



Componentes principales

Integrantes

Juan Pablo Oriana - 60621

Tomas Cerdeira - 60051

Santiago Garcia Montagner - 60352



Objetivos del trabajo

- Calcular las componentes principales de un dataset e interpretar la primera componente.
- Datos:

Country	Area	GDP	Inflation	Life.expect	Military	Pop.growth
Austria	83871	41600	3.5	79.91	0.8	0.03
Belgium	30528	37800	3.5	79.65	1.3	0.06
Bulgaria	110879	13800	4.2	73.84	2.6	-0.8
Croatia	56594	18000	2.3	75.99	2.39	-0.09
Czech Republic	78867	27100	1.9	77.38	1.15	-0.13
Denmark	43094	37000	2.8	78.78	1.3	0.24
Estonia	45228	20400	5	73.58	2	-0.65



Objetivo de componentes principales

- Disminuir la dimensionalidad del conjunto de entrada de tal manera que se mantenga la varianza/dispersión en cada dimensión
- No tomar en cuenta información redundante/ altamente correlacionada



Componentes principales



Implementación

```
from sklearn.decomposition import PCA
if __name__ == "__main__":
    pca = PCA()
    data_standardized, countries_data, labels = getInputsStandard('europe.csv')
    components = pca.fit_transform(data_standardized)

    principal_df = pd.DataFrame(data=components
                                , columns=['principal component ' + str(i) for i in range(7)])
    np.set_printoptions(precision=6, suppress=True)
    accum_variance = 0
    for idx, variance in enumerate(pca.explained_variance_ratio_):
        accum_variance += variance
```

Estandarización de los datos

Tenemos las variables $\{X_1, \dots, X_p\}$, cada una posee n registros.

- La media de X_i : $\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_i^j$
- La desviación estándar de X_i : $s_i = \frac{1}{n-1} \sum_{j=1}^n (X_i^j - \bar{X}_i)^2$
- Las variables estandarizadas:

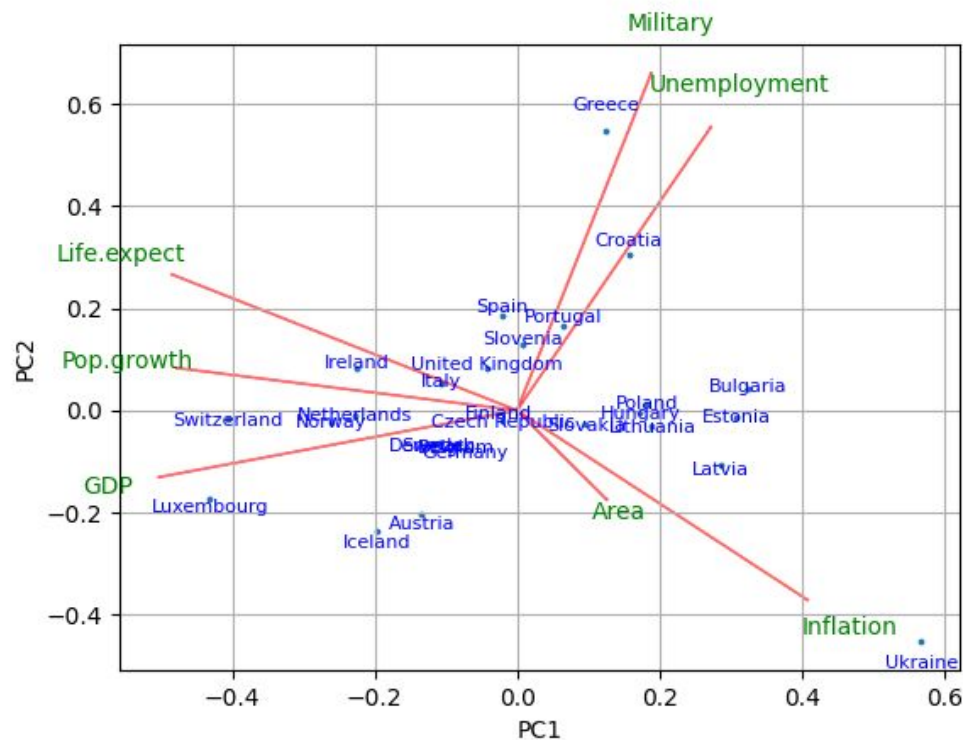
$$\tilde{X}_i = \frac{X_i - \bar{X}_i}{s_i}$$

Country	Area	GDP	Inflation	Life.expect	Military	Pop.growth	Unemployment
Austria	83871	41600	3,5	79,91	0,8	0,03	4,2
Belgium	30528	37800	3,5	79,65	1,3	0,06	7,2
Bulgaria	110879	13800	4,2	73,84	2,6	-0,8	9,6
Croatia	56594	18000	2,3	75,99	2,39	-0,09	17,7
Czech Republic	78867	27100	1,9	77,38	1,15	-0,13	8,5
Denmark	43004	37000	2,8	78,78	1,2	0,14	6,1



Country	Area	GDP	Inflation	Life.expect	Military	Pop.growth	Unemployment
Austria	-0,49868	0,671577	0,11238454	0,560493	-1,00589	-0,1736033	-1,223083601
Belgium	-0,82092	0,409546	0,11238454	0,478967	-0,38194	-0,1138382	-0,581766357
Bulgaria	-0,33553	-1,24538	0,613006582	-1,34283	1,240314	-1,8271036	-0,068712562
Croatia	-0,66346	-0,95577	-0,745824675	-0,66867	0,978257	-0,4126636	1,662843996
Czech Republic	-0,52891	-0,32828	-1,031894413	-0,23282	-0,56913	-0,4923503	-0,303862218
Denmark	-0,74501	0,354382	-0,388237502	0,206167	-0,38194	0,2447522	-0,816916013
Estonia	0,73312	0,70038	1,185146050	1,43436	0,40158	1,5383783	0,55133744

Resultados obtenidos



Resultados obtenidos



```
[[ 0.124874 -0.500506 0.406518 -0.482873 0.188112 -0.475704 0.271656]
[-0.172872 -0.13014 -0.369657 0.265248 0.658267 0.082622 0.553204]
[ 0.898297 0.083956 0.198195 0.246082 0.243679 0.163697 0.0005 ]
[ 0.04485 -0.084255 0.164686 0.026771 -0.562375 0.392463 0.701968]
[-0.324017 0.390632 0.689501 -0.101787 0.368148 0.347868 0.010159]
[ 0.190118 0.638657 -0.323867 -0.606434 0.035596 -0.120856 0.259705]
[ 0.066643 -0.397408 -0.2267 -0.507031 0.13731 0.671147 -0.244662]]
```

Correlación acumulada:

component 1: 0.4610236684445143
component 2: 0.6306127267691592
component 3: 0.7824970888162958
component 4: 0.8925479431570047
component 5: 0.9579548886840197
component 6: 0.9820511591223072
component 7: 1.0

**Gracias por su
atención.**

