

1. (2 points) Elaborar um programa Python para o método  $k$ -means para dados aleatórios.

1. Criar vetores  $x$  e  $y$  aleatórios

```
from sklearn.datasets import make_blobs
```

```
X_random, y_random = make_blobs(n_samples=200, centers=5, random_state=1)
```

2. Visualize os dados gerados

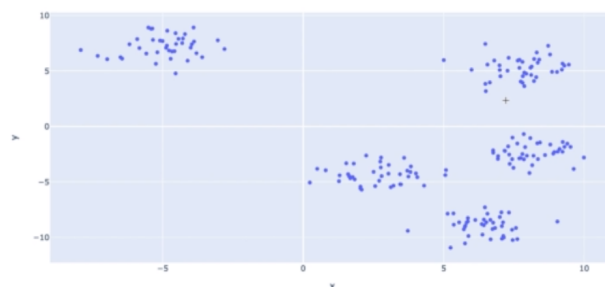
```
X_random
```

```
In [22]: X_random
Out[22]:
array([[ -8.64514414,  -7.50820548],
       [ -7.61753676,  -6.41777403],
       [ -7.59131756,  -8.3112429 ],
       [ -9.88764212,  -5.69286755],
       [ -9.21793491,  -7.2606315 ],
       [ 11.56488794,  -3.76631689],
       [ -1.35372355,   7.6634408 ],
       [-10.41485937,  -6.91043432],
       [ -7.86047666,  -6.13375427],
       [ -2.34045816,   7.82185362],
```

```
y_random
```

```
In [23]: y_random
Out[23]:
array([3, 1, 2, 3, 3, 0, 4, 3, 2, 4, 3, 3, 4, 4, 4, 0, 1, 3, 2, 3, 2, 0,
       0, 1, 0, 4, 4, 3, 4, 3, 4, 0, 4, 2, 2, 0, 0, 0, 3, 4, 1, 1, 0, 1,
       3, 3, 4, 2, 1, 3, 2, 3, 0, 1, 0, 2, 2, 3, 2, 2, 4, 0, 1, 3, 2, 1,
       1, 2, 0, 3, 2, 4, 0, 3, 2, 3, 4, 3, 1, 3, 1, 4, 2, 0, 4, 0, 1, 2,
       2, 1, 4, 1, 1, 1, 4, 0, 1, 1, 4, 0, 1, 0, 0, 3, 3, 0, 0, 1, 1, 4,
       3, 1, 2, 0, 1, 2, 3, 0, 0, 1, 0, 4, 0, 1, 2, 4, 2, 2, 2, 0, 2, 2,
       2, 0, 3, 3, 4, 4, 0, 0, 1, 4, 2, 3, 2, 4, 4, 4, 4, 4, 1, 2, 0, 1,
       1, 1, 4, 4, 3, 1, 1, 1, 2, 1, 0, 1, 2, 1, 1, 2, 0, 4, 0, 3, 3, 3,
       3, 1, 4, 3, 3, 4, 0, 0, 0, 3, 3, 2, 0, 4, 4, 4, 4, 2, 3, 2, 2, 2,
       3, 2])
```

3. Plote um gráfico dos dados para ver sua distribuição



4. Cria o  $k$ -means para o conjunto de dados para 5 *clusters*

5. Faça o ajuste do  $k$ -means (fit)

6. Aplique o *k*-means para X\_random

```
In [27]: rotulos
Out[27]:
array([4, 0, 0, 2, 1, 3, 4, 2, 2, 4, 1, 0, 4, 3, 2, 4, 3, 3, 0, 2, 3, 2,
       2, 3, 1, 4, 4, 2, 3, 3, 1, 0, 3, 3, 3, 2, 1, 4, 1, 0, 0, 4, 2, 2,
       1, 1, 4, 4, 3, 1, 1, 2, 3, 2, 1, 4, 4, 4, 1, 4, 0, 1, 1, 1, 4, 3,
       2, 1, 2, 0, 0, 2, 1, 0, 4, 4, 4, 4, 4, 4, 3, 4, 3, 1, 3, 3,
       2, 2, 1, 1, 4, 1, 0, 2, 4, 0, 1, 2, 0, 1, 4, 3, 4, 0, 3, 0, 2, 3,
       3, 4, 0, 0, 2, 2, 1, 2, 1, 4, 1, 0, 0, 3, 0, 3, 3, 2, 2, 3, 0, 2,
       2, 4, 3, 0, 1, 1, 0, 2, 4, 2, 0, 2, 2, 1, 1, 3, 2, 0, 2, 2, 3, 1,
       1, 4, 3, 1, 0, 2, 3, 1, 0, 3, 2, 3, 1, 0, 3, 2, 3, 3, 0, 2, 1, 2,
       2, 1, 0, 0, 2, 3, 3, 0, 2, 0, 2, 1, 4, 3, 2, 1, 0, 4, 1, 4, 3, 4,
       4, 0], dtype=int32)
```

## 7. Observar os centroides

```
In [28]: centroides
Out[28]:
array([[ -8.70220268,  -0.29907586],
       [  5.74652688,  -2.56757892],
       [-4.93477311,  -1.67205983],
       [-0.94888237,   0.62597036],
       [-6.44876426,   0.49442082]])
```

8. Plote o gráfico dos *clusters* com os centroides