



PME 3573 – Introdução à Ciência de Dados para Engenheiros – 2023

Exercício E2

- Esta atividade deve ser realizada individualmente.
- Esta atividade é baseada no arquivo de dados [BH_03_AIR_1.csv \(link\)](#), que registra séries temporais das coordenadas cartesianas dos alvos refletivos no ensaio de decaimento em ar de um longo tubo flexível, engastado em sua extremidade superior e munido de um pequeno lastro metálico em sua extremidade inferior ([vídeo do ensaio](#)).
- Para entrega, o aluno deverá elaborar um único relatório sucinto com os resultados das duas partes do exercício.
- A entrega deve ser realizada pelo Moodle da disciplina em um único arquivo em formato PDF identificado como “E02-NOME.pdf”, onde “NOME” deve ser substituído pelo nome do aluno que elaborou o relatório.
- O relatório deve conter o código utilizado pelo aluno em suas análises.
- É recomendado o uso do formato *R Markdown* para a elaboração do relatório.

Parte 1 (aulas 07, 08 e 09)

- (a) Tomando como base o arquivo [BH_03_MLR.R \(link\)](#) disponibilizado na pasta “Aula 09” do ambiente Moodle da disciplina, elabore um código, utilizando funções intrínsecas (*built-in functions*) da linguagem de programação escolhida, que ajuste uma função de interpolação linear que tenha como entradas (vetor de atributos) as coordenadas cartesianas de todos os alvos refletivos em dois frames subsequentes (f e $f + 1$) e produza como saída o valor da coordenada y , no frame $f + 20$, do alvo refletivo L03, localizado na extremidade inferior do lastro.

Para tal, dentre as n linhas do arquivo de dados, selecione um subconjunto de n_t linhas iniciais para o “treinamento” (ajuste de parâmetros) de sua função de interpolação e utilize as linhas $n_t + 1$ até n apenas para o teste do seu modelo.

Analise e interprete (utilizando gráficos ou ainda outros recursos que julgar conveniente) o erro entre os valores efetivamente observados no ensaio para a variável de saída e as previsões obtidas pelo modelo ajustado. Compare o erro obtido com a ordem de grandeza das amplitudes de oscilação observadas para o alvo de interesse na direção y .

- (b) Refaça o código do item anterior implementando o [método recursivo de mínimos quadrados](#) para o ajuste dos parâmetros do modelo em vez de funções intrínsecas da linguagem de programação escolhida.

Avalie o efeito do tamanho n_t do conjunto de treinamento na convergência dos parâmetros do modelo.

Repita os procedimentos de análise realizados no item anterior, ressaltando eventuais diferenças observadas nos resultados obtidos pelos dois códigos.



Parte 2 (aulas 10 e 11)

O ensaio completo, como pudemos observar é composto por 7 sub-ensaios nos quais se observam as oscilações naturalmente amortecidas do tubo BH-3 logo após uma excitação por meio de um impacto aplicado à extremidade inferior do lastro. No exemplo mostrado no arquivo [BH_03_AULA11.R \(link\)](#), que pode ser tomado como base para a realização desta Parte 2, selecionamos como “conjunto de treinamento” um conjunto de dados contido no quarto sub-ensaio, ignorando um curto intervalo de tempo sucessivo ao impacto (que pode conter efeitos transitórios muito rápidos e não tão bem representados nas medidas) e o trecho final, imediatamente anterior ao próximo impacto, em que as oscilações são de baixa amplitude (e a ordem de grandeza dos deslocamentos pode começar a ficar próxima demais do ruído de medição). O “conjunto de teste”, por sua vez contido no sexto sub-ensaio, foi selecionado respeitando as mesmas premissas.

Naturalmente, poderíamos ter identificado intervalos deste tipo nos 7 sub-ensaios. Nesta Parte 2, o objetivo é adaptar o código [BH_03_AULA11.R](#) para que, uma vez que você selecione estes 7 intervalos de dados (um interno a cada sub-ensaio), você deverá para pelo menos um dos algoritmos de **PCR** apresentados:

- (1) Faça o levantamento dos erros quadráticos médios (LSE) cometidos na aproximação, em função do número μ de variáveis latentes adotadas, quando se escolhe cada um destes 7 conjuntos como “conjunto de teste” e usando a união dos 6 conjuntos remanescentes como “conjunto de treinamento”. Assim, cada valor de μ terá a ele associado 7 valores de LSE (cada valor de LSE, por sua vez, corresponde à adoção de um dos sete conjuntos como “conjunto de teste”).
- (2) Proponha uma forma gráfica que julgar adequada para representar os resultados de LSE em função de μ , em particular, destacando para cada valor de μ a média dos 7 valores de LSE a ele associados, de forma a ressaltar qual o valor de μ a média dos LSEs é a menor possível (se houver mínimo) ou, eventualmente, qual valor de μ acima do qual a média dos LSEs se torna estacionária.
- (3) Para o valor “ótimo” de μ selecionado a partir do gráfico construído no item anterior, desta vez usando todos os 7 conjuntos de dados como “conjunto de treinamento”, obtenha as estimativas de funções de forma modal para os μ primeiros modos. Compare os resultados com os apresentados originalmente no código [BH_03_AULA11.R](#).

Repita o procedimento descrito nos passos (1) a (3) acima, desta vez para um dos algoritmos de **PLSR** apresentados.

Observação: o conjunto de procedimentos descritos nos passos (1) e (2) acima para a obtenção de um valor μ ótimo para o tamanho do sistema de ordem reduzida é denominado **validação cruzada** (*cross-validation*).