
Subject: Curious Cluster Analysis Conundrum

Posted by [jack.connerney](#) on Wed, 17 Jul 2013 19:03:14 GMT

[View Forum Message](#) <> [Reply to Message](#)

I'm using

```
iwts = CLUST_WTS(ibza,N_CLUSTERS=2)
result = CLUSTER(ibza, iwts, n_clusters=2)
```

to perform a cluster analysis on a 2 by n rows array (first two components of variable "zeros", 20 rows), and, contrary to my expectation, I find that the cluster analysis gives different results when the same program is run on precisely the same data a second time - that is, the cluster is not recognized the first time the pro is run, it is recognized the second time the pro is run, and yet not recognized again the third and forth time the pro is run.

Shouldn't the result of the same computation be the same each time?

Here's the output from two successive runs of the pro - the cluster is recognized the second run (array printed with cluster designation in third column; the cluster we want to identify is marked "1" in the second run).

```
IDL> zcan2,FNAME='3D-021877F0E7-2013-007T11.07.47.sts',IB_DZ,OB_D Z,ZQ,
      LAG=2,/SECONDS,HODO=4,/ZC,/VERBOSE,/VVERBOSE,SPINS=2
```

```
set =  0   0   30 zeros = -1.965 -2.002 -0.774 zq=  0.61
set =  1  30  60 zeros = -1.997 -2.013 -0.764 zq=  0.65
set =  2  60  90 zeros = -1.947 -1.951 -0.625 zq=  0.55
set =  3  90 120 zeros = -1.991 -1.978 -0.641 zq=  0.68
set =  4 120 150 zeros = -1.918 -1.985 -0.484 zq=  2.32
set =  5 150 180 zeros = -1.998 -1.960 -0.224 zq=  0.72
set =  6 180 210 zeros = -2.002 -2.007 -0.040 zq=  1.39
set =  7 210 240 zeros = -1.970 -1.976 -0.385 zq=  2.24
set =  8 240 270 zeros = -1.992 -1.985 -0.236 zq=  1.84
set =  9 270 300 zeros = -2.033 -1.976 -0.484 zq=  0.65
set = 10 300 330 zeros = -1.971 -1.980  0.018 zq=  3.10
set = 11 330 360 zeros = -2.037 -1.932 -0.726 zq=  0.91
set = 12 360 390 zeros = -2.041 -1.949 -0.356 zq=  1.26
set = 13 390 420 zeros = -2.084 -1.890 -0.133 zq=  1.27
set = 14 420 450 zeros = -2.107 -1.894 -0.186 zq=  1.34
set = 15 450 480 zeros = -2.065 -1.905 -0.239 zq=  0.79
set = 16 480 510 zeros = -2.084 -1.877 -0.507 zq=  1.06
set = 17 510 540 zeros = -2.077 -1.878 -0.082 zq=  1.79
set = 18 540 570 zeros = -2.079 -1.884 -0.521 zq=  1.32
set = 19 570 599 zeros = -2.091 -1.899 -0.240 zq=  1.84
```

```
-1.965 -2.002   0
-1.997 -2.013   0
-1.947 -1.951   0
-1.991 -1.978   0
```

```

-1.918 -1.985 0
-1.998 -1.960 0
-2.002 -2.007 0
-1.970 -1.976 0
-1.992 -1.985 0
-2.033 -1.976 0
-1.971 -1.980 0
-2.037 -1.932 0
-2.041 -1.949 0
-2.084 -1.890 0
-2.107 -1.894 0
-2.065 -1.905 0
-2.084 -1.877 0
-2.077 -1.878 0
-2.079 -1.884 0
-2.091 -1.899 0

```

```

IDL> zcan2,FNAME='3D-021877F0E7-2013-007T11.07.47.sts',IB_DZ,OB_D Z,ZQ,
      LAG=2,/SECONDS,HODO=4,/ZC,/VERBOSE,/VVERBOSE,SPINS=2

```

```

set = 0 0 30 zeros = -1.965 -2.002 -0.774 zq= 0.61
set = 1 30 60 zeros = -1.997 -2.013 -0.764 zq= 0.65
set = 2 60 90 zeros = -1.947 -1.951 -0.625 zq= 0.55
set = 3 90 120 zeros = -1.991 -1.978 -0.641 zq= 0.68
set = 4 120 150 zeros = -1.918 -1.985 -0.484 zq= 2.32
set = 5 150 180 zeros = -1.998 -1.960 -0.224 zq= 0.72
set = 6 180 210 zeros = -2.002 -2.007 -0.040 zq= 1.39
set = 7 210 240 zeros = -1.970 -1.976 -0.385 zq= 2.24
set = 8 240 270 zeros = -1.992 -1.985 -0.236 zq= 1.84
set = 9 270 300 zeros = -2.033 -1.976 -0.484 zq= 0.65
set = 10 300 330 zeros = -1.971 -1.980 0.018 zq= 3.10
set = 11 330 360 zeros = -2.037 -1.932 -0.726 zq= 0.91
set = 12 360 390 zeros = -2.041 -1.949 -0.356 zq= 1.26
set = 13 390 420 zeros = -2.084 -1.890 -0.133 zq= 1.27
set = 14 420 450 zeros = -2.107 -1.894 -0.186 zq= 1.34
set = 15 450 480 zeros = -2.065 -1.905 -0.239 zq= 0.79
set = 16 480 510 zeros = -2.084 -1.877 -0.507 zq= 1.06
set = 17 510 540 zeros = -2.077 -1.878 -0.082 zq= 1.79
set = 18 540 570 zeros = -2.079 -1.884 -0.521 zq= 1.32
set = 19 570 599 zeros = -2.091 -1.899 -0.240 zq= 1.84

```

```

-1.965 -2.002 0
-1.997 -2.013 0
-1.947 -1.951 0
-1.991 -1.978 0
-1.918 -1.985 0
-1.998 -1.960 0
-2.002 -2.007 0
-1.970 -1.976 0

```

```
-1.992 -1.985 0
-2.033 -1.976 0
-1.971 -1.980 0
-2.037 -1.932 1
-2.041 -1.949 0
-2.084 -1.890 1
-2.107 -1.894 1
-2.065 -1.905 1
-2.084 -1.877 1
-2.077 -1.878 1
-2.079 -1.884 1
-2.091 -1.899 1
```

IDL>

So, I'm thinking that CLUSTER uses some kind of random seed, and sometimes it works, sometimes not?

Subject: Re: Curious Cluster Analysis Conundrum
Posted by [Bill Nel](#) on Wed, 17 Jul 2013 19:44:50 GMT
[View Forum Message](#) <> [Reply to Message](#)

On Wednesday, July 17, 2013 3:03:14 PM UTC-4, jack.co...@nasa.gov wrote:

```
> I'm using
>
> iwts = CLUST_WTS(ibza,N_CLUSTERS=2)
> result = CLUSTER(ibza, iwts, n_clusters=2)
> ...
>
> So, I'm thinking that CLUSTER uses some kind of random seed, and sometimes it works,
sometimes not?
```

Yep, see the documentation for CLUST_WT :-)

Note: Because the initial clusters are chosen randomly, your results may differ slightly each time the CLUST_WTS routine is invoked, even for the same input data. For data with well-defined clusters the differences should be slight. For randomly-scattered data (no distinguishable clusters), the results may be significantly different, which may indicate that k-means clustering is not appropriate for your data.

Subject: Re: Curious Cluster Analysis Conundrum
Posted by [jack.connerney](#) on Wed, 17 Jul 2013 19:54:25 GMT
[View Forum Message](#) <> [Reply to Message](#)

Yes, thanks, that makes sense. It actually works pretty well most of the time. As an experiment, I copied the demonstration array and did a cluster analysis repeatedly; got the cluster identified

twice in 27 attempts. It's a pretty obvious cluster if you plot it out, given a two state a priori, I'd have thought cluster would find it without sooo much trouble.

Given time it might be fun to come up with a more successful algorithm...

Subject: Re: Curious Cluster Analysis Conundrum
Posted by [Fabzi](#) on Thu, 18 Jul 2013 14:48:26 GMT
[View Forum Message](#) <> [Reply to Message](#)

On 07/17/2013 09:54 PM, jack.connerney@nasa.gov wrote:

>

> Given time it might be fun to come up with a more successful algorithm...

There are plenty of other algorithms... All with their own strengths and weaknesses.

To be sure that CLUST_WTS converges you should increase the number of iterations with N_ITERATIONS. This will reduce the random factor considerably.
