DimensionsReductionSurvey-12VARIAB_11CDR

February 18, 2022

[1]:
```
#package prince https://github.com/MaxHalford/prince
#MCA multiple correspondance analysis
#three or more categorical features
```

[1]:
```
#Importing the necessary package
import pandas as pd
import numpy as np
from prince import MCA#Dataset preparation with only numerical features
df = pd.read_csv('AASER_11CDR.csv')
print(len(df))
print(len(df.columns))
df
```

```
11
19
```

[1]:
```
    CDR_name  com  dem rest Java  api  gui flat  gdl form  aql term open  \
0    EHRBase   no   no  yes  yes  yes  yes  yes   no   no  yes   no  yes
1     Better  yes   no  yes  yes  yes  yes  yes  yes  yes  yes  yes   no
2     Base24  yes  yes   no   no   no  yes   no   no  yes   no  yes   no
3       Cabo   no   no  yes   no  dev  yes   no   no   no   no   no  yes
4   ArenaEHR  yes   no  yes   no  yes  yes  yes  yes  yes  yes  yes   no
5     eWeave  yes  yes   no   no   no  yes   no   no  yes  yes  yes   no
6    EHRCare  yes  yes  yes  yes   no   no  yes   no  yes  yes   no   no
7     Clever  yes  yes  yes  yes   no  yes  yes  yes  yes  yes   no   no
8      EHRDB  yes   no  yes  yes  yes  yes  yes   no  yes  yes   no   no
9        RHP  yes  yes  yes  yes   no  yes  yes   no  yes   no  yes   no
10    EHRNet  yes   no  yes   no   no  yes  yes   no  dev  yes  yes   no

   archet temp   fi   fe extr oauth2
0      no   no  dev  dev   no    yes
1     yes  yes  dev  yes   no    yes
2      no   no  yes  yes   no    yes
3      no   no  ext  ext   no     no
4      no   no  yes  yes none    yes
5      no  yes   no   no   no     no
6      no   no   no   no   no     no
```

```
7     yes  yes   no   no   yes    yes
8      no   no  dev  dev   dev    yes
9     yes  yes   no   no    no     no
10     no   no  yes  yes    no     no
```

[564]:
```python
df=df.set_index('CDR_name')
#df.drop(['EHRNet','RHP'],axis=0, inplace = True)
#df.drop('RHP',axis=0, inplace=True)
#df.drop('EHRNet',axis=0, inplace=True)
df
```

[564]:
```
          com  dem rest Java  api  gui flat  gdl form  aql term open archet  \
CDR_name
EHRBase    no   no  yes  yes  yes  yes  yes   no   no  yes   no  yes     no
Better    yes   no  yes  yes  yes  yes  yes  yes  yes  yes  yes   no    yes
Base24    yes  yes   no   no   no  yes   no   no  yes   no  yes   no     no
Cabo       no   no  yes   no  dev  yes   no   no   no   no   no  yes     no
ArenaEHR  yes   no  yes   no  yes  yes  yes  yes  yes  yes  yes   no     no
eWeave    yes  yes   no   no   no  yes   no   no  yes  yes  yes   no     no
EHRCare   yes  yes  yes  yes   no   no  yes   no  yes  yes   no   no     no
Clever    yes  yes  yes  yes   no  yes  yes  yes  yes  yes   no   no    yes
EHRDB     yes   no  yes  yes  yes  yes  yes   no  yes  yes   no   no     no
RHP       yes  yes  yes  yes   no  yes  yes   no  yes   no  yes   no    yes
EHRNet    yes   no  yes   no   no  yes  yes   no  dev  yes  yes   no     no

         temp   fi   fe  extr oauth2
CDR_name
EHRBase    no  dev  dev    no    yes
Better    yes  dev  yes    no    yes
Base24     no  yes  yes    no    yes
Cabo       no  ext  ext    no     no
ArenaEHR   no  yes  yes  none    yes
eWeave    yes   no   no    no     no
EHRCare    no   no   no    no     no
Clever    yes   no   no   yes    yes
EHRDB      no  dev  dev   dev    yes
RHP       yes   no   no    no     no
EHRNet     no  yes  yes    no     no
```

[565]:
```python
#replace all dev with yes
df.replace("dev","yes",inplace=True)
#df.replace("dev","no",inplace=True)
df.replace("none","no",inplace=True)
df.replace("no","n",inplace=True)
df.replace("yes","y",inplace=True)
df.replace("ext","e",inplace=True)
df
```

[565]:
```
         com dem rest Java api gui flat gdl form aql term open archet temp fi  \
CDR_name
EHRBase    n   n    y    y   y   y    y   n    n   y    n    y      n    n  y
Better     y   n    y    y   y   y    y   y    y   y    y    n      y    y  y
Base24     y   y    n    n   n   y    n   n    y   n    y    n      n    n  y
Cabo       n   n    y    n   y   y    n   n    n   n    n    y      n    n  e
ArenaEHR   y   n    y    n   y   y    y   y    y   y    y    n      n    n  y
eWeave     y   y    n    n   n   y    n   n    y   y    y    n      n    y  n
EHRCare    y   y    y    y   n   n    y   n    y   y    n    n      n    n  n
Clever     y   y    y    y   n   y    y   y    y   y    n    n      y    y  n
EHRDB      y   n    y    y   y   y    y   n    y   y    n    n      n    n  y
RHP        y   y    y    y   n   y    y   n    y   n    y    n      y    y  n
EHRNet     y   n    y    n   n   y    y   n    y   y    y    n      n    n  y

         fe extr oauth2
CDR_name
EHRBase   y    n      y
Better    y    n      y
Base24    y    n      y
Cabo      e    n      n
ArenaEHR  y    n      y
eWeave    n    n      n
EHRCare   n    n      n
Clever    n    y      y
EHRDB     y    y      y
RHP       n    n      n
EHRNet    y    n      n
```

[566]:
```
#df.drop(['Java','form','com','term','archet','temp','extr','gdl','fi'],axis=1,␣
 ↪inplace = True)
#df.drop(['Java','term','archet','temp','extr','gdl'],axis=1, inplace = True)
#df.drop(['form','gdl','temp','archet'],axis=1, inplace = True)
#df.drop(['Java','term','archet','temp','extr','gdl'],axis=1, inplace = True)
df.drop(['Java','term','archet','temp','extr','gdl','oauth2'],axis=1, inplace =␣
 ↪True)
df
```

[566]:
```
         com dem rest api gui flat form aql open fi fe
CDR_name
EHRBase    n   n    y   y   y    y    n   y    y  y  y
Better     y   n    y   y   y    y    y   y    n  y  y
Base24     y   y    n   n   y    n    y   n    n  y  y
Cabo       n   n    y   y   y    n    n   n    y  e  e
ArenaEHR   y   n    y   y   y    y    y   y    n  y  y
eWeave     y   y    n   n   y    n    y   y    n  n  n
EHRCare    y   y    y   n   n    y    y   y    n  n  n
Clever     y   y    y   n   y    y    y   y    n  n  n
```

```
EHRDB      y   n    y   y   y    y    y   y    n  y  y
RHP        y   y    y   n   y    y    y   n    n  n  n
EHRNet     y   n    y   n   y    y    y   y    n  y  y
```

[567]:

```
df.to_csv('pippo1.csv')
```

[568]:

```
mca = MCA(n_components = 2, n_iter = 3, random_state = 101)
mca.fit(df)
df_mca = mca.transform(df)
df_mca
```

[568]:

```
                 0         1
EHRBase   0.964511 -0.384791
Better    0.065072 -0.656480
Base24   -0.267734  0.412437
Cabo      1.686646  0.882755
ArenaEHR  0.065072 -0.656480
eWeave   -0.618756  0.650747
EHRCare  -0.736238  0.285389
Clever   -0.586371  0.191964
EHRDB     0.065072 -0.656480
RHP      -0.509214  0.426850
EHRNet   -0.128061 -0.495913
```

[569]:

```
mca.explained_inertia_ #variance explained
```

[569]:

```
[0.41568284407988304, 0.2584786474008659]
```

[570]:

```
round(sum(mca.explained_inertia_)*100,1)
```

[570]:

```
67.4
```

[571]:

```
 mca.eigenvalues_
```

[571]:

```
[0.49126154300349817, 0.30547476511011423]
```

[572]:

```
mca.column_coordinates(df)
```

[572]:

```
               0         1
com_n   1.891251  0.450485
com_y  -0.420278 -0.100108
dem_n   0.646386 -0.593268
dem_y  -0.775663  0.711922
rest_n -0.632394  0.961814
rest_y  0.140532 -0.213736
api_n  -0.676837  0.443725
api_y   0.812205 -0.532470
```

```
gui_n  -1.050417  0.516356
gui_y   0.105042 -0.051636
flat_n  0.380537  1.173601
flat_y -0.142701 -0.440100
form_n  1.891251  0.450485
form_y -0.420278 -0.100108
aql_n   0.432633  1.038568
aql_y  -0.162237 -0.389463
open_n -0.420278 -0.100108
open_y  1.891251  0.450485
fi_e    2.406399  1.597176
fi_n   -0.874082  0.703346
fi_y    0.181655 -0.735093
fe_e    2.406399  1.597176
fe_n   -0.874082  0.703346
fe_y    0.181655 -0.735093
```

[573]:
```
#The result is like the PCA or CA result, two principal components with SVD␣
 ↪result as the values. Just like previous techniques, we could plot the␣
 ↪coordinates into a two-dimension graph.
mca.column_coordinates(df)
ax=mca.plot_coordinates(X␣
 ↪=df,figsize=(8,8),show_row_points=True,show_row_labels=True,
                     show_column_points=False, show_column_labels=False,
                    row_points_size=30, column_points_size=30)
```

```
<IPython.core.display.Javascript object>
```

```
<IPython.core.display.HTML object>
```

[574]:
```
ax.set_title('CDR in Principal Coordinates')
```

[574]:
```
Text(0.5, 1.0, 'CDR in Principal Coordinates')
```

[575]:
```
ax.get_figure().savefig('mca_coordinates_11cdr.svg')
ax.get_figure().savefig('mca_coordinates_11cdr.png')
```

[576]:
```
mca = MCA(n_components = 3, n_iter = 3, random_state = 101)
mca.fit(df)
df_mca = mca.transform(df)
df_mca
```

[576]:
```
                 0         1         2
EHRBase   0.964511 -0.384791 -0.242599
Better    0.065072 -0.656480  0.094008
Base24   -0.267734  0.412437  0.948881
```

```
Cabo       1.686646  0.882755 -0.111395
ArenaEHR   0.065072 -0.656480  0.094008
eWeave    -0.618756  0.650747  0.350136
EHRCare   -0.736238  0.285389 -0.800813
Clever    -0.586371  0.191964 -0.343925
EHRDB      0.065072 -0.656480  0.094008
RHP       -0.509214  0.426850 -0.190005
EHRNet    -0.128061 -0.495913  0.107695
```

[577]:
```
print(df_mca.iloc[0,2])
index=df_mca.index
print(index[0])
```

```
-0.24259857398808438
EHRBase
```

[578]:
```
mca.explained_inertia_ #variance explained
```

[578]:
```
[0.41568284407988343, 0.2584786474008658, 0.14831007260136259]
```

[579]:
```
round(sum(mca.explained_inertia_)*100,1)
```

[579]:
```
82.2
```

[580]:
```
 mca.eigenvalues_
```

[580]:
```
[0.4912615430034986, 0.3054747651101141, 0.17527554034706488]
```

[581]:
```
mca.column_coordinates(df)
```

[581]:
```
               0         1         2
com_n   1.891251  0.450485 -0.422771
com_y  -0.420278 -0.100108  0.093949
dem_n   0.646386 -0.593268  0.014222
dem_y  -0.775663  0.711922 -0.017067
rest_n -0.632394  0.961814  1.551401
rest_y  0.140532 -0.213736 -0.344756
api_n  -0.676837  0.443725  0.028651
api_y   0.812205 -0.532470 -0.034381
gui_n  -1.050417  0.516356 -1.912803
gui_y   0.105042 -0.051636  0.191280
flat_n  0.380537  1.173601  0.945575
flat_y -0.142701 -0.440100 -0.354591
form_n  1.891251  0.450485 -0.422771
form_y -0.420278 -0.100108  0.093949
aql_n   0.432633  1.038568  0.515519
aql_y  -0.162237 -0.389463 -0.193320
```

```
open_n -0.420278 -0.100108  0.093949
open_y  1.891251  0.450485 -0.422771
fi_e    2.406399  1.597176 -0.266076
fi_n   -0.874082  0.703346 -0.587953
fi_y    0.181655 -0.735093  0.436314
fe_e    2.406399  1.597176 -0.266076
fe_n   -0.874082  0.703346 -0.587953
fe_y    0.181655 -0.735093  0.436314
```

[583]:

```
%matplotlib notebook
import matplotlib.pyplot as plt

from mpl_toolkits.mplot3d import Axes3D


from matplotlib import interactive,pyplot
from mpl_toolkits.mplot3d import Axes3D
from numpy.random import rand
from pylab import figure


m=rand(3,3) # m is an array of (x,y,z) coordinate triplets

fig = figure()
ax = fig.add_subplot(projection='3d')
labels=['EHRBase', 'Better', 'Base24', 'Cabo', 'ArenaEHR', 'eWeave',␣
 ↪'EHRCare','Clever', 'EHRDB','RHP','EHRNet']
colors=['black', 'black', 'red','green',␣
 ↪'black','red','red','red','black','red','black']
markers=['o','p','>','*','H','^','<','v','D','1','2']
pippo=['EHRBase', 'Better,ArenaEHR,EHRDB', 'Base24', 'Cabo',␣
 ↪'Better,ArenaEHR,EHRDB', 'eWeave', 'EHRCare','Clever',␣
 ↪'Better,ArenaEHR,EHRDB','RHP','EHRNet']
#colors=['black', 'black', 'red','green',␣
 ↪'black','red','red','red','black','red']
#markers=['o','p','>','*','H','^','<','v','D','1']

for i in range(len(df_mca)): #plot each point + its index as text above
    ax.scatter(df_mca.iloc[i,0],df_mca.iloc[i,1],df_mca.
 ↪iloc[i,2],color=colors[i],marker=markers[i],s=30,label=labels[i])
#    ax.text(df_mca.iloc[i,0],df_mca.iloc[i,1],df_mca.iloc[i,2],  '%s' %␣
 ↪(pippo[i]), size=10, zorder=1,  color='k')

ax.set_xlabel(f'component 0 {round(mca.explained_inertia_[0]*100,1)}%',␣
 ↪fontsize=14)
```

```
ax.set_ylabel(f'component 1 {round(mca.explained_inertia_[1]*100,1)}%',␣
 ↪fontsize=14)
ax.set_zlabel(f'component 2 {round(mca.explained_inertia_[2]*100,1)}%',␣
 ↪fontsize=14)

#ax.set_title('CDRs in 3 Principal Components')
#ax.legend(loc="best",ncol=3,fontsize='small')
plt.show()
```

```
<IPython.core.display.Javascript object>


<IPython.core.display.HTML object>
```

[584]:
```
ax.get_figure().savefig('3d_mca_coordinates_11cdr_t.svg')
ax.get_figure().savefig('3d_mca_coordinates_11cdr_t.png')
```

[556]:
```
df_mca.index
```

[556]:
```
Index(['EHRBase', 'Better', 'Base24', 'Cabo', 'ArenaEHR', 'eWeave', 'EHRCare',
       'Clever', 'EHRDB', 'RHP', 'EHRNet'],
      dtype='object')
```

[557]:
```
mca.explained_inertia_
```

[557]:
```
[0.41568284407988343, 0.2584786474008658, 0.14831007260136259]
```

[558]:
```
#newindex=['Base24','eWeave','EHR_Care','Clever','Cabolabs','EHRBASE,','ArenaEHR','Better','EHR
 ↪DB']
#cat={'0':
 ↪['Base24','eWeave','EHR_Care','Clever','Cabolabs','EHRBASE,','ArenaEHR','Better','EHR␣
 ↪DB'],'1':
 ↪['Base24','eWeave','EHR_Care','Clever','Cabolabs','EHRBASE,','ArenaEHR','Better','EHR␣
 ↪DB'],'2':
 ↪['Base24','eWeave','EHR_Care','Clever','Cabolabs','EHRBASE,','ArenaEHR','Better','EHR␣
 ↪DB']}
#cat
df2_mca=df_mca.
 ↪reindex(['Base24','eWeave','EHRCare','Clever','RHP','Cabo','EHRBase','ArenaEHR','Better','EHR
#df2_mca=df_mca.
 ↪reindex(['Base24','eWeave','EHRCare','Clever','RHP','Cabo','EHRBase','ArenaEHR','Better','EHR
df2_mca
```

[558]:
```
                0         1         2
Base24  -0.267734  0.412437  0.948881
eWeave  -0.618756  0.650747  0.350136
EHRCare -0.736238  0.285389 -0.800813
```

```
Clever   -0.586371  0.191964 -0.343925
RHP      -0.509214  0.426850 -0.190005
Cabo      1.686646  0.882755 -0.111395
EHRBase   0.964511 -0.384791 -0.242599
ArenaEHR  0.065072 -0.656480  0.094008
Better    0.065072 -0.656480  0.094008
EHRDB     0.065072 -0.656480  0.094008
EHRNet   -0.128061 -0.495913  0.107695
```

[559]:
```python
symbols=['triangle-right','triangle-up','triangle-left','triangle-down','1','star','circle','ci
s=dict(zip(df2_mca.index,symbols))
colors=['red', 'red', 'red','red',␣
 ↪'red','green','black','black','black','black','black']
c=dict(zip(df2_mca.index,colors))
#c=dict(zip(''*len(colors),colors))
for i,ss in enumerate(s):
    print(ss,s[ss],c[ss])
```

```
Base24 triangle-right red
eWeave triangle-up red
EHRCare triangle-left red
Clever triangle-down red
RHP 1 red
Cabo star green
EHRBase circle black
ArenaEHR circle-x black
Better circle-cross black
EHRDB circle-cross-open black
EHRNet circle-dot black
```

[36]:
```python
#symbols=['square','circle','triangle-right','star','circle','triangle-up','triangle-left','tri
#s=dict(zip(df2_mca.index,symbols))
#colors=['red','black',,'green','black','#00AAFF','#AA00FF','orange','black']
#c=dict(zip(df2_mca.index,colors))
#s
#c
```

[37]:
```python
for a in df_mca.columns:
    print(a)
```

```
0
1
2
```

[309]:
```python
df_mca
```

```
[309]:                   0         1         2
        EHRBase   0.944200 -0.397299 -0.217138
        Better    0.091081 -0.680729  0.034501
        Base24   -0.247969  0.187577  0.999693
        Cabo      1.574615  0.981279 -0.059103
        ArenaEHR  0.091081 -0.680729  0.034501
        eWeave   -0.620275  0.637505  0.373042
        EHRCare  -0.730518  0.401874 -0.756321
        Clever   -0.553912  0.073379 -0.221952
        EHRDB     0.091081 -0.680729  0.034501
        RHP      -0.513183  0.485590 -0.166703
        EHRNet   -0.126202 -0.327716 -0.055020
```

```
[560]: import plotly.express as px

        labels = {
            str(i): f'Comp {str(i+1)} {round(var*100,1)}%'
            for i, var in enumerate(mca.explained_inertia_)
        }

        fig = px.scatter_matrix(
            df2_mca,
            labels=labels,
            dimensions=range(3),
        #     category_orders=cat,
            color=df2_mca.index,
            color_discrete_map=c,
            symbol=df2_mca.index,
            symbol_map=s,
            height=800, width=800,
            size=[15]*11,size_max=12
        )
        fig.update_traces(diagonal_visible=True)
        fig.show()
```

```
[561]: fig.write_image('scatterplot3dcomponents_11CDR.svg')
        fig.write_image('scatterplot3dcomponents_11CDR.png')
```

```
[562]: df2=df.
         ↪reindex(['Base24','eWeave','EHRCare','Clever','RHP','Cabo','EHRBase','ArenaEHR','Better','EH
        df2
```

```
[562]:          com dem rest api gui flat form aql open fi fe
        CDR_name
        Base24     y   y    n   n   y    n    y   n    n  y  y
        eWeave     y   y    n   n   y    n    y   y    n  n  n
        EHRCare    y   y    y   n   n    y    y   y    n  n  n
```

```
Clever     y   y    y   n   y    y    y   y    n  n  n
RHP        y   y    y   n   y    y    y   n    n  n  n
Cabo       n   n    y   y   y    n    n   n    y  e  e
EHRBase    n   n    y   y   y    y    n   y    y  y  y
ArenaEHR   y   n    y   y   y    y    y   y    n  y  y
Better     y   n    y   y   y    y    y   y    n  y  y
EHRDB      y   n    y   y   y    y    y   y    n  y  y
EHRNet     y   n    y   n   y    y    y   y    n  y  y
```

[ ]: