

Introdução a Modelagem e Aprendizado

ELTDI - DAELN UTFPR-CT

Prof. Heitor S. Lopes (2024) - Exercício #2a

Objetivo: utilizar métodos de regressão e classificação para a análise de dados de engenharia

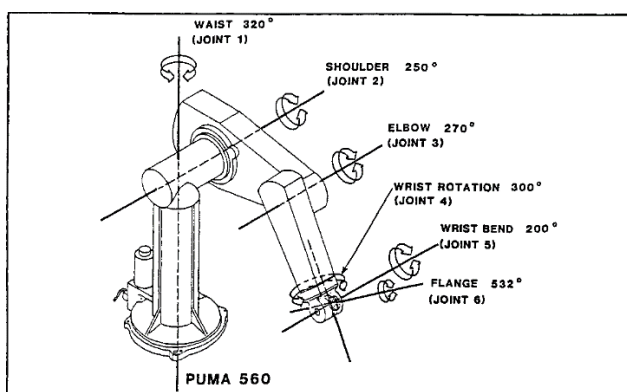
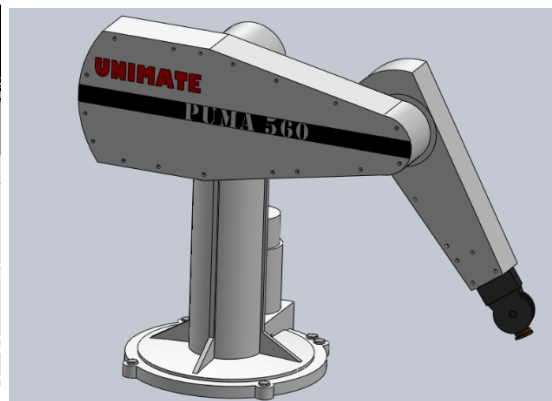
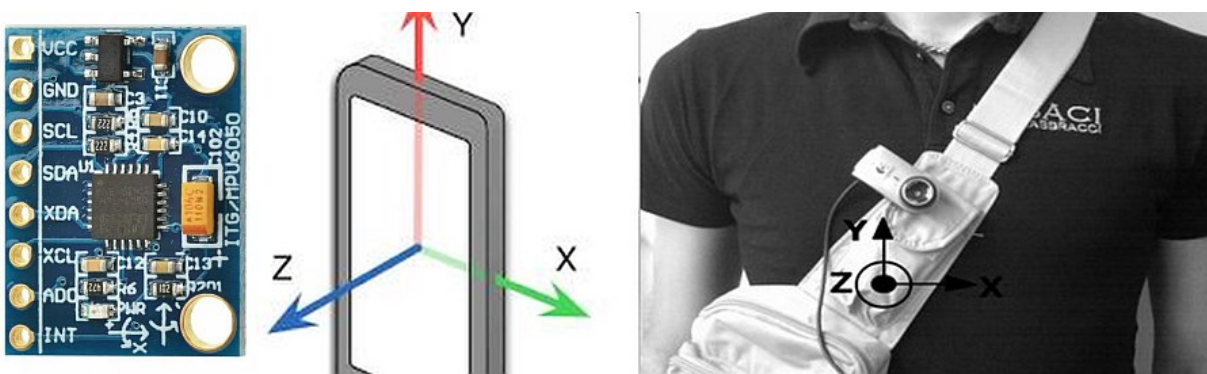


Figure 1.2 Degrees of Joint Rotation (Adapted from [68]).



Parte I:

- O *dataset* `puma32h` contém dados da dinâmica do braço robótico Puma 560, um equipamento comumente utilizado em automação industrial. Embora os dados sejam sinteticamente gerados a partir de simulações da dinâmica, eles são bastante realistas. O objetivo é prever a aceleração angular de um ponto na garra do robô (variável `thetadd6`) a partir de 32 informações dos diversos *links* do robô, sendo posições angulares, velocidades e torques.
- Importe os dados no Orange com o *widget* `CSV File Import`. Utilize os *widget* `Data Sampler` para dividir o *dataset* de entrada em duas partes: treinamento (2/3) e teste (1/3), e o *widget* `Select Columns` para definir a variável-meta (`thetadd6`).
- Como a variável-meta é numérica, trata-se de um problema de **regressão**. Utilize os métodos kNN, SVM, RN (Rede Neural) e LR (Linear Regression) para encontrar um modelo com o menor RMSE (*Root-means Square Error*) e máximo R2 (coeficiente de correlação) possíveis. Para tanto, ajuste iterativamente todos os parâmetros de controle dos métodos, tendo como referencial os resultados com os parâmetros *default*.
- Apresente o método (e seus parâmetros de controle) que obteve o melhor resultado.
- Explique por quê alguns dos métodos obtiveram resultados muito ruins.



Parte 2: Análise de sinais para a classificação de atividades físicas

- a. Sensores (acelerômetro, giroscópio, magnetômetro) de 3 eixos $\{x,y,z\}$ foram colocados em um indivíduo, localizados no peito (*chest*), tornozelo esquerdo (*left-ankle - lak*) e braço inferior direito (*right-lower-arm - rla*), bem como dois eletrodos de Eletrocardiografia (ECG1 e ECG2) para registrar sinais durante a realização das seguintes tarefas: T1: *Standing still* (1 min); T2: *Sitting and relaxing* (1 min); T3: *Lying down* (1 min); T4: *Walking* (1 min); T5: *Climbing stairs* (1 min); T6: *Waist bends forward* (20x); T7: *Frontal elevation of arms* (20x); T8: *Knees bending (crouching)* (20x); T9: *Cycling* (1 min); T10: *Jogging* (1 min); T11: *Running* (1 min); T12: *Jump front & back* (20x). Adicionalmente, a “tarefa” T0 é a classe “nula”.
- b. O objetivo final é identificar a tarefa que o indivíduo realiza com base nos sinais. Observe que há algumas atividades que são bastante similares ou “sobrepostas” a outras e também que a quantidade de amostras é diferente para cada atividade.
- c. O dataset fornecido é para apenas **uma** pessoa e tem 161280 linhas. Divida aleatoriamente (com o *widget “data sampler”*) em um conjunto de treinamento (2/3) e um conjunto de teste 1/3), mantendo a proporção das classes (estratificado).
- d. A primeira parte considera a variável-meta como discreta (**valores inteiros**). Isto é feito na importação dos dados do arquivo CSV. Logo, trata-se de um problema de **classificação**. Use o conjunto de treinamento para treinar os modelos dos métodos Regressão **Logística**, KNN, SVM e Rede Neural. O conjunto de teste serve para verificar a qualidade dos modelos. Use os parâmetros *default* dos métodos. **ATENÇÃO**: o tempo de processamento pode ser bastante significativo, dependendo do computador utilizado.
- e. Compare a qualidade das soluções obtidas pelos classificadores, identificando qual método é mais adequado para identificar as tarefas. Apresente a métrica de qualidade adequada, considerando que as classes são desbalanceadas, bem como a matriz de confusão. Identifique quais as Tarefas que são mais difíceis de identificar e justifique a resposta.
- f. A segunda parte considera que a variável-meta seja um **número real** (ajuste isto na importação do CSV). Logo, trata-se de um problema de **regressão**. Utilize os métodos Regressão **Linear**, KNN, SVM e RN (rede neural) e reporte os resultados (RMSE e coeficiente R2).

- g. Compare a qualidade das soluções obtidas pelos regressores, identificando qual método é mais adequado para identificar as tarefas. Justifique as possíveis discrepâncias dos resultados.
- h. Adicionalmente, discuta a possibilidade de embarcar em um microcontrolador os métodos utilizados para realizar uma tarefa similar.