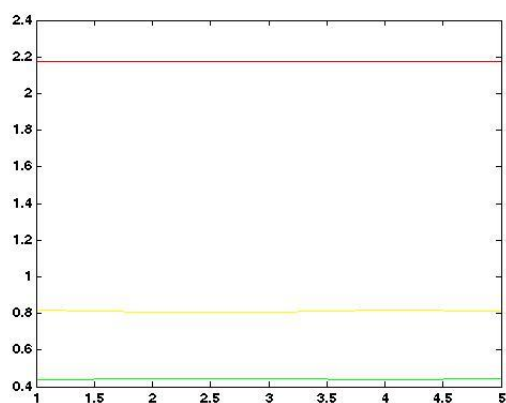
شکل ۱۴-۱

نمودار حاصل از توابع ارزیابی برای جدول ۱۱-۱

شاخص دیوس بلدین با رنگ قرمز

شاخص RMSSDT با رنگ سبز

شاخص RS با رنگ زرد



جدول ۱۲-۱

اجرای الگوریتم SOM برای ایجاد ۵ خوشه روی مجموعه داده ای iris

تعداد اجرا	تعداد عناصر در هر خوشه					نسل
۱	۲۴	۲۸	۱۵	۳۵	۴۸	۵۰
۲	۱۵	۳۵	۴۸	۲۸	۲۴	۵۰
۳	۲۴	۴۸	۳۵	۲۸	۱۵	۵۰
۴	۲۸	۲۴	۴۸	۱۵	۳۵	۵۰
۵	۲۴	۴۸	۲۸	۳۵	۱۵	۵۰

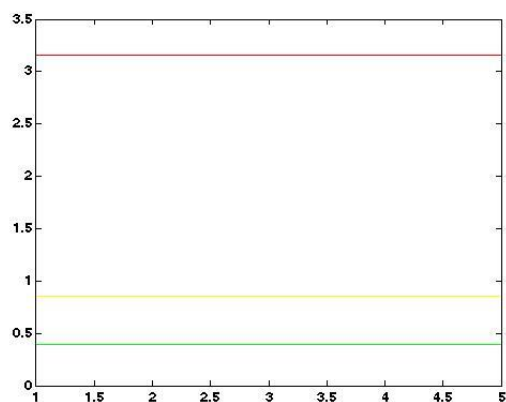
شکل ۱۵-۱

نمودار حاصل از توابع ارزیابی برای جدول ۱۲-۱

شاخص دیوس بلدین با رنگ قرمز

شاخص RMSSDT با رنگ سبز

شاخص RS با رنگ زرد



اکنون در اینجا این الگوریتم های خوشه بندی گفته شده برای مجموعه داده ای ایریس ارزیابی می شود.

با توجه جدول های ۱-۱ تا ۴-۱ و شکل های ۴-۱ تا ۷-۱، از بین ۲۰ اجرای الگوریتم کاتا میانگین (Kmeans) نتایج زیر حاصل می شود:

مقدار شاخص	شماره اجرا	بهترین تعداد خوشه	شاخص دیویس بلدین
$1.319197e + 00$	۳	۲	شاخص دیویس بلدین
$4.250844e - 01$	۵	۵	شاخص RMSSDT
$8.277923e - 01$	۴	۵	شاخص RS

جدول ۱۶-۱ ارزیابی اجرای الگوریتم Kmeans با توجه به جدول های ۱-۱ تا ۴-۱ و شکل های ۴-۱ تا ۷-۱ روی مجموعه داده ای iris

با توجه جدول های ۵-۱ تا ۸-۱ و شکل های ۸-۱ تا ۱۱-۱، از بین ۲۰ اجرای الگوریتم سلسله مراتبی پایین به بالا (Hierarchical) نتایج زیر حاصل می شود :

مقدار شاخص	شماره اجرا	بهترین تعداد خوشه	شاخص دیویس بلدین
$1.618472e + 00$	۱	۲	شاخص دیویس بلدین
$4.674539e - 01$	۱	۵	شاخص RMSSDT
$7.902860e - 01$	۱	۵	شاخص RS

جدول ۱۷-۱ اجرای الگوریتم Hierarchical با توجه به جدول های ۵-۱ تا ۸-۱ و شکل های ۸-۱ تا ۱۱-۱ روی مجموعه داده ای iris

با توجه جدول های ۹-۱ تا ۱۲-۱ و شکل های ۱۲-۱ تا ۱۵-۱، از بین ۲۰ اجرای الگوریتم نگاشت خود سازمان دهی شده (SOM) نتایج زیر حاصل می شود:

مقدار شاخص	شماره اجرا	بهترین تعداد خوشه	شاخص دیویس بلدین
$1.319197e + 00$	۱	۲	شاخص دیویس بلدین
$3.935253e - 01$	۲	۵	شاخص RMSSDT
$8.524132e - 01$	۲	۵	شاخص RS

جدول ۱۸-۱ اجرای الگوریتم SOM با توجه به جدول های ۹-۱ تا ۱۲-۱ و شکل های ۱۲-۱ تا ۱۵-۱ روی مجموعه داده ای iris

برای الگوریتم کاتا میانگین از آنجا که هم اشاره شد ممکن الگوریتم متوقف نشود به صورت دستی شرط پایان در صورتی که خودش متوقف نشود، پس از ۲۰۰۰ بار تکرار گذاشته شد. شرط توقف الگوریتم نگاشت خود سازمان دهی شده زمانی است که وزن ها پس از یک نسل تغییر نکنند.

با اجرای الگوریتم کاتا میانگین، روی مجموعه داده ست ایمیج، نتایج جدول های ۱-۱۹ تا ۱-۲۳ و شکل های ۱-۱۶ تا ۱-۲۰، زیر حاصل شده است. با اجرای الگوریتم سلسله مراتبی پایین به بالا روی مجموعه داده ای ست ایمیج نتایج جدول های ۱-۲۴۱۸ تا ۱-۲۸ و شکل های ۱-۲۱ تا ۱-۲۵، زیر حاصل شده است. با اجرای الگوریتم نگاشت خود سازمان دهی روی مجموعه داده ای ست ایمیج نتایج جدول های ۱-۲۹ تا ۱-۳۳ و شکل های ۱-۲۶ تا ۱-۳۰، زیر حاصل شده است.

نسل	تعداد عناصر در هر خوشه				تعداد اجرا
۲۰۰۰	۱۸۱۰	۶۰۱	۲۷۱۹	۱۳۰۵	۱
۲۰۰۰	۲۱۹۲	۱۸۵۹	۱۷۸۵	۵۹۹	۲
۲۰۰۰	۲۷۱۵	۱۸۱۰	۱۳۰۹	۶۰۱	۳
۲۰۰۰	۲۱۶۷	۱۹۱۱	۱۷۵۷	۶۰۰	۴
۲۰۰۰	۶۰۰	۲۱۶۷	۱۹۱۱	۱۷۵۷	۵

جدول ۱۹-۱ اجرای الگوریتم Kmeans برای ایجاد ۴ خوشه روی مجموعه داده ای satimage

نسل	تعداد عناصر در هر خوشه					تعداد اجرا
۲۰۰۰	۱۵۶۰	۶۰۰	۱۰۱۸	۱۶۰۰	۱۶۵۷	۱
۲۰۰۰	۱۵۶۰	۱۶۰۶	۱۰۱۷	۶۰۰	۱۶۵۲	۲
۲۰۰۰	۱۱۵۴	۵۷۳	۲۰۱۹	۱۰۴۶	۱۶۴۳	۳
۲۰۰۰	۱۵۶۰	۱۶۰۲	۱۰۱۷	۶۰۰	۱۶۵۶	۴
۲۰۰۰	۱۶۵۹	۶۰۰	۱۰۲۰	۱۵۹۶	۱۵۶۰	۵

جدول ۲۰-۱ اجرای الگوریتم Kmeans برای ایجاد ۵ خوشه روی مجموعه داده ای satimage

نسل	تعداد عناصر در هر خوشه						تعداد اجرا
۲۰۰۰	۱۳۶۹	۱۵۱۱	۹۶۷	۱۱۵۵	۸۶۵	۵۶۸	۱
۲۰۰۰	۱۰۰۵	۲۸۲	۱۶۵۳	۱۱۴۹	۳۴۱	۲۰۰۵	۲
۲۰۰۰	۱۵۱۱	۵۶۸	۸۶۳	۱۳۶۵	۱۱۵۶	۹۷۲	۳
۲۰۰۰	۹۷۲	۵۶۸	۱۵۰۹	۸۶۳	۱۳۶۵	۱۱۵۸	۴
۲۰۰۰	۱۳۶۸	۱۱۵۶	۹۷۲	۸۶۴	۵۶۸	۱۵۰۷	۵

جدول ۲۱-۱ اجرای الگوریتم Kmeans برای ایجاد ۶ خوشه روی مجموعه داده ای satimage

نسل	تعداد عناصر در هر خوشه							تعداد اجرا
۲۰۰۰	۹۵۶	۹۵۹	۵۶۸	۱۲۳۴	۷۳۳	۱۱۲۴	۸۶۱	۱
۲۰۰۰	۸۶۷	۹۵۳	۱۱۴۹	۹۸۵	۵۶۸	۸۷۳	۱۰۴۰	۲
۲۰۰۰	۱۱۰۴	۵۶۷	۱۳۹۲	۷۷۹	۱۳۴۹	۵۹۶	۶۴۸	۳
۲۰۰۰	۹۸۵	۹۵۳	۵۶۸	۸۷۳	۸۶۷	۱۱۴۹	۱۰۴۰	۴
۲۰۰۰	۳۴۳	۹۷۰	۲۷۵	۱۳۶۵	۱۱۵۳	۸۱۴	۱۵۱۵	۵

جدول ۲۲-۱ اجرای الگوریتم Kmeans برای ایجاد ۷ خوشه روی مجموعه داده ای satimage

نسل	تعداد عناصر در هر خوشه							تعداد اجرا
۲۰۰۰	۸۳۷	۴۱۲	۹۱۹	۸۲۰	۵۶۹	۱۰۳۶	۹۴۷ ۸۹۵	۱
۲۰۰۰	۱۱۵۳	۹۸۵	۳۳۴	۹۵۴	۱۰۳۶	۸۶۵	۲۸۴ ۸۲۴	۲
۲۰۰۰	۱۰۳۰	۵۶۶	۶۳۹	۷۷۱	۱۰۵۴	۹۷۸	۸۴۵ ۵۵۲	۳
۲۰۰۰	۹۵۳	۱۱۰۸	۷۸۵	۴۹۵	۱۲۲۸	۱۴۷	۹۵۳ ۷۶۶	۴
۲۰۰۰	۱۰۳۷	۶۴۸	۵۴۳	۸۰۳	۸۴۶	۱۰۲۷	۹۶۴ ۵۶۷	۵

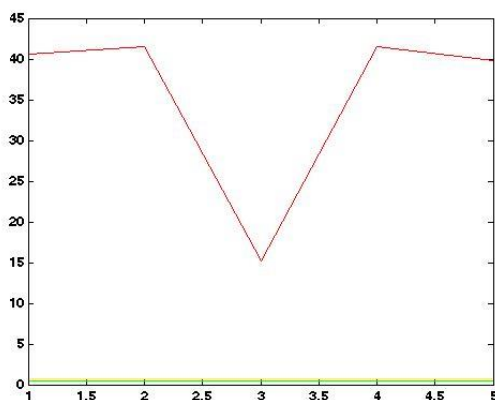
جدول ۲۳-۱ اجرای الگوریتم Kmeans برای ایجاد ۸ خوشه روی مجموعه داده ای satimage

تعداد عناصر در هر خوشه					تعداد اجرا
۱۸۲۳	۳۴۱	۵۱۶	۳۷۵۵	۱	
۱۸۲۳	۳۴۱	۵۱۶	۳۷۵۵	۲	
۱۸۲۳	۳۴۱	۵۱۶	۳۷۵۵	۳	
۱۸۲۳	۳۴۱	۵۱۶	۳۷۵۵	۴	
۱۸۲۳	۳۴۱	۵۱۶	۳۷۵۵	۵	

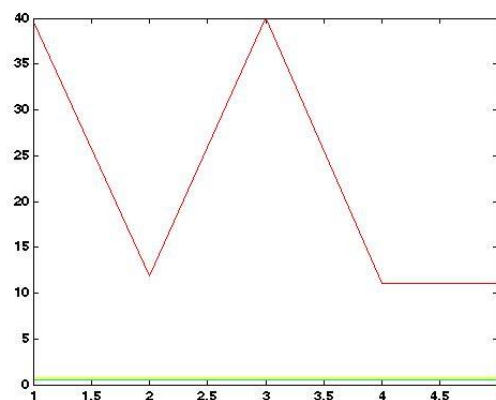
جدول ۲۴-۱ اجرای الگوریتم Hierarchical برای ایجاد ۴ خوشه روی مجموعه داده ای satimage

تعداد عناصر در هر خوشه					تعداد اجرا
۲۱	۱۸۲۳	۳۴۱	۵۱۶	۳۷۳۴	۱
۲۱	۱۸۲۳	۳۴۱	۵۱۶	۳۷۳۴	۲
۲۱	۱۸۲۳	۳۴۱	۵۱۶	۳۷۳۴	۳
۲۱	۱۸۲۳	۳۴۱	۵۱۶	۳۷۳۴	۴
۲۱	۱۸۲۳	۳۴۱	۵۱۶	۳۷۳۴	۵

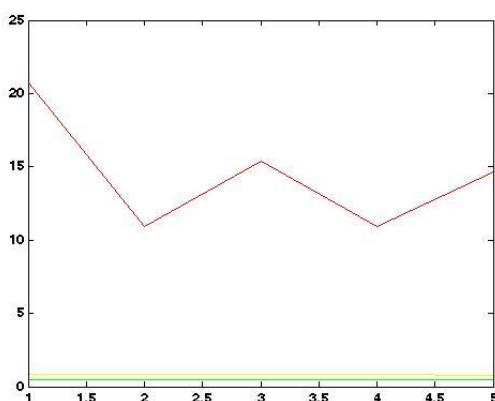
جدول ۲۵-۱ اجرای الگوریتم Hierarchical برای ایجاد ۵ خوشه روی مجموعه داده ای satimage



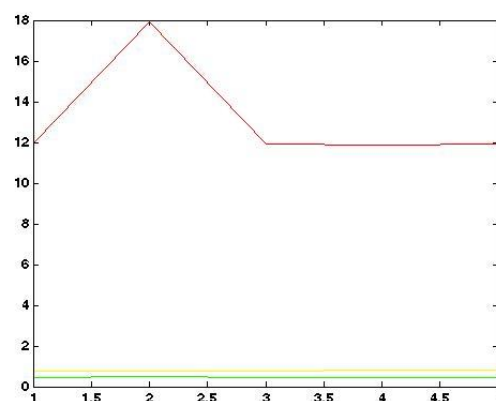
شکل ۱۷-۱ نمودار حاصل از توابع ارزیابی برای جدول ۱۴-۱؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد



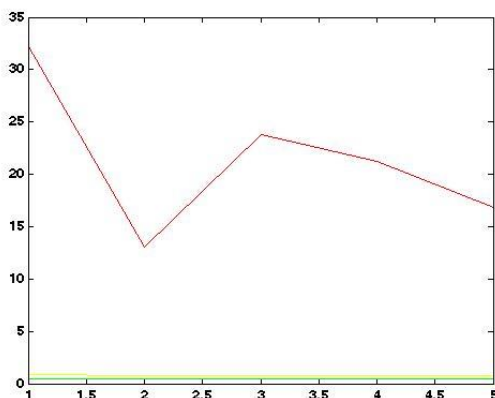
شکل ۱۶-۱ نمودار حاصل از توابع ارزیابی برای جدول ۱۳-۱؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد



شکل ۱۹-۱ نمودار حاصل از توابع ارزیابی برای جدول ۱۶-۱؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد



شکل ۱۸-۱ نمودار حاصل از توابع ارزیابی برای جدول ۱۵-۱؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد



شکل ۲۰-۱ نمودار حاصل از توابع ارزیابی برای جدول ۱۷-۱؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد

تعداد اجرا	تعداد عناصر در هر خوشه					
۱	۵۴۳	۲۱	۱۸۲۳	۳۴۱	۳۱۹۱	۵۱۶
۲	۵۴۳	۲۱	۱۸۲۳	۳۴۱	۳۱۹۱	۵۱۶
۳	۵۴۳	۲۱	۱۸۲۳	۳۴۱	۳۱۹۱	۵۱۶
۴	۵۴۳	۲۱	۱۸۲۳	۳۴۱	۳۱۹۱	۵۱۶
۵	۵۴۳	۲۱	۱۸۲۳	۳۴۱	۳۱۹۱	۵۱۶

جدول ۱-۲۶ اجرای الگوریتم Hierarchical برای ایجاد ۶ خوشه روی مجموعه داده ای satimage

تعداد اجرا	تعداد عناصر در هر خوشه						
۱	۲۱۵	۵۴۳	۲۱	۱۸۲۳	۳۰۱	۳۴۱	۳۱۹۱
۲	۲۱۵	۵۴۳	۲۱	۱۸۲۳	۳۰۱	۳۴۱	۳۱۹۱
۳	۲۱۵	۵۴۳	۲۱	۱۸۲۳	۳۰۱	۳۴۱	۳۱۹۱
۴	۲۱۵	۵۴۳	۲۱	۱۸۲۳	۳۰۱	۳۴۱	۳۱۹۱
۵	۲۱۵	۵۴۳	۲۱	۱۸۲۳	۳۰۱	۳۴۱	۳۱۹۱

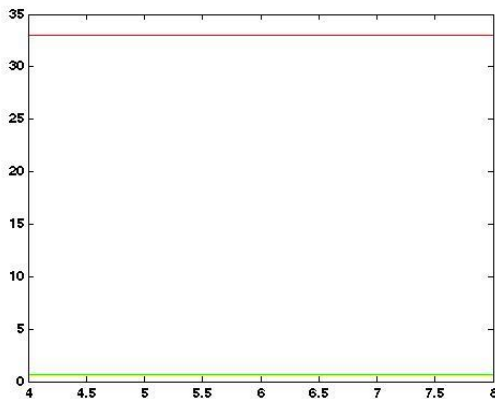
جدول ۱-۲۷ اجرای الگوریتم Hierarchical برای ایجاد ۷ خوشه روی مجموعه داده ای satimage

تعداد اجرا	تعداد عناصر در هر خوشه							
۱	۱۹	۲۱۵	۵۴۳	۲۱	۱۸۲۳	۳۱۷۲	۳۰	۳۴۱
۲	۱۹	۲۱۵	۵۴۳	۲۱	۱۸۲۳	۳۱۷۲	۳۰	۳۴۱
۳	۱۹	۲۱۵	۵۴۳	۲۱	۱۸۲۳	۳۱۷۲	۳۰	۳۴۱
۴	۱۹	۲۱۵	۵۴۳	۲۱	۱۸۲۳	۳۱۷۲	۳۰	۳۴۱
۵	۱۹	۲۱۵	۵۴۳	۲۱	۱۸۲۳	۳۱۷۲	۳۰	۳۴۱

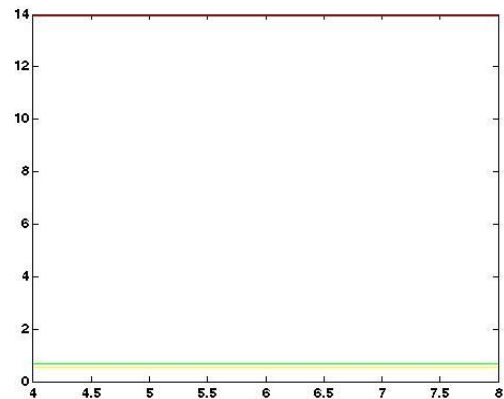
جدول ۱-۲۸ اجرای الگوریتم Hierarchical برای ایجاد ۸ خوشه روی مجموعه داده ای satimage

تعداد اجرا	نسل	تعداد عناصر در هر خوشه			
۱	۹	۲۱۱۰	۵۹۶	۱۹۰۵	۱۸۲۴
۲	۹	۱۹۰۵	۵۹۶	۲۱۱۰	۱۸۲۴
۳	۹	۲۱۱۰	۱۹۰۵	۵۹۶	۱۸۲۴
۴	۹	۲۱۱۰	۱۸۲۴	۱۹۰۵	۵۹۶
۵	۹	۲۱۱۰	۱۸۲۴	۵۹۶	۱۹۰۵

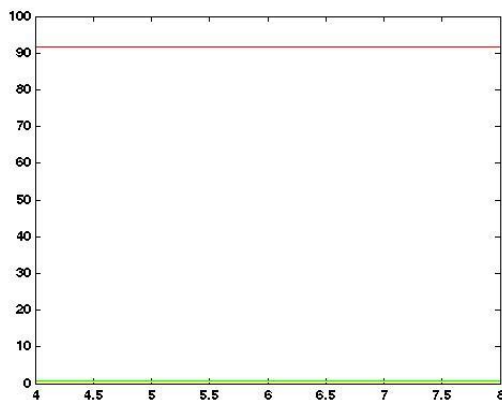
جدول ۱-۲۹ اجرای الگوریتم SOM برای ایجاد ۴ خوشه روی مجموعه داده ای satimage



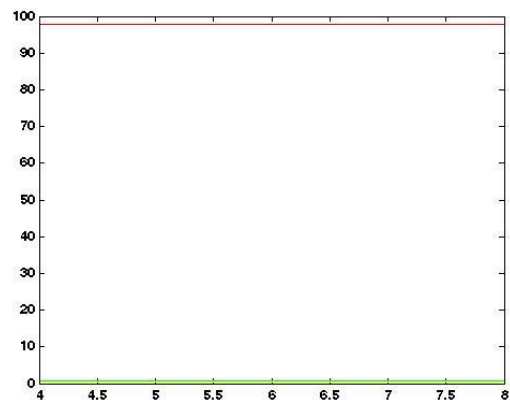
شکل ۱-۲۲ نمودار حاصل از توابع ارزیابی برای جدول ۱-۲۵؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد



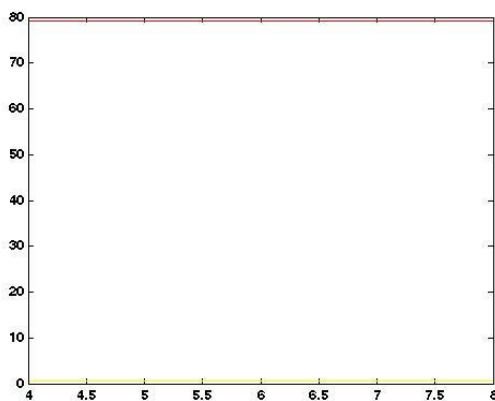
شکل ۱-۲۱ نمودار حاصل از توابع ارزیابی برای جدول ۱-۲۴؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد



شکل ۱-۲۴ نمودار حاصل از توابع ارزیابی برای جدول ۱-۲۷؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد



شکل ۱-۲۳ نمودار حاصل از توابع ارزیابی برای جدول ۱-۲۶؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد



شکل ۱-۲۵ نمودار حاصل از توابع ارزیابی برای جدول ۱-۲۸؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد

نسل	تعداد عناصر در هر خوشه						تعداد اجرا
۹	۱۱۳۶	۱۶۴۱	۱۵۰۳	۱۵۶۶	۵۸۹	۱	۱
۹	۱۶۴۱	۱۵۰۳	۵۸۹	۱۵۶۶	۱۱۳۶	۲	۲
۹	۱۶۴۱	۱۵۶۶	۱۱۳۶	۵۸۹	۱۵۰۳	۳	۳
۹	۱۵۶۶	۱۵۰۳	۵۸۹	۱۶۴۱	۱۱۳۶	۴	۴
۹	۱۵۰۳	۱۶۴۱	۱۱۳۶	۵۸۹	۱۵۶۶	۵	۵

جدول ۱-۳۰ اجرای الگوریتم SOM برای ایجاد ۵ خوشه روی مجموعه داده ای satimage

نسل	تعداد عناصر در هر خوشه						تعداد اجرا
۹	۹۲۵	۱۳۷۰	۱۱۹۹	۵۶۴	۱۴۹۱	۸۸۶	۱
۹	۱۴۹۱	۹۲۵	۵۶۴	۸۸۶	۱۱۹۹	۱۳۷۰	۲
۹	۱۴۹۱	۹۲۵	۱۱۹۹	۸۸۶	۵۶۴	۱۳۷۰	۳
۹	۸۸۶	۱۱۹۹	۱۴۹۱	۱۳۷۰	۵۶۴	۹۲۵	۴
۹	۹۲۵	۱۱۹۹	۸۸۶	۱۱۹۹	۱۴۹۱	۱۳۷۰	۵

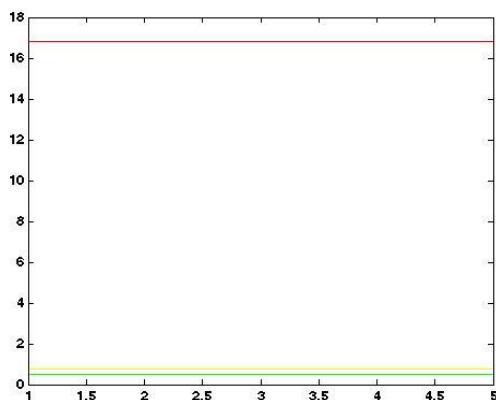
جدول ۱-۲۱ اجرای الگوریتم SOM برای ایجاد ۶ خوشه روی مجموعه داده ای satimage

نسل	تعداد عناصر در هر خوشه							تعداد اجرا
۱۰	۹۲۰	۸۶۰	۱۲۸۲	۵۶۶	۱۱۷۷	۹۵۳	۶۷۷	۱
۱۰	۱۲۸۲	۱۱۷۷	۶۷۷	۵۶۶	۹۲۰	۸۶۰	۹۵۳	۲
۱۰	۱۱۷۷	۶۷۷	۸۶۰	۱۲۸۲	۹۵۳	۹۲۰	۵۶۶	۳
۱۰	۱۱۷۷	۱۲۸۲	۹۵۳	۶۷۷	۹۲۰	۸۶۰	۵۶۶	۴
۱۰	۱۱۷۷	۹۲۰	۹۵۳	۸۶۰	۱۲۸۲	۵۶۶	۶۷۷	۵

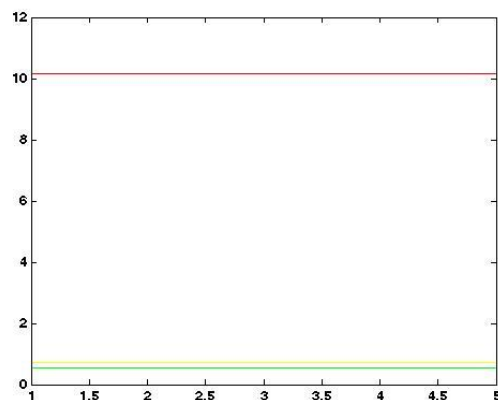
جدول ۱-۳۲ اجرای الگوریتم SOM برای ایجاد ۷ خوشه روی مجموعه داده ای satimage

نسل	تعداد عناصر در هر خوشه								تعداد اجرا
۱۵	۳۲۷	۹۶۶	۱۱۹۴	۷۷۶	۱۴۷۷	۱۰۴	۲۱۷	۱۳۷۴	۱
۱۵	۱۴۷۷	۹۶۶	۳۲۷	۱۰۴	۲۱۷	۱۱۹۴	۱۳۷۴	۷۷۶	۲
۱۵	۲۱۷	۷۷۶	۱۱۹۴	۳۲۷	۱۳۷۴	۱۴۷۷	۹۶۶	۱۰۴	۳
۱۵	۱۴۷۷	۲۱۷	۱۱۹۴	۹۶۶	۷۷۶	۳۲۷	۱۳۷۴	۱۰۴	۴
۱۵	۱۱۹۴	۷۷۶	۱۴۷۷	۱۰۴	۹۶۶	۳۲۷	۲۱۷	۱۳۷۴	۵

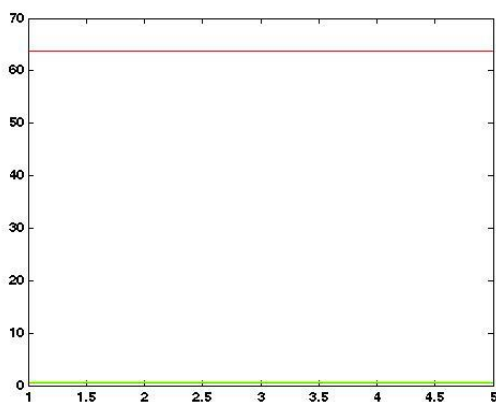
جدول ۱-۳۳ اجرای الگوریتم SOM برای ایجاد ۸ خوشه روی مجموعه داده ای satimage



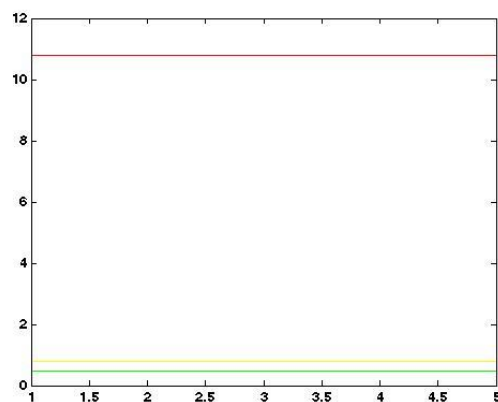
شکل ۱-۲۷ نمودار حاصل از توابع ارزیابی برای جدول ۱-۳۰؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد



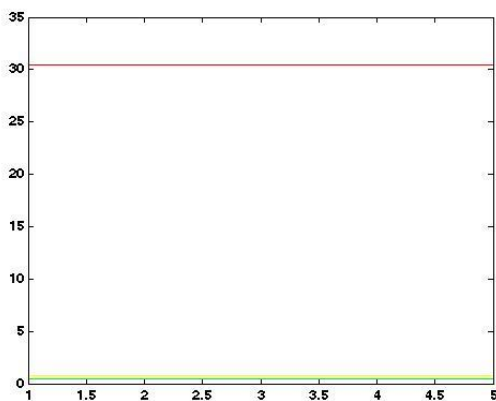
شکل ۱-۲۶ نمودار حاصل از توابع ارزیابی برای جدول ۱-۲۹؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد



شکل ۱-۲۹ نمودار حاصل از توابع ارزیابی برای جدول ۱-۳۲؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد



شکل ۱-۲۸ نمودار حاصل از توابع ارزیابی برای جدول ۱-۳۱؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد



شکل ۱-۳۰ نمودار حاصل از توابع ارزیابی برای جدول ۱-۳۳؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد

اکنون در اینجا این الگوریتم های خوشه بندی گفته شده برای مجموعه داده ای ست ایمپج ارزیابی می شود.

با توجه جدول های ۱۹-۱ تا ۲۳-۱ و شکل های ۱۶-۱ تا ۲۰-۱، از بین ۲۵ اجرای الگوریتم کا تا میانگین (Kmeans) نتایج زیر حاصل می شود:

مقدار شاخص	شماره اجرا	بهترین تعداد خوشه	
$1.096199e + 01$	۴	۷	شاخص دیویس بلدین
$4.260680e - 01$	۱	۸	شاخص RMSSDT
$8.196510e - 01$	۱	۸	شاخص RS

جدول ۳۴-۱ ارزیابی اجرای الگوریتم Kmeans با توجه به جدول های ۱۹-۱ تا ۲۳-۱ و شکل های ۱۶-۱ تا ۲۰-۱ روی مجموعه داده ای satimage

با توجه جدول های ۲۴-۱ تا ۲۸-۱ و شکل های ۲۱-۱ تا ۲۵-۱، از بین ۲۵ اجرای الگوریتم سلسله مراتبی پایین به بالا (Hierarchical) نتایج زیر حاصل می شود :

مقدار شاخص	شماره اجرا	بهترین تعداد خوشه	
۱۳.۹۱۸۶	۱	۴	شاخص دیویس بلدین
۰.۶۵۲۲	۱	۸	شاخص RMSSDT
۰.۵۷۷۳	۱	۸	شاخص RS

جدول ۳۵-۱ اجرای الگوریتم Hierarchical با توجه به جدول های ۲۴-۱ تا ۲۸-۱ و شکل های ۲۱-۱ تا ۲۵-۱ روی مجموعه داده ای satimage

با توجه جدول های ۲۹-۱ تا ۳۳-۱ و شکل های ۲۶-۱ تا ۳۰-۱، از بین ۲۵ اجرای الگوریتم نگاشت خود سازمان دهی شده (SOM) نتایج زیر حاصل می شود:

مقدار شاخص	شماره اجرا	بهترین تعداد خوشه	
$1.016877e + 01$	۱	۴	شاخص دیویس بلدین
$4.377424e - 01$	۳	۸	شاخص RMSSDT
$8.096324e - 01$	۳	۸	شاخص RS

جدول ۳۶-۱ اجرای الگوریتم SOM با توجه به جدول های ۲۹-۱ تا ۳۳-۱ و شکل های ۲۶-۱ تا ۳۰-۱ روی مجموعه داده ای satimage

۶. نتیجه گیری

با توجه به جداول ۱-۱ تا ۱۵-۱ و شکل های ۱-۴ تا ۱۵-۱ بهترین خوشه بندی ها از بین سه الگوریتم بیان شده روی مجموعه داده ای ایریس در جدول ۱-۳۷ آمده است.

الگوریتم	مقدار شاخص	شماره اجرا	بهترین تعداد خوشه	
SOM و Kmeans	$1.319197e + 00$	۱	۲	شاخص دیویس بلدین
SOM	$1.319197e + 00$	۲	۵	شاخص RMSSDT
SOM	$8.524132e - 01$	۲	۵	شاخص RS

جدول ۱-۳۷ بهترین نتایج از بین ۱۲ اجرای الگوریتم های خوشه بندی برای مجموعه داده ای iris

با توجه به جداول ۱-۱۹ تا ۱-۳۳ و شکل های ۱-۱۶ تا ۱-۳۰ بهترین خوشه بندی ها از بین سه الگوریتم بیان شده روی مجموعه داده ای ست ایمیچ در جدول ۱-۳۸ آمده است.

الگوریتم	مقدار شاخص	شماره اجرا	بهترین تعداد خوشه	
Hierarchical	۱۳.۹۱۸۶	۱	۴	شاخص دیویس بلدین
Kmeans	$4.260680e - 01$	۱	۸	شاخص RMSSDT
SOM	$8.096324e - 01$	۳	۸	شاخص RS

جدول ۱-۳۸ بهترین نتایج از بین ۲۰ اجرای الگوریتم های خوشه بندی برای مجموعه داده ای satimage

