Lista de exercícios Matemática Computacional
Parte B – Prof. Dr. Reinaldo Rosa - 2020

Denis M. A. Eiras

Exercício 2 - Descrição

Repita o exercício anterior considerando, entretanto, o algoritmo colorednoise.py. Neste caso, considere N= 8192 e diversifique os dados em 3 famílias: white noise, pink noise e red noise. Produza pelo menos 20 sinais para cada família. Total do Grupo ColorNoise: 60.

Exercício 2 – Detalhes da implementação

Para a resolução do Exercício 2, foi implementado o programa principal exercicio2.py para executar o exercício por completo, que utiliza as funções dos programas do exercício 1: exercicio1_2.py e exercicio1_3.py.

O exercício 2 funciona de maneira análoga ao exercício1, mas implementa a geração de sinais *colored noise* em uma função própria, onde é possível utilizar os parâmetros que determinam, a quantidade de elementos (famílias), número de sinais e betas.

Adicionalmente, foi implementado um gráfico que exibe a densidade do Power Spectrum, caso seja desejada a exibição (pode ser configurado).

Mais detalhes podem ser consultados na própria documentação do programa.

Análise

Algumas execuções do exercicio2.py mostraram que, os histogramas com dados do tipo *White Noise* e *Pink Noise* aproxima-se de uma gaussiana, como exibem as figuras 1.a, 1.b.

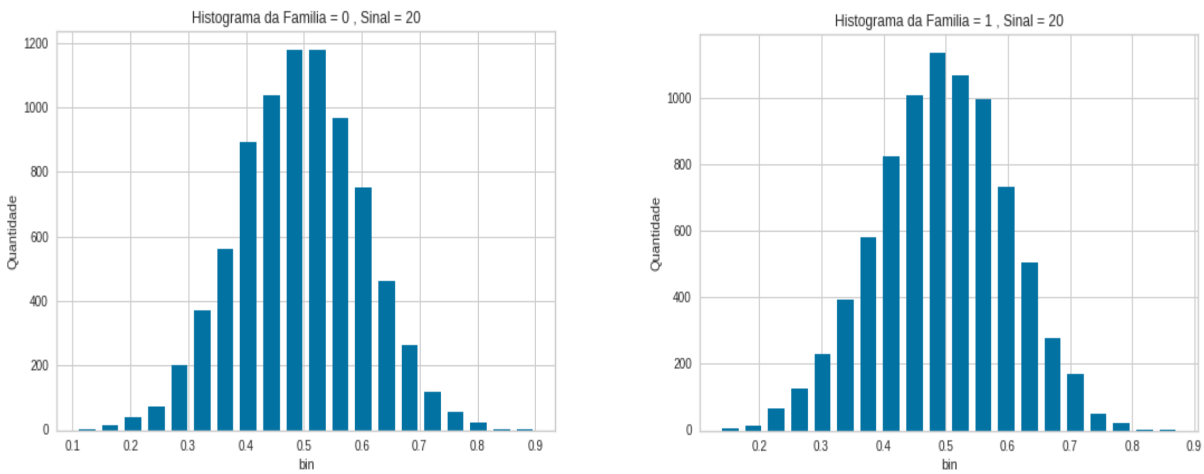

Figura 1. a) Família *White Noise*. b) Família *Pink Noise*.

A figura 2.a, exibe um histograma do ruído *Red Noise* um pouco menos próximo da curva normal, indicando uma maior quantidade de valores com maior variabilidade.

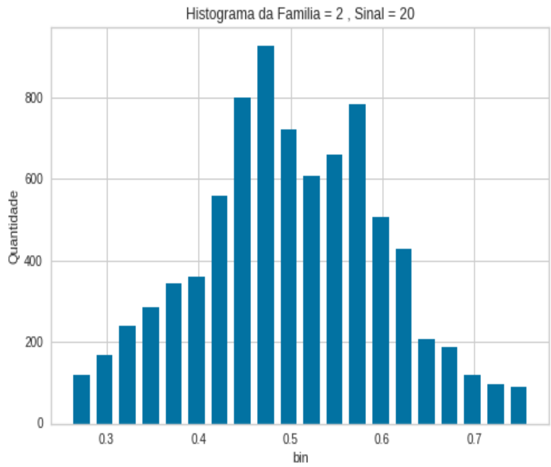

Figura 2. a) Família *Red Noise*

Adicionalmente, foi implementado o gráfico da frequência de log por log (Power Spectrum), onde é possível verificar a inclinação do ruído. O espectro de potência do ruído gerado é proporcional a S (f) = (1 / f) ** beta.

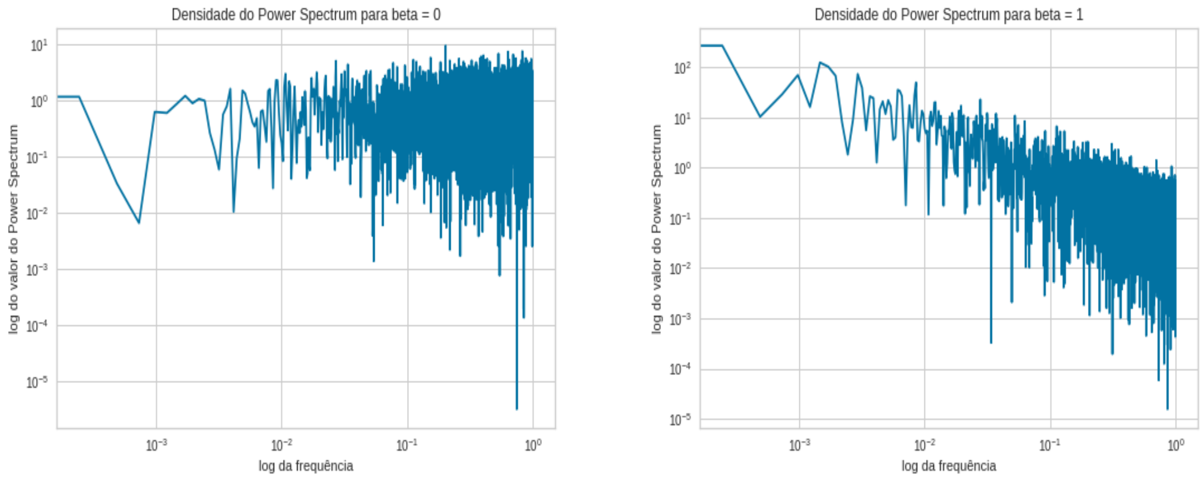

Figura 3. Densidade do Power Spectum. a) Beta = 0 b) Beta = 1.

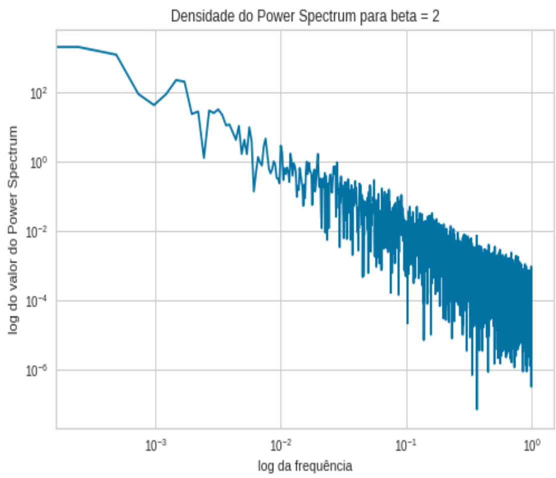

Figura 4. a) Densidade do Power Spectum para Beta = 2

Para verificar se há classes nos espaço de parâmetros composto por variância, assimetria e curtose, o algoritmo K-means foi executado com k entre 2 e 7.

A primeira estratégia para identificar o melhor k, foi analisar o gráficos gerados pelo K-means, para cada k, que exibem as áreas dos momentos estatísticos na diagonal principal e o agrupamento dos pontos para cada par de momento estatístico.

Observando todas as figuras, a figura 5 é a que exibe um melhor agrupamento dos pontos, principalmente considerando a assimetria e curtose. O agrupamento dos pontos entre variância e curtose, ou variância e assimetria, mostra que, devido à variância, alguns pontos estão próximos do centróide de outro agrupamento (*outliers*), o que pode ser confirmado com a intersecção dos gráficos da variância, no canto superior esquerdo da figura 5.

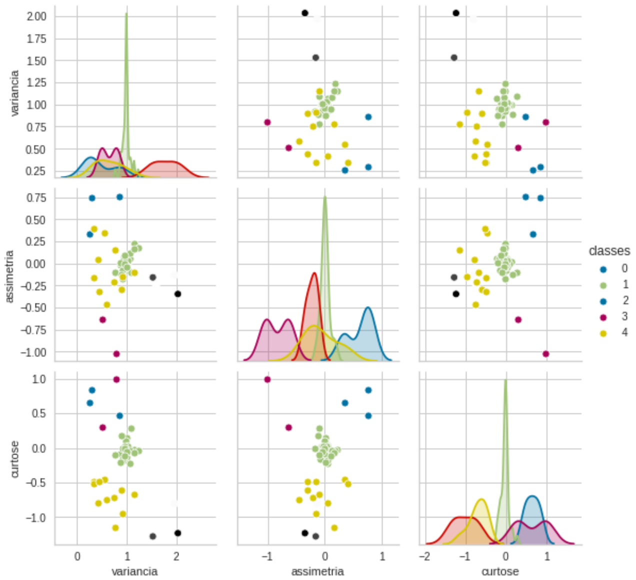

Figura 5. Agrupamentos K-means

Utilizando o método do cotovelo implementado, na figura 6.a, e a implementação do pacote yellowbrick, na figura 6.b, percebe-se uma quebra da curva para k =5 ligeiramente maior que para os outros valores de k, confirmando a escolha do k = 5 da figura 5. O pacote yellowbrick não mostrou automaticamente qual o melhor k nesta execução.

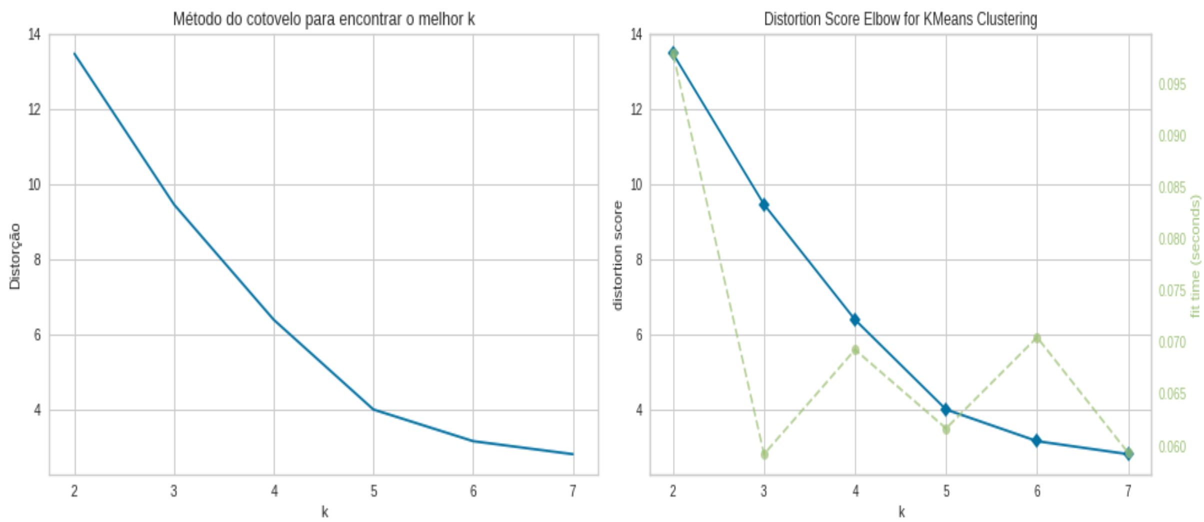

Figura 6. Métodos do cotovelo. a) implementação do autor b) utilização do pacote yellowbrick

A observação das figuras do método da silhueta também indicaram uma maior média para k = 5, sem apresentar *outliers*, como se pode observar na figura 7.

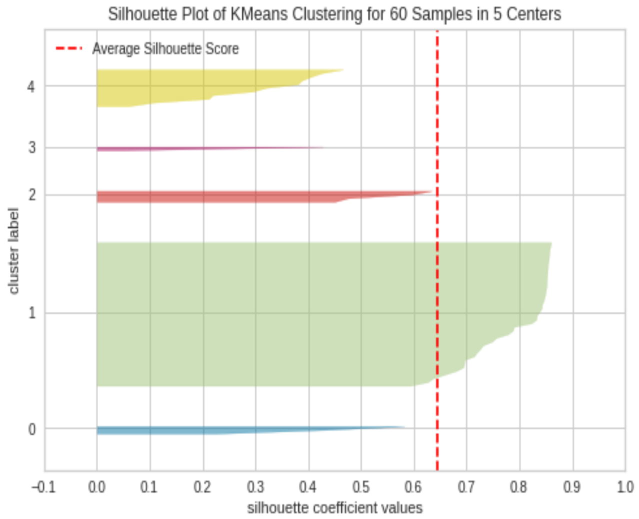

Figura 7. Método da silhueta do pacote yellowbrick.

Exercício 2 – Conclusão

O método k-means não identificou claramente uma divisão de classes bem definida para um determinado k, como ocorreu para o exercício 1, utilizando o método do cotovelo, mas o método da Silhueta ajudaram a confirma o melhor k como sendo igual a 5.

Sendo 3 a quantidade de famílias utilizadas, uma para cada Beta, pode-se verificar que, o valor de k para a melhor classificação foi igual a 5, diferente do número de famílias beta utilizadas (3). Portanto o k-means utilizado no espaço de parâmetros de variância, assimetria e curtose não é adequado para classisficar as 3 famílias utilizadas.

Os momentos estatísticos assimetria e curtose serviram como bons agrupadores, diferentemente da variância, que apresentou alguns *outliers*.