****

**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**

**810890001**

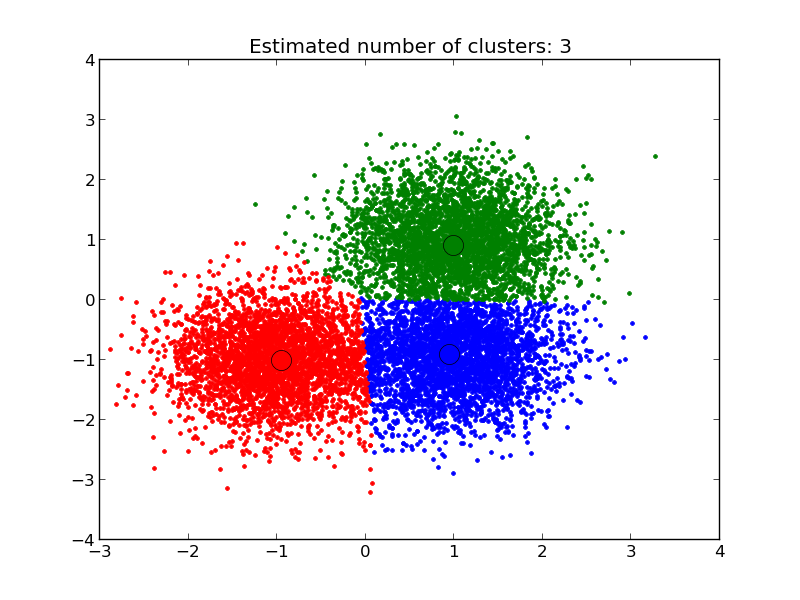
***چکیده***

در این پروژه ، از روش های ***خوشه بندی*** [[1]](#footnote-1) 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***[[2]](#footnote-2)، ***سلسله مراتبی پایین با بالا***[[3]](#footnote-3) و ***نگاشت خود سازمان دهی شده***[[4]](#footnote-4) شبیه سازی شده اند. در الگوریتم خوشه بندی نگاشت خود سازمان دهی شده از روش یادگیری البته به صورت کورکورانه[[5]](#footnote-5) است.

1. ***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 1-1 یک دسته از داده ها را در یک فضای دوبعدی نشان می دهد. در این شکل سه خو شه وجود دارد.



**شکل 1-1 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[[6]](#footnote-6) استفاده می کنند. همه ی روش ها از معیار شباهت برای تشخیص نمونه هایی که در یک خوشه استفاده می کنند. در واقع این 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 اندازگیری فاصله دو نمونه وجود دارد؛ به عنوان مثال می توان به فرمول های فاصله ی مینکاسکی[[7]](#footnote-7) و منهتن[[8]](#footnote-8) اشاره کرد که به ترتیب در رابطه ها ی 1-1 و 1-2 آمده اند.

(1-1)

(1-2)

که و دو نمنونه 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 1-3 فرمول این فاصله را نشان می دهد.

(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[[9]](#footnote-9) و ست ایمیج[[10]](#footnote-10) اجرا می شوند. برای این کار الگوریتم ها را برای تعداد خوشه های 3 تا 8 و به ازای هر کدام از این خوشه ها 5 با الگوریتم اجرا می شود.

1. ***الگوریتم ها***
2. ***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 دوبه 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 مجددا تکرار می شود. این فرایند تا زمانیکه به تعداد خوشه های مورد نظر برسد تکرار می شود.

1. ***خوشه بندی کا تا میانگین***

در این روش ابتدا به تعداد خوشه ها از بین نمونه به صورت تصادفی مرکز[[11]](#footnote-11) انتخاب می شود (هر یک از این مرکز ها به عنوان یک خوشه در نظر گرفته می شود). سپس فا صله ی هر نمونه تا مرکز ها محاصبه می شود؛ و هر نمونه به خوشه ی مرکزی که فا صله 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 محاصبه 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.

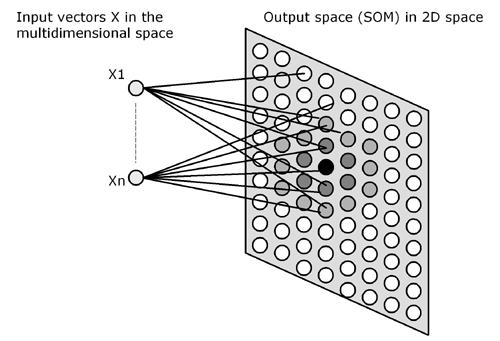
در این الگوریتم خلاف الگوریتم خوشه بندی سلسله مراتبی ممکن است شرایطی ایجاد شود که الگوریتم هیچگاه متوقف نشود ( در هر بار تکرار مرکز خت تغییر کنند). لذا در اینجا در شبیه سازی ای الگوریتم، برای اجتناب از این مشکل یک شرط تعداد تکرار می گذاریم که اگر پس از 1000 باز تکرار مرکز ها هم چنان تغییر می کنند متوقف شود.

از معایب این روش می توان به موارد زیر اشاره کرد:

1. به انتخاب اولیه مرکز ها بشدت وابسته است
2. سربار محاصباتی بالا دارد
3. کا بهینه باید پیدا شود
4. در خیلی مواقع الگوریتم پایان پذیر نیست
5. ***خوشه بندی نگاشت خود سازمان دهی شده***

این روش اولین بار توسط مالسبورگ[[12]](#footnote-12) در سال 1973 ارائه شده و توسط کوهونن[[13]](#footnote-13) در سال 1982 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 کاملا متصل تشکیل شده است : لایه ورودی ، لایه خروجی. شکل 1-2 ساختار این شبکه عصبی را نشان می دهد.



**شکل 1-2 ساختار شبکه عصبی نگاشت خود سازمان دهی شده**

تعداد نرون های لایه ورودی برابر با ابعاد فضای مسئله (تعداد ویژیگی 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 اتصلات 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 محاصبه 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 1-4 به روز می شود،

|  |  |
| --- | --- |
| (1-4) |  |

که نرخ آموزش(به صورت پویا)، ي 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 نیز به سمت صفر میل کی کند. این فرایند تا زمانیکه همه نمونه ها خوشه بندی شوند ادامه پیدا می کند (شکل 1-3).

|  |  |
| --- | --- |
| تابع همسایگی |  |

**شکل 1-3 نمودار نرخ آموزش و تابع همسایگی؛ شعاع تابع همسایگی در هر تکرار کم می شود.**

1. ***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.

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

* : تعداد خوشه ها
* : تعداد بعد ها
* : فاصله ی بین دو نمونه
* : میانگین j امین بعد
* : که یک بردار ستونی است.
* : تعداد عناصر در امین خوشه و بعد
* : تعداد عناصر در امین بعد در کل داده
* : مرکز امین خوشه
* : امین خوشه
* : تعداد عناصر در خوشه ی ام

1. ***شاخص دویس - بولدین***[[14]](#footnote-14)

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



معمولا به صورت زیر تعریف می شود:

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

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.

1. ***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.

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.

1. ***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

RMSSDT\_and\_RS : شاخص های RS و RMSSTD

1. ***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 2 تا 5اجرا می شود. از طرفی به ازای هر تعداد خوشه مشخص هر الگوریتم 5 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 4 تا 8اجرا می شود. از طرفی به ازای هر تعداد خوشه مشخص هر الگوریتم 5 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 1000 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 1-1 تا 1-4 و شکل های 1-4 تا 1-7، زیر حاصل شده است. با اجرای الگوریتم سلسله مراتبی پایین به بالا روی مجموعه داده ای ایریس نتایج جدول های 1-5 تا 1-8 و شکل های 1-8 تا -1-11، زیر حاصل شده است. با اجرای الگوریتم نگاشت خود سازمان دهی روی مجموعه داده ای ایریس نتایج جدول های 1-9 تا 1-11 و شکل های 1-12 تا 1-15، زیر حاصل شده است.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | تعداد اجرا | تعداد عناصر در هر خوشه | نسل | | 1 | 100 50 | 1000 | | 2 | 50 100 | 1000 | | 3 | 100 50 | 1000 | | 4 | 50 100 | 1000 | | 5 | 50 100 | 1000 | | **جدول 1-1**  **اجرای الگوریتم kmeans برای ایجاد 2 خوشه روی مجموعه داده ای iris** |
|  | **شکل1-4**  **نمودار حاصل از توابع ارزیابیبرای جدول 1-1**  **شاخص دیوس بلدین با رنگ قرمز**  **شاخص RMSSDT با رنگ سبز**  **شاخص RS با رنگ زرد** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | تعداد اجرا | تعداد عناصر در هر خوشه | نسل | | 1 | 46 55 49 | 1000 | | 2 | 59 50 41 | 1000 | | 3 | 47 50 53 | 1000 | | 4 | 56 44 50 | 1000 | | 5 | 37 100 13 | 1000 | | **جدول 1-2**  **اجرای الگوریتم kmeans برای ایجاد3 خوشه روی مجموعه داده ای iris** |
|  | **شکل1-5**  **نمودار حاصل از توابع ارزیابی برای جدول 1-2**  **شاخص دیوس بلدین با رنگ قرمز**  **شاخص RMSSDT با رنگ سبز**  **شاخص RS با رنگ زرد** |
| |  |  |  | | --- | --- | --- | | تعداد اجرا | تعداد عناصر در هر خوشه | نسل | | 1 | 59 15 35 41 | 1000 | | 2 | 27 50 36 37 | 2 | | 3 | 53 49 20 28 | 5 | | 4 | 44 50 45 11 | 2 | | 5 | 49 44 30 27 | 3 | | **جدول 1-3**  **اجرای الگوریتم kmeans برای ایجاد4خوشه روی مجموعه داده ای iris** |
|  | **شکل1-6**  **نمودار حاصل از توابع ارزیابی برای جدول 1-3**  **شاخص دیوس بلدین با رنگ قرمز**  **شاخص RMSSDT با رنگ سبز**  **شاخص RS با رنگ زرد** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | تعداد اجرا | تعداد عناصر در هر خوشه | نسل | | 1 | 22 23 49 18 38 | 1000 | | 2 | 24 17 49 11 49 | 1000 | | 3 | 35 9 26 31 49 | 1000 | | 4 | 21 18 27 49 35 | 1000 | | 5 | 33 33 49 19 16 | 1000 | | **جدول 1-4**  **اجرای الگوریتم kmeans برای ایجاد5خوشه روی مجموعه داده ای iris** |
|  | **شکل1-7**  **نمودار حاصل از توابع ارزیابی برای جدول 1-4**  **شاخص دیوس بلدین با رنگ قرمز**  **شاخص RMSSDT با رنگ سبز**  **شاخص RS با رنگ زرد** |
| |  |  | | --- | --- | | تعداد اجرا | تعداد عناصر در هر خوشه | | 1 | 79 71 | | 2 | 79 71 | | 3 | 79 71 | | 4 | 79 71 | | 5 | 79 71 | | **جدول 1-5**  **اجرای الگوریتم Hierarchical برای ایجاد2خوشه روی مجموعه داده ای iris** |
|  | **شکل1-8**  **نمودار حاصل از توابع ارزیابی برای جدول 1-5**  **شاخص دیوس بلدین با رنگ قرمز**  **شاخص RMSSDT با رنگ سبز**  **شاخص RS با رنگ زرد** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | تعداد اجرا | تعداد عناصر در هر خوشه | | 1 | 13 79 58 | | 2 | 13 79 58 | | 3 | 13 79 58 | | 4 | 13 79 58 | | 5 | 13 79 58 | | **جدول 1-6**  **اجرای الگوریتم Hierarchical برای ایجاد3خوشه روی مجموعه داده ای iris** |
|  | **شکل1-9**  **نمودار حاصل از توابع ارزیابی برای جدول 1-6**  **شاخص دیوس بلدین با رنگ قرمز**  **شاخص RMSSDT با رنگ سبز**  **شاخص RS با رنگ زرد** |
| |  |  | | --- | --- | | تعداد اجرا | تعداد عناصر در هر خوشه | | 1 | 13 37 21 79 | | 2 | 13 37 21 79 | | 3 | 13 37 21 79 | | 4 | 13 37 21 79 | | 5 | 13 37 21 79 | | **جدول 1-7**  **اجرای الگوریتم Hierarchical برای ایجاد4خوشه روی مجموعه داده ای iris** |
|  | **شکل1-10**  **نمودار حاصل از توابع ارزیابی برای جدول 1-7**  **شاخص دیوس بلدین با رنگ قرمز**  **شاخص RMSSDT با رنگ سبز**  **شاخص RS با رنگ زرد** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | تعداد اجرا | تعداد عناصر در هر خوشه | | 1 | 1 13 36 21 79 | | 2 | 1 13 36 21 79 | | 3 | 1 13 36 21 79 | | 4 | 1 13 36 21 79 | | 5 | 1 13 36 21 79 | | **جدول 1-8**  **اجرای الگوریتم Hierarchical برای ایجاد 5 خوشه روی مجموعه داده ای iris** |
|  | **شکل1-11**  **نمودار حاصل از توابع ارزیابی برای جدول 1-8**  **شاخص دیوس بلدین با رنگ قرمز**  **شاخص RMSSDT با رنگ سبز**  **شاخص RS با رنگ زرد** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | تعداد اجرا | تعداد عناصر در هر خوشه | نسل | | 1 | 50 100 | 22 | | 2 | 100 50 | 22 | | 3 | 50 100 | 22 | | 4 | 50 100 | 22 | | 5 | 50 100 | 22 | | **جدول 1-9**  **اجرای الگوریتم SOM برای ایجاد 2 خوشه روی مجموعه داده ای iris** |
|  | **شکل1-12**  **نمودار حاصل از توابع ارزیابی برای جدول 1-9**  **شاخص دیوس بلدین با رنگ قرمز**  **شاخص RMSSDT با رنگ سبز**  **شاخص RS با رنگ زرد** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | تعداد اجرا | تعداد عناصر در هر خوشه | نسل | | 1 | 50 56 44 | 24 | | 2 | 56 50 44 | 24 | | 3 | 44 50 56 | 24 | | 4 | 50 56 44 | 24 | | 5 | 44 50 56 | 24 | | **جدول 1-10**  **اجرای الگوریتم SOM برای ایجاد 3 خوشه روی مجموعه داده ای iris** |
|  | **شکل1-13**  **نمودار حاصل از توابع ارزیابی برای جدول 1-10**  **شاخص دیوس بلدین با رنگ قرمز**  **شاخص RMSSDT با رنگ سبز**  **شاخص RS با رنگ زرد** |
| |  |  |  | | --- | --- | --- | | تعداد اجرا | تعداد عناصر در هر خوشه | نسل | | 1 | 49 49 28 24 | 38 | | 2 | 49 49 24 28 | 38 | | 3 | 49 28 24 49 | 38 | | 4 | 49 24 28 49 | 38 | | 5 | 28 24 49 49 | 38 | | **جدول 1-11**  **اجرای الگوریتم SOM برای ایجاد 4 خوشه روی مجموعه داده ای iris** |
|  | **شکل1-14**  **نمودار حاصل از توابع ارزیابی برای جدول 1-11**  **شاخص دیوس بلدین با رنگ قرمز**  **شاخص RMSSDT با رنگ سبز**  **شاخص RS با رنگ زرد** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | تعداد اجرا | تعداد عناصر در هر خوشه | نسل | | 1 | 24 28 15 35 48 | 50 | | 2 | 15 35 48 28 24 | 50 | | 3 | 24 48 35 28 15 | 50 | | 4 | 28 24 48 15 35 | 50 | | 5 | 24 48 28 35 15 | 50 | | **جدول 1-12**  **اجرای الگوریتم SOM برای ایجاد 5 خوشه روی مجموعه داده ای iris** |
|  | **شکل1-15**  **نمودار حاصل از توابع ارزیابی برای جدول 1-12**  **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 1-1 تا 1-4 و شکل های 1-4 تا 1-7، از بین 20 اجرای الگوریتم کا تا میانگین (Kmeans) نتایج زیر حاصل می شود:

|  |  |  |  |
| --- | --- | --- | --- |
| مقدار شاخص | شماره اجرا | بهترین تعداد خوشه |  |
|  | 3 | 2 | **شاخص دیویس بلدین** |
|  | 5 | 5 | **شاخص RMSSDT** |
|  | 4 | 5 | **شاخص RS** |

**جدول 1-16 ارزیابی اجرای الگوریتم Kmeans با توجه به جدول های 1-1 تا 1-4 و شکل های 1-4 تا 1-7 روی مجموعه داده ای iris**

با توجه جدول های 1-5 تا 1-8 و شکل های 1-8 تا 1-11، از بین 20 اجرای الگوریتم سلسله مراتبی پایین به بالا(Hierarchical )نتایج زیر حاصل می شود :

|  |  |  |  |
| --- | --- | --- | --- |
| مقدار شاخص | شماره اجرا | بهترین تعداد خوشه |  |
|  | 1 | 2 | **شاخص دیویس بلدین** |
|  | 1 | 5 | **شاخص RMSSDT** |
|  | 1 | 5 | **شاخص RS** |

**جدول 1-17 اجرای الگوریتم Hierarchical با توجه به جدول های 1-5 تا 1-8 و شکل های 1-8 تا 1-11 روی مجموعه داده ای iris**

با توجه جدول های 1-9 تا 1-12 و شکل های 1-12 تا 1-15، از بین 20 اجرای الگوریتم نگاشت خود سازمان دهی شده(SOM) نتایج زیر حاصل می شود:

|  |  |  |  |
| --- | --- | --- | --- |
| مقدار شاخص | شماره اجرا | بهترین تعداد خوشه |  |
|  | 1 | 2 | **شاخص دیویس بلدین** |
|  | 2 | 5 | **شاخص RMSSDT** |
|  | 2 | 5 | **شاخص RS** |

**جدول 1-18 اجرای الگوریتم SOM با توجه به جدول های 1-9 تا 1-12 و شکل های 1-12 تا 1-15 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 2000 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 1-91 تا 1-23 و شکل های 1-16 تا 1-20، زیر حاصل شده است. با اجرای الگوریتم سلسله مراتبی پایین به بالا روی مجموعه داده ای ست ایمیج نتایج جدول های 1-2418 تا 1-28 و شکل های 1-21 تا 1-25، زیر حاصل شده است. با اجرای الگوریتم نگاشت خود سازمان دهی روی مجموعه داده ای ست ایمیج نتایج جدول های 1-29 تا 1-33 و شکل های 1-26 تا 1-30، زیر حاصل شده است.

|  |  |  |
| --- | --- | --- |
| نسل | تعداد عناصر در هر خوشه | تعداد اجرا |
| 2000 | 1810 601 2719 1305 | 1 |
| 2000 | 2192 1859 1785 599 | 2 |
| 2000 | 2715 1810 1309 601 | 3 |
| 2000 | 2167 1911 1757 600 | 4 |
| 2000 | 600 2167 1911 1757 | 5 |

**جدول 1-19 اجرای الگوریتم Kmeans برای ایجاد 4 خوشه روی مجموعه داده ای satimage**

|  |  |  |
| --- | --- | --- |
| نسل | تعداد عناصر در هر خوشه | تعداد اجرا |
| 2000 | 1560 600 1018 1600 1657 | 1 |
| 2000 | 1560 1606 1017 600 1652 | 2 |
| 2000 | 1154 573 2019 1046 1643 | 3 |
| 2000 | 1560 1602 1017 600 1656 | 4 |
| 2000 | 1659 600 1020 1596 1560 | 5 |

**جدول 1-20 اجرای الگوریتم Kmeans برای ایجاد 5 خوشه روی مجموعه داده ای satimage**

|  |  |  |
| --- | --- | --- |
| نسل | تعداد عناصر در هر خوشه | تعداد اجرا |
| 2000 | 1369 1511 967 1155 865 568 | 1 |
| 2000 | 1005 282 1653 1149 341 2005 | 2 |
| 2000 | 1511 568 863 1365 1156 972 | 3 |
| 2000 | 972 568 1509 863 1365 1158 | 4 |
| 2000 | 1368 1156 972 864 568 1507 | 5 |

**جدول 1-21 اجرای الگوریتم Kmeans برای ایجاد 6 خوشه روی مجموعه داده ای satimage**

|  |  |  |
| --- | --- | --- |
| نسل | تعداد عناصر در هر خوشه | تعداد اجرا |
| 2000 | 956 959 568 1234 733 1124 861 | 1 |
| 2000 | 867 953 1149 985 568 873 1040 | 2 |
| 2000 | 1104 567 1392 779 1349 596 648 | 3 |
| 2000 | 985 953 568 873 867 1149 1040 | 4 |
| 2000 | 343 970 275 1365 1153 814 1515 | 5 |

**جدول 1-22 اجرای الگوریتم Kmeans برای ایجاد7 خوشه روی مجموعه داده ای satimage**

|  |  |  |
| --- | --- | --- |
| نسل | تعداد عناصر در هر خوشه | تعداد اجرا |
| 2000 | 837 412 919 820 569 1036 895 947 | 1 |
| 2000 | 1153 985 334 954 1036 865 824 284 | 2 |
| 2000 | 1030 566 639 771 1054 978 552 845 | 3 |
| 2000 | 953 1108 785 495 1228 147 766 953 | 4 |
| 2000 | 1037 648 543 803 846 1027 567 964 | 5 |

**جدول 1-23 اجرای الگوریتم Kmeans برای ایجاد 8 خوشه روی مجموعه داده ای satimage**

|  |  |
| --- | --- |
| تعداد عناصر در هر خوشه | تعداد اجرا |
| 1823 341 516 3755 | 1 |
| 1823 341 516 3755 | 2 |
| 1823 341 516 3755 | 3 |
| 1823 341 516 3755 | 4 |
| 1823 341 516 3755 | 5 |

**جدول 1-24 اجرای الگوریتم Hierarchical برای ایجاد 4 خوشه روی مجموعه داده ای satimage**

|  |  |
| --- | --- |
| تعداد عناصر در هر خوشه | تعداد اجرا |
| 21 1823 341 516 3734 | 1 |
| 21 1823 341 516 3734 | 2 |
| 21 1823 341 516 3734 | 3 |
| 21 1823 341 516 3734 | 4 |
| 21 1823 341 516 3734 | 5 |

**جدول 1-25 اجرای الگوریتم Hierarchical برای ایجاد 5 خوشه روی مجموعه داده ای satimage**

|  |  |
| --- | --- |
|  |  |
| **شکل1-16 نمودار حاصل از توابع ارزیابی برای جدول 1-13؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد** | **شکل1-17 نمودار حاصل از توابع ارزیابی برای جدول 1-14؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد** |
|  |  |
| **شکل1-18 نمودار حاصل از توابع ارزیابی برای جدول 1-15؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد** | **شکل1-19 نمودار حاصل از توابع ارزیابی برای جدول 1-16؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زر** |
| **شکل1-20 نمودار حاصل از توابع ارزیابی برای جدول 1-17؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد** |  |

|  |  |
| --- | --- |
| تعداد عناصر در هر خوشه | تعداد اجرا |
| 543 21 1823 341 3191 516 | 1 |
| 543 21 1823 341 3191 516 | 2 |
| 543 21 1823 341 3191 516 | 3 |
| 543 21 1823 341 3191 516 | 4 |
| 543 21 1823 341 3191 516 | 5 |

**جدول 1-26 اجرای الگوریتم Hierarchical برای ایجاد 6 خوشه روی مجموعه داده ای satimage**

|  |  |
| --- | --- |
| تعداد عناصر در هر خوشه | تعداد اجرا |
| 215 543 21 1823 301 341 3191 | 1 |
| 215 543 21 1823 301 341 3191 | 2 |
| 215 543 21 1823 301 341 3191 | 3 |
| 215 543 21 1823 301 341 3191 | 4 |
| 215 543 21 1823 301 341 3191 | 5 |

**جدول 1-27 اجرای الگوریتم Hierarchical برای ایجاد 7 خوشه روی مجموعه داده ای satimage**

|  |  |
| --- | --- |
| تعداد عناصر در هر خوشه | تعداد اجرا |
| 19 215 543 21 1823 3172 30 341 | 1 |
| 19 215 543 21 1823 3172 30 341 | 2 |
| 19 215 543 21 1823 3172 30 341 | 3 |
| 19 215 543 21 1823 3172 30 341 | 4 |
| 19 215 543 21 1823 3172 30 341 | 5 |

**جدول 1-28 اجرای الگوریتم Hierarchical برای ایجاد 8 خوشه روی مجموعه داده ای satimage**

|  |  |  |
| --- | --- | --- |
| نسل | تعداد عناصر در هر خوشه | تعداد اجرا |
| 9 | 2110 596 1905 1824 | 1 |
| 9 | 1905 596 2110 1824 | 2 |
| 9 | 2110 1905 596 1824 | 3 |
| 9 | 2110 1824 1905 596 | 4 |
| 9 | 2110 1824 596 1905 | 5 |

**جدول 1-29 اجرای الگوریتم SOM برای ایجاد 4 خوشه روی مجموعه داده ای satimage**

|  |  |
| --- | --- |
|  |  |
| **شکل1-21 نمودار حاصل از توابع ارزیابی برای جدول 1-24؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زر**د | **شکل1-22 نمودار حاصل از توابع ارزیابی برای جدول 1-25؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد** |
|  |  |
| **شکل1-23 نمودار حاصل از توابع ارزیابی برای جدول 1-26؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد** | **شکل1-24 نمودار حاصل از توابع ارزیابی برای جدول 1-27؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زر** |
| **شکل1-25 نمودار حاصل از توابع ارزیابی برای جدول 1-28؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد** |  |

|  |  |  |
| --- | --- | --- |
| نسل | تعداد عناصر در هر خوشه | تعداد اجرا |
| 9 | 1136 1641 1503 1566 589 | 1 |
| 9 | 1641 1503 589 1566 1136 | 2 |
| 9 | 1641 1566 1136 589 1503 | 3 |
| 9 | 1566 1503 589 1641 1136 | 4 |
| 9 | 1503 1641 1136 589 1566 | 5 |

**جدول 1-30 اجرای الگوریتم SOM برای ایجاد 5 خوشه روی مجموعه داده ای satimage**

|  |  |  |
| --- | --- | --- |
| نسل | تعداد عناصر در هر خوشه | تعداد اجرا |
| 9 | 925 1370 1199 564 1491 886 | 1 |
| 9 | 1491 925 564 886 1199 1370 | 2 |
| 9 | 1491 925 1199 886 564 1370 | 3 |
| 9 | 886 1199 1491 1370 564 925 | 4 |
| 9 | 925 1199 886 564 1491 1370 | 5 |

**جدول 1-31 اجرای الگوریتم SOM برای ایجاد 6 خوشه روی مجموعه داده ای satimage**

|  |  |  |
| --- | --- | --- |
| نسل | تعداد عناصر در هر خوشه | تعداد اجرا |
| 10 | 920 860 1282 566 1177 953 677 | 1 |
| 10 | 1282 1177 677 566 920 860 953 | 2 |
| 10 | 1177 677 860 1282 953 920 566 | 3 |
| 10 | 1177 1282 953 677 920 860 566 | 4 |
| 10 | 1177 920 953 860 1282 566 677 | 5 |

**جدول 1-32 اجرای الگوریتم SOM برای ایجاد 7 خوشه روی مجموعه داده ای satimage**

|  |  |  |
| --- | --- | --- |
| نسل | تعداد عناصر در هر خوشه | تعداد اجرا |
| 15 | 327 966 1194 776 1477 104 217 1374 | 1 |
| 15 | 1477 966 327 104 217 1194 1374 776 | 2 |
| 15 | 217 776 1194 327 1374 1477 966 104 | 3 |
| 15 | 1477 217 1194 966 776 327 1374 104 | 4 |
| 15 | 1194 776 1477 104 966 327 217 1374 | 5 |

**جدول 1-33 اجرای الگوریتم SOM برای ایجاد 8 خوشه روی مجموعه داده ای satimage**

|  |  |
| --- | --- |
|  |  |
| **شکل1-26 نمودار حاصل از توابع ارزیابی برای جدول 1-29؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زر**د | **شکل1-27 نمودار حاصل از توابع ارزیابی برای جدول 1-30؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد** |
|  |  |
| **شکل1-28 نمودار حاصل از توابع ارزیابی برای جدول 1-31؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زرد** | **شکل1-29 نمودار حاصل از توابع ارزیابی برای جدول 1-32؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زر** |
| **شکل1-30 نمودار حاصل از توابع ارزیابی برای جدول 1-33؛ شاخص دیوس بلدین با رنگ قرمز؛ شاخص RMSSDT با رنگ سبز؛ شاخص RS با رنگ زر** |  |

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

با توجه جدول های 1-19 تا 1-23 و شکل های 1-16 تا 1-20، از بین 25 اجرای الگوریتم کا تا میانگین (Kmeans) نتایج زیر حاصل می شود:

|  |  |  |  |
| --- | --- | --- | --- |
| مقدار شاخص | شماره اجرا | بهترین تعداد خوشه |  |
|  | 4 | 7 | **شاخص دیویس بلدین** |
|  | 1 | 8 | **شاخص RMSSDT** |
|  | 1 | 8 | **شاخص RS** |

**جدول 1-34 ارزیابی اجرای الگوریتم Kmeans با توجه به جدول های 1-19 تا 1-23 و شکل های 1-16 تا 1-20 روی مجموعه داده ای satimage**

با توجه جدول های 1-24 تا 1-28 و شکل های 1-21 تا 1-25، از بین 25 اجرای الگوریتم سلسله مراتبی پایین به بالا(Hierarchical )نتایج زیر حاصل می شود :

|  |  |  |  |
| --- | --- | --- | --- |
| مقدار شاخص | شماره اجرا | بهترین تعداد خوشه |  |
|  | 1 | 4 | **شاخص دیویس بلدین** |
|  | 1 | 8 | **شاخص RMSSDT** |
|  | 1 | 8 | **شاخص RS** |

**جدول 1-35** اجرای **الگوریتم Hierarchical با توجه به جدول های 1-24 تا 1-82 و شکل های 1-21 تا 1-25 روی مجموعه داده ای satimage**

با توجه جدول های 1-29 تا 1-33 و شکل های 1-26 تا 1-30، از بین 25 اجرای الگوریتم نگاشت خود سازمان دهی شده(SOM) نتایج زیر حاصل می شود:

|  |  |  |  |
| --- | --- | --- | --- |
| مقدار شاخص | شماره اجرا | بهترین تعداد خوشه |  |
|  | 1 | 4 | **شاخص دیویس بلدین** |
|  | 3 | 8 | **شاخص RMSSDT** |
|  | 3 | 8 | **شاخص RS** |

**جدول 1-36 اجرای الگوریتم SOM با توجه به جدول های 1-29 تا 1-33 و شکل های 1-26 تا 1-30 روی مجموعه داده ای satimage**

1. ***نتیجه گیری***

با توجه به جداول 1-1 تا 1-15 و شکل های 1-4 تا 1-15 بهترین خوشه بندی ها از بین سه الگوریتم بیان شده روی مجموعه داده ای ایریس در جدول 1-37 آمده است.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| الگوریتم | مقدار شاخص | شماره اجرا | بهترین تعداد خوشه |  |
| SOM و Kmeans |  | 1 | 2 | **شاخص دیویس بلدین** |
| SOM |  | 2 | 5 | **شاخص RMSSDT** |
| SOM |  | 2 | 5 | **شاخص RS** |

**جدول 1-37 بهترین نتایج از بین 12 اجرای الگوریتم های خوشه بندی برای مجموعه داده ای iris**

با توجه به جداول 1-91 تا 1-33 و شکل های 1-16 تا 1-30 بهترین خوشه بندی ها از بین سه الگوریتم بیان شده روی مجموعه داده ای ست ایمیج در جدول 1-38 آمده است.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| الگوریتم | مقدار شاخص | شماره اجرا | بهترین تعداد خوشه |  |
| Hierarchical |  | 1 | 4 | **شاخص دیویس بلدین** |
| Kmeans |  | 1 | 8 | **شاخص RMSSDT** |
| SOM |  | 3 | 8 | **شاخص RS** |

**جدول 1-38 بهترین نتایج از بین 20 اجرای الگوریتم های خوشه بندی برای مجموعه داده ای satimage**

1. clustering [↑](#footnote-ref-1)
2. kmeans [↑](#footnote-ref-2)
3. Hierarchical bottom-up [↑](#footnote-ref-3)
4. Self-Organizing Mapps ( SOM) [↑](#footnote-ref-4)
5. unsupervised [↑](#footnote-ref-5)
6. clusters [↑](#footnote-ref-6)
7. Minkowski [↑](#footnote-ref-7)
8. Manhattan [↑](#footnote-ref-8)
9. iris [↑](#footnote-ref-9)
10. satimage [↑](#footnote-ref-10)
11. center [↑](#footnote-ref-11)
12. C.von der Malsburg [↑](#footnote-ref-12)
13. Kohonen [↑](#footnote-ref-13)
14. Davis Bouldin Index [↑](#footnote-ref-14)