

Relatório Ficha 3 – Unsupervised Learning Exercise
Carlos Guerra – nº78299 METI

1) Corresponde ao excerto de código da função `run()` ou a função `run_2()`, ambas dão o mesmo resultado (os pontos do ficheiro txt disponibilizado no e-learning e lido através da função `numpy.loadtxt()`). Optei por esta solução porque na minha opinião era a maneira mais fiável de conseguir reproduzir os valores várias vezes para obter os gráficos juntamente com um valor de seed de 2 (`random.seed(2)`).

a. Pontos gerados são os que correspondem à seguinte figura:

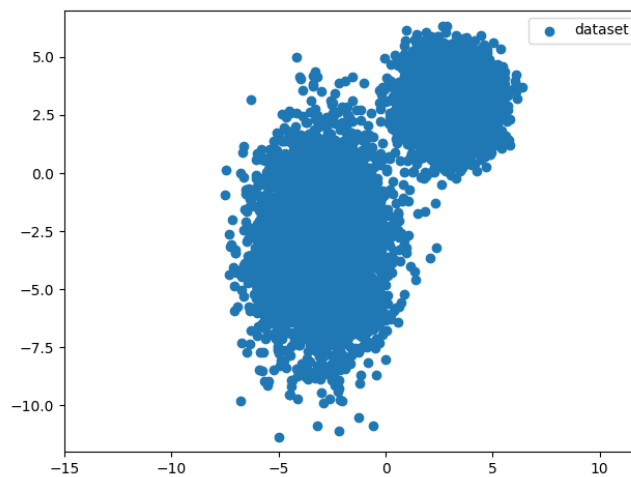


Figura 1-Full dataset de pontos

2) Para correr a função da questão 2 é necessário correr a função `run()`. As posições de `r1` e `r2` mudando para cada exemplo são as seguintes:

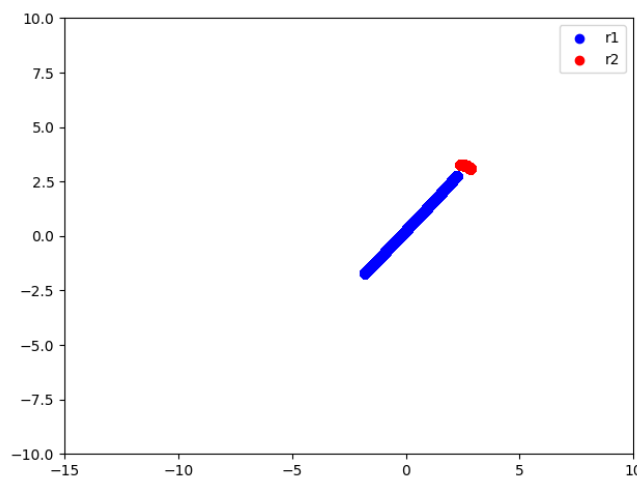


Figura 2- Evolução dos valores de r1 e r2

No gráfico seguinte, realizei um histograma de forma a poder analisar os valores de x e y ao longo do tempo para r1 e r2:

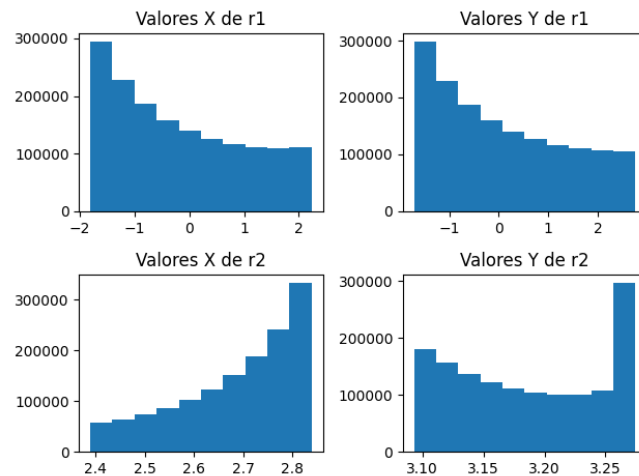


Figura 3- Histograma com os valores mais comuns de x e y para r1 e r2

Ao analisar a figura 3 reparamos que os valores mais comuns de x e y para r1 e r2 correspondem com os valores da média dos pontos para cada ser, ou seja, o mean1 e o mean2. Podemos também verificar um valor de desvio padrão de à volta de 1.75 o que é um pouco alto, significando que ainda existe alguma variação dentro dos valores.

Para mudar como se vai alterando os valores de r1 e r2 é apenas necessário correr a função `run_2()`, que mais uma vez vai buscar os valores ao ficheiro de texto. Obtendo assim como gráficos:

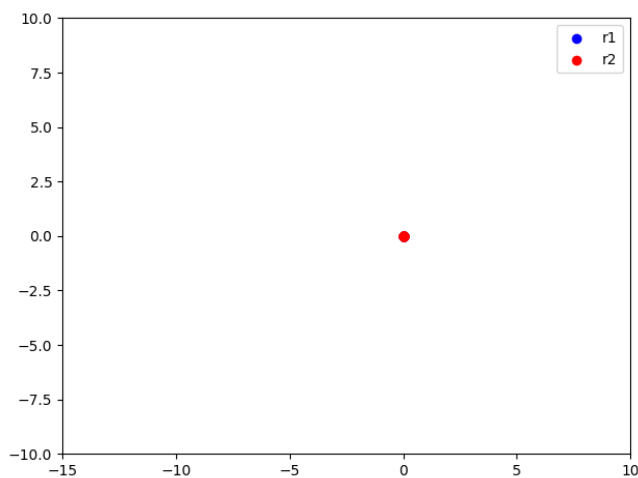


Figura 4- Variação dos valores, com escala

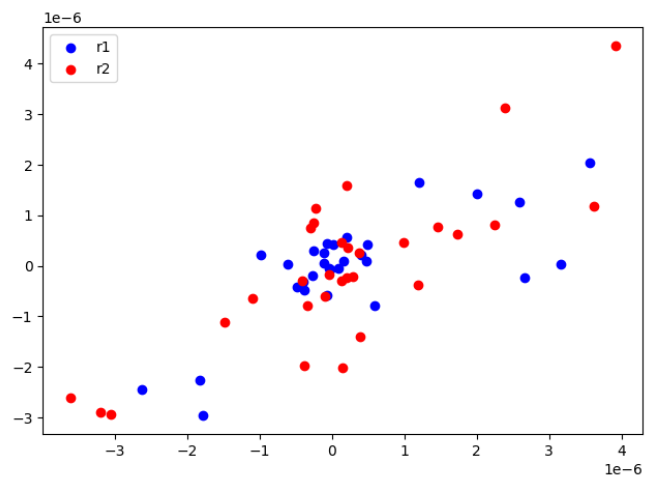


Figura 5- Variação dos valores, sem escala

É de notar que ambas as figuras correspondem ao gráfico obtido no entanto a figura 4 tem sobre ela um limite dos valores de y e x o que leva a parecer que os pontos convergem no mesmo sitio, no entanto ao retirar esses limites e obtendo assim a figura 5 é possível observar que existe sim uma variação dos valores, tendo obtido um valor de desvio de $1.4418646298808802 \times 10^{-6}$ algo que é bastante pequeno e que confirma que a variação entre os valores foi muito pequena.

Comparando com os valores do outro método de alterar os valores podemos é de notar a grande diferença entre os valores do desvio.

3) Para correr a função associada a esta pergunta é necessário correr a função `run_3()` obtendo assim o gráfico:

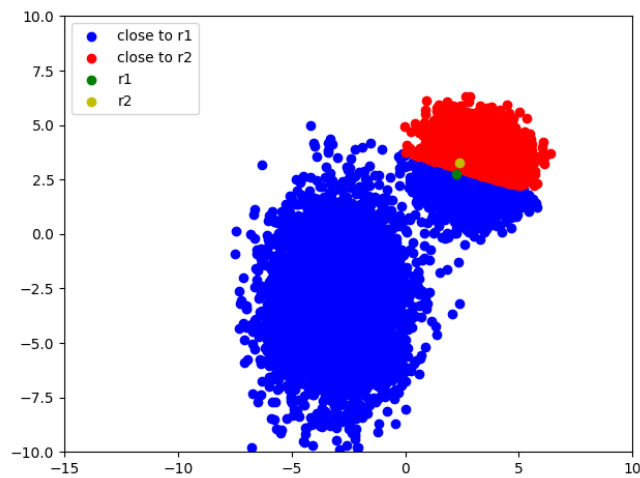
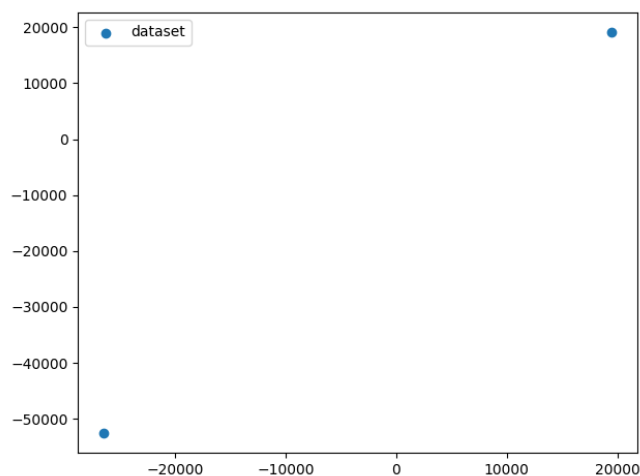


Figura 6- Pontos próximos de r1 e r2

4) Mais uma vez para obter os resultados é necessário correr a função `run_4()`, obtendo assim este gráfico:



Na minha opinião este gráfico não é um bom indicativo de qualquer um dos clusters, não só pelos seus valores muito altos, pelo facto de termos escolhido 2 pontos próximos de si e fazendo a sua média sem nenhum valor de referencia, ou seja, o que ia acontecendo é que ao fazer o valor da média os valores de x e y iam subindo de tal maneira que apenas sobraram estes 2 pontos que em nada representam os 2 clusters