

Inteligência Artificial

Instrutores

Ph.D. Professor Aluisio Igor Rego
Fontes



Capacitação Tecnológica
em Indústria 4.0 e Cidades
Inteligentes



Redução de Dimensionalidade Atividade

Exemplo resolvido - Câncer de Mama

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
```

Importando PCA, e StandardScaler que escala os dados de forma que a PCA entenda.

```
from sklearn.datasets import load_breast_cancer
cancer = load_breast_cancer()
```

load_breast_cancer é uma função do scikit-learn que importa um *Dataset* relacionado a diagnósticos de câncer de mama, nele além de informações sobre os nódulos existe a classificação entre maligno e benigno.

Exemplo resolvido - Câncer de Mama

```
[ ] print(cancer['data'])
```

```
[[1.799e+01 1.038e+01 1.228e+02 ... 2.654e-01 4.601e-01 1.189e-01]
 [2.057e+01 1.777e+01 1.329e+02 ... 1.860e-01 2.750e-01 8.902e-02]
 [1.969e+01 2.125e+01 1.300e+02 ... 2.430e-01 3.613e-01 8.758e-02]
 ...
 [1.660e+01 2.808e+01 1.083e+02 ... 1.418e-01 2.218e-01 7.820e-02]
 [2.060e+01 2.933e+01 1.401e+02 ... 2.650e-01 4.087e-01 1.240e-01]
 [7.760e+00 2.454e+01 4.792e+01 ... 0.000e+00 2.871e-01 7.039e-02]]
```

Está em formato de numpy array

```
[ ] print(cancer['target_names'])
```

```
['malignant' 'benign']
```

Mas já se pode notar as classificações e colunas

```
[ ] print(cancer['feature_names'])
```

```
['mean radius' 'mean texture' 'mean perimeter' 'mean area'
 'mean smoothness' 'mean compactness' 'mean concavity'
 'mean concave points' 'mean symmetry' 'mean fractal dimension'
 'radius error' 'texture error' 'perimeter error' 'area error'
 'smoothness error' 'compactness error' 'concavity error'
 'concave points error' 'symmetry error' 'fractal dimension error'
 'worst radius' 'worst texture' 'worst perimeter' 'worst area'
 'worst smoothness' 'worst compactness' 'worst concavity'
 'worst concave points' 'worst symmetry' 'worst fractal dimension']
```

Exemplo resolvido - Câncer de Mama

```
df_cancer = pd.DataFrame(np.c_[cancer['data'], cancer['target']], columns =
np.append(cancer['feature_names'], ['target']))
```

E assim se transforma em *dataframe*

```
df_cancer.shape
```

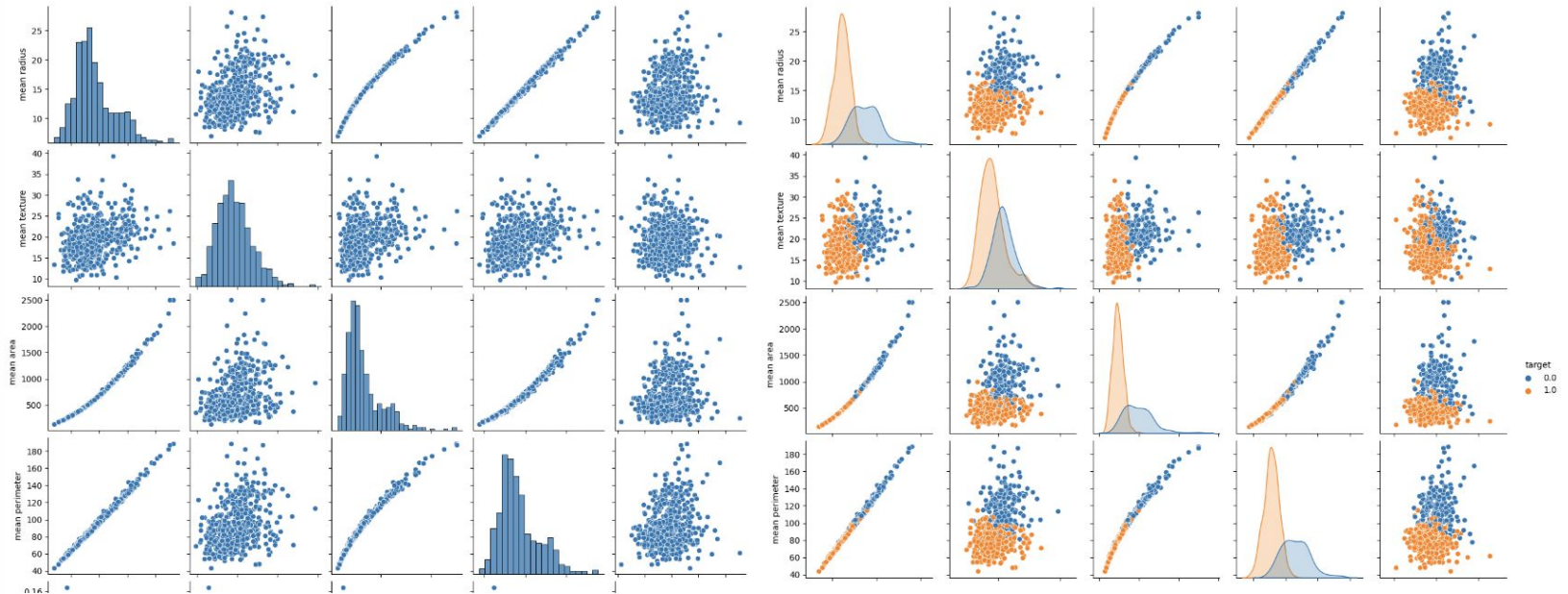
```
(569, 31)
```

```
df_cancer.head()
```

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mean concavity	mean concave points	mean symmetry	mean fractal dimension	...
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.3001	0.14710	0.2419	0.07871	...
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.0869	0.07017	0.1812	0.05667	...
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.1974	0.12790	0.2069	0.05999	...
3	11.42	20.38	77.58	386.1	0.14250	0.28390	0.2414	0.10520	0.2597	0.09744	...
4	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.1980	0.10430	0.1809	0.05883	...

```
5 rows x 31 columns
```

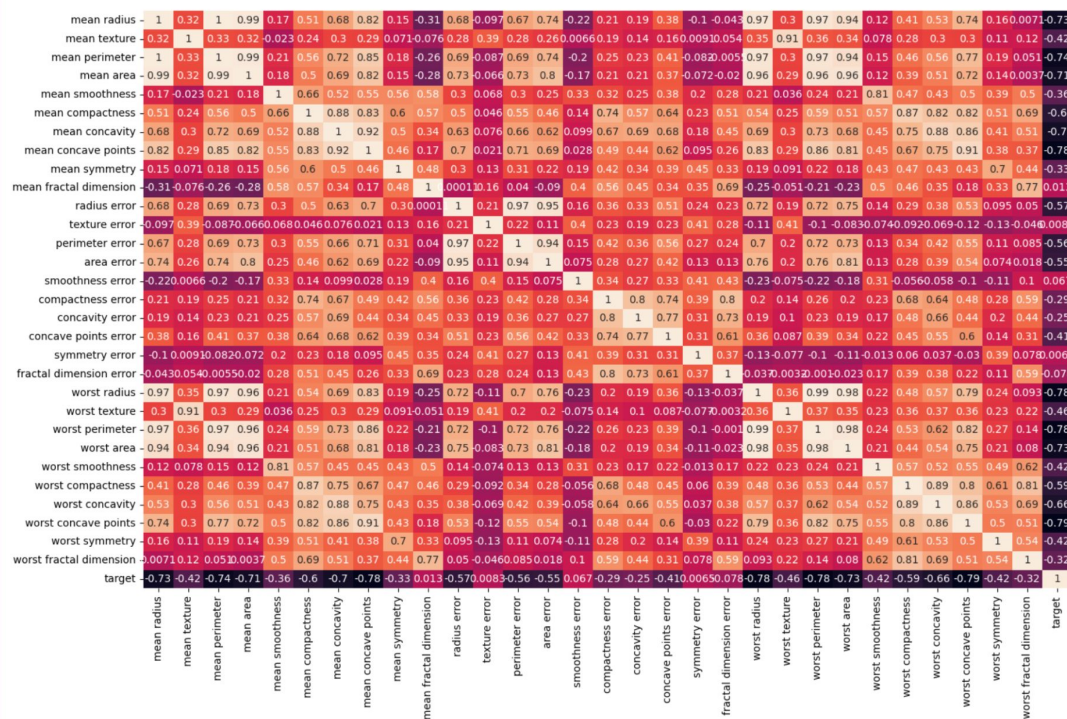

Exemplo resolvido - Câncer de Mama



Nota-se a presença forte de padrões e correlações

Maligno = 1 (Laranja)
Benigno = 0 Azul

Exemplo resolvido - Câncer de Mama



Plotando o HeatMap nota-se ainda mais a presença de padrões, então vamos explorar a possibilidade de reduzir essas 31 dimensões utilizando PCA para fazer uma classificação.

Aplicar PCA

```
X = df_cancer.drop('target', axis=1)
y = df_cancer['target']

scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

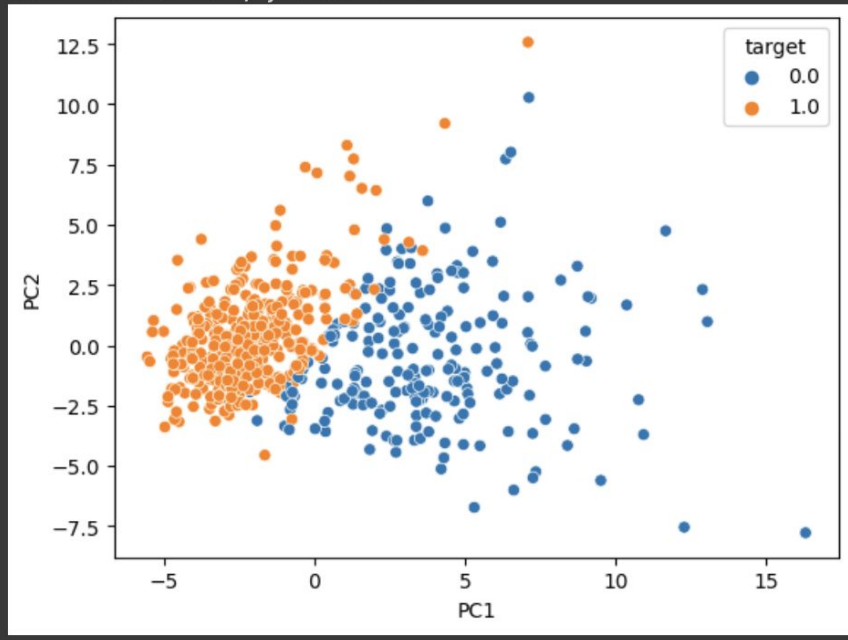
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X_scaled)

df_pca = pd.DataFrame(data=X_pca, columns=['PC1', 'PC2'])
df_pca['target'] = y
```


Resultado

```
sns.scatterplot(x='PC1', y='PC2', hue='target', data=df_pca)
```

<Axes: xlabel='PC1', ylabel='PC2'>



De 31 colunas reduziu-se para 2 e agora pode-se obser dois clusters bem definidos.

Atividade

Faça a redução de dimensionalidade do dataset sobre o preço das casas na cidade de Melbourne - Austrália.

PCA é muito utilizado em problemas financeiros como o deste dataset, que possui muitas tabelas correlacionadas

