



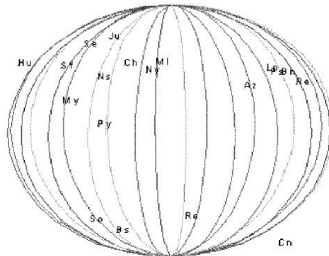
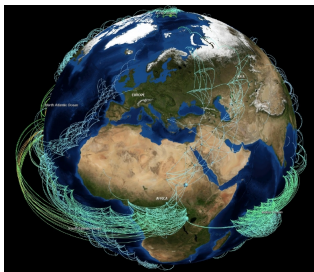
مقیاس گذاری چند بعدی و کاربردها

محمد آرشی

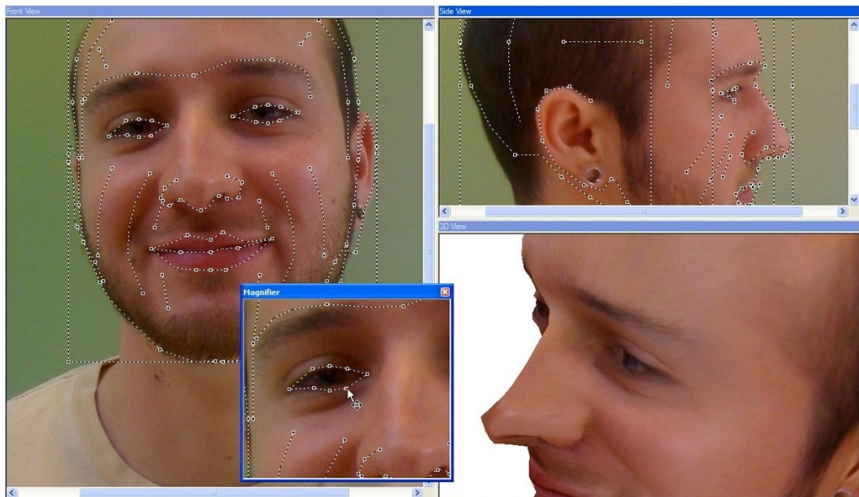
۲۱ بهمن ۱۳۹۴

مقیاس‌گذاری چند بعدی

مقیاس‌گذاری چند بعدی (MDS)، روشی برای تولید یک طرح (بازیابی مکان اشیا) با بعدی کمتر از بعد داده‌ها با استفاده از ماتریس فاصله حاصل از ماتریس داده‌های چند متغیره X است.



شکل: تعیین مکان شهرها با استفاده از روش مقیاس‌گذاری چندبعدی



مقیاس گذاری چند بعدی

انواع مقیاس گذاری:
► مقیاس گذاری کلاسیک

مقیاس گذاری چند بعدی

انواع مقیاس گذاری:

- ▶ مقیاس گذاری کلاسیک
- ▶ مقیاس گذاری متوسط کريتچلی

مقیاس گذاری چند بعدی

انواع مقیاس گذاری:

- ▶ مقیاس گذاری کلاسیک
- ▶ مقیاس گذاری متوسط کريتچلی
- ▶ مقیاس گذاری نامتری

مقیاس‌گذاری چند بعدی

انواع مقیاس‌گذاری:

- ▶ مقیاس‌گذاری کلاسیک
- ▶ مقیاس‌گذاری متوسط کريتچلی
- ▶ مقیاس‌گذاری نامتری
- ▶ مقیاس‌گذاری تک بعدی

مقیاس گذاری چند بعدی

انواع مقیاس گذاری:

- ▶ مقیاس گذاری کلاسیک
- ▶ مقیاس گذاری متوسط کريتچلی
- ▶ مقیاس گذاری نامتری
- ▶ مقیاس گذاری تک بعدی
- ▶ مقیاس گذاری متری کمترین توان های دوم

مقیاس‌گذاری چند بعدی

انواع مقیاس‌گذاری:

- ▶ مقیاس‌گذاری کلاسیک
- ▶ مقیاس‌گذاری متوسط کریتچلی
- ▶ مقیاس‌گذاری نامتری
- ▶ مقیاس‌گذاری تک بعدی
- ▶ مقیاس‌گذاری متری کمترین توان‌های دوم
- ▶ مقیاس‌گذاری بیزی



روش یافتن مختصات واقعی بر اساس
فاصله اقلیدسی اولین بار توسط شونبرگ
(۱۹۳۵) مطرح شد.



یانگ و هولدر (۱۹۳۸) نشان دادند که
چطور با ماتریس فاصله بین نقاط در یک
فضای اقلیدسی می‌توان مختصات نقاط را
به طوری که فاصله‌ها حفظ شوند یافت.



گرینا کر و آندرهیل (۱۹۸۲) از مقیاس‌گذاری
چندبعدی برای بروز کردن نقشه شهرهای کشور
آفریقای جنوبی، بر اساس زمان پرواز بین
فرودگاه‌های شهرهای آفریقای جنوبی استفاده
کردند.



اه و رافتری (۲۰۰۱) برای اولین بار
مقیاس‌گذاری چندبعدی را از دیدگاه بیزی
بررسی کردند. و در سال (۲۰۱۲) آه این
مطالعات را تعمیم داد.

مقیاس‌گذاری کلاسیک

در پی بازیابی پیکره اشیا با یک مدل هندسی ساده است.
چنین مدلی با یک مجموعه از نقاط $x_1, x_2, \dots, x_n$ در p بعد نشان داده می‌شود.
هدف:

تعیین بعد مناسب p از مدل
 n مختصات p بعدی $x_1, x_2, \dots, x_n$ است (به طوری که مدل، یک برازش خوب
برای مجاورت‌های مشاهده شده را نتیجه دهد).
حال سوالی که پیش می‌آید این است که چگونه p و مقادیر مختصات
 $x_1, x_2, \dots, x_n$ از ماتریس مجاورت مشاهده شده را برآورد کنیم؟
مقیاس‌گذاری چند بعدی کلاسیک پاسخ این سوال را به درستی می‌دهد.

چگونگی تعیین مختصات‌های جدید

بر اساس

$$X = (x'_1, \dots, x'_n)' \sim n \times p$$

$$x_i = (x_{i1}, \dots, x_{ip})$$

فاصله اقلیدسی بین شی i و j :

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + \dots + (x_{ip} - x_{jp})^2}$$

ماتریس مجاورت D :

$$D = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ & d_{22} & \dots & d_{2n} \\ & & \ddots & \vdots \\ & & & d_{nn} \end{bmatrix}$$

چگونگی تعیین مختصات‌های جدید

مقیاس‌گذاری چندبعدي کلاسیک:

با داشتن فاصله مشاهده شده D (بر اساس مشاهدات p بعدی) چگونه می‌توانیم X را (در صورت امکان با بعدی کمتر) بازیابی کنیم.
فرض کنید:

$$B = XX' \sim n \times n$$

$$B = (b_{ij}), \quad b_{ij} = \sum_{k=1}^p x_{ik}x_{jk}$$

هدف برقرار کردن ارتباط بین مولفه‌های B و D

چگونگی تعیین مختصات‌های جدید

$$b_{ij} = -\frac{1}{\sqrt{2}}(d_{ij}^2 - d_{i.}^2 - d_{.j}^2 + d_{..}^2)$$

چگونگی تعیین مختصات‌های جدید

$$b_{ij} = -\frac{1}{\sqrt{2}}(d_{ij}^{\sqrt{2}} - d_{i.}^{\sqrt{2}} - d_{.j}^{\sqrt{2}} + d_{..}^{\sqrt{2}})$$

که در آن

$$d_{i.}^{\sqrt{2}} = \frac{1}{n} \sum_{j=1}^n d_{ij}^{\sqrt{2}}, \quad d_{.j}^{\sqrt{2}} = \frac{1}{n} \sum_{i=1}^n d_{ij}^{\sqrt{2}}, \quad d_{..}^{\sqrt{2}} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_{ij}^{\sqrt{2}}$$

چگونگی تعیین مختصات‌های جدید

$$B = V_1 \Lambda_1 V_1'$$

که در آن

$p : V_1$ بردار ویژه اول

$p : \Lambda_1$ مقدار ویژه غیر صفر

$$X = V_1 \Lambda_1^{\frac{1}{2}}$$

چگونگی تعیین مختصات‌های جدید

$$B = V_1 \Lambda_1 V_1'$$

که در آن
 $p : V_1$ بردار ویژه اول
 $p : \Lambda_1$ مقدار ویژه غیر صفر

$$X = V_1 \Lambda_1^{\frac{1}{2}}$$

که در آن $\Lambda_1^{\frac{1}{2}} = \text{diag} \left(\lambda_1^{\frac{1}{2}}, \dots, \lambda_p^{\frac{1}{2}} \right)$

چگونگی تعیین مختصات‌های جدید

$$B = V_1 \Lambda_1 V_1'$$

که در آن

$V_1 : p$ بردار ویژه اول

$\Lambda_1 : p$ مقدار ویژه غیر صفر

$$X = V_1 \Lambda_1^{\frac{1}{2}}$$

$$\Lambda_1^{\frac{1}{2}} = \text{diag} \left(\lambda_1^{\frac{1}{2}}, \dots, \lambda_p^{\frac{1}{2}} \right) \quad \text{که در آن}$$

XX' باز یابی کردیم

چگونگی تعیین بعد

استفاده از همه‌ی p بعد منجر به بازیابی کامل ماتریس فاصله اقلیدسی اولیه خواهد شد در حالی که هدف ما استفاده از p بعد نمی‌باشد و به دنبال کاهش بعد هستیم. بنابراین فرض می‌کنیم m بعد ($m < p$) بهترین برازش به وسیله‌ی m بردار ویژه B متناظر با m مقدار ویژه بزرگتر باشد. مناسب بودن m به وسیله معیار زیر داوری می‌شود به طوری که مقادیر P_m بزرگتر یا مساوی $۰/۸$ یک برازش معقول را پیشنهاد می‌کند، که در آن

$$P_m = \frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^n \lambda_i}$$

$$P_m^{(1)} = \frac{\sum_{i=1}^m |\lambda_i|}{\sum_{i=1}^n |\lambda_i|}$$

$$P_m^{(2)} = \frac{\sum_{i=1}^m \lambda_i^2}{\sum_{i=1}^n \lambda_i^2}$$

جدول: مقادیر ویژه

$2.956972e - 02$	$6.990457e - 02$	$1.492622e - 01$	$2.626003e - 01$	$7.359910e - 01$
$-2.849924e - 02$	$-1.279569e - 02$	$-1.139451e - 02$	$1.387779e - 16$	$1.931184e - 02$
	$-1.097833e - 01$	$-7.406143e - 02$	$-5.255450e - 02$	$-4.251502e - 02$

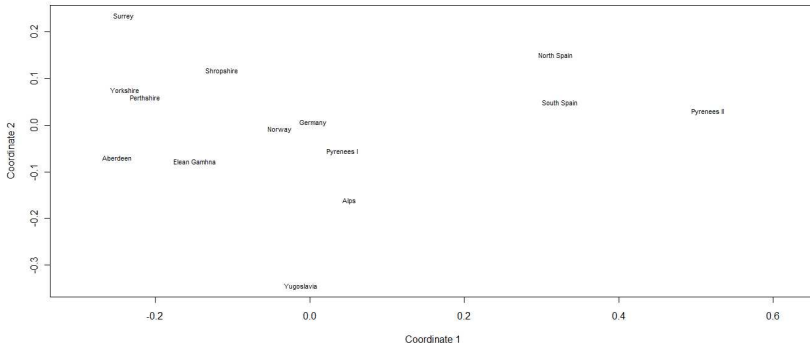
همانطور که مشاهده می‌شود، برخی مقادیر ویژه منفی هستند. لذا برای تعیین بعد مناسب که یک نمایش خوب از ماتریس فاصله‌ی مشاهده شده را داشته‌باشیم، مقادیر معیارهای $P_m^{(1)}$ و $P_m^{(2)}$ را محاسبه می‌کنیم.

جدول: مقادیر $P_m^{(1)}$

0.7925199	0.7925199	0.7804367	0.7619353	0.7181970	0.6248056	0.4605000
1.0000000	0.9313100	0.8849708	0.8520881	0.8254870	0.8076554	0.7996493

جدول: مقادیر $P_m^{(2)}$

0.9649496	0.9649496	0.9643865	0.9630662	0.9556875	0.9220468	0.8179213
1.0000000	0.9818014	0.9735191	0.9693486	0.9666193	0.9653929	0.9651457



شکل: نمایش دو بعدی از مقیاس‌گذاری چند بعدی برای ماتریس فاصله‌ی موش‌های صحرایی-آبی

فواصل هوایی بین شهرها

هارتیگان (۱۹۷۵)، فواصل هوایی بین ۳۰ شهر اصلی جهان را که در جدول نشان داده شده است، را ارائه نمود. شهرها روی سطح زمین به صورت سه بعدی قرار گرفته و هواپیماها در بالای سطح زمین (روی سطح یک کره) حرکت می‌کنند. بنابراین فواصل هوایی دقیقاً فواصل اقلیدسی نیستند و ممکن است انتظار داشته باشیم که بعد x_i ها بین ۲ و ۳ باشد. برای داده‌های فواصل هوایی، مقیاس‌گذاری کلاسیک و بیزی را بکار بردیم.

بعد	<i>STRESS</i>
۱	0.872769
۲	0.524379
۳	0.409542
۴	0.557581
۵	0.5626275

$$Stress = \sqrt{\frac{\sum_{i < j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i < j} d_{ij}^2}}$$

محاسبات مقیاس‌گذاری چندبعدی فاصله‌های ناقلیدسی در R

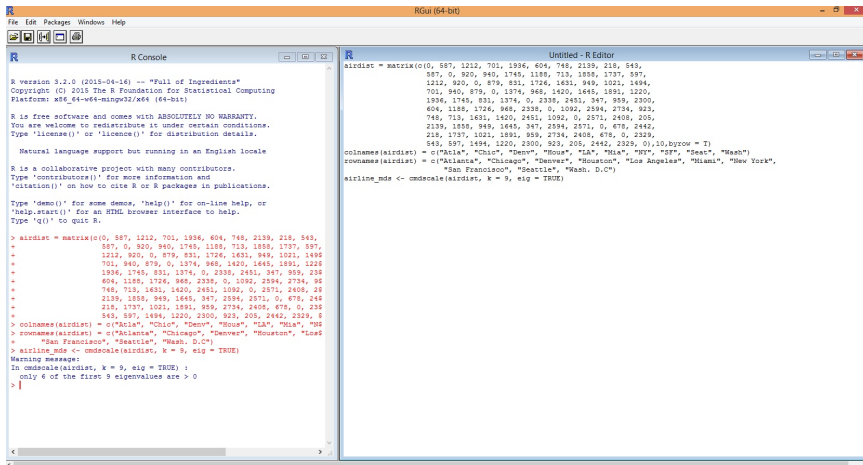
فاصله‌های بین فرودگاه‌های آمریکا

فاصله‌های بین ده فرودگاه در آمریکا را به صورت زیر در نظر بگیرید.

	Atla	Chic	Denv	Hous	LA	Mia	NY	SF	Seat	Wash
Atlanta	-	587	1212	701	1936	604	748	2139	218	543
Chicago	587	-	920	940	1745	1188	713	1858	1737	597
Denver	1212	920	-	879	831	1726	1631	949	1021	1494
Houston	701	940	879	-	1374	968	1420	1645	1891	1220
Los Angeles	1936	1745	831	1374	-	2338	2451	347	959	2300
Miami	604	1188	1726	968	2338	-	1092	2594	2734	923
New York	748	713	1631	1420	2451	1092	-	2571	2408	205
San Francisco	2139	1858	949	1645	347	2594	2571	-	678	2442
Seattle	218	1737	1021	1891	959	2734	2408	678	-	2329
Wash. D. C.	543	597	1494	1220	2300	923	205	2442	2329	-

فاصله ی بین دو فرودگاه، فاصله بین دو نقطه از سطح کره زمین است. بنابراین یک فاصله اقلیدسی نیست.

با تابع `cmdscale` از نرم افزار R می توان مقیاس کلاسیک را برای این فاصله ها به کار برد و مقادیر ویژه ی متناظر با آن را به دست آورد.



```
R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> airdist = matrix(c(0, 587, 1212, 701, 1936, 604, 748, 2139, 218, 543,
+ 587, 0, 920, 940, 1745, 1188, 713, 1858, 1737, 597,
+ 1212, 920, 0, 878, 831, 1726, 1631, 949, 1021, 1495
+ 701, 940, 878, 0, 1374, 968, 1420, 1645, 1891, 1228
+ 1936, 1745, 831, 1374, 0, 2338, 2451, 347, 959, 2300,
+ 604, 1188, 1726, 968, 2338, 0, 1092, 2594, 2734, 923,
+ 748, 713, 1631, 1420, 2451, 1092, 0, 2571, 2408, 205,
+ 2139, 1858, 949, 1645, 347, 2594, 2571, 0, 678, 2442,
+ 218, 1737, 1021, 1891, 959, 2734, 2408, 678, 0, 2329,
+ 543, 597, 1494, 1220, 2300, 923, 205, 2442, 2329, 0), 10, byrow = T)
> colnames(airdist) = c("Atlanta", "Chico", "Denver", "Hous", "LA", "Mia", "NY", "SF", "Seac", "Wash")
> rownames(airdist) = c("Atlanta", "Chicago", "Denver", "Houston", "Los Angeles", "Miami", "New York",
+ "San Francisco", "Seattle", "Wash. D.C")
> airline_mds <- cmdscale(airdist, k = 9, eig = TRUE)

Warning message:
In cmdscale(airdist, k = 9, eig = TRUE) :
  only 6 of the first 9 eigenvalues are > 0
> |
```

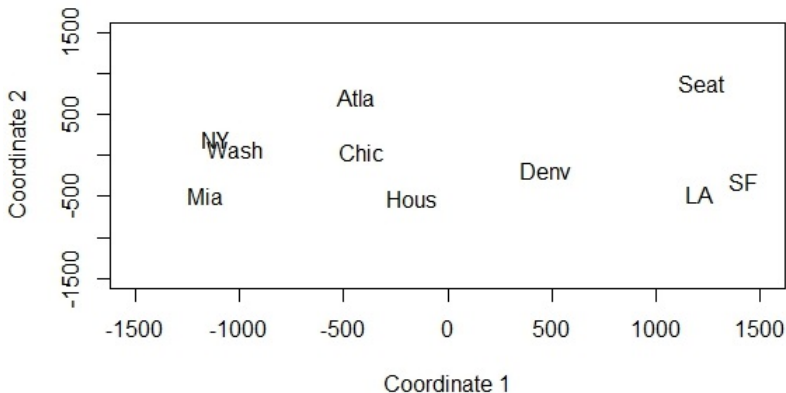
مقادیر ویژه برابر خواهد بود

```
> airline_mds <- cmdscale(airdist, k = 9, eig = TRUE)
Warning message:
In cmdscale(airdist, k = 9, eig = TRUE) :
  only 6 of the first 9 eigenvalues are > 0
> (lam <- airline_mds$eig)
[1] 9.212993e+06 2.200398e+06 1.082981e+06 3.343205e+03
[5] 9.361601e+02 2.019022e+02 -5.823877e-09 -2.147508e+03
[9] -1.014155e+04 -1.723147e+06
```

با توجه به اینکه فاصله‌ها غیر اقلیدسی هستند، مشاهده می‌شود برخی از مقادیر ویژه منفی هستند و بنابراین باید تعیین کنیم چه تعداد مختصات برای نشان دادن ماتریس فاصله مورد نیاز هستند.

```
> cumsum(abs(lam)) / sum(abs(lam))  
[1] 0.6471485 0.8017111 0.8777829 0.8780178 0.8780835 0.8780977  
[7] 0.8780977 0.8782486 0.8789609 1.0000000  
> cumsum(lam^2) / sum(lam^2)  
[1] 0.9042867 0.9558698 0.9683651 0.9683652 0.9683652 0.9683652  
[7] 0.9683652 0.9683652 0.9683663 1.0000000
```

این مقادیر نشان می‌دهند که دو مختص اول نمایشی مناسب از فاصله‌های مشاهده شده را ارائه می‌دهند.



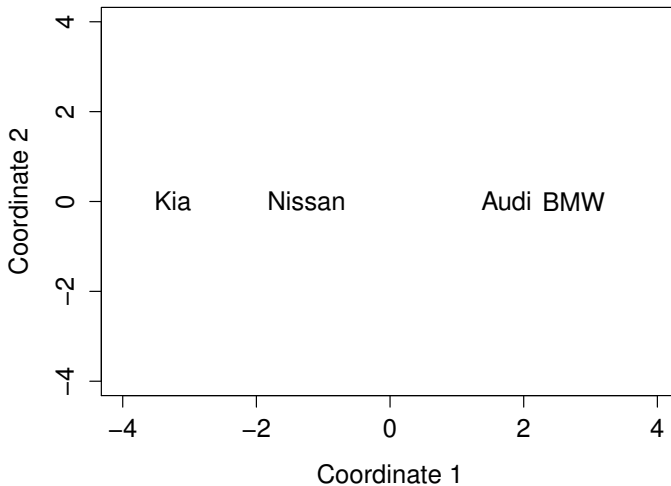
در این نمایش (نقشه) دو بعدی محل جغرافیایی شهرها بر اساس مقیاس گذاری دوبعدی فواصل هوایی به خوبی بازیابی شده است.

مقیاس‌گذاری چندبعدی نامتری

مقایسه‌ی کیفیت اتومبیل‌ها

	Nissan	Kia	BMW	Audi
Nissan	-	-	-	-
Kia	2	-	-	-
BMW	4	6	-	-
Audi	3	5	1	-

جدول: ماتریس عدم تشابه برچسب‌های اتومبیل‌ها



از صبر و حوصله شما سپاسگزارم.