

Tworzenie rozszerzenia do SAS Enterprise Miner z wykorzystaniem pakietu IML

Wojciech Bogucki
Michał Pastuszka



Plan

- Wprowadzenie: klasteryzacja spektralna
- Tworzenie pakietu SAS/IML
- Tworzenie rozszerzenia do Enterprise Miner
- Prezentacja działania

Wstęp

- Rozwiązanie powstało w ramach pracy inżynierskiej związanej z klasteryzacją spektralną
- Pytanie: jak dostarczyć algorytm, by jego użycie było jak najprostsze?
- Odpowiedź: IML Packages + Enterprise Miner Extension Node

Wprowadzenie: Klasteryzacja spektralna

- Rodzaj klasteryzacji opierający się na spektralnej teorii grafów
- Zyskał popularność dzięki solidnym podstawom matematycznym, prostej implementacji i konkurencyjnym wynikom
- Szeroko używany w wielu obszarach, takich jak statystyka, uczenie maszynowe, rozpoznawanie wzorców, eksploracja danych i analiza obrazów

An abstract graphic featuring a large white semi-circle on the right side of the frame. On the left, a solid blue circle is partially visible. Above this blue circle, four short, blue, curved dashed lines are arranged in an arc. The background is a solid blue color.

IML Packages

SAS/IML[®] Software

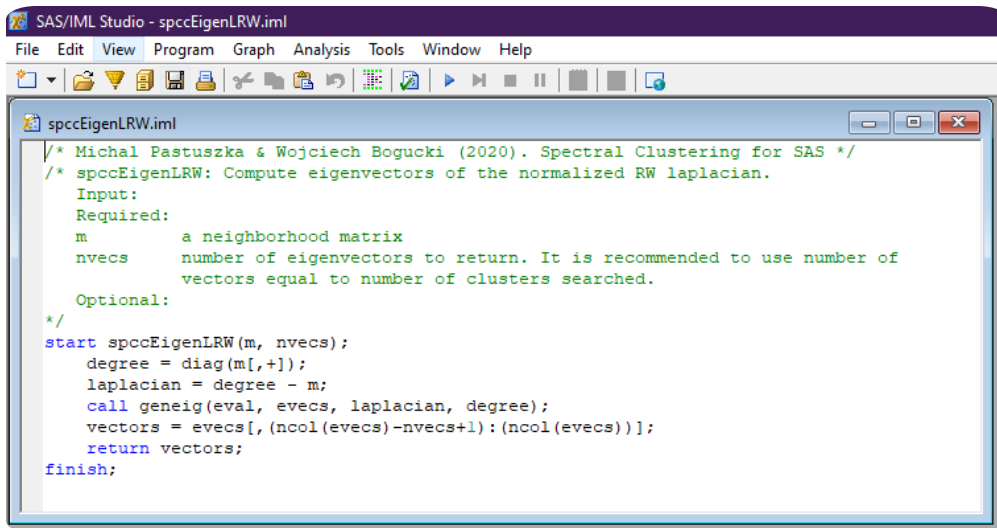
An interactive matrix programming language for creating custom analyses
with integration to R



SAS IML



SAS IML



```
/* Michal Pastuszka & Wojciech Bogucki (2020). Spectral Clustering for SAS */
/* spccEigenLRW: Compute eigenvectors of the normalized RW laplacian.
Input:
Required:
m          a neighborhood matrix
nvecs      number of eigenvectors to return. It is recommended to use number of
           vectors equal to number of clusters searched.
Optional:
*/
start spccEigenLRW(m, nvecs);
  degree = diag(m[,+]);
  laplacian = degree - m;
  call geneig(eval, evecs, laplacian, degree);
  vectors = evecs[, (ncol(evecs)-nvecs+1):(ncol(evecs))];
  return vectors;
finish;
```

- Język i narzędzie do operacji na macierzach, zapewniający zestaw funkcji z zakresu statystyki, analizy i wizualizacji danych
- Przypomina składnią języki takie jak Matlab i R, pozwala na integrację z R
- Może być wykonywany z poziomu 4GL procedurą **proc iml;**

IML Packages



Mechanizm dodany stosunkowo niedawno, bo w wersji 14.1 (2016r.)



Pozwala na dzielenie się kodem, wraz z dokumentacją, programami i danymi za pomocą ustrukturyzowanego archiwum zip



Wykorzystuje komendę `packages` pozwalającą na instalowanie i wczytywanie zawartości pakietu



Komenda
packages

package install '{ścieżka do pliku zip}';

- instaluje pakiet umieszczając zawartość archiwum w odpowiedniej lokalizacji. Zależnie od środowiska, w którym wywołana jest komenda pakiet może zostać zainstalowany dla wszystkich lub dla danego użytkownika

```
7      proc iml;  
NOTE: IML Ready  
8  
8      ! package install '/folders/myfolders/spectralclust.zip';  
9      quit;  
NOTE: Exiting IML.  
NOTE: PROCEDURE IML zajęła (całkowity czas przetwarzania):  
      real time      0.60 seconds  
      cpu time       0.25 seconds
```

package load {nazwa pakietu};

- wczytuje pakiet, wykonując kod zawarty w plikach źródłowych. Najczęściej są to definicje funkcji, do których otrzymujemy dostęp po wykonaniu polecenia

```
10      proc iml;
NOTE: IML Ready
11
11      ! package load spectralclust;
NOTE: Module SPCCEIGENL defined.
NOTE: Module SPCCEIGENLRW defined.
NOTE: Module SPCCGAUSSNEIGH defined.
NOTE: Module SPCCSPECTRALIZE defined.
NOTE: Module SPCKNNNEIGH defined.
NOTE: Module SPCCMUTKNNNEIGH defined.
NOTE: Module SPCCFASTCLUS defined.
NOTE: Module SPCCEIGENLSYM defined.
NOTE: Module SPCCNEGEUCLIDNEIGH defined.
12      quit;
NOTE: Exiting IML.
NOTE: PROCEDURE IML zajęła (całkowity czas przetwarzania):
      real time          0.09 seconds
      cpu time           0.03 seconds
```

package help {nazwa pakietu};

- wyświetla w konsoli dokumentację pakietu w postaci tekstowej*

```
13      proc iml;
NOTE: IML Ready
14      ! package help spectralclust;
Spectralclust Package

Description: Spectral clustering for SAS

Spectral clustering is a type of clustering algorithm that takes use of spectral graph theory.
Names used are based on "A Tutorial on Spectral Clustering" (von Luxburg, 2007).

Recommended use is to pass the data to the spccSpectralize function, save it to a dataset, and exec
algorithm, such as the fastclus procedure, on the created dataset.

Module Syntax:

spccSpectralize(m, nclus, laplacian='normalizedRW',
               neighborhood_fun='gaussian', sigma=1, neighborhood_type='knn', k=10);
  creates eigenvectors of a graph laplacian matrix given a dataset m
  for use in clustering
  returns a matrix of nclus eigenvectors corresponding to the smallest eigenvalues of the laplacian
  each row of the matrix represents a corresponding observation from the original dataset

spccEigenL(m, nvecs);
  computes eigenvectors of the unnormalized laplacian of the neighborhood matrix m.
  returns nvecs number of eigenvectors corresponding to the smallest eigenvalues stored columnwise

spccEigenLRW(m, nvecs);
  computes eigenvectors of the normalized RW laplacian of the neighborhood matrix m.
  returns nvecs number of eigenvectors corresponding to the smallest eigenvalues stored columnwise
```

package list;

- wyświetla wszystkie dostępne pakiety

System SAS

Private Packages

Name	Version
polygon	1.0
spectralclust	0.5

Public Packages

Name	Version
(None)	

System Packages

Name	Version
AboveBelow	1.0
ListUtil	1.0
wavelet	1.0

package uninstall {nazwa pakietu};

- odinstalowuje pakiet usuwając odpowiednie pliki

```
94      proc iml;
NOTE: IML Ready
95
96      ! package uninstall spectralclust;
97      quit;
NOTE: Exiting IML.
NOTE: PROCEDURE IML zajęła (całkowity czas przetwarzania):
      real time          0.09 seconds
      cpu time           0.03 seconds
```

package info {nazwa pakietu};

- wyświetla metadane pakietu (krótki opis, wersja, lokalizacja instalacji itd.)

System SAS	
Package Information	
Name	spectralclust
Description	Spectral clustering for SAS
Author	Michał Pastuszka <pastuszkam@student.mini.pw.edu.pl> Wojciech Bogucki <boguckiw@student.mini.pw.edu.pl>
Collection	Private
Version	0.5
Requires IML	15.1
Directory	/folders/myfolders/sas/iml/packages/spectralclust

package libname {nazwa biblioteki}
{nazwa pakietu};

- tworzy bibliotekę zawierającą dane dostarczane z pakietem
- Wymaga załadowania pakietu poleceniem load

```
88
88      ! package libname demodata spectralclust;
89      quit;
NOTE: Exiting IML.
NOTE: PROCEDURE IML zajęła (całkowity czas przetwarzania):
      real time          0.10 seconds
      cpu time           0.05 seconds

90
91      data _NULL_;
92      set demodata.jain;
NOTE: Plik danych DEMODATA.JAIN.DATA ma format innego systemu operacyjnego
      będzie Cross Environment Data Access, co może wymagać dodatkowych
93      run;

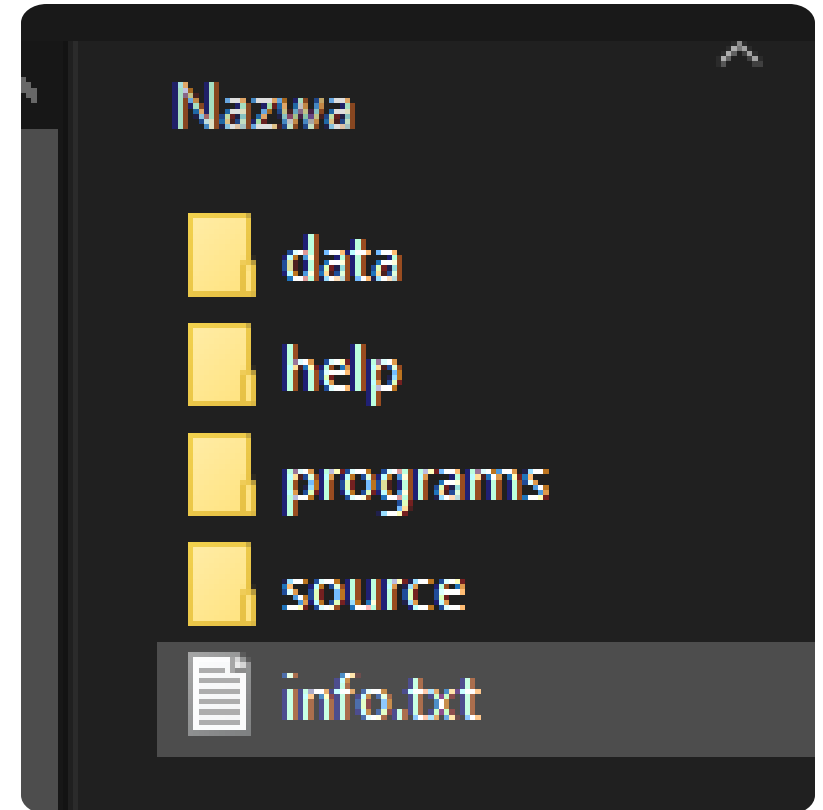
NOTE: There were 373 observations read from the data set DEMODATA.JAIN.
NOTE: Instrukcja DATA zajęła (całkowity czas przetwarzania):
      real time          0.01 seconds
      cpu time           0.02 seconds
```


The image features a large blue semi-circle on the right side, which serves as a background for the text. To the left of the blue shape, there is a purple circle, an orange square, and an orange line forming a corner. Several teal-colored dashed lines of varying lengths are scattered in the upper left and lower left areas. A small teal circle is partially visible at the top right edge.

Tworzenie pakietu



Struktura folderu


- Data - dane wczytywane poleceniem package libname
- Help - dokumentacja
- Programs - przykładowe programy lub dodatkowe narzędzia
- Source - kod pakietu uruchamiany poleceniem package load
- Info.txt - metadane



data



- Dane wczytywane poleceniem `package libname`
- Mogą to być na dane wykorzystywane w przykładach
- Pliki w formacie tabel SAS






Nazwa	Typ
 jain.sas7bdat	SAS Data Set

help


- Dokumentacja w formacie txt wyświetlana poleceniem `package help`
- Można dodatkowo umieścić tu bardziej kompletną dokumentację w formacie pdf lub html
- Pliki muszą mieć nazwę zgodną z nazwą pakietu






Nazwa	Typ
 spectralclust.docx	Dokument programu Microsoft Word
 spectralclust.pdf	Microsoft Edge PDF Document
 spectralclust.txt	Dokument tekstowy

programs

- Programy dostarczane z pakietem (inne niż sam kod pakietu)
- Nie ma wymagań co do ich formatu, nie są wykorzystywane przez mechanizm pakietów
- Mogą to być przykładowe programy wykorzystujące nasz pakiet bądź dodatkowe narzędzia













Nazwa	Typ
 Example_1.sas	SAS System Program
 Example_2.sas	SAS System Program




source

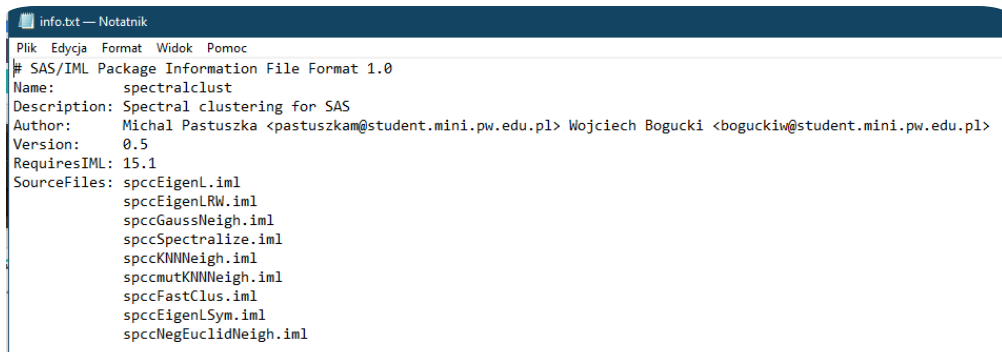
- Właściwy kod pakietu w postaci plików IML
- Najczęściej są to definicje funkcji
- Zaleca się umieszczanie funkcji w osobnych plikach o nazwach zgodnych z nazwą funkcji
- Dobrą praktyką jest rozpoczynanie nazwy funkcji od prefiksu identyfikującego pakiet



Nazwa	Typ
 spccEigenL.iml	SAS/IML Source Code
 spccEigenLRW.iml	SAS/IML Source Code
 spccEigenLSym.iml	SAS/IML Source Code
 spccFastClus.iml	SAS/IML Source Code
 spccGaussNeigh.iml	SAS/IML Source Code
 spccKNNNeigh.iml	SAS/IML Source Code
 spccMutKNNNeigh.iml	SAS/IML Source Code
 spccNegEuclidNeigh.iml	SAS/IML Source Code
 spccSpectralize.iml	SAS/IML Source Code



Info.txt



```
info.txt - Notatnik
Plik Edycja Format Widok Pomoc
# SAS/IML Package Information File Format 1.0
Name: spectralclust
Description: Spectral clustering for SAS
Author: Michał Pastuszka <pastuszkam@student.mini.pw.edu.pl> Wojciech Bogucki <boguckiw@student.mini.pw.edu.pl>
Version: 0.5
RequiresIML: 15.1
SourceFiles: spccEigenL.iml
              spccEigenLRW.iml
              spccGaussNeigh.iml
              spccSpectralize.iml
              spccKNNNeigh.iml
              spccmutKNNNeigh.iml
              spccFastClus.iml
              spccEigenLSym.iml
              spccNegEuclidNeigh.iml
```

- Zawiera metadane wyświetlane poleceniem `package info`
- Dodatkowo zawiera listę plików źródłowych - tylko pliki tu wymienione zostaną wykonane w momencie ładowania pakietu



Tworzenie rozszerzenia do SAS Enterprise Miner

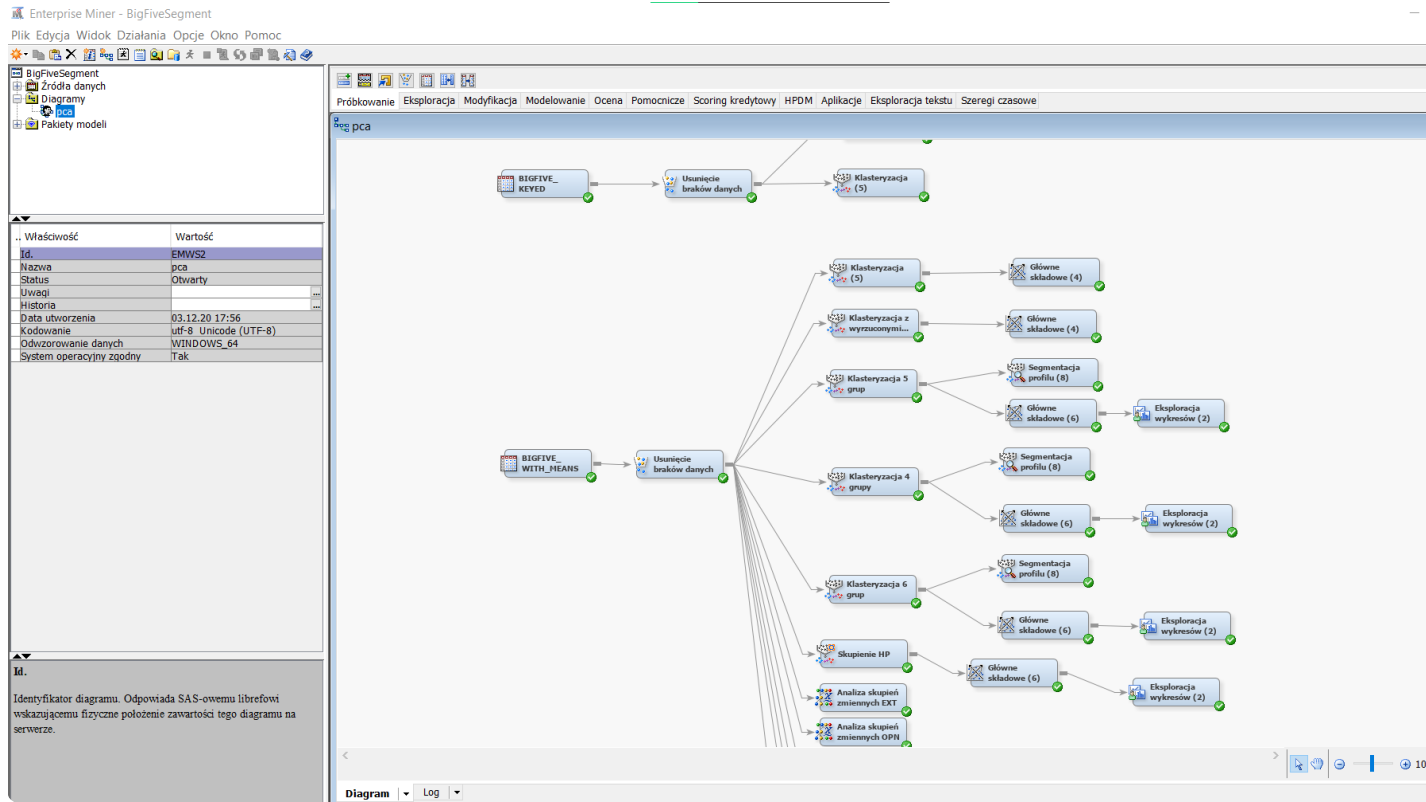
SAS® Enterprise Miner™

Create highly accurate analytical models that enable you
to predict with confidence



SAS Enterprise
Miner

SAS Enterprise Miner



- Graficzne narzędzie do modelowania i eksploracji danych
- Pozwala na szybkie testowanie wielu ścieżek przetwarzania danych i porównywanie wyników
- Zapewnia przejrzystość wykonywanych operacji



Extension Node

- Mechanizm pozwalający na rozszerzenie funkcjonalności programu przez użytkownika
- Może zostać w pełni zintegrowany z programem i łączyć się z innymi węzłami w diagramie

Extension Node – potrzebne pliki



Katalog SAS z plikami z kodem SAS-owym zawierającymi program do wykonania



Plik XML definiujący właściwości węzła



Pliki GIF – ikony węzła

main.sas

- Zawiera główne makro `%main`
- Używa polecenia `%include` do wczytania zawartości innych plików
- Uruchamia pozostałe pliki z zależności od stanu węzła w diagramie
- Do rozpoznawania stanu węzła używana jest makrozmienna `&EM_ACTION`

```
%macro main;
    %if %upcase(&EM_ACTION) = CREATE %then %do;
        filename temp catalog 'sashelp.specc.spcl_create.source';
        %include temp;
        filename temp;
        %create;
    %end;

    %else
    %if %upcase(&EM_ACTION) = TRAIN %then %do;
        filename temp catalog 'sashelp.specc.spcl_train.source';
        %include temp;
        filename temp;
        %train;
    %end;

    %if %upcase(&EM_ACTION) = SCORE %then %do;
        filename temp catalog 'sashelp.specc.spcl_score.source';
        %include temp;
        filename temp;
        %score;
    %end;

    %if %upcase(&EM_ACTION) = REPORT %then %do;
        filename temp catalog 'sashelp.specc.spcl_report.source';
        %include temp;
        filename temp;
        %report;
    %end;

%mend main;
%main;
```

create.sas

- Zawiera makro **%create**
- Uruchamiany, gdy węzeł zostaje dodany do diagramu
- Inicjalizacja właściwości/parametrów (%EM_PROPERTY)
- Rejestracja zbiorów/plików/katalogów (%EM_REGISTER)

```

%macro create;
    %EM_PROPERTY(NAME=ClusterNum,          VALUE=2);
    %EM_PROPERTY(NAME=Neighborhood,        VALUE=complete);
    %EM_PROPERTY(NAME=Laplacian,           VALUE=normalizedRW);
    %EM_PROPERTY(NAME=NeighFun,            VALUE=gaussian);
    %EM_PROPERTY(NAME=Sigma,               VALUE=1.0);
    %EM_PROPERTY(NAME=K,                   VALUE=5);
    %EM_PROPERTY(NAME=MaxIter,             VALUE=10);
    %EM_PROPERTY(NAME=Distance,            VALUE=EUCLIDEAN);
    %EM_PROPERTY(NAME=Seed,                VALUE=12346);
    %EM_PROPERTY(NAME=Impute,              VALUE=none);
    %EM_PROPERTY(NAME=Standard,            VALUE=N);
    %EM_PROPERTY(NAME=ClusRole,            VALUE=SEGMENT, ACTION=SCORE);

    %EM_REGISTER(key=import_data,          type=DATA);
    %EM_REGISTER(key=vectors,              type=DATA);
    %EM_REGISTER(key=indvectors,           type=DATA);
    %EM_REGISTER(key=OUTSTAT,              type=DATA);
    %EM_REGISTER(key=graph_table,          type=DATA);
    %EM_REGISTER(key=MODELINFO,            type=DATA);
    %EM_REGISTER(key=CLUSTERSUM,           type=DATA);
    %EM_REGISTER(key=ITERSTAT,             type=DATA);
    %EM_REGISTER(key=OVERALLVARSTAT,       type=DATA);
    %EM_REGISTER(key=CLUSTERBASEDVARSTAT, type=DATA);

%mend create;

```


train.sas

- Odpowiada za przetworzenie danych wejściowych i przeprowadzenie klasteryzacji spektralnej
- Zawiera główne makro **%train** oraz pomocnicze makro **%getNObs**
- Wczytuje zarejestrowane wcześniej zbiory makrem **%EM_GETNAME**

Makro pomocnicze

```
%macro getNObs(inds, nobs);  
/*macro for setting macrovariable nobs with number of observations in  
dataset inds*/  
    %global &nobs;  
    data _null_;  
        set &inds end=eof;  
        if eof then call symputx("&nobs", _N_);  
    run;  
    quit;  
%mend getNObs;
```

A co gdyby zrobić to sprytniej?

```
%macro getNObs(inds, nobs);  
%global &nobs;  
  data _null_;  
    if 0 then set &inds nobs=n;  
    call symputx("&nobs", n);  
  stop;  
run;  
quit;  
%mend getNObs;
```

```
NOTE: View EMWS5.IDS2_DATA.VIEW used (Total process time):  
real time          0.03 seconds  
user cpu time      0.00 seconds  
system cpu time    0.01 seconds  
memory             87987.06k  
OS Memory          96256.00k  
Timestamp          04/21/2021 11:26:48 AM  
Step Count          1 Switch Count  3
```

```
NOTE: DATA statement used (Total process time):  
real time          0.05 seconds  
user cpu time      0.00 seconds  
system cpu time    0.03 seconds  
memory             87987.06k  
OS Memory          96256.00k  
Timestamp          04/21/2021 11:26:48 AM  
Step Count          1 Switch Count  4
```

Number of observations in dataset: 9007199254740992

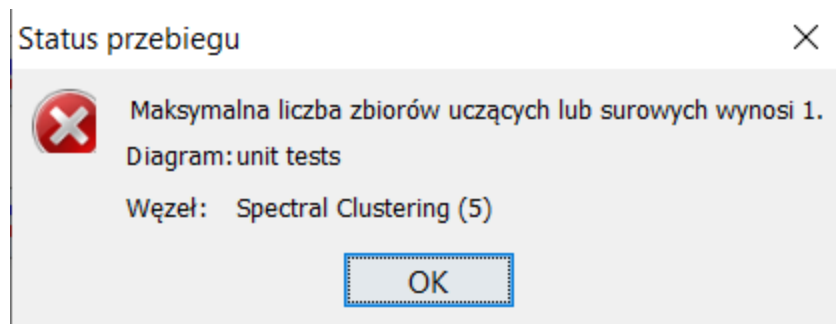
Walidacje – mechanizm wyjątków

- Sprawdzenie czy istnieją dane

```
%if (^%sysfunc(exist(&EM_IMPORT_DATA)) and  
    ^%sysfunc(exist(&EM_IMPORT_DATA, VIEW)))  
    or "&EM_IMPORT_DATA" eq "" %then %do;  
    %let EM_EXCEPTIONSTRING = exception.server.IMPORT.NOTRAIN,1;  
    %goto doenda;  
%end;
```

<reszta kodu makra>

```
%doenda:  
%mend train;
```

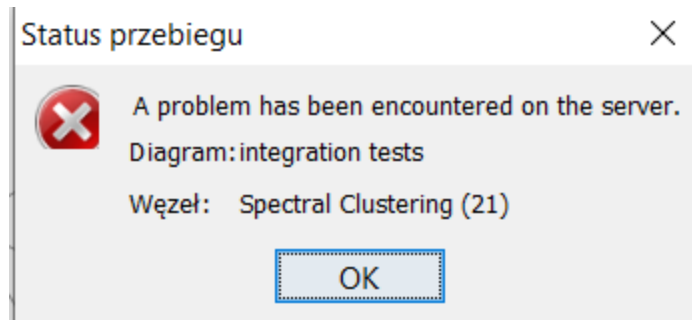


```
*-----*  
*  
* ERROR: Must have at least 1 Train or Raw data set(s)  
* Please report unresolved problems to Technical Support.  
*  
*-----*
```

Walidacje

- Sprawdzenie czy istnieje chociaż jedna zmienna przedziałowa, porządkowa lub binarna

```
%if (%EM_INTERVAL_INPUT %EM_ORDINAL_INPUT %EM_BINARY_INPUT eq ) %then %do;  
    %let EMEXCEPTIONSTRING = ERROR;  
    %put &em_codebar;  
    %put Error: Must use at least one interval, ordinal or binary input;  
    %put &em_codebar;  
    %goto doenda;  
%end;
```



```
*-----*  
Error: Must use at least one interval, ordinal or binary input  
*-----*
```

Użycie pakietu IML

```
proc iml;

    package load spectralclust;
    varnames = { %EM_INTERVAL_INPUT %EM_ORDINAL_INPUT %EM_BINARY_INPUT };
    print 'Used variables:';
    print varnames;

    use &em_user_import_data;
    read all var varnames into m;
    close &em_user_import_data;

    m = spccSpectralize(m, &EM_PROPERTY_ClusterNum , "&EM_PROPERTY_Laplacian.",
        "&EM_PROPERTY_NeighFun.", &EM_PROPERTY_Sigma ,
        "&EM_PROPERTY_Neighborhood.", &EM_PROPERTY_K );

    create &em_user_vectors from m;
        append from m;
    close &em_user_vectors;
    CALL symput("nvar", putn(ncol(m), "BEST6."));

quit;
```

	COL1	COL2
1	0.0949537218	0.0188684256
2	0.0949492513	0.0188684256
3	0.0947392604	0.0188684256
4	0.0947156138	0.0188684256
5	0.0947023317	0.0188684256
6	0.0946817826	0.0188684256
7	0.0943418065	0.0188684256
8	0.0943406896	0.0188684256
9	0.0943284186	0.0188684256
10	0.0942987893	0.0188684256

Szybka klasteryzacja - HPCLUS

- Używa algorytmu k-średnich do klasteryzacji
- Korzysta z wielu wątków procesora
- Tworzy wiele zbiorów zawierających statystyki klasteryzacji

```
ods listing exclude Standardization ;
filename flowtemp "&em_file_emflowscorecode";
ods output PerformanceInfo = _tmp_outperformanceinfo
      ModelInfo = &em_user_modelinfo
      NObs = _tmp_outnobs
      ClusterSum = &em_user_clustersum
      IterStats = &em_user_ITERSTAT
      DescStats = &em_user_OVERALLVARSTAT
      WithinClusStats = &em_user_CLUSTERBASEDVARSTAT
      Timing = _tmp_outtiming;
```

```
proc hpclus data = &em_user_vectors
      maxclusters = &EM_PROPERTY_ClusterNum
      maxiter = &EM_PROPERTY_MaxIter
      outstat = &EM_USER_OUTSTAT
      distance = &EM_PROPERTY_Distance
      Seed = &EM_PROPERTY_Seed;
  input %DO i=1 %to &nvar; COL&i %END;;
  score out=_tmp_out_score;
  code file=flowtemp;
  &hpdm_performance.;
run;
```

	Cluster ID	Distance to Centroid
1	1	0.029612897
2	1	0.0296084265
3	1	0.0293984356
4	1	0.029374789
5	1	0.0293615069
6	1	0.0293409578
7	1	0.0290009817
8	1	0.0289998648
9	1	0.0289875938
10	1	0.0289579645

Wyjściowe zbiory danych

```
data &em_export_train;  
    set &em_user_import_data;  
    set _tmp_out_score(drop=_DISTANCE_);  
run;
```

	VAR1	VAR2	VAR3	Cluster ID
1	0.85	17.45	2	1
2	0.75	15.6	2	1
3	3.3	15.45	2	1
4	5.25	14.2	2	1
5	4.9	15.65	2	1
6	5.35	15.85	2	1
7	5.1	17.9	2	1
8	4.6	18.25	2	1
9	4.05	18.75	2	1
10	3.4	19.7	2	1

```
data &em_user_vectors;  
    set &em_user_vectors;  
    set _tmp_out_score;  
run;
```

	COL1	COL2	Cluster ID	Distance to Centroid
1	0.0949537218	0.0188684256	1	0.029612897
2	0.0949492513	0.0188684256	1	0.0296084265
3	0.0947392604	0.0188684256	1	0.0293984356
4	0.0947156138	0.0188684256	1	0.029374789
5	0.0947023317	0.0188684256	1	0.0293615069
6	0.0946817826	0.0188684256	1	0.0293409578
7	0.0943418065	0.0188684256	1	0.0290009817
8	0.0943406896	0.0188684256	1	0.0289998648
9	0.0943284186	0.0188684256	1	0.0289875938
10	0.0942987893	0.0188684256	1	0.0289579645

score.sas

- Zawiera makro **%score** odpowiadające za zmianę metadanych zbioru wyjściowego
- Uruchamiany po zakończeniu makra **%train**

Zmiana metadanych

```
%macro score;
```

```
    filename _FHPTS "&EM_FILE_CDELTA_TRAIN";  
    data _null_;  
        file _FHPTS;  
        put 'if upcase(NAME) eq "_DISTANCE_" then Role="REJECTED";'  
        put "if upcase(NAME) eq '_CLUSTER_ID_' then do;";  
        put "    Role='&em_property_clusrole';";  
        put "    Level='NOMINAL';";  
        put "end;";  
    run;
```

```
    filename _FHPTS;
```

```
%mend score;
```

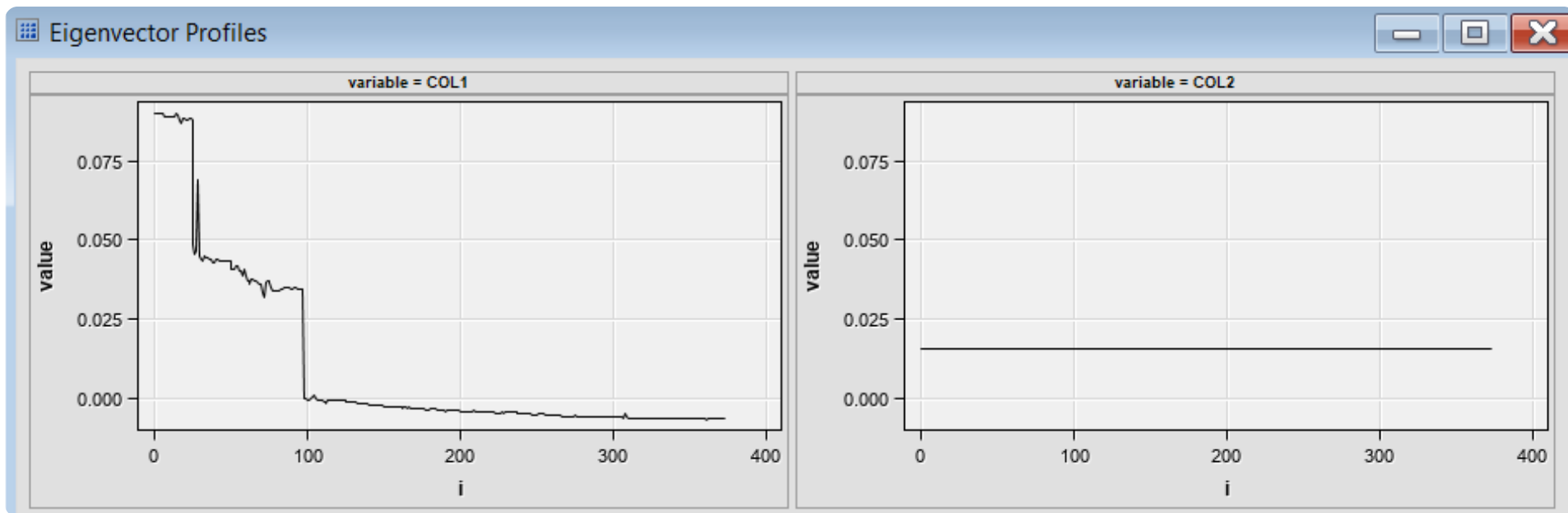
Nazwa	Użycie	Raport	Rola próby	Rola	Poziom
VAR1	Domyślne	Nie	Domyślnie	Wejście	Przedziałowa
VAR2	Domyślne	Nie	Domyślnie	Wejście	Przedziałowa
VAR3	Domyślne	Nie	Domyślnie	Zmienna celu	Przedziałowa
_CLUSTER_ID_	Domyślne	Nie	Domyślnie	Segment	Nominalna

report.sas

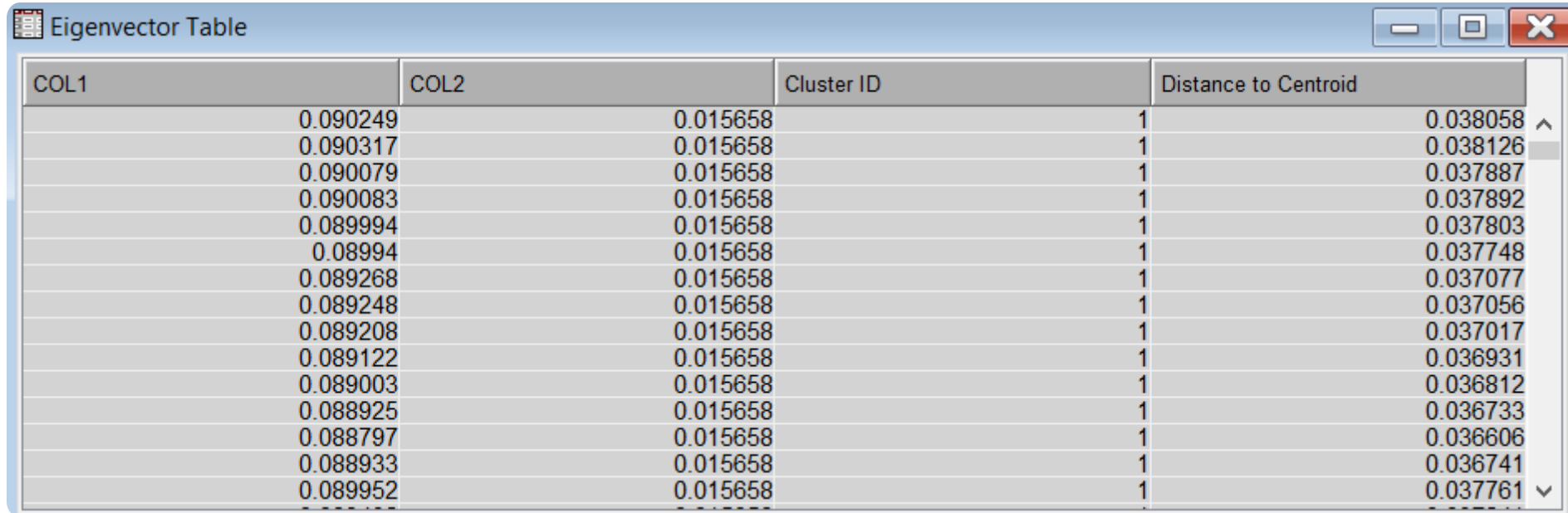
- Zawiera makro **%report** odpowiadające za generowanie raportów, które wyświetlają się następnie w oknie Rezultaty
- Uruchamiany po zakończeniu makra **%score**
- Do tworzenia wizualizacji używane jest makro %EM_REPORT

Tworzenie raportów

```
%em_report(key=indvectors,  
viewtype=Lattice,  
latticetype=Lineplot,  
block=eigenvectors,  
views=3,  
autodisplay=Y,  
x=i,  
y=value,  
latticex=variable,  
equalizerow=Y,  
description=Eigenvector Profiles);
```



Tworzenie raportów

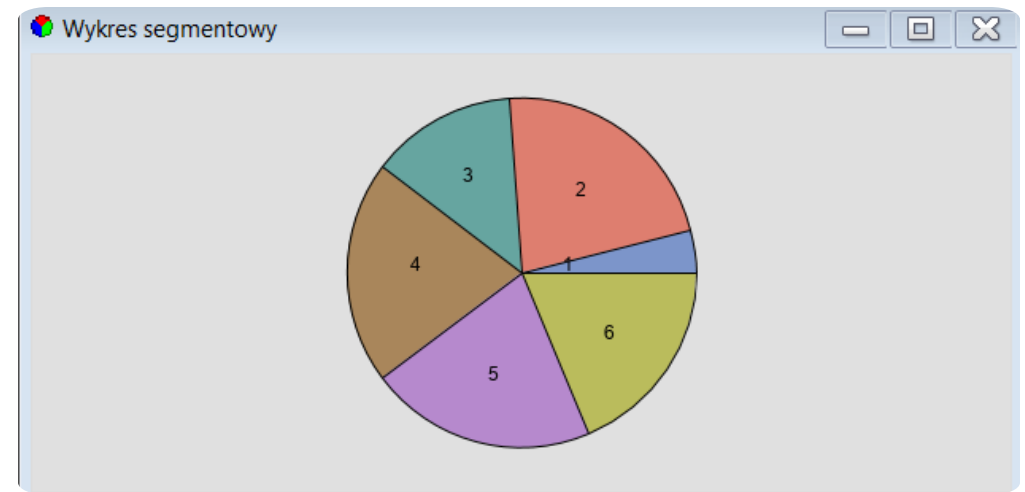


COL1	COL2	Cluster ID	Distance to Centroid
0.090249	0.015658	1	0.038058
0.090317	0.015658	1	0.038126
0.090079	0.015658	1	0.037887
0.090083	0.015658	1	0.037892
0.089994	0.015658	1	0.037803
0.08994	0.015658	1	0.037748
0.089268	0.015658	1	0.037077
0.089248	0.015658	1	0.037056
0.089208	0.015658	1	0.037017
0.089122	0.015658	1	0.036931
0.089003	0.015658	1	0.036812
0.088925	0.015658	1	0.036733
0.088797	0.015658	1	0.036606
0.088933	0.015658	1	0.036741
0.089952	0.015658	1	0.037761

```
%em_report( key=vectors,  
viewtype=Data,  
block=eigenvectors,  
views=1,  
autodisplay=Y,  
description=Eigenvector Table);
```

Tworzenie raportów

```
%EM_REPORT(key=CLUSTERSUM,  
viewtype=PIE,  
x=Cluster,  
freq=Frequency,  
block=PLOT,  
autodisplay=Y,  
description=segmentplotlabel )
```



Tworzenie katalogu

```
%macro CreateSources(dirPath, libraryname, catname);  
  
    libname &libraryname "&dirPath";  
    filename src_cat catalog "&libraryname..&catname..spcl.source";  
    filename mydata "&dirPath.\main.sas";  
    data _null_;  
        file src_cat;  
        infile mydata;  
        input;  
        put _infile_;  
    run;  
    <wczytanie reszty plików>  
  
    proc catalog cat=&libraryname..&catname.;  
        copy out=sashelp.&catname.;  
    run;  
%mend CreateSources;
```

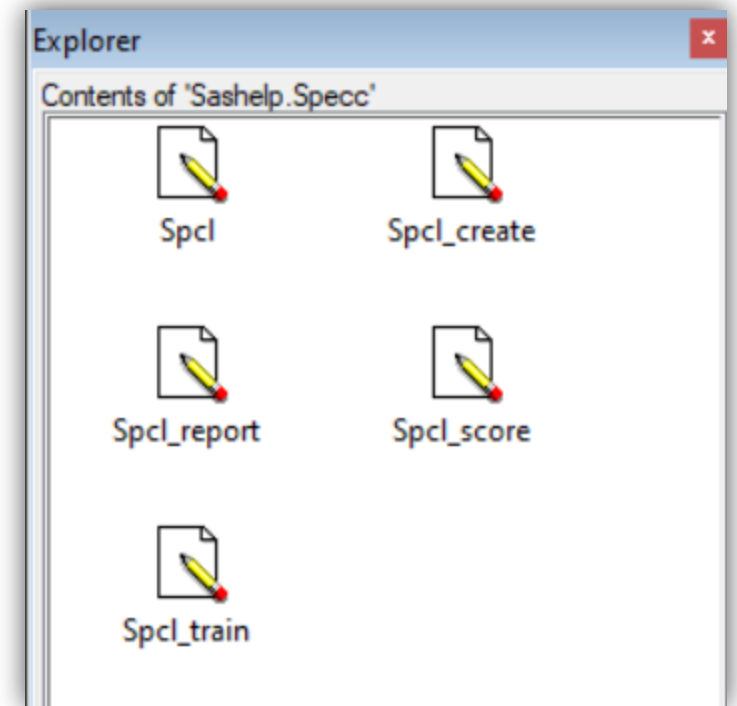

Tworzenie katalogu

```
%macro createCatalog(userPath);  
    %CreateSources(&userPath\SpectralClusteringSAS\SAS EM\SpecClust,  
    specc, specc)  
    LIBNAME specc CLEAR;  
%mend createCatalog;
```

Tworzenie katalogu - wywołanie

```
%include "Path_to_repository\SpectralClusteringSAS\SAS  
EM\SpecClust\create_catalog.sas";
```

```
%createCatalog(Path_to_repository);
```



.. Właściwość	Wartość
Ogólne	
Id. węzła	Specc
Importowane dane	...
Eksportowane dane	...
Uwagi	...
Uczenie	
Variables	...
Standardization	Nie
Imputation	None
Number of Clusters	2
Laplacian Type	Random walk normalized
Similarity Function	Gaussian Function
Sigma	1.0
Similarity Graph Type	Complete
K	5
HPCLUS options	
Maximal number of iterations	10
Distance	Euclidean
Seed	12346
Wynik punktowy	
Role of cluster variable	Segment
Status	
Utwórz	05.01.21 17:01
Id. przebiegu	0f90264c-29fe-4b51-9225-d2fd25
Ostatni błąd	
Ostatni status	Wykonany
Czas ostatniego uruchomienia	05.01.21 17:02
Czas trwania przebiegu	0 godz. 0 min. 9,72 sek.
Host sieci gridowej	
Węzeł dodany przez użytkownika	Nie

Właściwości węzła

- Daje użytkownikowi możliwość podania hiperparametrów do algorytmu
- Umożliwia ograniczenie wartości możliwych do podania
- Wygląd okna zdefiniowany jest w pliku XML

Okno dialogowe

```
<Property
  type="String"
  name="VariableSet"
  displayName="Variables"
  description="Variable Properties">
  <Control>
    <Dialog
      class="com.sas.analytics.eminer.visuals.VariablesDialog"
      showValue="N" />
    </Control>
  </Property>
```

Zmienne - Specc2

(brak) ☐ nie równe Zastosuj Przywróć

Kolumny: ☐ Etykieta ☐ Eksploracja ☐ Podstawowe ☐ Statystyki

Nazwa	Użycie	Rola	Poziom
VAR1	Domyślne	Wejście	Przedziałowa
VAR2	Domyślne	Wejście	Przedziałowa
VAR3	Domyślne	Zmienna celu	Przedziałowa

Eksploruj... Uaktualnij ścieżkę OK Anuluj

Przykładowa definicja hiperparametru

```
<Property
  type="int"
  name="ClusterNum"
  displayName="Number of Clusters"
  description="Indicates the number of clusters to be detected in data"
  initial="2" >
  <Control>
    <Range min="2" excludeMin="N" />
  </Control>
</Property>
```

Uczenie	
Variables	...
Standardization	Nie
Imputation	None
Number of Clusters	2
Laplacian Type	Random walk normalized
Similarity Function	Gaussian Function
Sigma	1.25
Similarity Graph Type	Complete
K	5
HPCLUS options	
Maximal number of iterations	10
Distance	Euclidean
Seed	12346
Wynik punktowy	
Role of cluster variable	Segment
Status	
Utwórz	31.12.20 14:08
Id. przebiegu	30c5e426-d7fc-4583-a38e-4ce45
Ostatni błąd	
Ostatni status	Wykonany
Czas ostatniego uruchomienia	18.01.21 23:54
Czas trwania przebiegu	0 godz. 0 min. 10,00 sek.
Host sieci gridowej	
Węzeł dodany przez użytkownika	Nie

Number of Clusters
Indicates the number of clusters to be detected in data

Układ hiperparametrów

Właściwość	Wartość
Ogólne	
Id. węzła	Specc
Importowane dane	...
Eksportowane dane	...
Uwagi	...
Uczenie	
Variables	...
Standardization	Nie
Imputation	None
Number of Clusters	2
Laplacian Type	Random walk normalized
Similarity Function	Gaussian Function
Sigma	1.25
Similarity Graph Type	Complete
K	5
<input type="checkbox"/> HPCLUS options	
Maximal number of iterations	10
Distance	Euclidean
Seed	12346
Wynik punktowy	
Role of cluster variable	Segment

```
<Views>
  <View name="Train">
    <PropertyRef nameref="VariableSet"/>
    <PropertyRef nameref="Standard"/>
    <PropertyRef nameref="Impute"/>
    <PropertyRef nameref="ClusterNum"/>
    <PropertyRef nameref="Laplacian"/>
    <PropertyRef nameref="NeighFun"/>
    <PropertyRef nameref="Sigma"/>
    <PropertyRef nameref="Neighborhood"/>
    <PropertyRef nameref="K"/>
    <Group
      name="HPCLUS"
      displayName="HPCLUS options"
      description="Options passed to HPCLUS
        procedure which executes k-means clustering algorithm">
      <PropertyRef nameref="MaxIter"/>
      <PropertyRef nameref="DISTANCE"/>
      <PropertyRef nameref="Seed"/>
    </Group>
  </View>
  <View name="Score">
    <PropertyRef nameref="CLUSROLE"/>
  </View>
</Views>
```



Prezentacja działania

1. *Journal of the American Medical Association*, 2000; 283: 2689-2695.

Id.

Identyfikator diagramu. Odpowiada SAS-owemu librefowi wskazującemu fizyczne położenie zawartości tego diagramu na serwerze.

Example2

Diagram Log

Przydatne linki

- Writing Packages: A New Way to Distribute and Use SAS/IML® Programs
<https://support.sas.com/resources/papers/proceedings16/SAS4201-2016.pdf>
- SAS/IML User's Guide - Packages
https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.3/implug/implug_packages_sect001.htm
- SAS Enterprise Miner Extension Nodes: Developer's guide
<https://documentation.sas.com/doc/pl/emxndg/15.1/titlepage.htm>
- Procedura HPCLUS
https://documentation.sas.com/doc/en/emhpprcref/14.2/emhpprcref_hpclus_overview01.htm
- Kod naszego rozwiązania: <https://github.com/Pastuszka/SpectralClusteringSAS>



Dziękujemy za
uwagę