



Cálculo Numérico – Prova 2

Projeto C – Avaliação ortopédica na coluna vertebral

Contexto: Condições na coluna vertebral, como hérnia de disco e espondilolistese, afetam não apenas a mobilidade e a qualidade de vida dos pacientes, mas também representam um desafio significativo para a medicina diagnóstica, dado que fatores como idade, peso corporal, níveis de atividade e predisposições genéticas contribuem para a evolução dos quadros clínicos.

Com base na análise de dados retrospectivos de pacientes, como imagens de exames, histórico de dor e outros parâmetros clínicos, podemos estabelecer uma relação entre múltiplas variáveis que podem influenciar o agravamento da patologia, como o aumento da dor ou a perda de função motora. Além disso, os modelos podem prever a necessidade de intervenções mais invasivas, como cirurgias, ao correlacionar padrões históricos de progressão com a severidade do quadro. Isso oferece uma forma de suporte decisório para os médicos, pois com as previsões derivadas desses modelos, é possível agir preventivamente, monitorando mais de perto os casos que podem piorar e implementando tratamentos conservadores para retardar a evolução dos sintomas.

Neste estudo, você será responsável pela análise de dados clínicos de pacientes com patologias da coluna vertebral, focando em como variáveis como raio e ângulo pélvico, ângulo de lordose lombar e outras, que estão associadas ao desfecho clínico de cada paciente. O objetivo não é prever a ocorrência inicial da patologia, mas entender como esses diferentes parâmetros podem ser usados para entender a atual condição da coluna, auxiliando, assim, na tomada de decisões clínicas para um tratamento mais direcionado e eficaz.

Dados: Um conjunto de dados foi obtido durante um estudo com pacientes que apresentam condições de coluna vertebral, como espondilolistese e deformidades na pélvis e coluna lombar.

Cada coluna na tabela representa uma medida clínica específica relacionada à biomecânica da coluna. Os dados estão no arquivo "column_2C_cleaned.csv," e incluem um total de 6 variáveis biomecânicas relacionadas à postura e ao alinhamento da coluna. Quando os dados são carregados no arquivo p2-c.m na linha 9, as colunas seguem a ordem abaixo, sendo a última coluna a variável de classificação que indica o diagnóstico de espondilolistese.

As colunas são:

1 – Pelvic Incidence – Incidência pélvica

2 – Pelvic Tilt – Inclinação pélvica



3 – Lumbar Lordosis Angle – Ângulo de lordose lombar

4 – Sacral Slope – Inclinação sacral

5 – Pelvic Radius – Raio da pélvis

6 – Degree of Spondylolisthesis – Grau de espondilolistese

7 – Class – Diagnóstico do paciente para a condição da coluna, sendo “0” para normal e “1” para anormal

Dúvidas sobre os dados podem ser consultadas em Neto & Barreto (2009).

Questionamentos: seu trabalho deve conter, obrigatoriamente, as análises que seguem, mas outras avaliações que o grupo julgar pertinentes podem ser consideradas. Utilize o Octave para responder:

Análise 1: Seleção das variáveis

- 1.1. Com base nos dados fornecidos, selecione duas variáveis que possam estar diretamente relacionados ao grau de espondilolistese. Justifique suas escolhas utilizando uma análise visual com gráficos adequados.
- 1.2. Após identificar esses pares, aplique análises numéricas para apoiar sua decisão sobre qual variável tem maior influência na identificação do grau de espondilolistese. Justifique seu raciocínio com base nos resultados das análises numéricas vistas em aula.

Análise 2: Análise de regressão polinomial

- 2.1. Tome agora a relação de todas as variáveis em relação ao grau de espondilolistese apenas dos casos anormais.

A partir de uma análise gráfica, qual variável é a melhor candidata para uma regressão polinomial? Isto é, aquela que possui uma maior concentração de dados em um formato que não seja linear. Justifique.

- 2.2. Faça uma regressão polinomial de menor grau n possível e $n \geq 1$ para todas as variáveis que preveem espondilolistese. Faça o plot de todas sobre os pontos utilizados.

Dica: dificilmente utilizamos polinômios de grau maior que 3.

- 2.3. Mostre, numericamente, qual polinômio é melhor para aquele conjunto de dados. Justifique.

Análise 3: Predição

- 3.1. Construa dois modelos de regressão linear simples usando as duas variáveis identificadas na Análise 1. Por exemplo, um modelo pode ser da forma $y_1 = a_{0,1} + a_{1,1}x_1$ e outro $y_2 = a_{0,2} + a_{1,2}x_2$ onde x_1 e x_2 foram obtidos na Análise 1 para a previsão de espondilolistese. Calcule as métricas de qualidade do ajuste, como S_r , r^2 e $s_{y/x}$ para cada modelo. Qual modelo apresentou melhor ajuste? Justifique.



- 3.2. Implemente um terceiro modelo de regressão linear múltipla do tipo $y_3 = a_{0,3} + a_{1,3}x_1 + a_{2,3}x_2$ que combina as duas métricas utilizadas em (3.1).

Isso significa que você deve encontrar os parâmetros $a_{0,3}$, $a_{1,3}$ e $a_{2,3}$ de tal maneira que a regressão $y = a_1x_1 + a_2x_2 + a_0$ seja o melhor possível. Calcule qual a soma dos resíduos S_r , r^2 e $s_{y/x}$ para o modelo y_3 . Compare-o com y_1 e y_2 . Qual é o melhor? Justifique.

- 3.3. Verificando as métricas do modelo obtido em (3.1), você diria que esse é um bom modelo de regressão para prever o paciente com a doença? Justifique sua resposta.

Atenção: essa não é uma prova de aprendizado de máquina, mas de cálculo numérico. Logo, não espero uma análise utilizando métricas rebuscadas, discussões sobre modelos mais robustos ou que seu modelo performe bem. Quero que seu grupo foque na utilização do modelo construído, usando a ferramenta Octave e a interpretação dos resultados obtidos segundo a teoria dada em aula.

Códigos: o arquivo que deve resolver seu projeto é dado "p2_c.m". Esse código já carrega os dados e faz o plot de dispersão (scatter) das métricas duas a duas. Você pode criar outros scripts .m caso precise, mas eu só devo precisar rodar "p2_c.m" para verificar os entregáveis do seu projeto.

Funções proibidas:

- polyfit, linsolve, regress, interp1, interp2, interp3, interpN, spline, fitlm, compact, fitrlinear, mvregress, mvregresslike, plsregress.
- Também é proibido resolver sistemas lineares com o método da inversa (função inv()) ou pelo operador barra invertida (\). Caso você precise, utilize a função lu() para obter a decomposição e implemente uma pequena função que resolve o sistema dado que você possui L e U.



Referências:

Neto, A. R. R., & Barreto, G. A. (2009). On the Application of Ensembles of Classifiers to the Diagnosis of Pathologies of the Vertebral Column: A Comparative Analysis. *IEEE Latin America Transactions*, 7(4), 487–496. <https://doi.org/10.1109/TLA.2009.5349049>

ANEXO

Formato dos dados:

	pelvic_incidence	pelvic_tilt	lumbar_lordosis_angle	sacral_slope	pelvic_radius	degree_spondylolisthesis	class
0	63.027817	22.552586	39.609117	40.475232	98.672917	-0.254400	Abnormal
1	39.056951	10.060991	25.015378	28.995960	114.405425	4.564259	Abnormal
2	68.832021	22.218482	50.092194	46.613539	105.985135	-3.530317	Abnormal
3	69.297008	24.652878	44.311238	44.644130	101.868495	11.211523	Abnormal
4	49.712859	9.652075	28.317406	40.060784	108.168725	7.918501	Abnormal

Figura 1. Cinco primeiras linhas dos dados em column_2C_cleaned.csv

Pairplots dos dados:

