



Trabalho 12

Nome da Atividade: K-means++

Nome e Matrícula: Lucas Lima do Nascimento - 12111ECP024

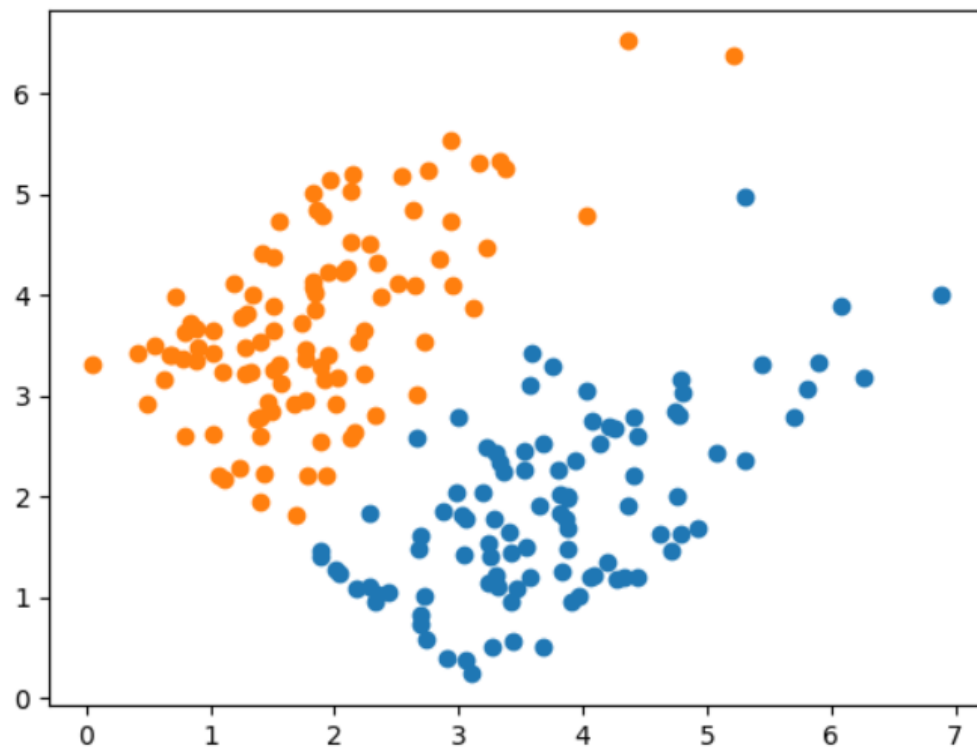
Introdução:

Nesse trabalho, fiz a implementação do algoritmo K-means++ para clusterização de dados. 200 dados foram fornecidos, tanto para X quanto para Y.

Resultados:

Como no trabalho anterior, obtive como valor mais adequado de $K = 2$, partirei do princípio de que esse é o melhor agrupamento possível para esse conjunto de dados.

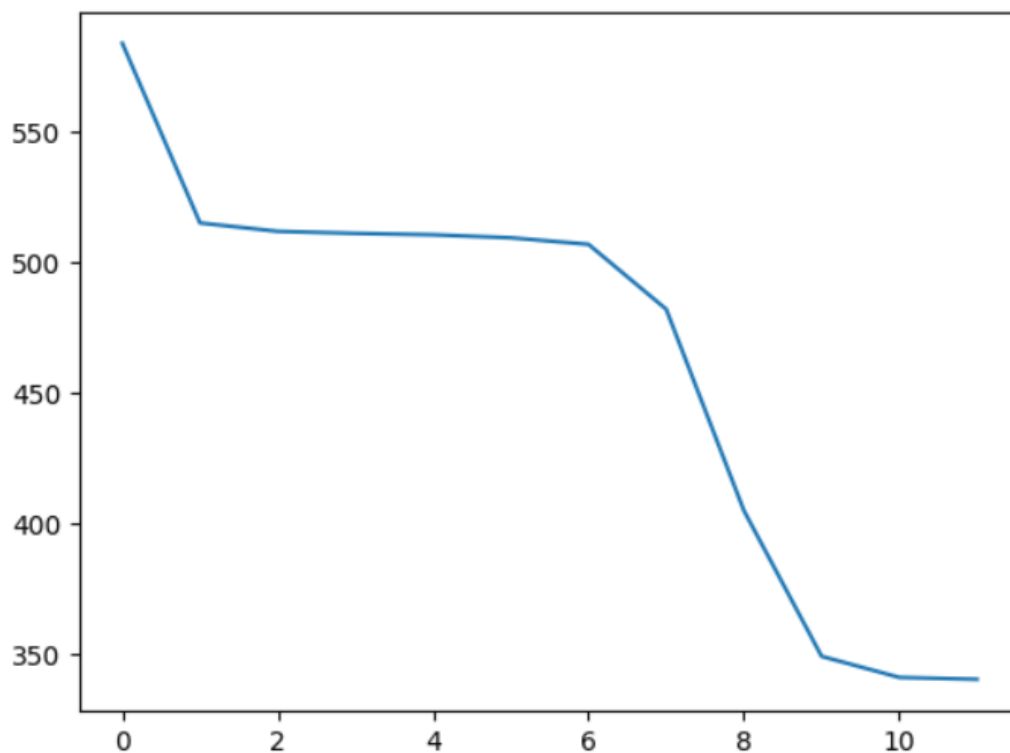
Em seguida, utilizei a biblioteca sklearn para implementar o algoritmo do K-means++, obtendo a seguinte clusterização:



Logo em seguida, para os dados acima, obtive esses valores de MSE:

```
Initialization complete  
Iteration 0, inertia 583.8033335282207.  
Iteration 1, inertia 515.1178271371449.  
Iteration 2, inertia 511.9148247334133.  
Iteration 3, inertia 511.1360873242597.  
Iteration 4, inertia 510.60274513004094.  
Iteration 5, inertia 509.447662236641.  
Iteration 6, inertia 506.9438881104418.  
Iteration 7, inertia 482.1341261458277.  
Iteration 8, inertia 405.1266322687888.  
Iteration 9, inertia 349.40922943993127.  
Iteration 10, inertia 341.35283273256323.  
Iteration 11, inertia 340.6367877954994.  
Converged at iteration 11: strict convergence.
```

Plotei eles para obter o gráfico da curva de erro quadrático:



No eixo Y, temos o valor da distância quadrática e no eixo X a iteração.

Código:

ML/kmeans++ at master · LLxD/ML

Contribute to LLxD/ML development by creating an account on GitHub.

 <https://github.com/LLxD/ML/tree/master/kmeans%2B%2B>