

## Exercício 5: PCA Clássica

Os dados provenientes do *dataset* iris possuem 3 classificações possíveis (setosa, versicolor e virginica), com 4 atributos cada, sendo constituída por 150 exemplares (50 de cada). Neste exercício, iremos reduzir o tamanho do *dataset* usando o algoritmo de PCA. Para isso, usamos a biblioteca *sklearn* para obter a base de dados iris.

```
iris = load_iris()
#x = matriz, y = classe
x, y = iris.data, iris.target
```

Uma vez que obtemos *dataset*, calculamos a matriz de covariância da mesma utilizando a função *cov* da biblioteca *numpy*. Após calcular a matriz de covariância, os autovalores e autovetores são encontrados, utilizando a função *eig* (também da *numpy*).

```
#Matriz de covariância
cov = np.cov(x.T)

#Covalor e Covetor
val, vet = eig(cov)
```

Com os autovalores e autovetores em mãos, é realizado uma ordenação desses vetores (de acordo com seus autovalores), e os maiores são escolhidos - os menores são descartados, uma vez que não contribuem quase nada para a base. Assim, agrupando os vetores escolhidos, temos a matriz resultado do algoritmo.

```
[ [ 0.52106591 -0.37741762]
  [-0.26934744 -0.92329566]
  [ 0.5804131  -0.02449161]
  [ 0.56485654 -0.06694199]]
```

Após agrupar os dados, realizamos um produto escalar entre a matriz original de exemplos (com 150 linhas) com a nova matriz, resultado nos “novo” grupo de dados (na verdade, apenas aplicamos a transformação nos dados).

```
[ [5.1 3.5 1.4 0.2] [-2.26470281 -0.4800266 ]
  [4.9 3. 1.4 0.2] [-2.08096115 0.67413356]
  [4.7 3.2 1.3 0.2] [-2.36422905 0.34190802]
  [4.6 3.1 1.5 0.2] [-2.29938422 0.59739451]
  [5. 3.6 1.4 0.2] [-2.38984217 -0.64683538]
  [5.4 3.9 1.7 0.4] [-2.07563095 -1.48917752]
  [4.6 3.4 1.4 0.3] [-2.44402884 -0.0476442 ]
  [5. 3.4 1.5 0.2] [-2.23284716 -0.22314807]
  [4.4 2.9 1.4 0.2] [-2.33464048 1.11532768]
  [4.9 3.1 1.5 0.1] [-2.18432817 0.46901356]
  [5.4 3.7 1.5 0.2] [-2.1663101 -1.04369065]
  [4.8 3.4 1.6 0.2] [-2.32613087 -0.13307834]
  [4.8 3. 1.4 0.1] [-2.2184509 0.72867617]
  [4.3 3. 1.1 0.1] [-2.6331007 0.96150673]
  [5.8 4. 1.2 0.2] [-2.1987406 -1.86005711]
  [5.7 4.4 1.5 0.4] [-2.26221453 -2.68628449]]
```

(Dados antigos x Dados novos)