

Reconhecimento de padrões em Biomecânica de Corrida usando Aprendizado de Máquina

Mônica de Oliveira

Professor Orientador: Talles Henrique de Medeiros

Professor Co-orientador: Rafael S. O. Martins

ICEA - UFOP

17 de julho de 2018



Universidade Federal
de Ouro Preto

Sumário

- 1 Introdução
- 2 Identificação do problema
- 3 Revisão Bibliográfica
- 4 Metodologias
- 5 Resultados
- 6 Conclusão e Trabalhos Futuros
- 7 Referências

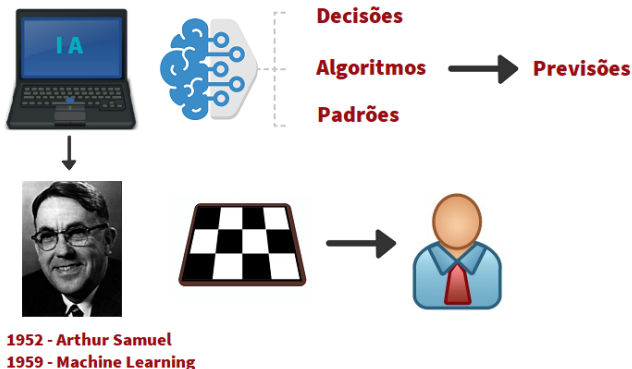


Universidade Federal
de Ouro Preto

Introdução

Aprendizado de Máquina

- Aprendizado de Máquina é um campo de pesquisa da Inteligência Artificial que estuda o desenvolvimento de métodos capazes de extrair conceitos (conhecimento) a partir de amostras de dados. (MITCHELL et al., 1997)



Introdução

Aprendizado de Máquina



- A análise da marcha biomecânica é comumente usada para analisar desempenho desportivo e avaliação de marcha patológica.



Identificação do problema

- (HINO et al., 2009) As lesões musculares.(lentidão na recuperação);
- Embora haja estudos aplicados a aprendizado de máquina em relação a classificação de marcha em atletas, eles geralmente usam poucas amostras e um conjuntos limitados de variáveis considerando apenas uma parte do ciclo de marcha atribuindo a maior parte das pesquisas as variáveis de centro de pressão.(MCGREGOR; BOLLT; FEDEL, 2014)



Objetivos Propostos

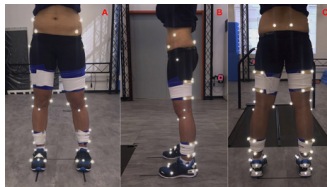
A proposta deste estudo basicamente consiste a princípio, aplicar técnicas de algoritmos de aprendizado de máquina supervisionado na base de dados, treinando-os com o conjunto de treinamento e depois realizando classificação do conjunto de testes para verificar qual algoritmo apresenta a melhor taxa de reconhecimento. Espera-se que o modelo que atingir maior grau de classificação possa contribuir ao máximo sob predições à lesões entre os atletas, melhorando assim tomada de decisões onde determinado grupo possa vir adquirir de acordo com os padrões encontrados.



Base de Dados

Disponível pela mega revista científica (PeerJ) pela Universidade Federal do ABC

■ Cinemática e Cinética



(a) Marcadores

	Time	Fx	Fy	Fz	COPx	COPy	COPz	Ty
1	1	-184.63	328.67	-72.78	2591.73	0	1134.51	-15305.20
2	2	-194.63	348.66	-66.45	2514.68	0	1123.29	9418.18
3	3	-96.66	400.85	-45.39	2546.33	0	1136.36	-29979.20
4	4	-47.38	491.49	-19.99	2493.38	0	1127.69	12225.30
5	5	-36.24	552.83	-20.07	2500.02	0	1129.17	21395.00

(b) Dados Força

Figura: Fonte: (FUKUSHI; DUARTE, 2017)



UFOP
Universidade Federal
de Ouro Preto

Revisão Bibliográfica

Revisão na literatura sobre Aprendizado de Máquina

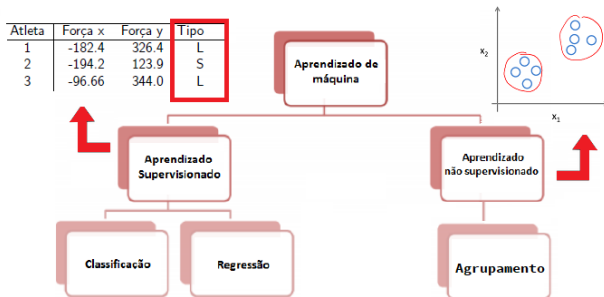


Figura: Hierarquia do Aprendizado de Máquina, FONTE: adaptado de Monard e Baranauskas (2003)



Revisão Bibliográfica

Algoritmos de Classificação: Knn - K vizinhos mais próximos

- Knn - K-vizinhos mais próximos (COVER; HART, 1967);
- Classificação;
- Aprendizado de máquina supervisionado;
- Ideia:
 - Espaço - entregue ponto de consulta;
 - Ponto de consulta/query point
 - Verificar quais são os k pontos mais próximos;
 - Distância;
 - Verifico Classes de vizinhos mais próximos;
 - votação;

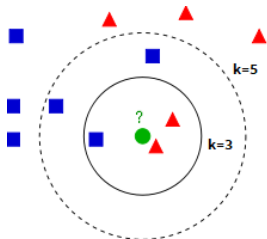


Figura: K vizinhos



Revisão Bibliográfica

Algoritmos de Classificação: Knn - K vizinhos mais próximos

- Dado um vetor de característica desconhecido N (Conjunto de dados para treinamento);
- Defina um valor de k;
- Calcule uma métrica de distância;

$$d(x, y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} \quad (1)$$

- Fora dos N vetores de treinamento, identifique os vizinhos mais próximos, independentemente dos rótulos das classes.
- Fora dessas k amostras, conte o número de vizinhos mais próximos que pertencem a cada classe k_1 com $i = 1, 2, \dots, M$. Assim temos: $\sum_i k_i = k$
- Atribua o rótulo à classe mais frequente na vizinhança. Classificando o novo elemento.



Revisão Bibliográfica

Escolha do K apropriado

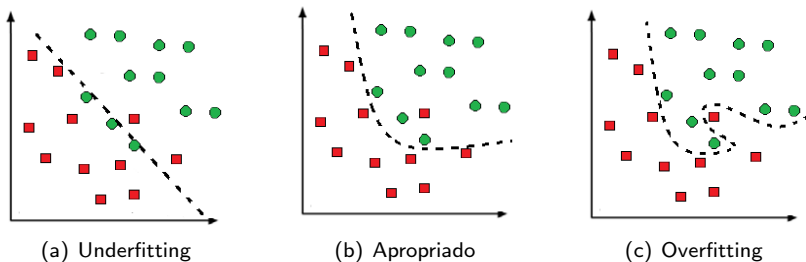


Figura: Escolha de um k apropriado: underfitting, K apropriado e overfitting e respectivamente



Revisão Bibliografica

Implementação Knn

```
#####
# Monica Oliveira - Knn Teste          #
#####

knn <- function(dataset, query, k=1){
  classeId = ncol(dataset)

  #distancia Euclidiana
  E = apply(dataset, 1, function(row){
    sqrt (sum ((query - as.numeric(row[1:(classeId-1)]))^2))
  })

  sort.list(E, dec =F)[1:k]
  classes = dataset[ids,classeId]
  U = unique(classes)
  R = rep(0, length(U))

  for(i in 1:length(U)){
    R[i] = sum(U[i] == classes)
  }
}
```

Figura: Implementação básica do Algoritmo Knn



Revisão Bibliográfica

Pesquisa Algoritmos padrões de projeto: SVM - Máquinas de Vetores Suporte

- Classificação;
- SVM - técnica embasada na Teoria de Aprendizado Estatístico (CORTES; VAPNIK, 1995);
- Objetivo: Encontrar Hiperplano de separação com margem máxima;

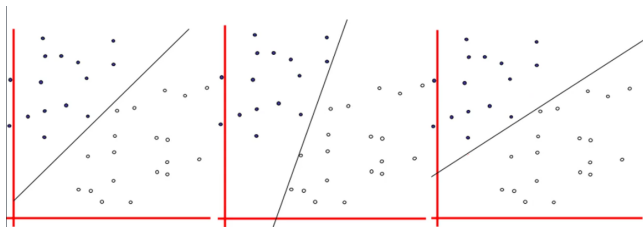


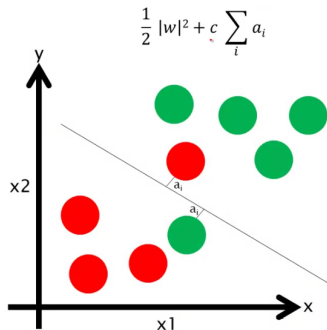
Figura: Hiperplanos Ótimos



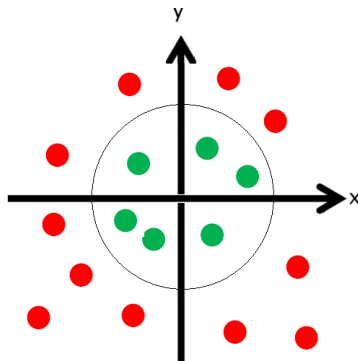
UFOP
Universidade Federal
de Ouro Preto

Revisão Bibliográfica

Pesquisa Algoritmos padrões de projeto: SVM - Máquinas de Vetores Suporte



(a) Linearmente separável



(b) Não linearmente separável

Figura: Projeção dos dados



Revisão Bibliográfica

SVM - Kernels

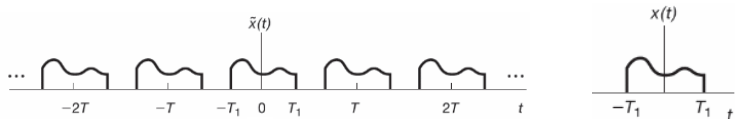
Tabela: Algumas das funções *Kernels* mais utilizadas

Tipo de Kernel	Função K (x_i, x_j)	Tipo do classificador
Polinomial	$((x_i \cdot x_j) + 1)^2$	Máquina de aprendizagem polinomial
Gaussiano (RBF)	$\exp(-\frac{\ x_i - x_j\ ^2}{2\sigma^2})$	Rede RBF
Linear	$x \times y$	classificador linear
Curva Sigmoide	$\tanh(\beta_0(x_i \cdot x_j)) + \beta_1$	Perceptron de duas camadas



Revisão Bibliográfica

Transformada de Fourier



Um sinal aperiódico pode ser visto como um sinal periódico com um período infinito. Quando o período $T \rightarrow \infty$, $x(t) \rightarrow \tilde{x}(t)$

A representação em série de Fourier do sinal periódico $\tilde{x}(t)$ é:

$$\tilde{x}(t) = \sum_{k=-\infty}^{\infty} a_k e^{jk\omega_0 t} \quad (2)$$

Sendo \tilde{x} :

$$a_k = \frac{1}{T} \int_T x(t) e^{-jk\omega_0 t} dt = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) e^{-jk\omega_0 t} dt$$



Revisão Bibliográfica

Transformada de Fourier

Considerando:

$$\tilde{x}(t) = \begin{cases} x(t), & \text{para } |t| \leq T/2 \\ 0, & \text{cc} \end{cases} \quad (4)$$

Podemos reescrever os limites de a_k , com $w = kw_0$ como:

$$a_k = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \tilde{x}(t) e^{-jkw_0 t} dt = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) e^{-jkw_0 t} dt = \frac{1}{T} \int_{-\infty}^{\infty} x(t) e^{-jkw_0 t} dt \quad (5)$$

Portanto a Transformada de Fourier é definida como:

$$X(jw) = \int_{-\infty}^{\infty} x(t) e^{-jw t} dt$$

Também denominada espectro de frequências de $x(t)$.



Metodologia

Ambiente de desenvolvimento e Base de Dados

Softwares utilizados para análise, observação, testes, etc

- RStudio;
- Matlab;

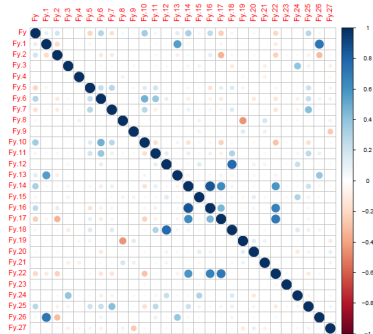
Base de dados:

- Mega revista científica PeerJ disponibilizada pela Universidade Federal do ABC;
- 28 Atletas, esteira instrumentada;
- ensaios de corrida foram realizados em uma esteira instrumentada a três velocidades de marcha distintas (2,5 m / s, 3,5 m / s e 4,5 m / s);
- sinais amostrados 300 Hz para cinética;

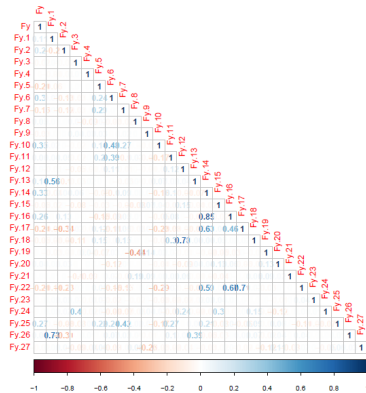


Metodologia

Observações e Testes (Pré-processamento e autoconhecimento dos dados.)



(c) Teste Correlação



(d) Mod Teste Correlação

Figura: Correlação dos dados de Força nos eixos x,y e z dos 28 atletas



Metodologia

Observações e Testes

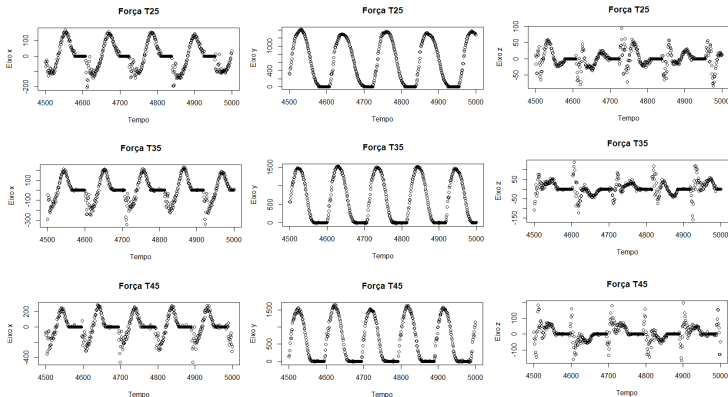


Figura: Dados brutos no domínio do tempo em relação aos eixos x, y e z do atributo Força de um sujeito aleatório para mostrar aperiodicidade



Metodologia

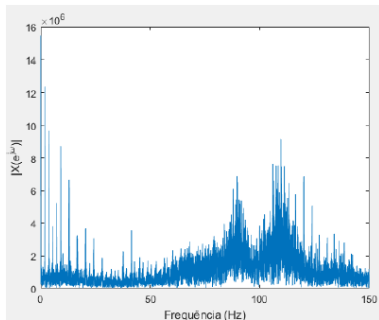
Observações e Testes

```
%x -> é o sinal
%Fs -> é a frequência de amostragem
%n_pontos -> é o número de pontos que se deseja calcular
%FFT (sugestão utilizar 512 ou 1024 ou em torno de isso)
%
%[f,mod_FFT] = calc_fft(x,Fs,n_pontos);
%
function [f,mod_FFT] = calc_fft(x,Fs,n_pontos)

FFT_X = fft(x,n_pontos);
f = 0:Fs/n_pontos:Fs - Fs/n_pontos;
mod_FFT = abs(FFT_X);
plot(f(1:n_pontos/2),mod_FFT(1:n_pontos/2));
xlabel('Frequência (Hz)');
ylabel('|X(e^{j\omega})|');

csvwrite('mod_FFT.csv', mod_FFT);
```

(a) Código TF - Matlab



(b) Comportamento TF (dado aleatório)

Figura: Transformada de Fourier



Metodologia

Features - Teia de Determinantes

Base de dados final: 139 colunas e 28 linhas.

- Média, var, sd, máximo (F_x , F_y , F_z);
- Tempo fora da esteira [segundos]
- Centro de Pressão nos 2 eixos: (COP_x , COP_z);
- Momento livre sobre o eixo vertical T_y ;
- TF da média, var, sd;
- TF das 2 COP;
- TF eixo vertical T_y ;
- Atleta lesionado [S/N];

(Nos tempos 2.5, 3.5 e 4.5 m/s)



Metodologia

Métodos de avaliação

- Matriz de Confusão
- Acurácia

Tabela: Exemplo matriz de confusão

	Valor Real	
	Positivo	Negativo
Previsto Positivo	VP	FP
Previsto Negativo	FN	VN



Metodologia

Resumo da divisão da base juntamente com os algoritmos kNN e SVM

```
#####
# Monica de Oliveira      #
# Universidade Federal de Ouro Preto  #
#####

# Leitura da base de dados
#base = read.csv('dadoAtletas.csv')

base<-dadoAtletas
dim(dadoAtletas)

# Divisão entre treinamento e teste
library(caTools)
set.seed(214)
divisao = sample.split(base$Lesao, SplitRatio = 0.7)
base_treinamento = subset(base, divisao == TRUE)
base_teste = subset(base, divisao == FALSE)

#####
# Algoritmo de Classificacao SVM
classificador1 = svm(formula = Lesao ~., data = base_treinamento,
                     type = 'C-classification', kernel = 'radial', cost = 10.0)
#####

# Algoritmo de Classificacao kNN
classificador2 = knn(train = base_treinamento[, -139] , test = base_teste[, -139],
                     cl = base_treinamento[, 139, drop = TRUE], k = 3)
```



Resultados

Algoritmo kNN

Tabela: Matriz de confusão kNN com $k = 1$

Previsto	Valor Real	
	0	1
0	4	2
1	3	0

Tabela: Matriz de confusão kNN com $k = 2$

Previsto	Valor Real	
	0	1
0	3	3
1	2	1

Tabela: Matriz de confusão kNN com $k = 3$

Previsto	Valor Real	
	0	1
0	4	2
1	2	1

K	Acurácia
1	44,4 %
2	44,4 %
3	55,5 %



Resultados

Algoritmo SVM

Tabela: Matriz de confusão Kernel linear

Previsto	Valor Real	
	0	1
0	6	1
1	3	0

Tabela: Matriz de confusão Kernel Polinomial

Previsto	Valor Real	
	0	1
0	4	2
1	1	2



Resultados

Algoritmo SVM

Tabela: Matriz de confusão Kernel Base Radial

Previsto	Valor Real	
	0	1
0	5	1
1	1	2

Tabela: Matriz de confusão Kernel Sigmoide

Previsto	Valor Real	
	0	1
0	4	2
1	2	1

Kernel	cost	Acurácia
Linear	1	60%
Polinomial	3	66,6%
Base Radial	5	77,7%
Sigmoide	3	55,5 %



Resultados

Características dos Atletas

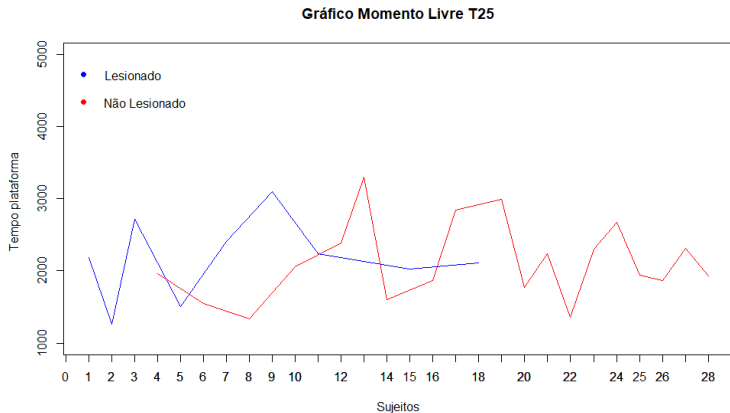


Figura: Média fora da Esteira: L = 2.1s NL = 2.3s



Resultados

Características dos Atletas

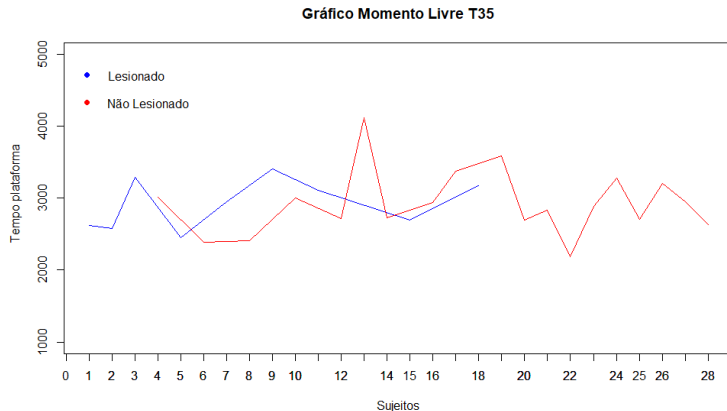


Figura: Média fora da Esteira: L = 2.7s NL = 2.9s



Resultados

Características dos Atletas

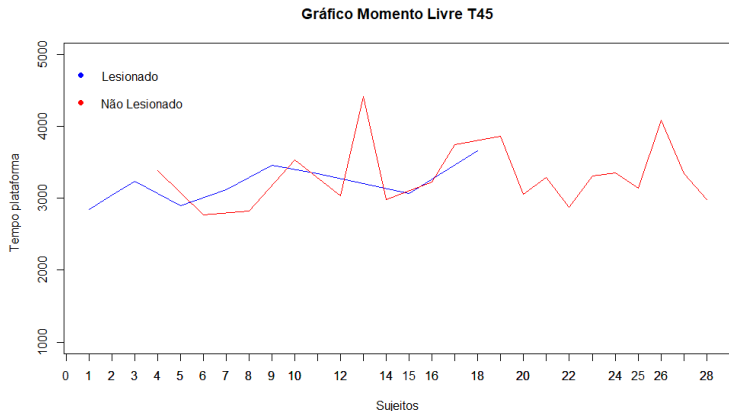


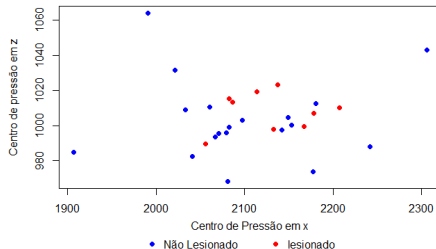
Figura: Média fora da Esteira: L = 3.0s NL = 3.3s



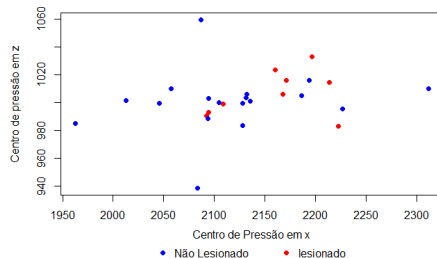
Resultados

Características dos Atletas

COP T25

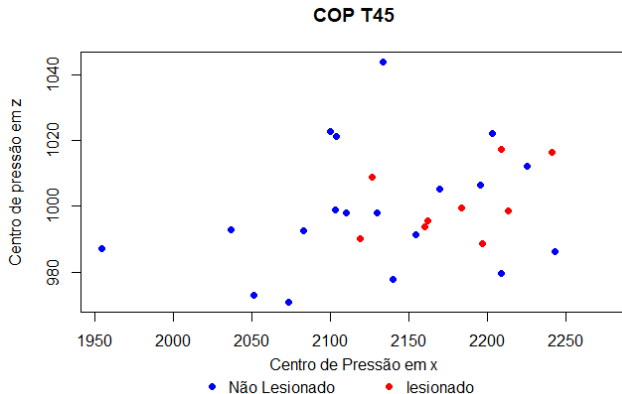


COP T35



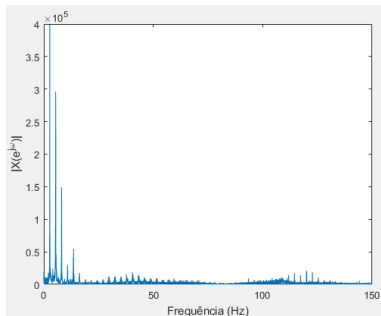
Resultados

Características dos Atletas

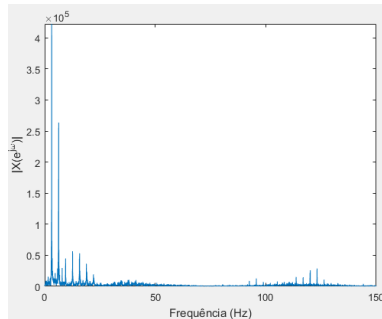


Resultados

Características dos Atletas



(a) Atletas Lesionados



(b) Atletas sem lesionados

Figura: Média da Transformada de Fourier das Forças no eixo x, 2.5 m/s 3.5 m/s e 4.5 m/s

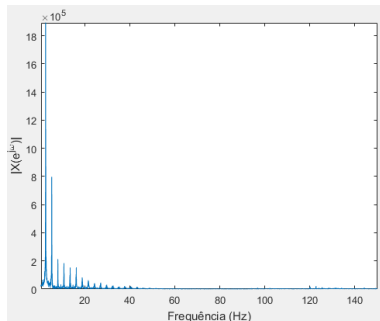


UFOP

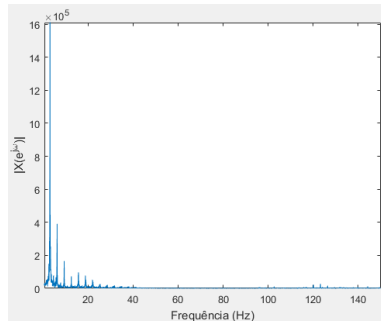
Universidade Federal
de Ouro Preto

Resultados

Características dos Atletas



(a) Atletas Lesionados



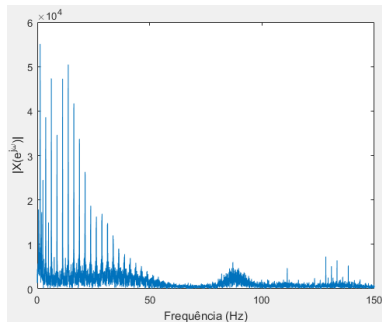
(b) Atletas sem lesão

Figura: Média da Transformada de Fourier das Forças no eixo y, 2.5 m/s 3.5 m/s e 4.5 m/s

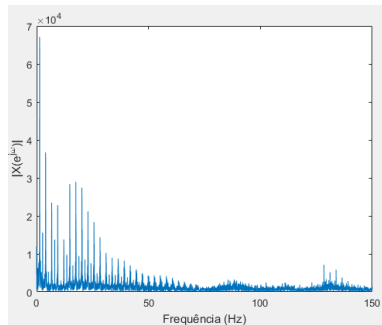


Resultados

Características dos Atletas



(a) Atletas Lesionados



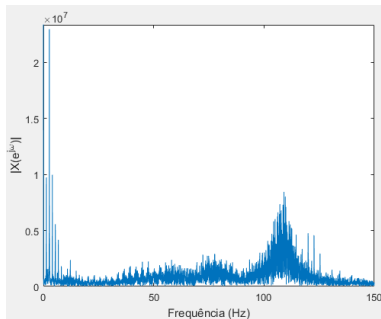
(b) Atletas sem lesão

Figura: Média da Transformada de Fourier das Forças no eixo z, 2.5 m/s 3.5 m/s e 4.5 m/s

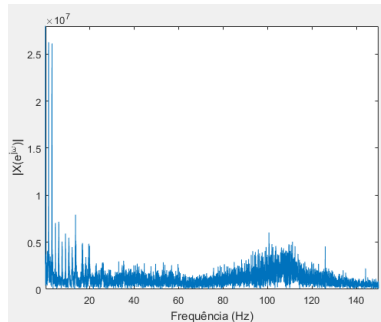


Resultados

Características dos Atletas



(a) Atletas Lesionados



(b) Atletas sem lesão

Figura: Média da Transformada de Fourier do momento livre sob o eixo vertical T_y nos tempos 2.5 m/s 3.5 m/s e 4.5 m/s


Conclusões e Trabalhos Futuros


Obrigado!





Universidade Federal
de Ouro Preto


Referências I

 CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, n. 3, p. 273–297, 1995.

 COVER, T.; HART, P. Nearest neighbor pattern classification. *IEEE transactions on information theory*, IEEE, v. 13, n. 1, p. 21–27, 1967.

 HINO, A. A. F. et al. Prevalência de lesões em corredores de rua e fatores associados. *Rev Bras Med Esporte*, v. 15, n. 1, p. 36–9, 2009.

 MCGREGOR, S. J.; BOLLT, E.; FEDEL, F. J. *Method of monitoring human body movement*. [S.l.]: Google Patents, 2014. US Patent 8,821,417.

 MITCHELL, T. M. et al. Machine learning. 1997. *Burr Ridge, IL: McGraw Hill*, v. 45, n. 37, p. 870–877, 1997.