

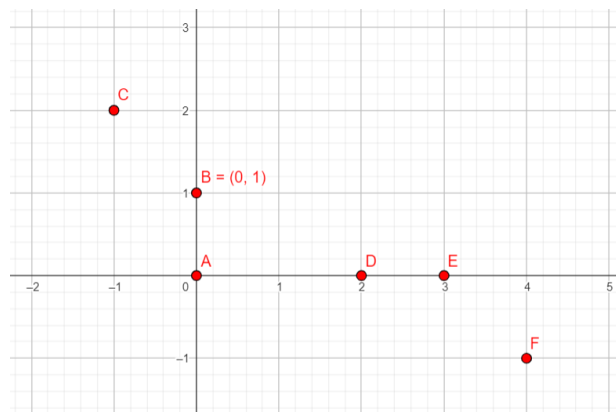
## سوال ۱: الگوریتم خوشه‌بندی (تحلیلی)

### قسمت الف: خوشه‌بندی به روش k-Means

همانطور که در صورت تمرین گفته شده است، فاصله بین دو نقطه  $x_i$  و  $x_j$  به صورت زیر بدست می‌آید.

$$D_{ij} = \text{dist}_{\text{eucl.}}(x_i, x_j)^2 = (x_{i,1} - x_{j,1})^2 + (x_{i,2} - x_{j,2})^2$$

تصویر ۱-۱ داده‌های صورت تمرین را در صفحه دو بعدی  $x_1 - x_2$  نشان می‌دهد.



تصویر ۱-۱: داده‌های اولیه در صفحه دو بعدی  $x_1 - x_2$

در ابتدا، دو نقطه  $A$  و  $F$  را به عنوان مرجع در نظر می‌گیریم. در این صورت ماتریس فواصل نقاط تا نقاط مرجع  $C_1$  و  $C_2$  مطابق

جدول ۱-۱ بدست می‌آید.

$$C_1 = A = (0,0)$$

$$C_2 = F = (4,-1)$$

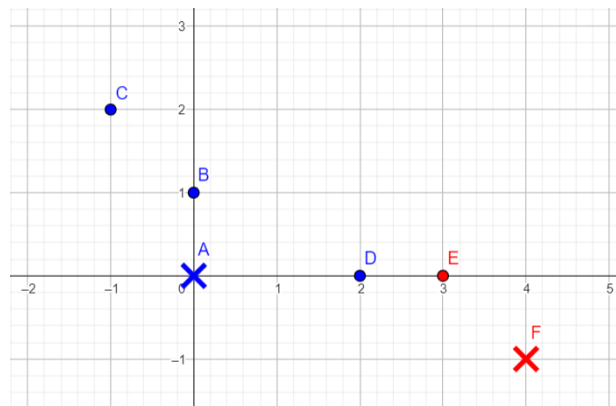
$x$	$A$	$B$	$C$	$D$	$E$	$F$
$d(x, C_1)$	0	1	5	4	9	17
$d(x, C_2)$	17	20	34	5	2	0

جدول ۱-۱: ماتریس فواصل نقاط  $A, B, C, D, E$  و  $F$  تا نقاط مرجع  $C_1 = A = (0,0)$  و  $C_2 = F = (4,-1)$

در نتیجه خوشه‌ها به صورت زیر خواهد بود.

$$C_1: \{A, B, C, D\}, \quad C_2: \{E, F\}$$

تصویر ۱-۲، خوشه‌ها را به همراه مراکز آن‌ها نشان می‌دهد.



تصویر ۱-۲: نقاط هر خوشه و مراکز خوشه‌ها

حال نقاط مرجع به شکل زیر تغییر می‌کند.

$$C_1 = \frac{A + B + C + D}{4} = \left(\frac{1}{4}, \frac{3}{4}\right)$$

$$C_2 = \frac{E + F}{2} = \left(\frac{7}{2}, -\frac{1}{2}\right)$$

در این صورت ماتریس فواصل نقاط تا نقاط مرجع  $C_1$  و  $C_2$  مطابق جدول ۱-۲ بدست می‌آید.

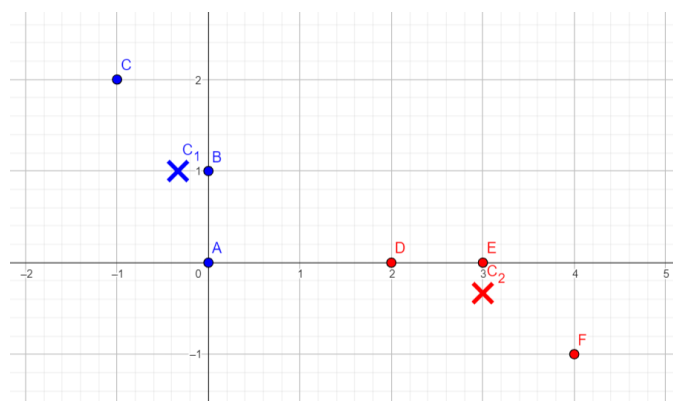
$x$	$A$	$B$	$C$	$D$	$E$	$F$
$d(x, C_1)$	0.625	0.125	3.125	3.625	8.125	17.125
$d(x, C_2)$	12.5	14.5	26.5	2.5	0.5	0.5

جدول ۱-۲: ماتریس فواصل نقاط  $A, B, C, D, E$  و  $F$  تا نقاط مرجع  $C_1 = (\frac{1}{4}, \frac{3}{4})$  و  $C_2 = (\frac{7}{2}, -\frac{1}{2})$

در نتیجه خوشه‌ها به صورت زیر خواهد بود.

$$C_1: \{A, B, C\}, \quad C_2: \{D, E, F\}$$

تصویر ۱-۳، خوشه‌ها را به همراه مراکز آن‌ها نشان می‌دهد.



تصویر ۱-۳: نقاط هر خوشه و مراکز خوشه‌ها

حال نقاط مرجع به شکل زیر تغییر می‌کند.

$$C_1 = \frac{A + B + C}{3} = \left(-\frac{1}{3}, 1\right)$$

$$C_2 = \frac{D + E + F}{3} = \left(3, -\frac{1}{3}\right)$$

در این صورت ماتریس فواصل نقاط تا نقاط مرجع  $C_1$  و  $C_2$  مطابق جدول ۳-۱ بدست می‌آید.

$x$	$A$	$B$	$C$	$D$	$E$	$F$
$d(x, C_1)$	1.1111	0.1111	1.4444	6.4444	12.1111	22.7778
$d(x, C_2)$	9.1111	10.7778	21.4444	1.1111	0.1111	1.4444

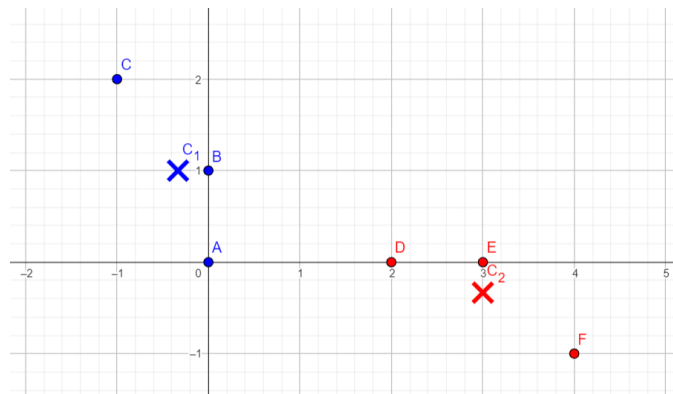
جدول ۳-۱: ماتریس فواصل نقاط  $A, B, C, D, E$  و  $F$  تا نقاط مرجع  $C_1 = (3, -\frac{1}{3})$  و  $C_2 = (-\frac{1}{3}, -1)$

در نتیجه خوشه‌ها به صورت زیر خواهد بود که نسبت به خوشه‌های قبلی تغییر نکرده است.

$$C_1 = \left(-\frac{1}{3}, 1\right), \quad C_1: \{A, B, C\}$$

$$C_2 = \left(3, -\frac{1}{3}\right), \quad C_2: \{D, E, F\}$$

تصویر ۴-۱، خوشه‌های نهایی را به همراه مراکز آن‌ها برای داده‌های مسئله را نشان می‌دهد.



تصویر ۴-۱: نقاط هر خوشه و مراکز خوشه‌ها

– نقطه  $x^* = (1, 1)$  در کدام خوشه قرار می‌گیرد؟

برای اینکه تشخیص دهیم،  $x^* = (1, 1)$  در کدام خوشه قرار می‌گیرد، باید فاصله آنرا تا مرکز هر خوشه بدست آوریم.

$$D(C_1, x^*) = \left(1 - \left(-\frac{1}{3}\right)\right)^2 + (1 - 1)^2 = \frac{16}{9}$$

$$D(C_2, x^*) = (1 - 3)^2 + \left(1 - \left(-\frac{1}{3}\right)\right)^2 = 8$$

## قسمت ب: خوشه بندی سلسله مراتبی

ب. ۱) پیوند واحد (Single Linkage): در این روش، از کمترین فاصله شروع کرده و خوشه بندی را آغاز می‌کنیم. در هر مرحله کمترین فاصله هر داده با خوشه را انتخاب کرده و جدول فواصل را به روز کرده و مرز خوشه بندی را مشخص می‌نمائیم. جدول ۱-۴، خوشه بندی به روش سلسله مراتبی پیوند واحد را مرحله به مرحله نشان می‌دهد.

	A	B	C	D	E	F
A	0					
B	0.12	0				
C	0.51	0.25	0			
D	0.84	0.16	0.14	0		
E	0.28	0.77	0.7	0.45	0	
F	0.34	0.61	0.93	0.2	0.67	0

الف) مرحله ۱

	A, B	C	D	E	F
A, B	0				
C	0.25	0			
D	0.16	0.14	0		
E	0.28	0.7	0.45	0	
F	0.34	0.93	0.2	0.67	0

ب) مرحله ۲

	A, B	C, D	E	F
A, B	0			
C, D	0.16	0		
E	0.28	0.45	0	
F	0.34	0.2	0.67	0

پ) مرحله ۳

	A, B, C, D	E	F
A, B, C, D	0		
E	0.28	0	
F	0.2	0.67	0

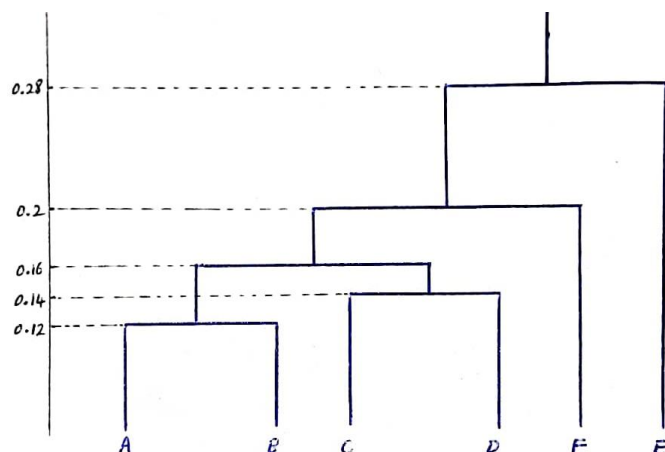
ت) مرحله ۴

	A, B, C, D, F	E
A, B, C, D, F	0	
E	0.28	0

(ث) مرحله ۵

جدول ۴-۱: مراحل خوشه بندی روش پیوند واحد (single linkage)

تصویر ۵-۱، نمودار درختی خوشه بندی را برای روش پیوند واحد (single linkage) نشان می‌دهد.



تصویر ۵-۱: نمودار درختی خوشه بندی به روش پیوند واحد (single linkage)

ب. ۲) پیوند کامل (Complete Linkage): در این روش، از کمترین فاصله شروع کرده و خوشه بندی را آغاز می‌کنیم. در هر مرحله بیشترین فاصله هر داده با خوشه را انتخاب کرده؛ سپس جدول فواصل را به روز کرده و مرز خوشه بندی را مشخص می‌نمائیم.

جدول ۵-۱، خوشه بندی به روش سلسه مراتبی پیوند کامل را مرحله به مرحله نشان می‌دهد.

	A	B	C	D	E	F
A	0					
B	0.12	0				
C	0.51	0.25	0			
D	0.84	0.16	0.14	0		
E	0.28	0.77	0.7	0.45	0	
F	0.34	0.61	0.93	0.2	0.67	0

(الف) مرحله ۱

	A, B	C	D	E	F
A, B	0				
C	0.25	0			

D	<u>0.84</u>	0.14	0		
E	0.77	<u>0.7</u>	0.45	0	
F	0.61	<u>0.93</u>	0.2	0.67	0

(پ) مرحله ۲

	A, B	C, D	E	F
A, B	0			
C, D	0.84	0		
E	<u>0.77</u>	0.7	0	
F	<b>0.61</b>	<u>0.93</u>	0.67	0

(پ) مرحله ۳

	A, B, F	C, D	F
A, B, F	0		
C, D	<u>0.93</u>	0	
E	0.77	<b>0.7</b>	0

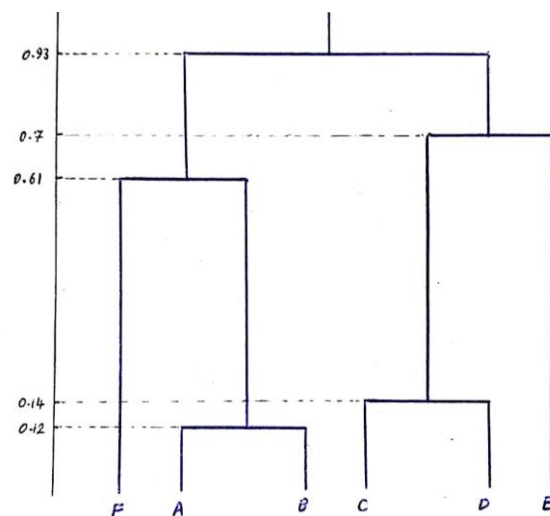
(ت) مرحله ۴

	A, B, F	C, D, E
A, B, F	0	
C, D, E	<b>0.93</b>	0

(ث) مرحله ۵

جدول ۱-۵: مراحل خوشه بندی روش پیوند کامل (complete linkage)

تصویر ۱-۶، نمودار درختی خوشه بندی را برای روش پیوند کامل (complete linkage) نشان می‌دهد.



تصویر ۱-۶: نمودار درختی خوشه بندی به روش پیوند کامل (complete linkage)

## ب. ۳) مقایسه:

در این قسمت می‌خواهیم دو عدد از ماتریس فواصل را طوری تغییر دهیم تا درخت نهایی الگوریتم Complete Linkage مشابه با درخت نهایی الگوریتم Single Linkage شود. به این ترتیب مراحل Complete Linkage را تا جایی که خوشه‌ها مشابه با Single Linkage است پیش می‌بریم و هر جا تغییر بین دو الگوریتم مشاهده شد، عددی از ماتریس را طوری تغییر می‌دهیم تا دو الگوریتم خروجی یکسانی داشته باشند.

جدول ۱-۶، خوشه بندی به روش سلسه مراتبی پیوند واحد را مرحله به مرحله تا وقتی مشابه با الگوریتم پیوند واحد است، نشان می‌دهد.

	A	B	C	D	E	F
A	0					
B	0.12	0				
C	0.51	0.25	0			
D	0.84	0.16	0.14	0		
E	0.28	0.77	0.7	0.45	0	
F	0.34	0.61	0.93	0.2	0.67	0

الف) مرحله ۱

	A, B	C	D	E	F
A, B	0				
C	0.25	0			
D	0.84	0.14	0		
E	0.77	0.7	0.45	0	
F	0.61	0.93	0.2	0.67	0

ب) مرحله ۲

	A, B	C, D	E	F
A, B	0			
C, D	0.84	0		
E	0.77	0.7	0	
F	0.61	0.93	0.67	0

پ) مرحله ۳

جدول ۱-۶: خوشه بندی به روش سلسه مراتبی پیوند واحد مرحله به مرحله تا زمانی که مشابه با الگوریتم پیوند واحد است

حال جدول مرحله ۳ را باید طوری تغییر دهیم تا مشابه با الگوریتم پیوند واحد، در این مرحله دو خوشه (A,B) و (C,D) جدا شوند. به این منظور عدد 0.84 را طوری به کوچکترین عدد ماتریس تبدیل می‌کنیم. البته باید توجه کنیم که این مقدار در مرحله ۲ به عنوان عدد کوچک انتخاب نشود. (همچنین در نظر می‌گیریم که بزرگترین آستانه در روش پیوند واحد 0.28 است و با انتخاب عددی بزرگتر از آن مطمئن هستیم که نتیجه آن روش نیز تغییر نخواهد کرد.)

$$\max\{0.14, 0.25, 0.28\} < Value_{(A,B),(C,D)} < 0.61$$

به عنوان مثال این مقدار را 0.5 در نظر می‌گیریم. در این صورت جدول مرحله به بعد مطابق با جدول ۷-۱ می‌شود.

	A, B	C, D	E	F
A, B	0			
C, D	0.5	0		
E	0.77	0.7	0	
F	0.61	0.93	0.67	0

الف) مرحله ۳

	A, B, C, D	E	F
A, B, C, D	0		
E	0.77	0	
F	0.93	0.67	0

ب) مرحله ۴

جدول ۷-۱: خوشه بندی به روش سلسه مراتبی پیوند کامل (برای ماتریس فواصل تغییر یافته)

حال جدول مرحله ۴ را طوری تغییر می‌دهیم که در پس آن خوشه F از (A,B,C,D) جدا شود. به این منظور عدد 0.93 را طوری به کوچکترین عدد ماتریس تبدیل می‌کنیم. البته باید توجه کنیم که این مقدار در مرحله ۳ به عنوان عدد کوچک انتخاب نشود.

$$\max\{0.5, 0.28, 0.2\} < Value_{(A,B,C,D),F} < 0.67$$

به عنوان مثال این مقدار را 0.6 در نظر می‌گیریم. در این صورت جدول مرحله به بعد مطابق با جدول ۸-۱ می‌شود.

	A, B, C, D	E	F
A, B, C, D	0		
E	0.77	0	
F	0.6	0.67	0

الف) مرحله ۴



	A, B, C, D, F	E
A, B, C, D, F	0	
E	0.77	0

(ب) مرحله ۵

جدول ۸-۱: خوشه بندی به روش سلسه مراتبی پیوند کامل (برای ماتریس فواصل تغییر یافته)

در نتیجه اگر دو عنصر ماتریس فواصل را مطابق با جدول ۹-۱ تغییر دهیم، الگوریتم پیوند واحد و کامل شکل درخت مشابهی در نهایت خواهد داشت.

	A	B	C	D	E	F
A	0					
B	0.12	0				
C	0.51	0.25	0			
D	0.5	0.16	0.14	0		
E	0.28	0.77	0.7	0.45	0	
F	0.34	0.61	0.6	0.2	0.67	0

جدول ۹-۱: ماتریس فواصل تغییر یافته برای یکسان شدن الگوریتم های پیوند واحد و کامل

**سوال ۲: پیاده‌سازی الگوریتم خوشه‌بندی****قسمت اول: خوشه‌بندی k-Means ساده**

در این قسمت دو مقدار فاصله درون کلاسی  $Inner\ distance$  و برون کلاسی  $Outer\ distance$  را به صورت زیر در نظر می‌گیریم.

$Inner\ distance$ : مجموع فواصل هر داده تا مرکز خوشه مربوط به خودش

$Outer\ distance$ : مجموع فواصل هر داده تا مراکز خوشه‌های غیر از خوشه خودش

به همین ترتیب مقدار تابع هزینه را به صورت زیر تعریف می‌کنیم و سعی می‌کنیم آن را با استفاده از الگوریتم خوشه‌بندی k-Means کمینه کنیم.

$$cost = Inner\ distance$$

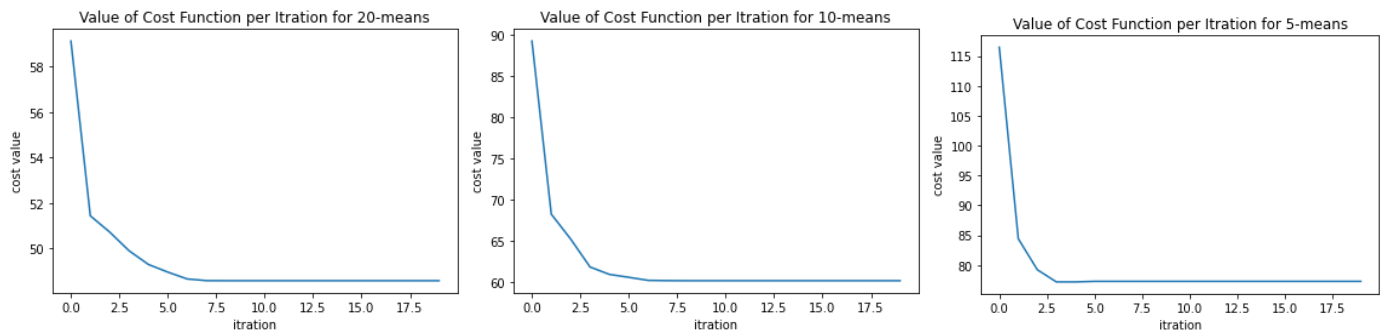
همچنین فاصله نقاط را با استفاده از نرم اقلیدسی به صورت زیر بدست می‌آوریم.

$$D_{ij} = dist_{eucl.}(x_i, x_j) = \sqrt{(x_{i,1} - x_{j,1})^2 + (x_{i,2} - x_{j,2})^2}$$

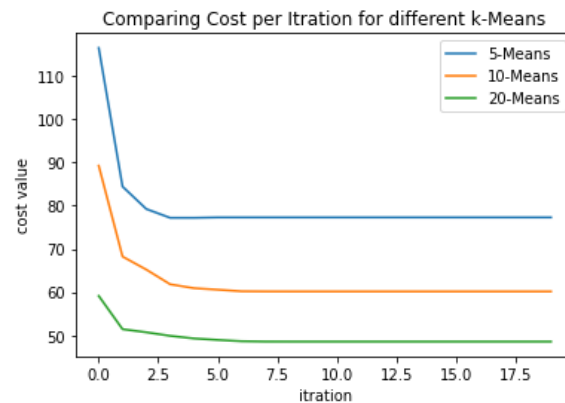
پیاده‌سازی: برای پیاده‌سازی این الگوریتم، ابتدا k داده را به صورت رندوم انتخاب کرده و به عنوان مراکز خوشه‌ها در نظر می‌گیریم. سپس در یک حلقه الگوریتم k-Means را اجرا می‌کنیم. به این منظور هر بار ماتریس فواصل داده از مراکز خوشه‌ها را بدست آورده و بر اساس کمترین فاصله تا مراکز خوشه‌ها، خوشه‌بندی را انجام می‌دهیم. سپس مراکز خوشه‌های جدید را با میانگین گرفتن از داده‌های هر خوشه بدست می‌آوریم. در تابع نوشته شده برای این قسمت، ورودی ITR به این معناست که اگر می‌خواهیم به تعداد Iteration های خاصی الگوریتم را تکرار کنیم، به آن مقدار دلخواه می‌دهیم و در غیر این صورت هنگامی که مراکز بهینه محاسبه می‌شوند، الگوریتم متوقف می‌شود.

**قسمت الف: تأثیر تعداد خوشه‌ها**

در این قسمت برای مقایسه تعداد خوشه‌ها در الگوریتم k-Means مقدار تابع cost را برای هر مقدار k دلخواه رسم می‌کنیم. تصویر ۱-۲ تغییر مقدار تابع هزینه برای مقادیر  $k = \{5, 10, 20\}$  نشان می‌دهد. همانطور که در این تصاویر مشاهده می‌شود، الگوریتم برای هر مقدار k بعد از حدود ۵ تکرار همگرا شده است. همچنین تصویر ۲-۲ نیز مقایسه سه نمودار را در یک نمودار نشان می‌دهد. همانطور که در این تصویر می‌بینیم، مقدار تابع هزینه برای k های بزرگتر کمتر شده است. این امر طبیعی است؛ زیرا با افزایش تعداد خوشه‌ها فاصله داده‌ها تا مراکز کمتر می‌شود. باید توجه داشته باشیم که این موضوع لزوماً این نتیجه را نمی‌دهد که الگوریتم به ازای k های بزرگتر بهتر عمل می‌کند.



تصویر ۱-۲: تابع هزینه (cost) برحسب iteration برای مقادیر مختلف k



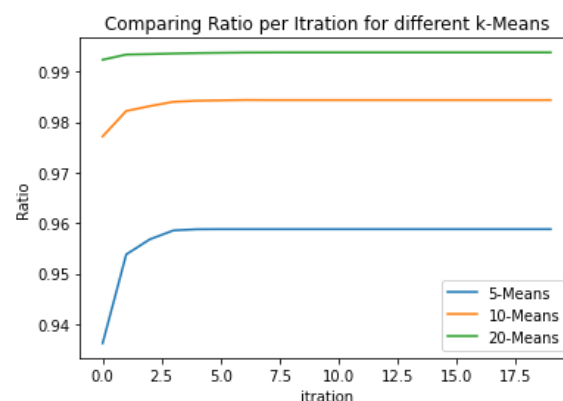
تصویر ۲-۲: مقایسه تابع هزینه (cost) برحسب iteration برای مقادیر مختلف k

## قسمت ب: تأثیر تکرار آزمایش

در این قسمت ابتدا تابع Ratio را به صورت زیر تعریف می‌کنیم.

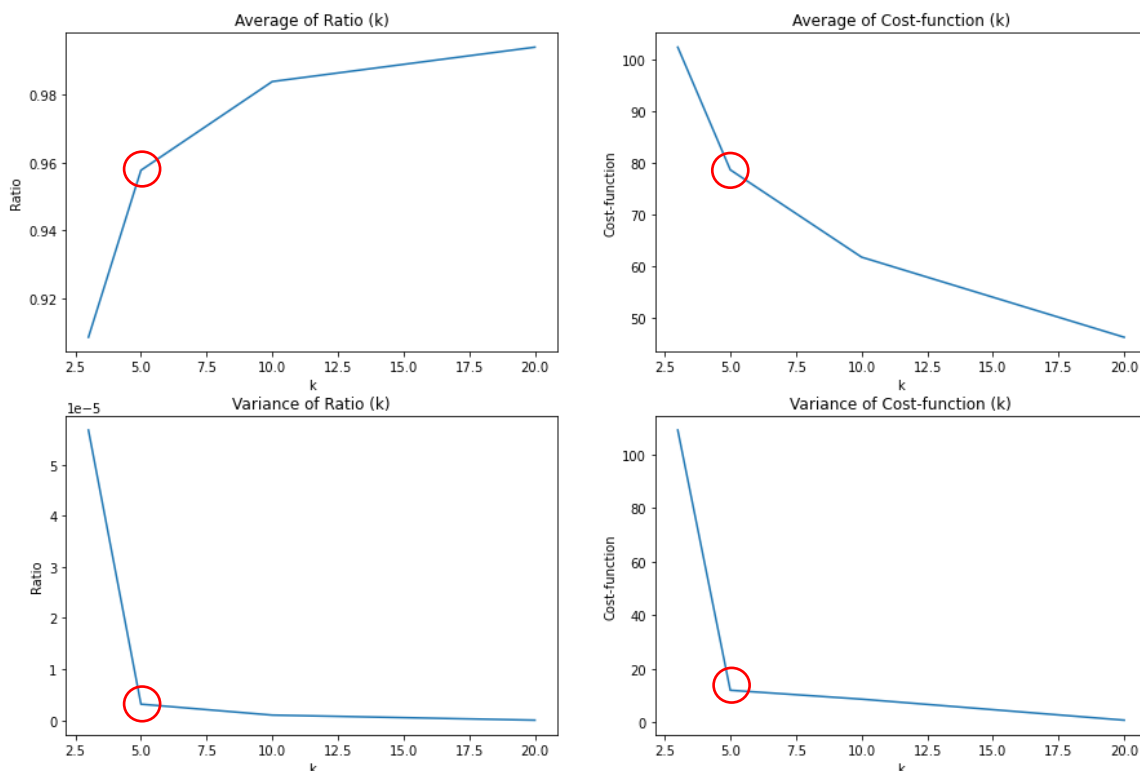
$$Ratio = \frac{Outer_{distance}}{Outer_{distance} + Inner_{distance}}$$

حال مشابه با قسمت قبل نمودار Ratio برحسب iteration را برای هر سه مقدار k مطابق با تصویر ۳-۲ رسم می‌کنیم. این تابع تقریباً مشابه با cost است با این تفاوت که با کم شدن فاصله درون کلاسی و هم‌چنین افزایش فاصله برون کلاسی مقدار Ratio بعد از چند تکرار به 1 میل خواهد کرد. باید توجه داشته باشیم که این موضوع لزوماً این نتیجه را نمی‌دهد که الگوریتم به ازای k های بزرگتر بهتر عمل می‌کند.



تصویر ۳-۲: مقایسه تابع Ratio برحسب iteration برای مقادیر مختلف k

حال می‌خواهیم الگوریتم را ۲۰ بار تکرار کنیم و سپس بین مقادیر نهایی Cost و Ratio هر بار، میانگین و واریانس را بدست آوریم. برای مقادیر  $k = \{3, 5, 10, 20\}$  این کار را انجام می‌دهیم. تصویر ۲-۴ میانگین و واریانس Cost و Ratio را برحسب مقادیر  $k$  نشان می‌دهد. این نمودار ها می‌تواند شهود بهتری برای اینکه بفهمیم کدام مقدار  $k$  بهتر عمل می‌کند به ما می‌دهد. همانطور که در تصویر مشاهده می‌شود، Elbow point نقطه  $k = 5$  است. البته نمی‌توان تشخیص داده که برای  $k < 5$  الگوریتم بهتر عمل می‌کند یا خیر.



تصویر ۲-۴: نمودار میانگین و واریانس Cost و Ratio را برحسب مقادیر  $k = \{3, 5, 10, 20\}$

Average and Variance of Ratio and Cost-function:

```
for k = 3: Average-Ratio = 0.9084789310929973, Variance-Ratio = 5.671693476168509e-05, Average-Cost = 102.43300960994789, Variance-Cost = 109.07631557113606
for k = 5: Average-Ratio = 0.9576949345028251, Variance-Ratio = 3.174166544977957e-06, Average-Cost = 78.71303130133334, Variance-Cost = 11.977328307782589
for k = 10: Average-Ratio = 0.983843216619799, Variance-Ratio = 1.0374128410723473e-06, Average-Cost = 61.82431957661487, Variance-Cost = 8.655162365805863
for k = 20: Average-Ratio = 0.9939752792758553, Variance-Ratio = 4.8088957622912295e-08, Average-Cost = 46.3281753129137, Variance-Cost = 0.8111262850881756
```

تصویر ۲-۵: مقادیر میانگین و واریانس Cost و Ratio را برحسب مقادیر  $k = \{3, 5, 10, 20\}$

## قسمت دوم: خوشه‌بندی k-Means هوشمند

در این قسمت مطابق با توضیحات صورت تمرین می‌خواهیم خوشه‌بندی را طوری انجام دهیم که شروط داده شده برقرار باشد. حال می‌خواهیم الگوریتمی که منجر به برقرار شدن شروط در انتها می‌شود را توضیح دهیم و آن را پیاده سازی کنیم.

## قسمت الف: طراحی و پیاده‌سازی الگوریتم

الگوریتم مورد نظر در این قسمت در کل مشابه با الگوریتم k-means خواهد بود، با این تفاوت که در پایان هر iteration توقف می‌کنیم و هر شرط را بررسی می‌کنیم.

در ابتدا براساس تعداد شرط‌های خواسته شده، به صورت رندوم از شروط داده شده انتخاب می‌کنیم. در مرحله توقف و بررسی شرط‌ها هر بار یک شرط را بررسی می‌کنیم. می‌دانیم شرط‌ها دو حالت دارند. اگر دو اندیس داده شده در یک خوشه باشند،  $state = 1$  و اگر هم خوشه نباشند،  $state = -1$  خواهد بود. به همین منظور هر حالت را جداگانه بررسی می‌کنیم و کاری می‌کنیم که برای خوشه‌بندی در نهایت شرط‌ها برقرار باشند.

**$state = 1$ :** در این حالت بررسی می‌کنیم که آیا دو اندیس داده شده در یک خوشه هستند یا خیر، اگر در یک خوشه نباشند، آنها را به این صورت هم‌خوشه می‌کنیم:

اگر خوشه مربوط به اندیس ۱ ( $ind1$ ) را cluster1 و خوشه مربوط به اندیس ۲ ( $ind2$ ) را cluster2 بنامیم، به دو صورت می‌توانیم دو اندیس را هم‌خوشه کنیم. یک حالت این است که  $ind2$  را وارد خوشه cluster1 کنیم و حالت دیگر این است که  $ind1$  را وارد خوشه cluster2 کنیم. برای اینکه بفهمیم کدام حالت بهتر است تابع هزینه یعنی مجموع فاصله تا مراکز را برای هر حالت بدست می‌آوریم و هرکدام هزینه کمتری داشت را انتخاب می‌کنیم. به این ترتیب دو اندیس هم‌خوشه می‌شوند.

$$cost1 = dist(ind1, cluster1) + dist(ind2, cluster1)$$

$$cost2 = dist(ind1, cluster2) + dist(ind2, cluster2)$$

**$state = 2$ :** در این حالت بررسی می‌کنیم که آیا دو اندیس داده شده در دو خوشه مختلف هستند یا خیر، اگر در یک خوشه باشند، خوشه‌های آنها را به این صورت متفاوت می‌کنیم:

ابتدا سعی می‌کنیم دومین مرکز خوشه نزدیک به هر اندیس را بدست آوریم. اگر خوشه دو اندیس را cluster0 و دومین خوشه مربوط به اندیس ۱ ( $ind1$ ) را cluster1 و دومین خوشه مربوط به اندیس ۲ ( $ind2$ ) را cluster2 بنامیم، به دو صورت می‌توانیم دو اندیس را وارد دو خوشه مختلف کنیم. یک حالت این است که  $ind2$  را وارد خوشه cluster2 کنیم و حالت دیگر این است که  $ind1$  را وارد خوشه cluster1 کنیم. برای اینکه بفهمیم کدام حالت بهتر است تابع هزینه یعنی مجموع فاصله تا مراکز را برای هر حالت بدست می‌آوریم و هرکدام هزینه کمتری داشت را انتخاب می‌کنیم. به این ترتیب دو اندیس هم‌خوشه می‌شوند.

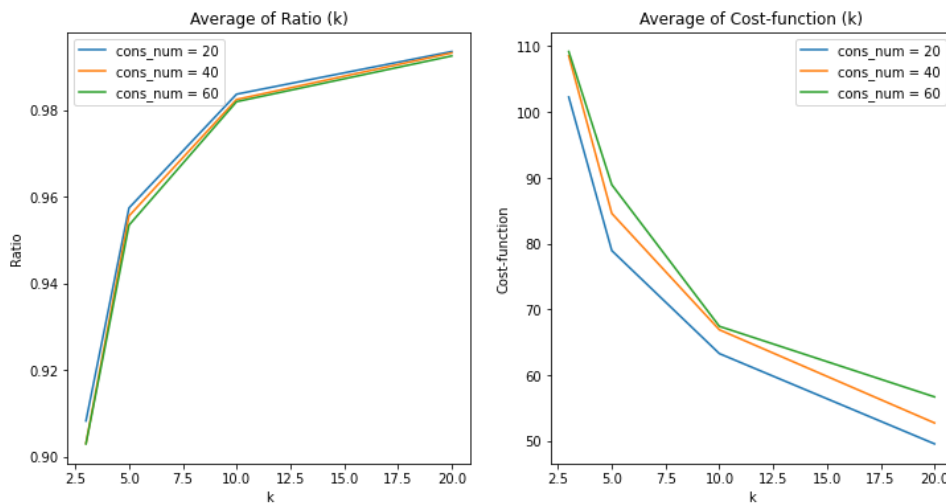
$$cost1 = dist(ind1, cluster0) + dist(ind2, cluster2)$$

$$\text{cost2} = \text{dist}(\text{ind1}, \text{cluster1}) + \text{dist}(\text{ind2}, \text{cluster0})$$

به همین ترتیب این الگوریتم را برای تمام شروط اجرا می‌کنیم. تابع `Constraints_Check()` برای بررسی و برقراری شروط نوشته شده است.

### قسمت ج: ارزیابی الگوریتم

**سوال ۱.** در این قسمت به ازای مقادیر  $k = \{3, 5, 10, 20\}$  و  $\text{size}_{\text{cons}} = \{20, 40, 60\}$  خوشه بندی ها را مقایسه می‌کنیم. به این منظور معیار ارزیابی را میانگین تابع `Ratio` و `Cost` برای تعداد ۱۰ تکرار در نظر گرفته و نمودارهای مربوط به هر حالت را رسم می‌کنیم. تصویر ۲-۶، نمودار میانگین `Ratio` و `Cost` را برحسب  $k$  برای تعداد شروط مختلف نشان می‌دهد.



تصویر ۲-۶: نمودار میانگین `Cost` و `Ratio` را برحسب مقادیر  $k = \{3, 5, 10, 20\}$  و  $\text{size}_{\text{cons}} = \{20, 40, 60\}$

همانطور که در تصویر بالا مشاهده می‌شود، با افزایش تعداد شروط دو تابع `Cost` و `Ratio` از حالت بهینه خارج می‌شود. البته باید توجه کنیم با توجه به اینکه از داده‌ها اطلاعات بیشتری داریم خوشه بندی دقیق تر شده است اما این نکته قابل ملاحظه است که لزوماً فاصله داده هر خوشه تا مرکز خوشه‌اش نباید کمترین فاصله تا مرکز خوشه‌ها را داشته باشد.

**سوال ۲.** حال برای  $k = 3$  می‌خواهیم دقت خوشه بندی را برای الگوریتم خوشه بندی ساده و هوشمند با تعداد شروط خواسته شده مقایسه کنیم. با توجه به اینکه از Label داده‌های می‌دانیم که ۵۰ داده اول و دوم و سوم هر کدام مربوط به یک کلاستر هستند، دقت خوشه بندی را بدست می‌آوریم. (دقت شود لزوماً شماره خوشه با Label داده یکسان نیست به همین خاطر بیشترین تکرار را در هر ۵۰ داده بدست آورده و دقت کل را محاسبه می‌کنیم. مطابق انتظار همانطور که در تصویر ۲-۷ مشاهده می‌شود، دقت خوشه بندی هوشمند نسبت به خوشه بندی ساده بیشتر است و هرچقدر تعداد شروط را بیشتر کنیم این دقت نیز افزایش می‌یابد.

```
Accuracy of 3-Means Simple and Intelligent Clustering:  
Simple Clustering: 88.6666666666667%  
Intelligent Clustering (20 Constraints): 90.0%  
Intelligent Clustering (40 Constraints): 90.6666666666667%  
Intelligent Clustering (60 Constraints): 92.6666666666667%
```

تصویر ۷-۲: دقت الگوریتم خوشه بندی ساده و خوشه بندی هوشمند با  $size_{cons} = \{20,40,60\}$  به ازای  $k = 3$

---