



LISTA DE EXERCÍCIOS (10/10) – Entregar em 25/5/2020

1-Simulação de Sinais Estocásticos com GRNG1.py com N valores de medidas.

1.1. Utilize o algoritmo e gere 10 sinais para cada família com N elementos:

N1: 64; N2:128; N3:256; N4:512; N5:1024; N6: 2048; N7:4096; N8: 8192

1.2. Escreva um algoritmo em Python que permita, tendo como entrada cada um dos sinais acima, obter sua forma normalizada entre 0 e 1, obter o respectivo Histograma e calcular os 4 momentos estatísticos respectivos.

1.3. Organize todos os dados num dataset (instancias x atributos) e tente agrupá-los com a técnica *K-means* para caracterizar, se houver, classes no espaço de parâmetros composto por variância, skewness e kurtosis.

2- 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*.

3- Repita o exercício anterior considerando, entretanto, o algoritmo *pmodel.py* .

Neste caso, considere N= 8192 e diversifique os dados em 3 famílias: ruído branco, Ruído rosa e ruído vermelho.

4- Espaço de Cullen-Frey e Distribuições de Probabilidades

4.1. Considere duas das séries das famílias N8 dos exercícios anteriores. Classifique as mesmas no espaço de Cullen-Frey.

4.2. Escreva (ou utilize) um algoritmo em Python que permita ajustar uma PDF (Gaussiana ou GEV) aos respectivos histogramas selecionadas em 2.1. Escreva (ou adapte um programa) que permita ler arquivo de uma série genérica com diferentes tamanho

5- Processos Estocásticos Canônicos: Caos Determinístico e Turbulência

5.1. A partir do Mapeamento Logístico e do Mapeamento de Henon gere 2 famílias de series Temporais com 5 séries em cada uma. Aplique as respectivas análises estatísticas dos exercícios 4.1 e 4.2.

5.2. A partir da 2ª lei de Newton construa a equação de Navier-Stokes na sua forma mais simples.

Exercício Bonus (+1 na nota final)

Considere o caso particular da Equação de Burgers Estocástica e gere uma família de Sinais estocásticos a partir da mesma (dica: artigo Flores et al.

<http://www.aims sciences.org/article/doi/10.3934/dcdss.2020224>

6- PSD & DFA:

- 6.1. Considere as séries temporais listadas na tabela `dataset_signal` e obtenha, para cada série, os valores respectivos dos seguintes atributos: S^2 , K , β (via PSD) e α (via DFA). Confira para todas as séries se β (via PSD) está bem ajustado a partir da fórmula WKP: $\beta = 2\alpha - 1$. Construa dois espaços de parâmetros EPSB-K-means: $S^2 \times K \times \beta$ e EDF-K-means: $S^2 \times K \times \alpha$.
- 6.2. Classifique, nos espaços de parâmetros do exercício anterior, as séries temporais: (a) ST-Sol3GHz, (b) ST-surftemp504 e (c) ST-OWS_NDC_Covid1 para um país de sua escolha. Discuta comparativamente as classificações nos dois espaços.
- 6.3. Aplique k-means para todas as séries ST-OWS_NDC_Covid1 considerando os seguintes espaços de atributos: $S^2 \times \alpha$ e $K \times \alpha$. Obtenha os melhores agrupamentos, identifique os grupos e discuta os resultados.

7- Singularity Multifractal Spectra (SMS), também conhecido como MDFA.

- 7.1. Considere o programa `mfdfa.py`. Aprimore o programa para que o mesmo calcule o índice $\Psi = \Delta\alpha/\alpha_{\max}$.
- 7.2. Obtenha o espectro de singularidade para todos os sinais do `dataset_signal` e inclua os valores de Ψ como mais um atributo.
- 7.3. Com base nos valores obtidos em 7.2., repita o exercício 6.3 no espaço $S^2 \times \Psi$.

8- Global Wavelet Spectrum

- 8.1. Utilize o `Waipy` para obter o GWS (Morlet) de todas as ST do exercício 6.2.
- 8.2. Repita 9.1. utilizando uma Db8.

9- Self-Organized Criticality (SOC).

- 10.1. Gere 100 ST (50 exo e 50 endo) com p-model para $N=8192$ e aplique `SOC.py` para as 50 de cada grupo.
- 10.2. Aplique o `SOC.py` para todas as ST do exercício 6.3.

11- EXERCÍCIO BONUS: +1.0 NA NOTA FINAL

Crie um programa em Python para *Preparação Generalizada de Dados (PGD)* e incorpore uma técnica de análise (estatística ou espectral) de sua escolha dentro do paradigma de Data Lake. Teste com alguns dados de seu interesse. Considere apenas os formatos `.txt` e `.csv`. Considere `Pchip` no módulo de normalização e uma RNR LSTM no módulo de previsão. Discuta a emulação deste programa como software embarcado em minicomputador (ex. Arduino ou RaspberryPi-4).

TRABALHO EM GRUPO – CAP 239 - 2020

1. Dataset: A partir do *dataset* da OWD (UK) construa um *dataset* com os dados indicados (no exercício2) relativos aos países e regiões selecionados para o seu grupo (ver tabela abaixo).

| Team | C1 | C2 | C3 | C4 | C5 | R1 | R2 |
|--------|--------|-----------|-------------|--------------|-------------|----|-------------------------|
| Red | Brazil | Australia | New Zealand | Singapore | Japan | RJ | DF |
| Blue | Brazil | Canada | Mexico | Cuba | Russia | MG | Niteroi |
| Green | Brazil | Portugal | Spain | France | Belgium | AM | Rocinha |
| Purple | Brazil | USA | Italy | China | South Korea | CE | Paraisópolis-Capital-SP |
| Orange | Brazil | India | Iran | South Africa | Egypt | SP | SJC |

2. Visualização, [Cullen-Frey](#), [Histogramas](#) e [PDFs](#): A partir do *dataset* do seu Team visualize, obtenha os respectivos Histogramas e PDFs para as seguintes variáveis: Número Total de Casos, Numero Total de Mortes, Numero Total De Testes, [Numero Diário de Casos \(NDC\)](#), [Numero Diário de Mortes \(NDM\)](#), [Numero Diário de Testes \(NDT\)](#). Identifique semelhanças e discrepâncias entre os países. Este exercício é apenas para os dados da OWD. Implemente ainda uma análise de regressão linear em Python entre as variáveis: NDC e NDT para os países que apresentam PDF próximas.

3. Previsão Diária: Adapte e Aplique o modelo IMCSF-COVID19.py para cálculo automático das curvas de g e s , para todos os países e regiões, até o dia 20/5/2020. Considere o primeiro dia com mais de 50 casos.

4. Com base na teoria da “*wisdom of the crowd*” adapte o modelo para prever o comportamento das curvas de g e s até uma data fixa escolhida como *input*. E com base na hipótese do máximo de 30 ± 8 semanas identifique o pico e o final (mínimo e máximo) da epidemia em cada país.

5. Desenvolva um modelo de interpolação aleatória de 23 pontos entre cada Medida diária, representando número de casos suspeitos por hora. Gere as séries e verifique se há assinatura de SOC a partir de todos os dados de flutuação. Aplique uma LSTM para prever o caso brasileiro (supervisão do Luis Arantes Filho).