

## Trabalho 3: Algoritmo para classificação de patologias da coluna vertebral

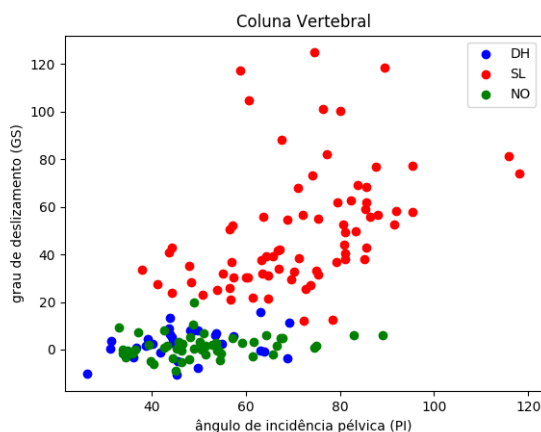
Até meados de 2005, a escassez de trabalhos voltados ao diagnóstico automático de patologias da coluna vertebral se devia em grande parte à ausência de atributos numéricos que descrevessem quantitativamente as patologias de interesse para o campo da Ortopedia, de modo a gerar um conjunto de dados adequado para o projeto de classificadores de padrões. No entanto, há poucos anos (2011, mais precisamente), um grupo de especialistas em Ortopedia e Biomecânica definiu um conjunto de informações (atributos) biomecânicas relacionadas a dores e deformidades da coluna vertebral.

O problema-alvo deste trabalho concentra-se na área de Ortopedia, e mais especificamente no diagnóstico de patologias da coluna vertebral, a saber, hérnia de disco e espondilolistese. O arquivo deste trabalho contém dados extraídos de 310 pacientes, a partir de radiografias panorâmicas sagistais em formato 30 x 90 cm.

Destes, 100 indivíduos são voluntários que não possuem patologias na coluna, doravante chamados de normais. Os dados restantes são obtidos a partir de radiografias de 60 pacientes operados de hérnias de disco (DH) e 150 indivíduos de espondilolistese (SL). O restante dos voluntários não possuem patologia (NO). Cada um dos 310 pacientes é descrito por seis atributos biomecânicos:

- ângulo de incidência pélvica (PI)
- ângulo de versão pélvica (PT)
- ângulo de lordose (LA)
- inclinação sacral (SS)
- raio pélvico (RP)
- grau de deslizamento (GS)

Esses dados estão disponíveis em dois arquivos **column\_3C.dat** e **column\_3C\_test.dat**. Cada arquivo possui as informações de 155 pacientes. A imagem a seguir representa o gráfico em que o eixo x representa o ângulo de incidência pélvica (PI) e o eixo y representa o grau de deslizamento (GS). Cada ponto representa as informações de um paciente. Além disso, de acordo com a legenda, cada cor representa um diagnóstico.



Exposto isso, este tema de trabalho compreende o desenvolvimento de um sistema de classificação de indivíduos em: normal (sem patologia), hérnia de disco ou espondilolistese. O sistema deve receber 6 atributos biomecânicos de um paciente e informar em qual condição ele se encontra. Para inferir esta condição, o sistema utilizará o algoritmo dos k vizinhos mais próximos. O algoritmo procura em um conjunto de registros (outros pacientes) os k pacientes mais similares que este paciente em questão. Ao extrair estes

k pacientes, o algoritmo computa a moda (moda é o valor que mais se repete em um conjunto/lista/vetor) e infere o diagnóstico com base nessa informação de moda.

Por exemplo, veja que um paciente com ângulo de incidência pélvica 80 e grau de deslizamento 60 fica próximo dos pacientes com espondilolistese. Dessa forma, podemos inferir que o diagnóstico do paciente é espondilolistese.

Primeiramente, o algoritmo deve perguntar os atributos biomecânicos de um novo paciente e localizar os k outros pacientes mais similares. Para recuperar um paciente similar basta computar a distância euclidiana entre os atributos biomecânicos do paciente em questão e os pacientes já registrados.

$$\text{distância euclidiana} = d((x_1, \dots, x_n), (y_1, \dots, y_n)) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Quanto menor a distância, mais similares os pacientes são. Ao encontrar os k outros pacientes mais similares, o algoritmo computa a moda da doença, ou seja, a doença que mais se repetir é o diagnóstico. Para testar seu software de diagnóstico, você pode usar o arquivo **column\_3C\_test.dat**. Para cada paciente no arquivo, você deve procurar os k mais similares do arquivo **column\_3C.dat** e atribuir um diagnóstico de acordo. Em seguida, você pode verificar se o seu programa atribuiu o diagnóstico correto.

No final, seu programa deverá salvar os resultados em um arquivo com as seguintes informações: atributos do paciente, classificação pelo algoritmo e classificação real. Além disso, a última linha do arquivo deve conter a porcentagem de diagnósticos corretos e porcentagem de diagnósticos errados do algoritmo.

**EXTRA:** Mostrar o gráfico a partir de dois atributos fornecidos como na figura acima. Você pode usar o módulo **matplotlib** (<https://matplotlib.org/>) para mostrar os gráficos.