

MFT 5725 – Pesquisa clínica: delineamento e condução,
processamento de sinais biológicos e tratamento estatístico
Programa de Ciências da Reabilitação



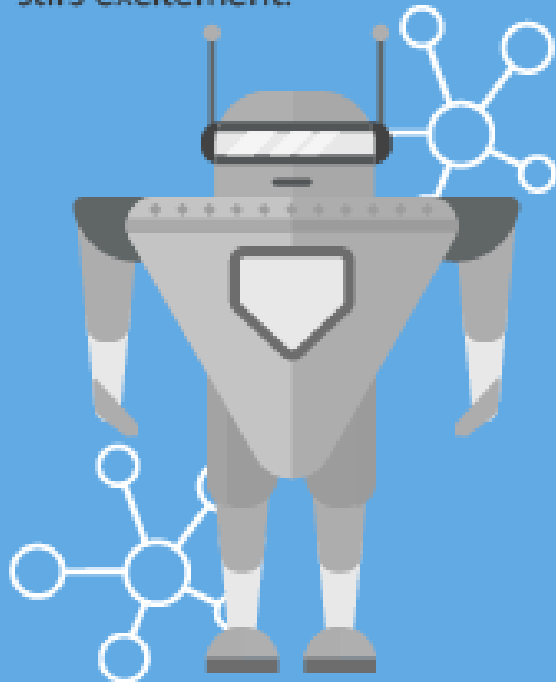
Machine Learning

Aplicações gerais e em pesquisa

Prof. Dr. Ricky Watari

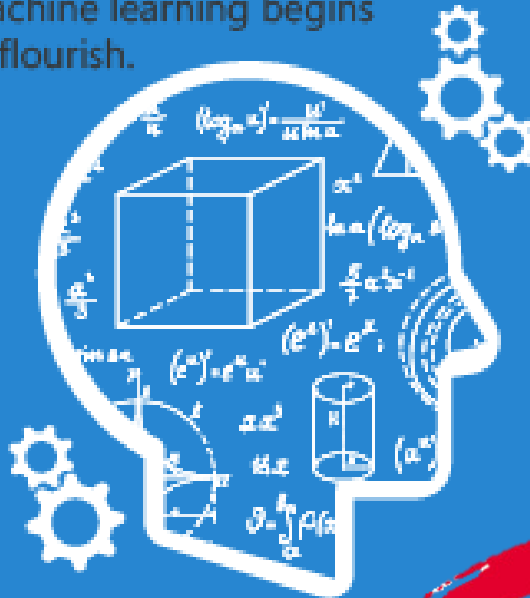
ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



MACHINE LEARNING

Machine learning begins to flourish.



DEEP LEARNING

Deep learning breakthroughs drive AI boom.



1950's 1960's 1970's 1980's 1990's 2000's 2010's



Machine Learning

“Campo de estudo que dá ao computador a habilidade de aprender sem ser explicitamente programado”



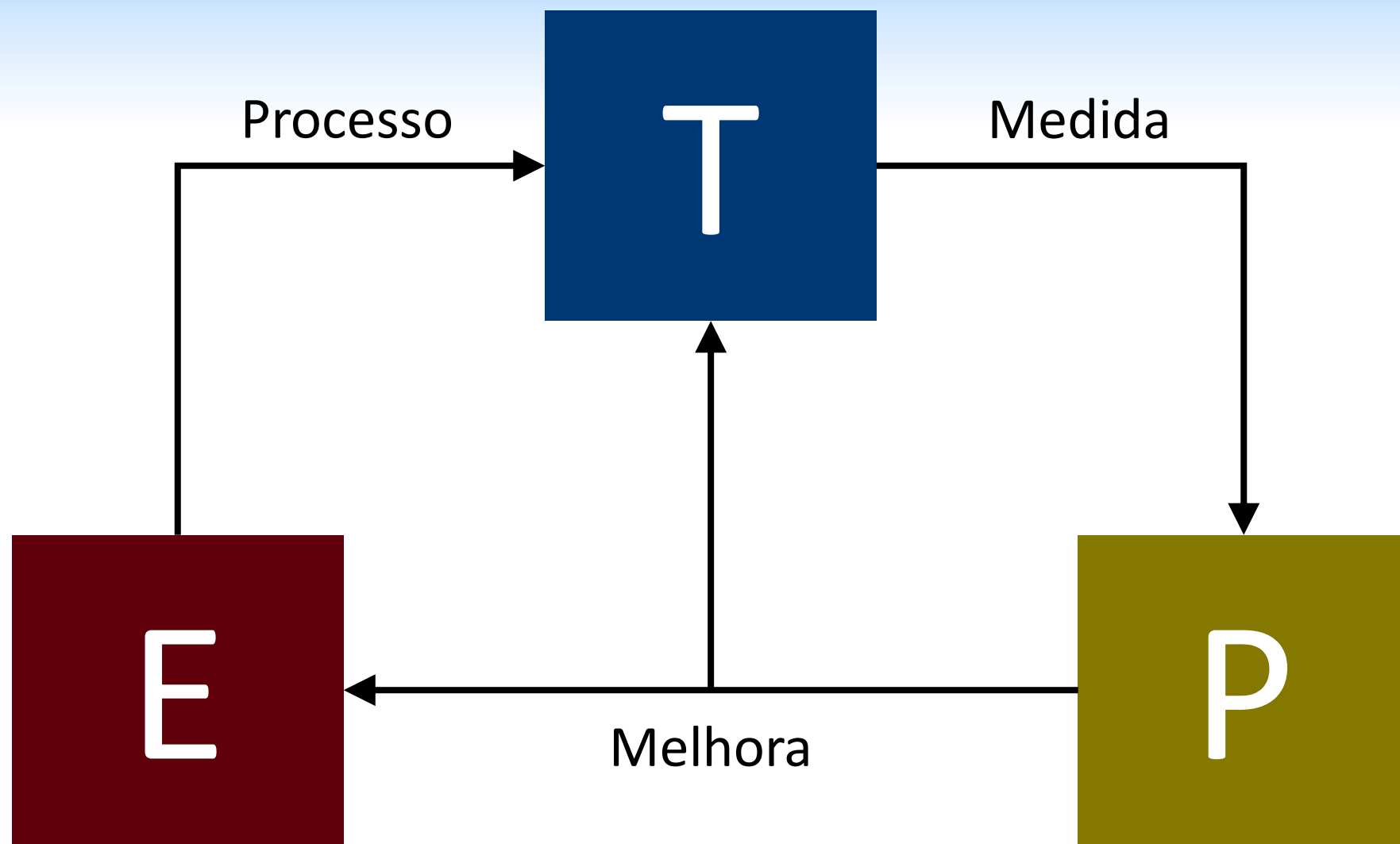
Arthur Samuel (1959)



Well-posed Learning Problem

“A computer program is said to learn from experience E with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with experience E .”

Tom Mitchell (1998)



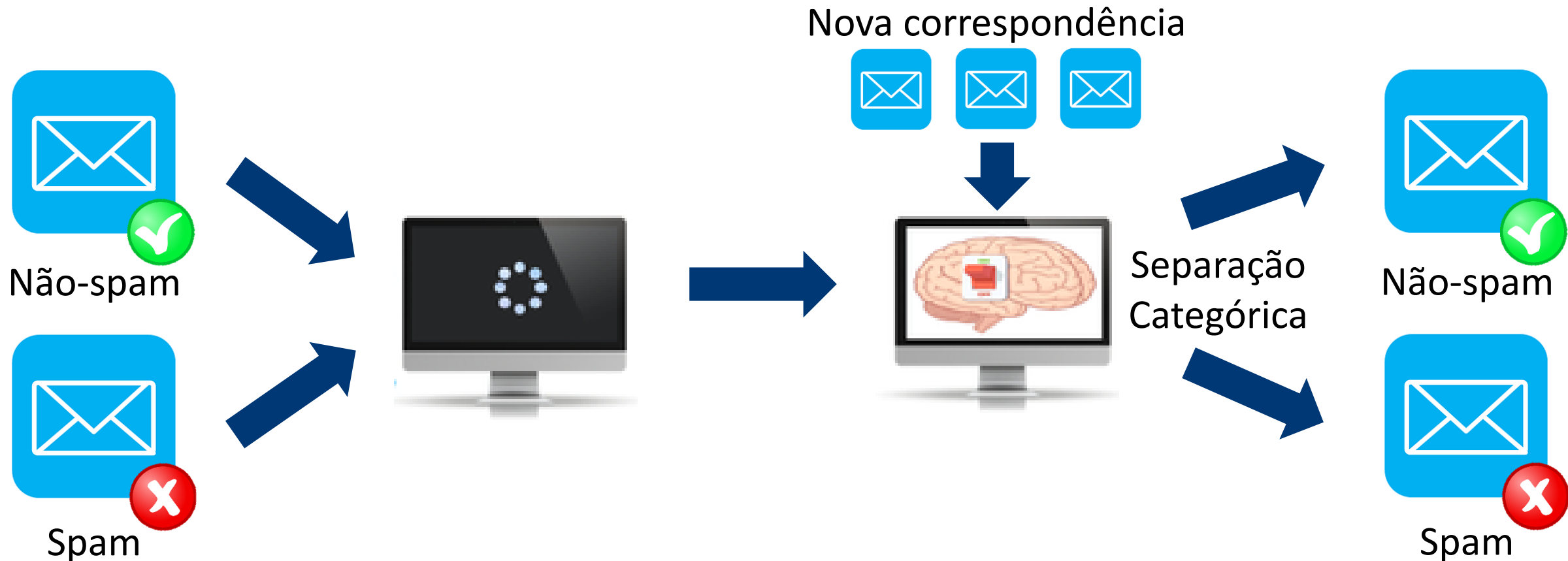


Tipos de Machine Learning

- Aprendizado Supervisionado
- Aprendizado Não-supervisionado
- Aprendizado por Reforço
- Sistemas de Recomendação

Aprendizado Supervisionado

- Mapeamento de uma entrada (input) a uma saída (output) baseado em exemplos de pares de entrada-saída.



Aprendizado Não-supervisionado

- Identificação de fatores comuns entre os dados, podendo o sistema reagir baseado na presença ou ausência de tais fatores em dados novos.



Aprendizado Não-supervisionado

- Identificação de fatores comuns entre os dados, podendo o sistema reagir baseado na presença ou ausência de tais fatores em dados novos.



Aprendizado Não-supervisionado

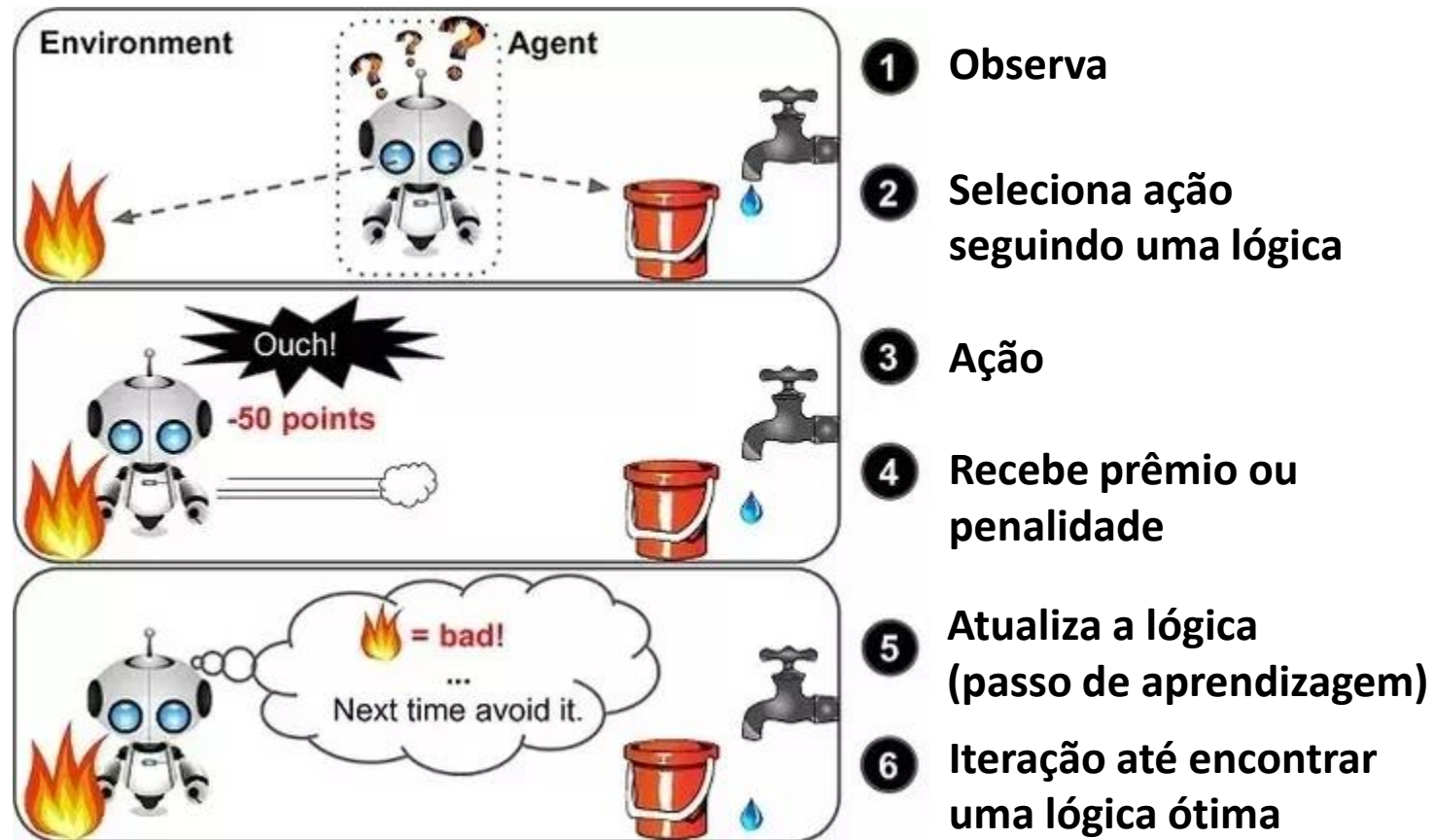
- Identificação de fatores comuns entre os dados, podendo o sistema reagir baseado na presença ou ausência de tais fatores em dados novos.

Sistemas de recomendações:



Reinforcement Learning

- Aprendizado pela interação com o ambiente (tentativa e erro), havendo premiação para decisões corretas e penalidades para as incorretas.



Machine Learning

```
graph TD; ML[Machine Learning] --> AS[Aprendizado Supervisionado]; ML --> ANS[Aprendizado Não-supervisionado]; AS --> R[Regressão]; AS --> C[Classificação]; ANS --> Cl[Clustering]; ANS --> RD[Redução de dimensão]; R --> RL[Regressão Linear]; C --> CL[Regressão Logística]; C --> SVM[SVM]; Cl --> k[k-means]; Cl --> H[Hierárquico]; RD --> PCA[PCA];
```

Aprendizado
Supervisionado

Aprendizado
Não-supervisionado

Regressão

Classificação

Clustering

Redução de
dimensão

Regressão
Linear

Regressão
Logística

SVM

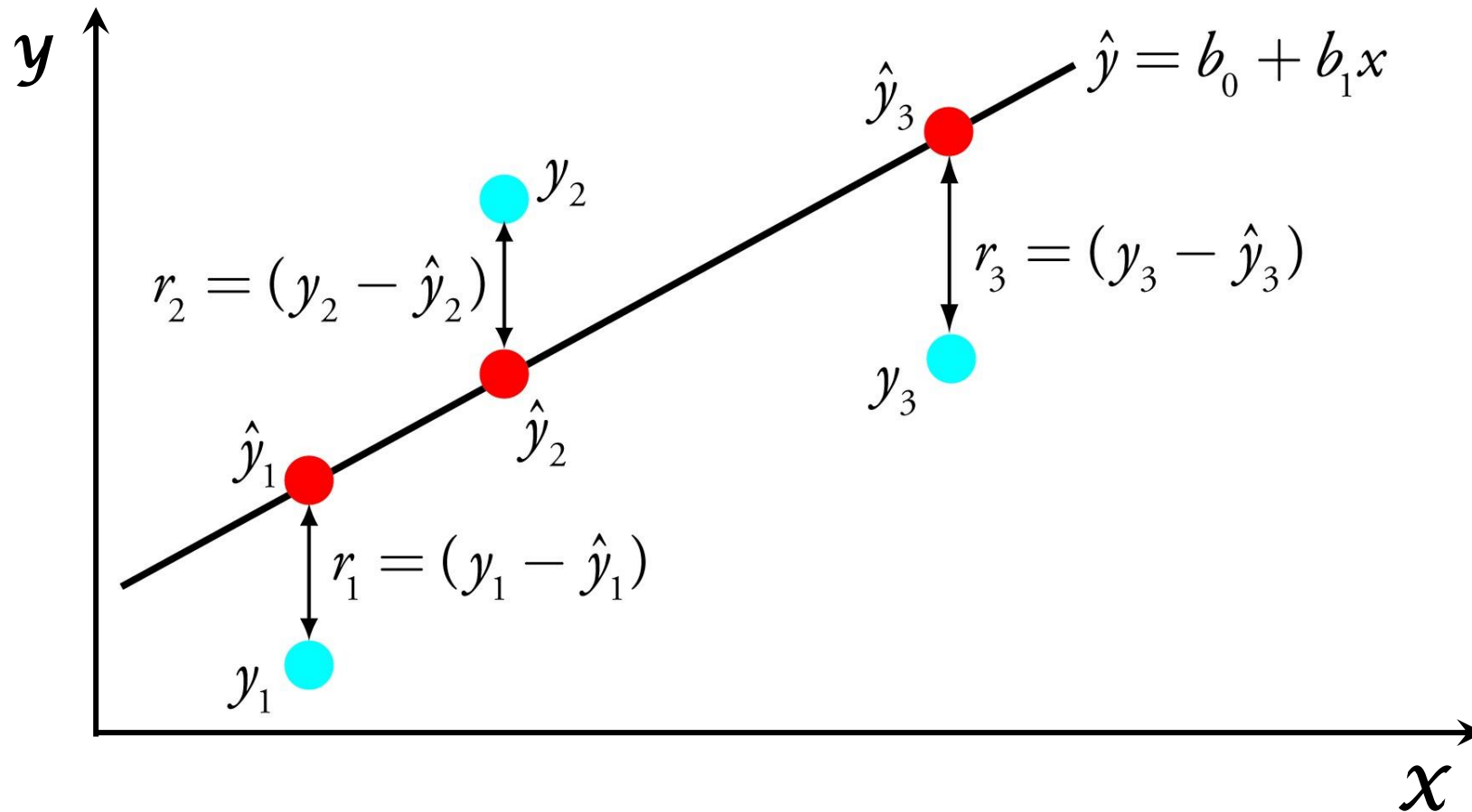
k-means

Hierárquico

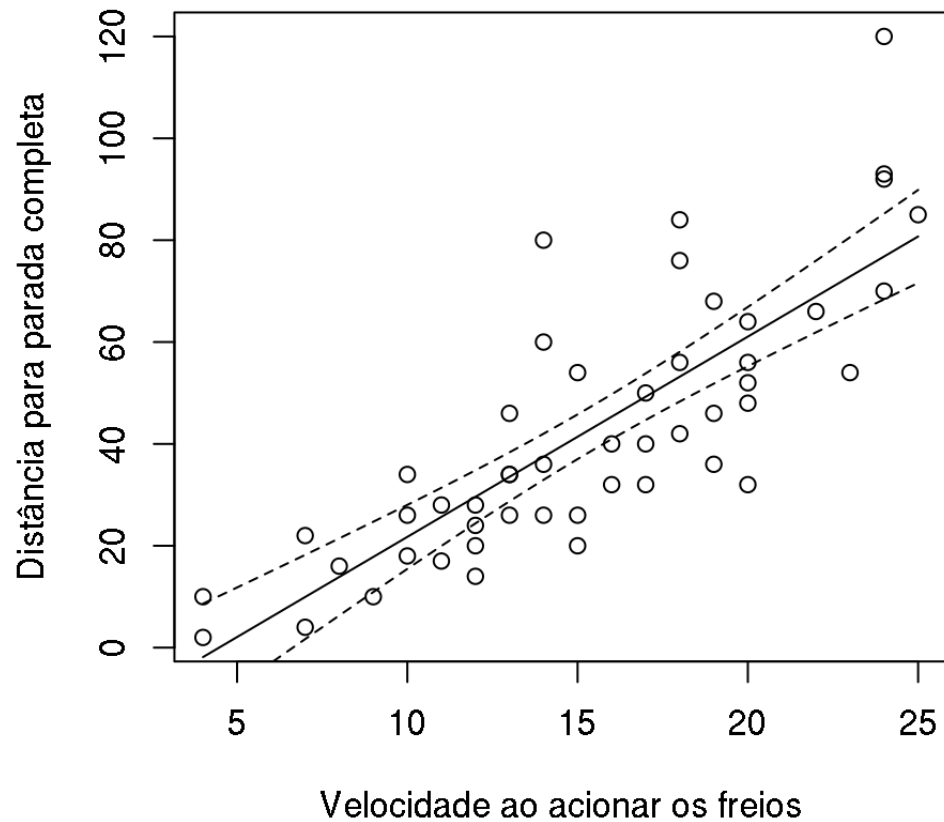
PCA

Regressão

- Predição de variáveis contínuas



Regressão linear simples



$$\text{distância} = -17,58 + 3,93 * \text{velocidade}$$

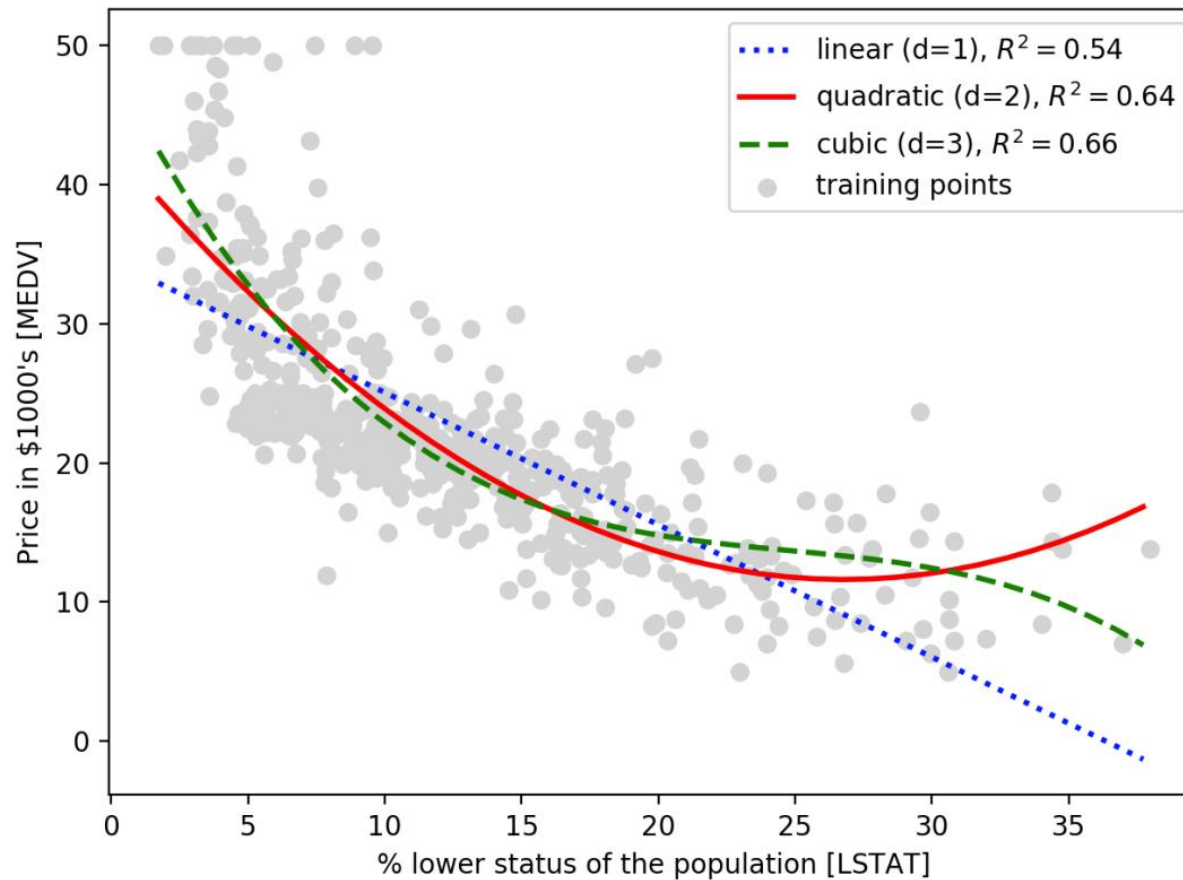
β

Two red arrows point down to the intercept (-17,58) and the slope (3,93) in the equation above. The intercept and slope are circled in red.

Regressão linear múltipla

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n$$

Modelos não-lineares





Improving the Prediction of Total Surgical Procedure Time Using Linear Regression Modeling

Eric R. Edelman^{1}, Sander M. J. van Kuijk², Ankie E. W. Hamaekers³,
Marcel J. M. de Korte³, Godefridus G. van Merode⁴ and Wolfgang F. F. A. Buhre³*

<https://doi.org/10.3389/fmed.2017.00085>



Variáveis
Preditivas
(2012-2015)

Tempo estimado de cirurgia

Idade do paciente

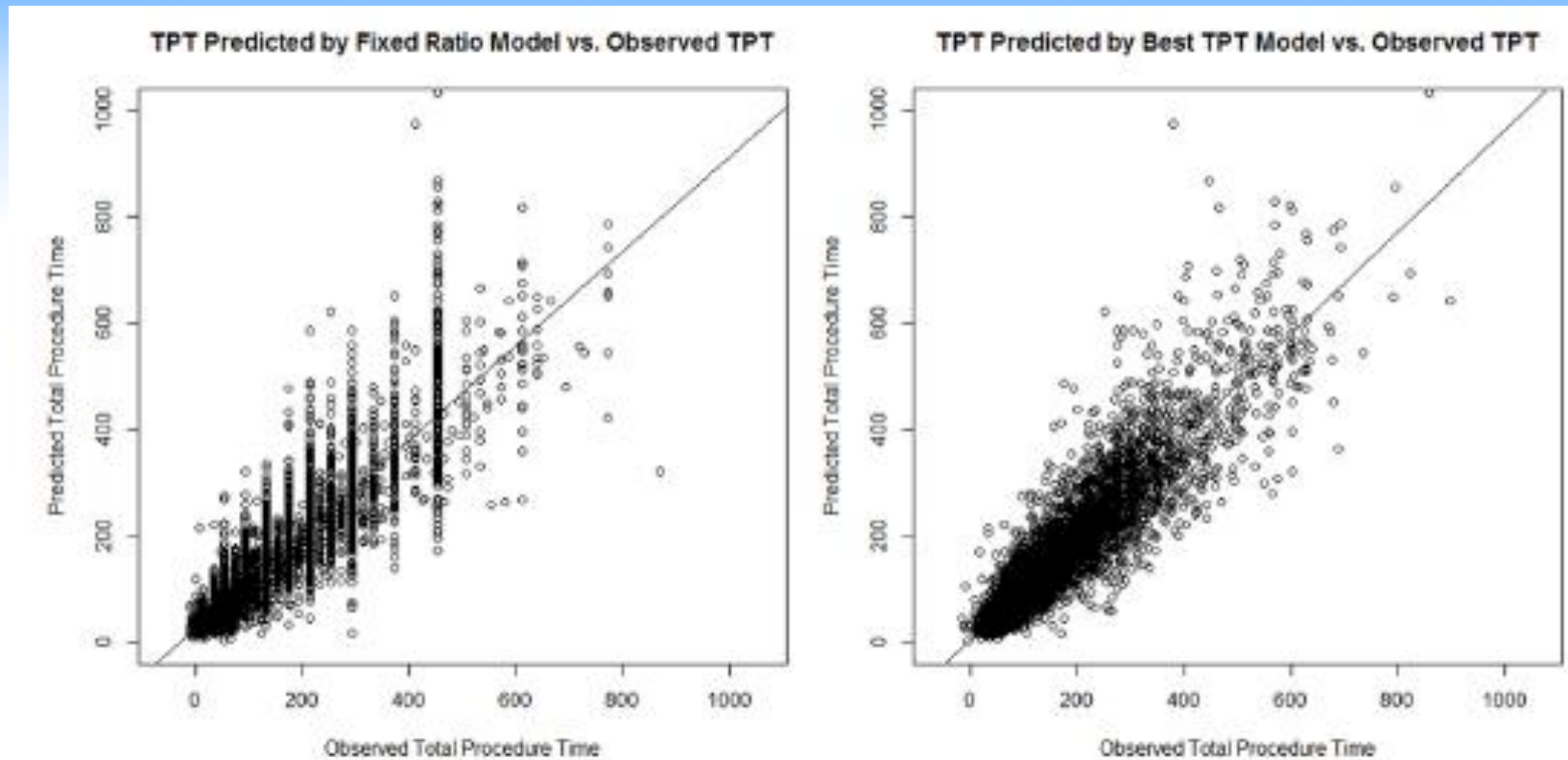
Tipo de cirurgia

Estado físico (Anestesiologia)

Tipo de anestesia

Regressão linear múltipla

Tempo total de cirurgia



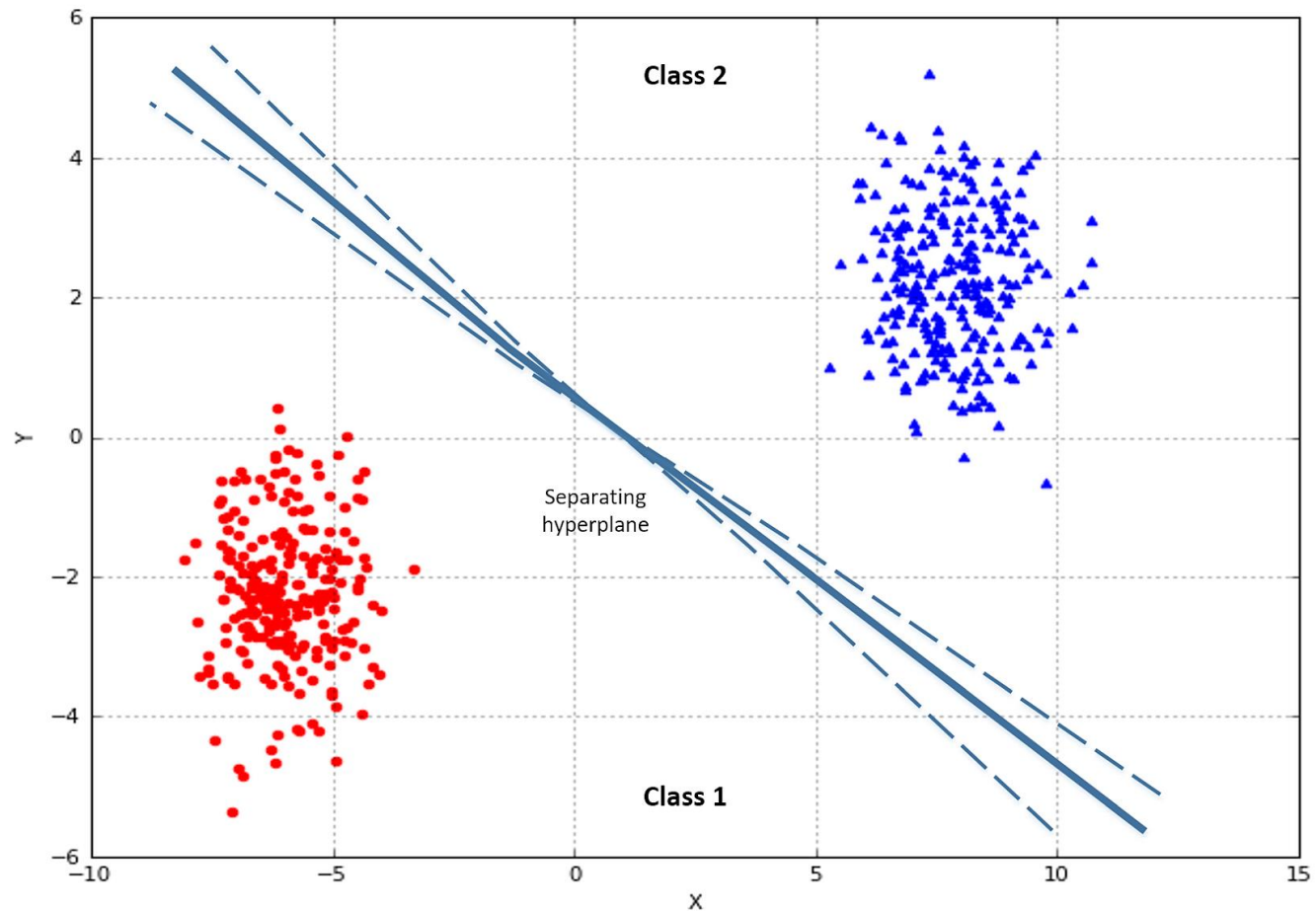
38,5
horas

Erro médio
(dados de 2016)

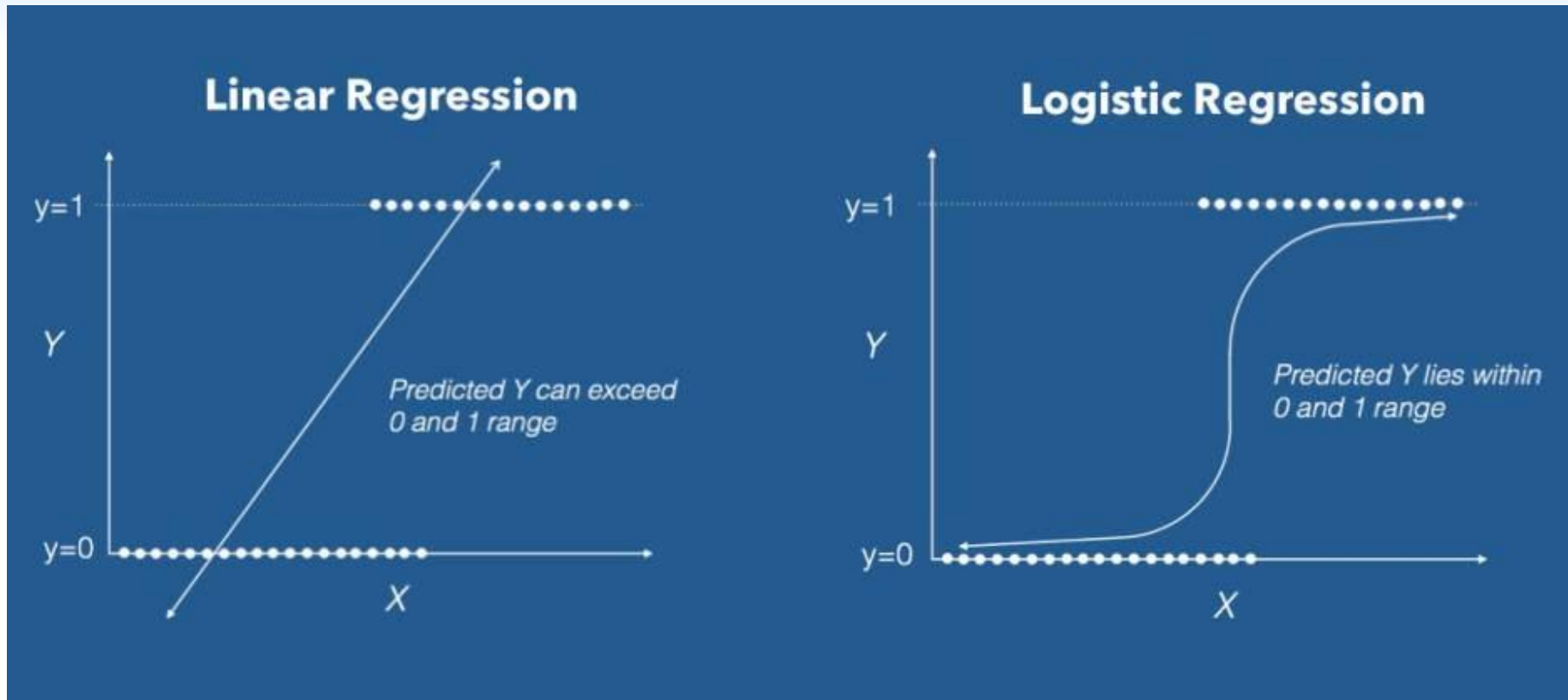
31,3
horas

Classificação

- Predição de variáveis categóricas (classes)



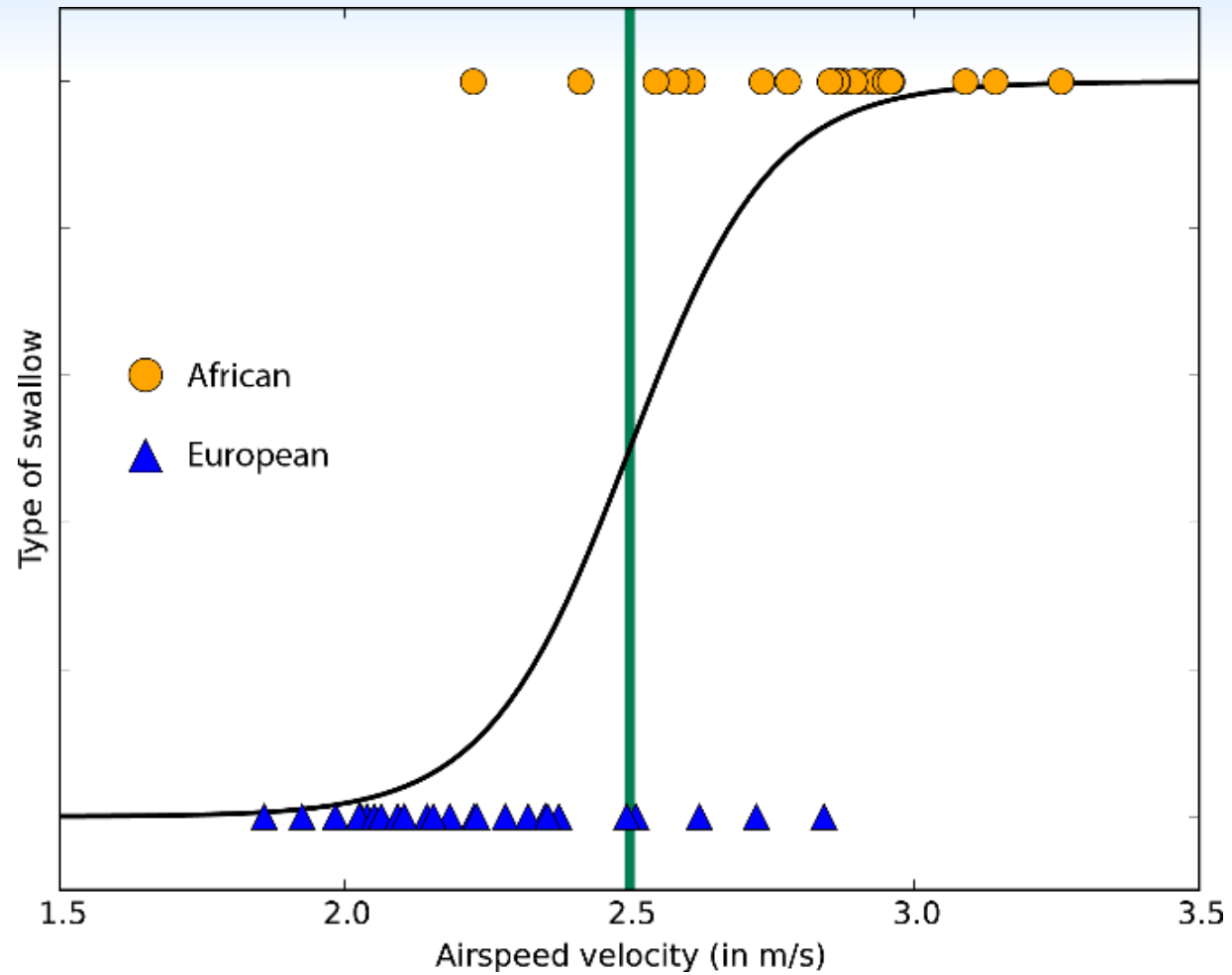
Regressão Logística



$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$$

$$P = \frac{e^{\text{logit}(P)}}{1 + e^{\text{logit}(P)}}$$

Regressão Logística



C. Walgaard, MD
H.F. Lingsma, PhD
L. Ruts, MD
P.A. van Doorn, MD
E.W. Steyerberg, PhD
B.C. Jacobs, MD

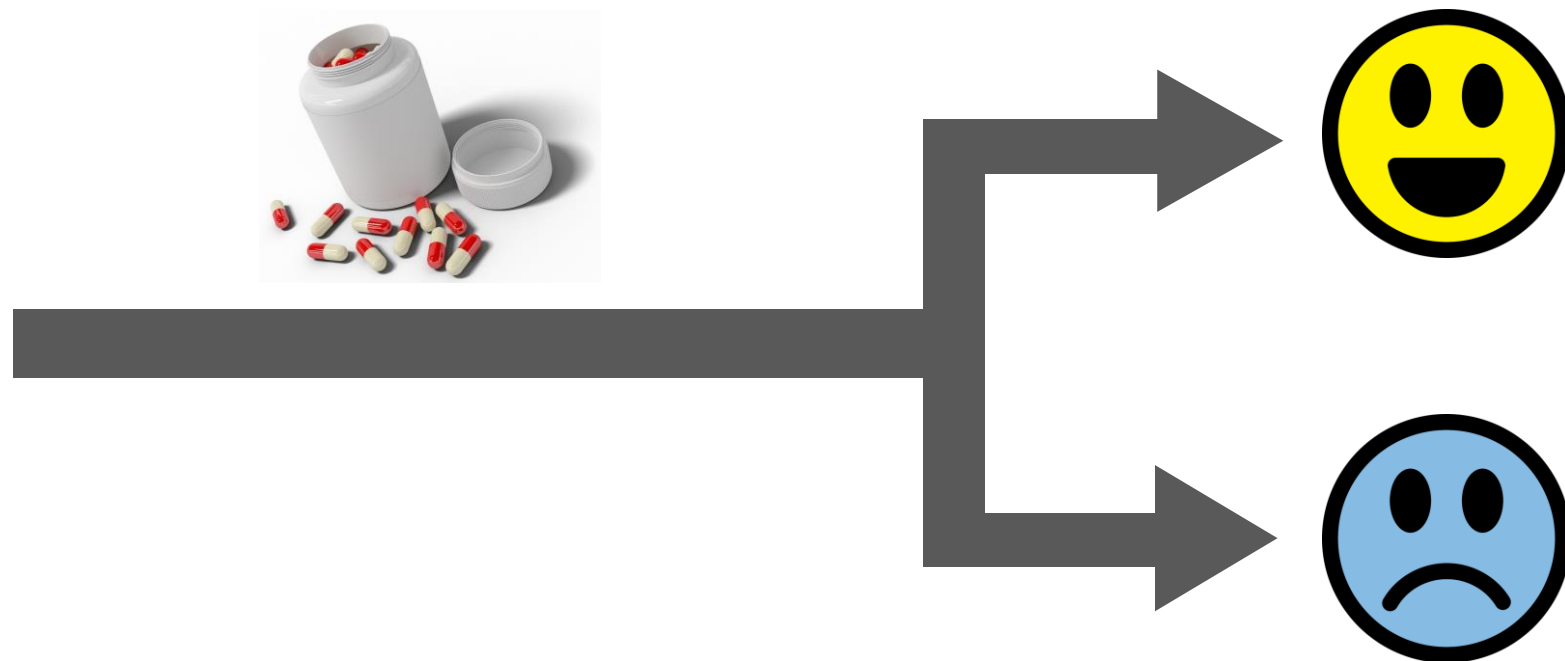
Early recognition of poor prognosis in Guillain-Barré syndrome



<https://doi.org/10.1212/WNL.0b013e3182104407>



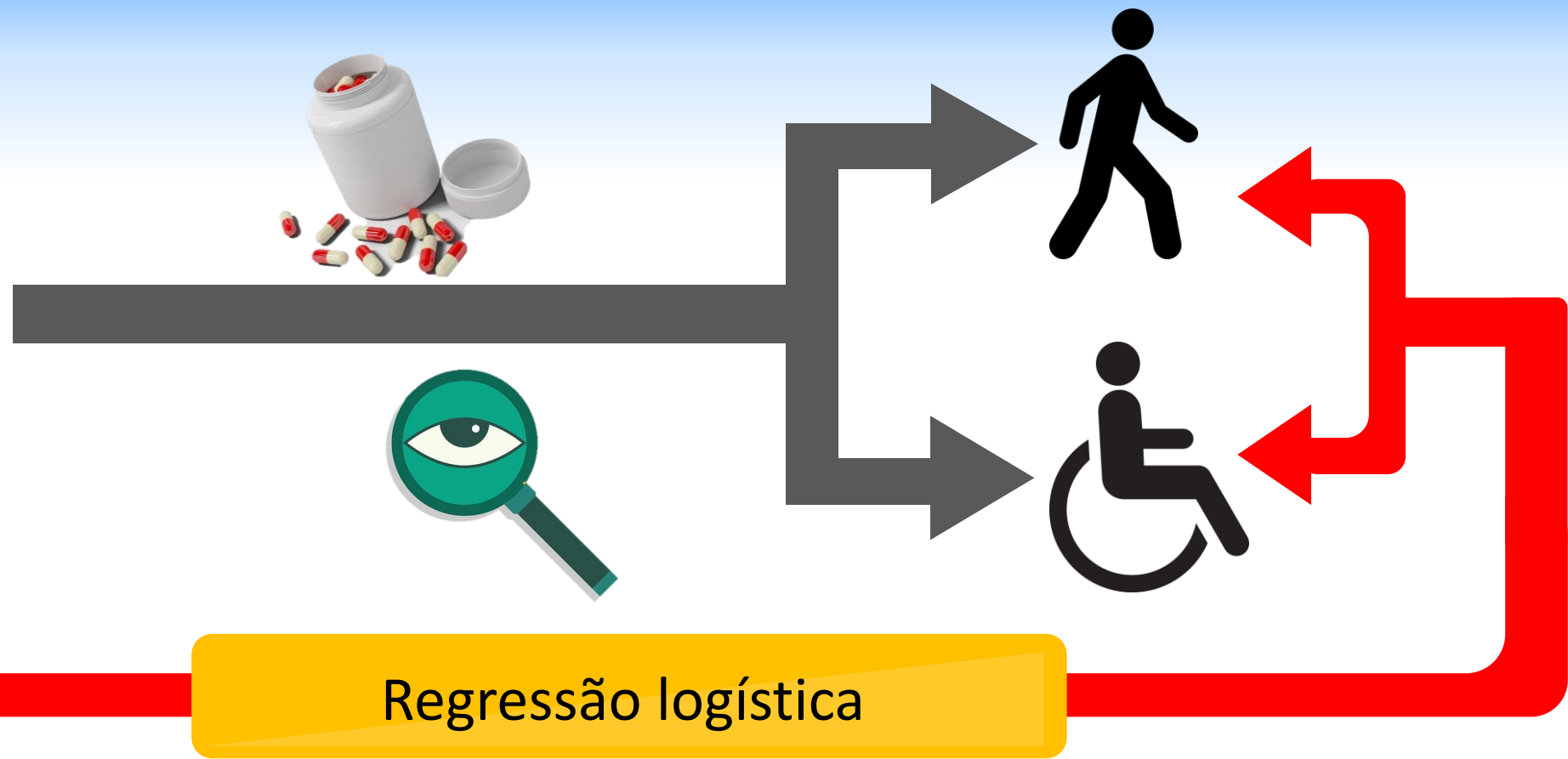
Guillain-Barré



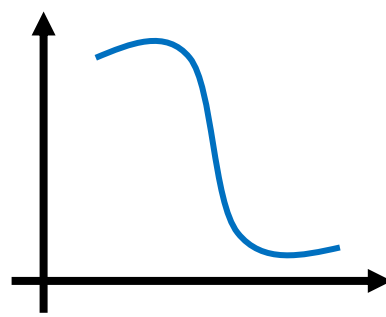


Guillain-Barré

n = 397



Regressão logística



Variáveis
Preditivas

Idade

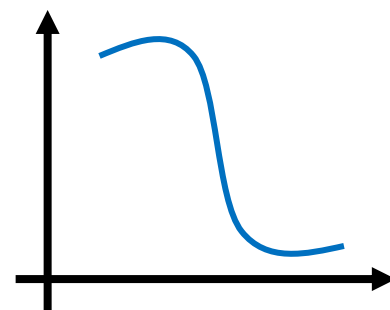
Diarreia

Escala MRC



Guillain-Barré

n = 191



Curva ROC (AUC):
0.87



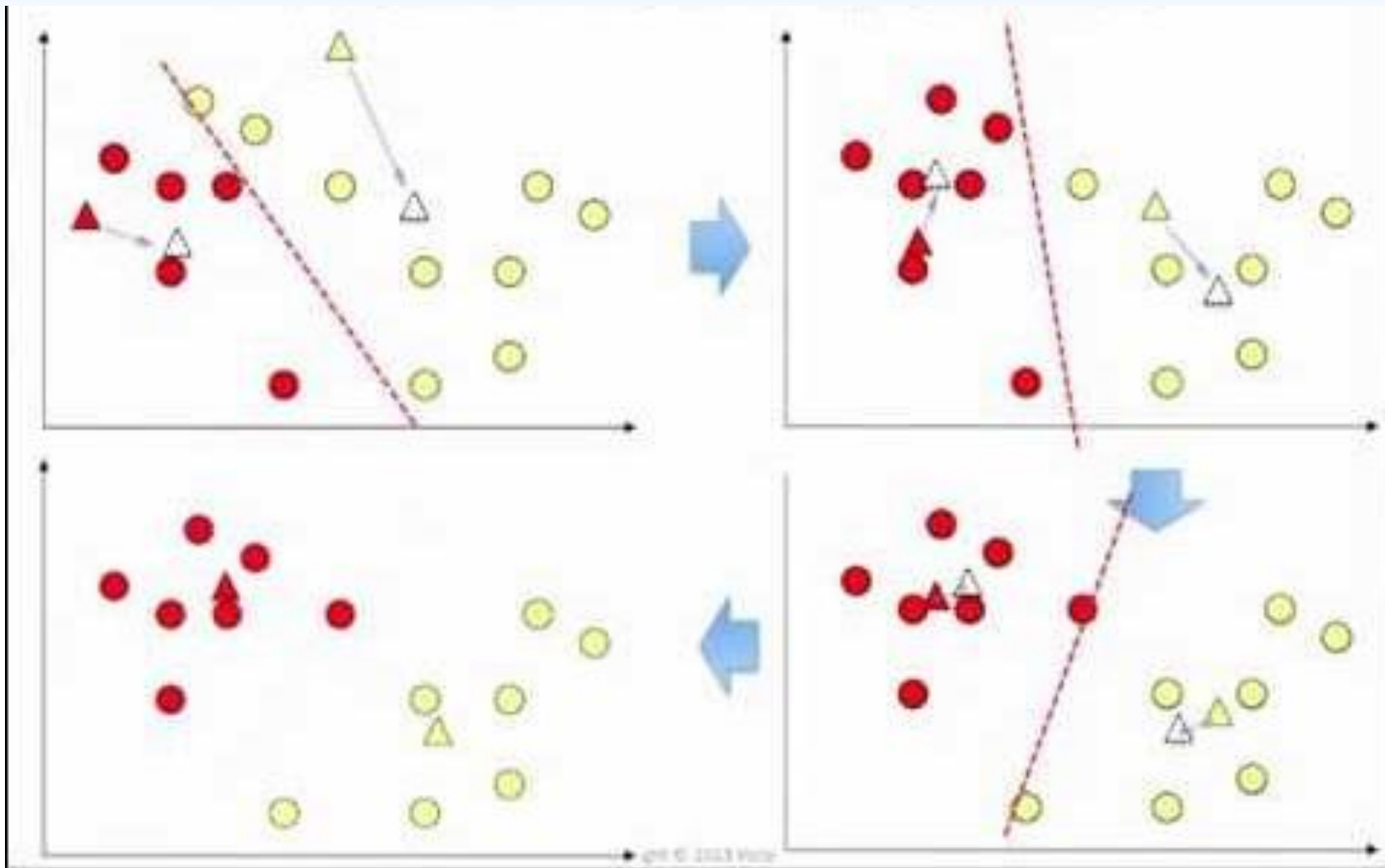
Clustering

- Partição de um conjunto de dados em grupos que são o mais similar possível entre seus elementos

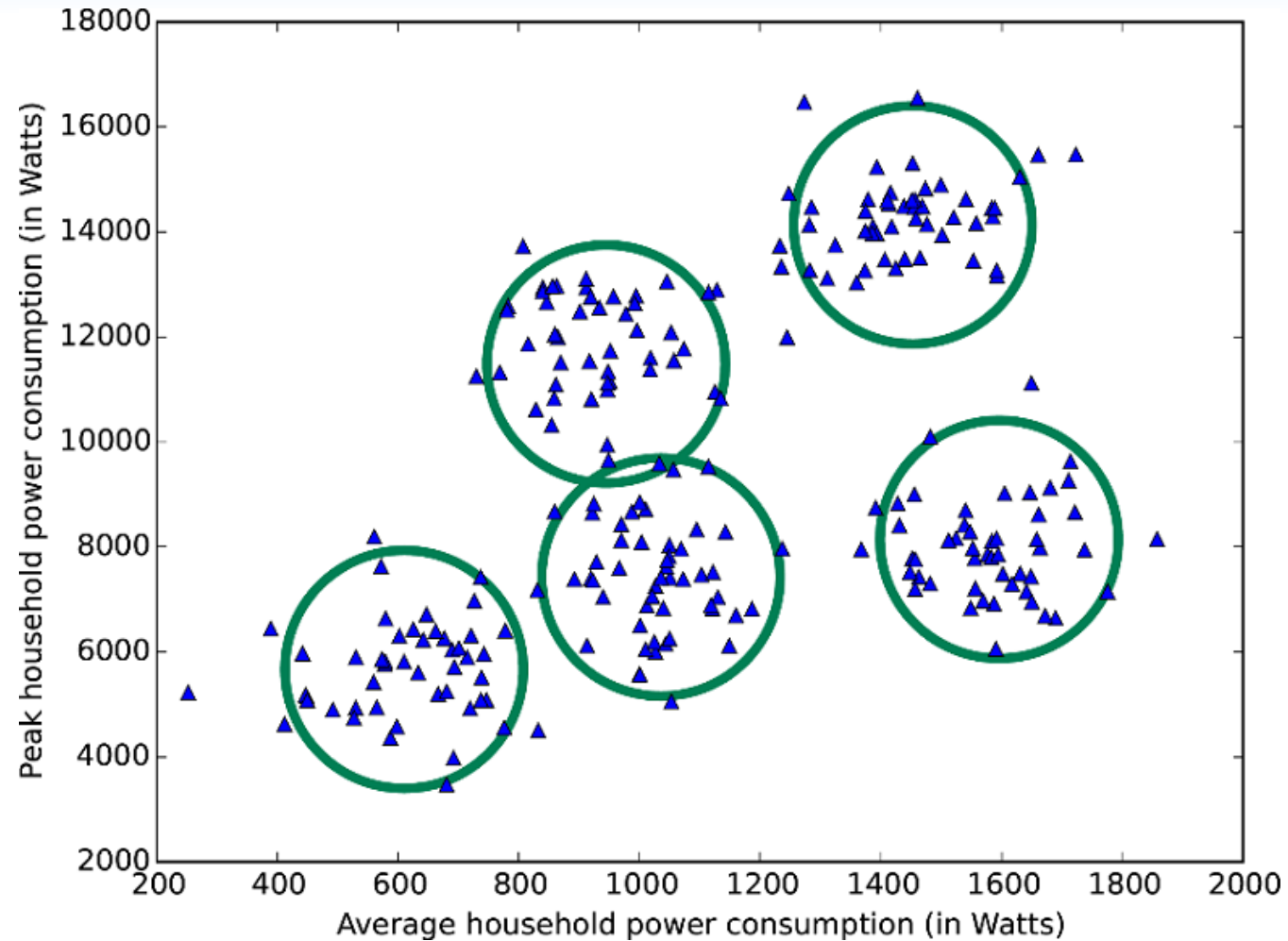


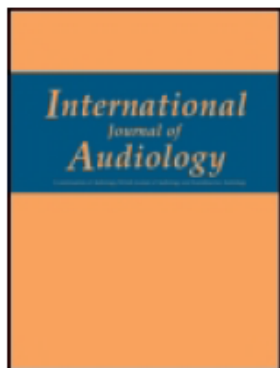
- Resumos de dados
- Detecção de padrões
- Detecção de outliers

k-means Clustering



k-means Clustering





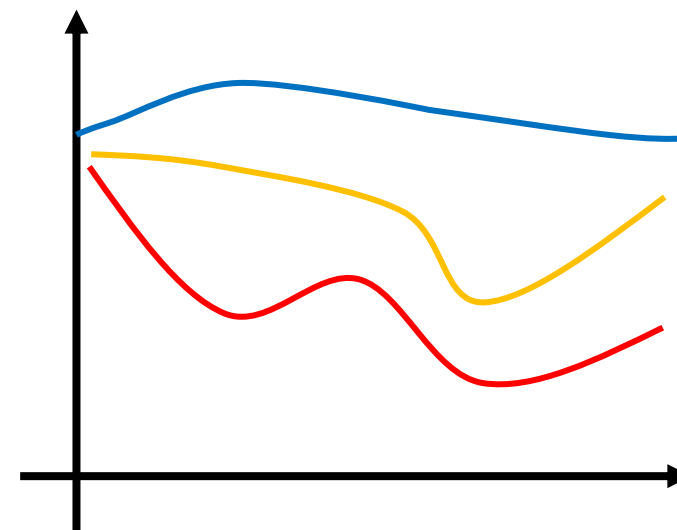
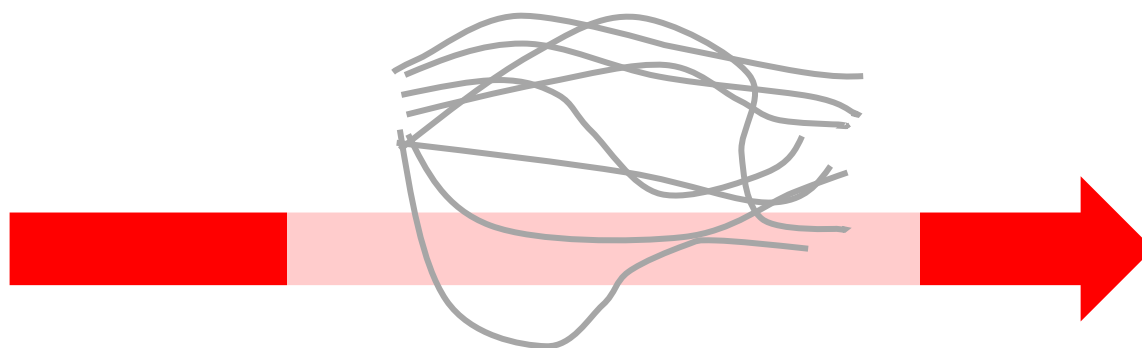
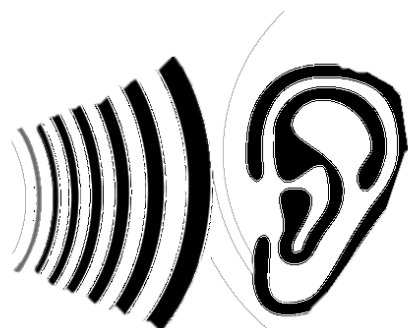
Original Article

<https://doi.org/10.3109/14992021003796887>

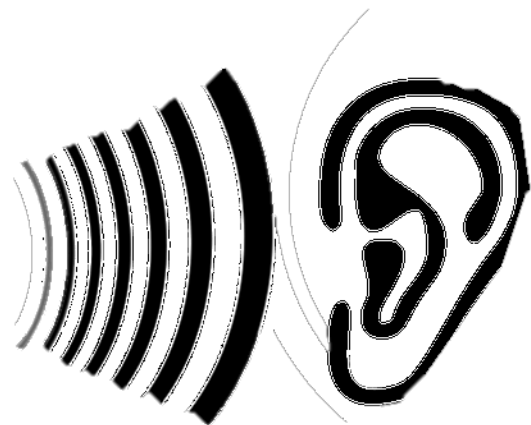
International Journal of Audiology 2010; 49: 628–633

*Cheng-Yung Lee**
Juen-Haur Hwang,†*
*Szu-Jen Hou**
Tien-Chen Liu‡

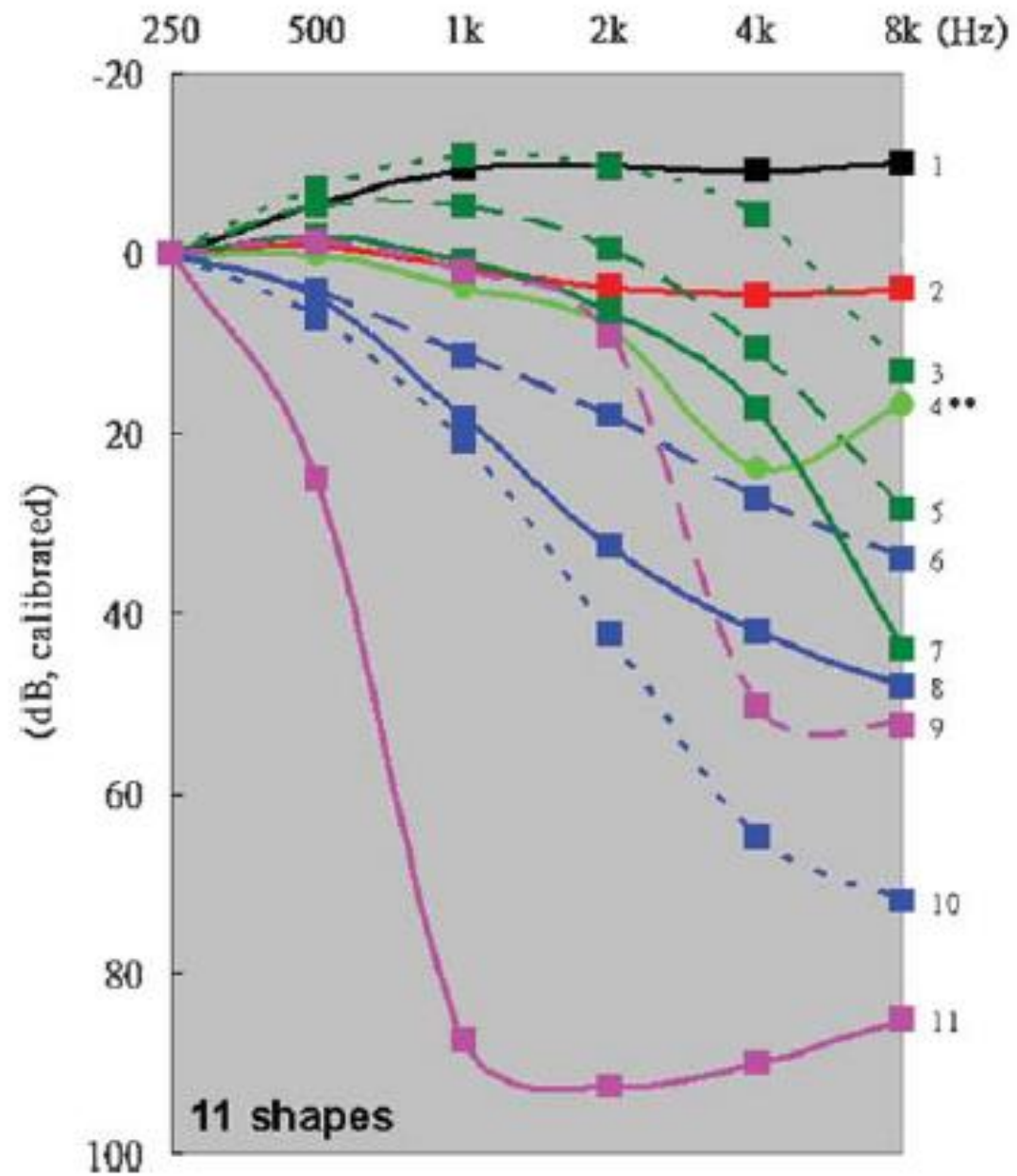
Using cluster analysis to classify audiogram shapes



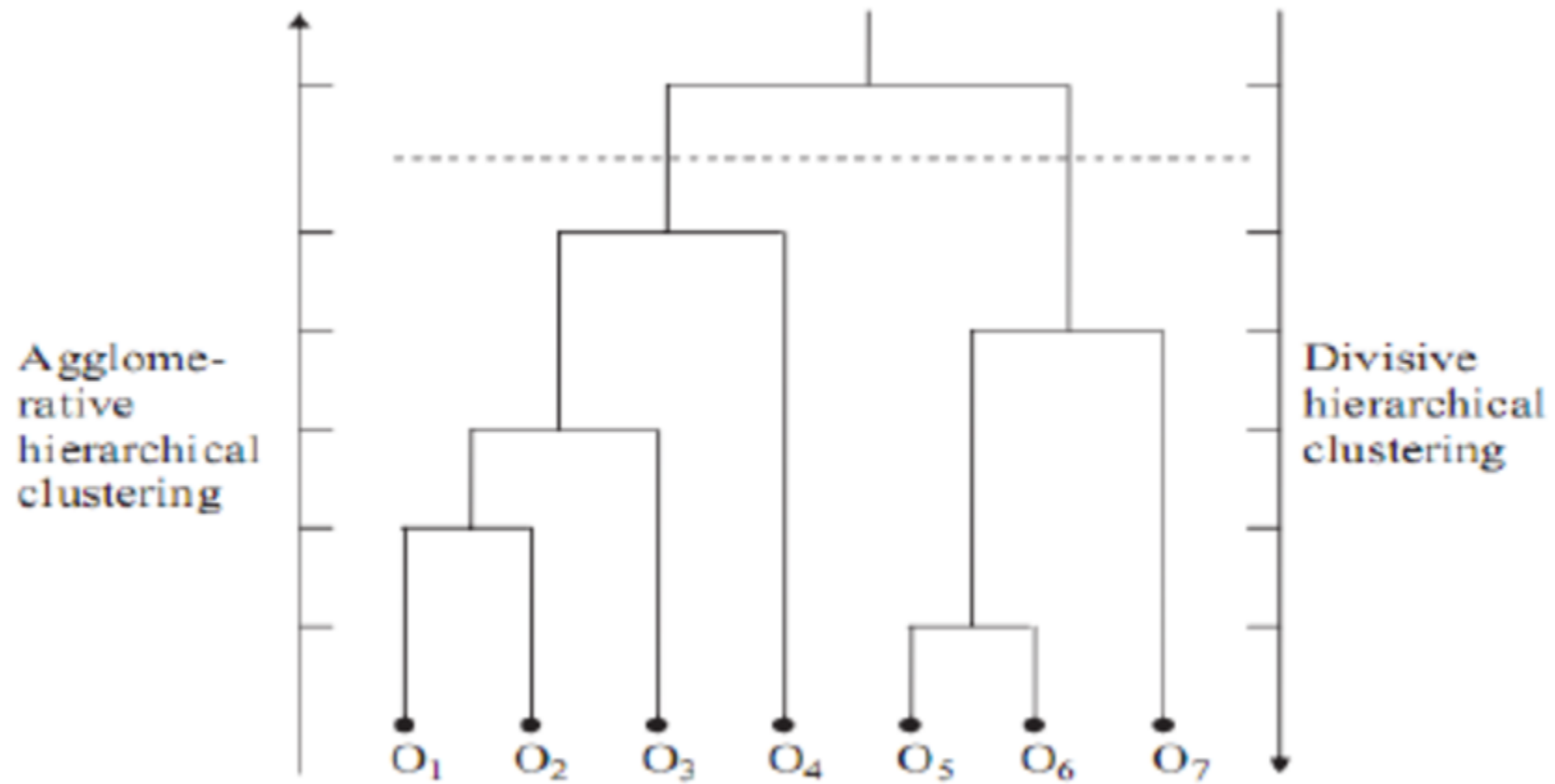
Análise de cluster k-means



n = 1633



Clustering hierárquico



Contents lists available at ScienceDirect

Parkinsonism and Related Disorders

journal homepage: www.elsevier.com/locate/parkreldis



Cortical atrophy patterns in early Parkinson's disease patients using hierarchical cluster analysis



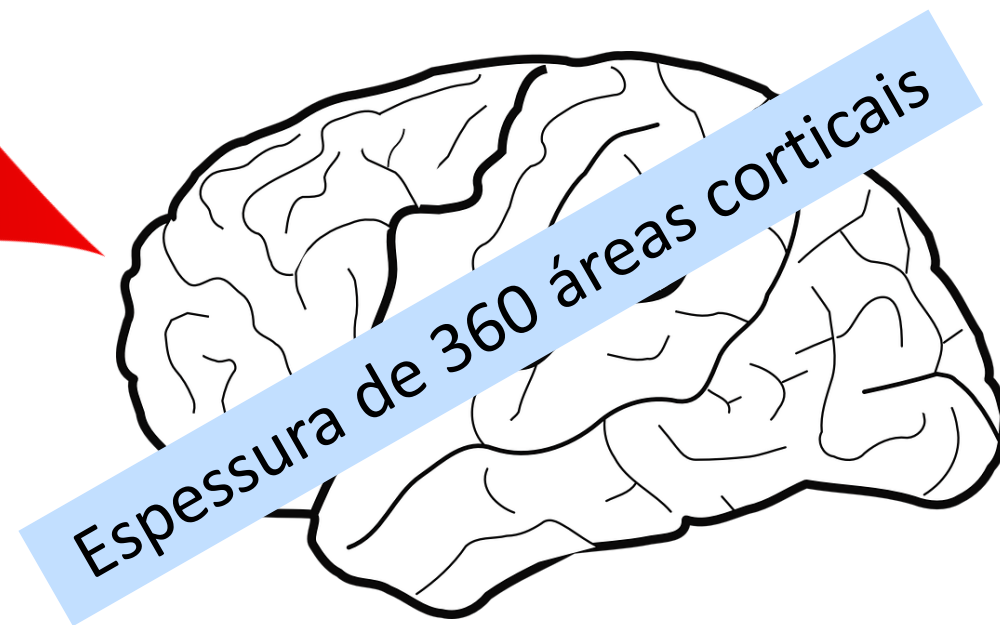
Carme Uribe ^a, Barbara Segura ^a, Hugo Cesar Baggio ^a, Alexandra Abos ^a,
Anna Isabel Garcia-Diaz ^a, Anna Campabadal ^{a, b}, Maria Jose Marti ^{b, c, d},
Francesc Valldeoriola ^{b, c, d}, Yaroslau Compta ^{b, c, d}, Eduard Tolosa ^{b, c, d},
Carme Junque ^{a, b, c, *}

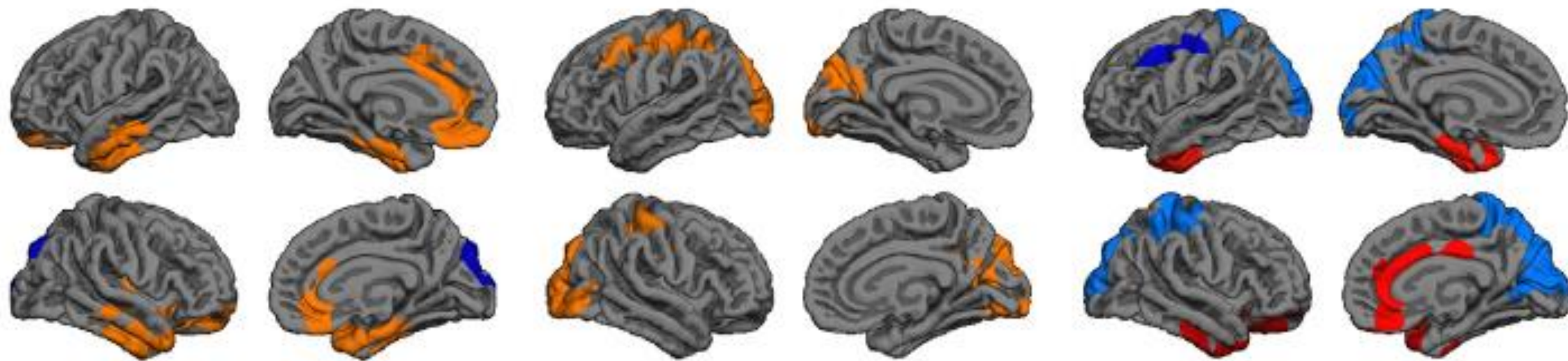
<https://doi.org/10.1016/j.parkreldis.2018.02.006>

Doença de Parkinson



n = 119







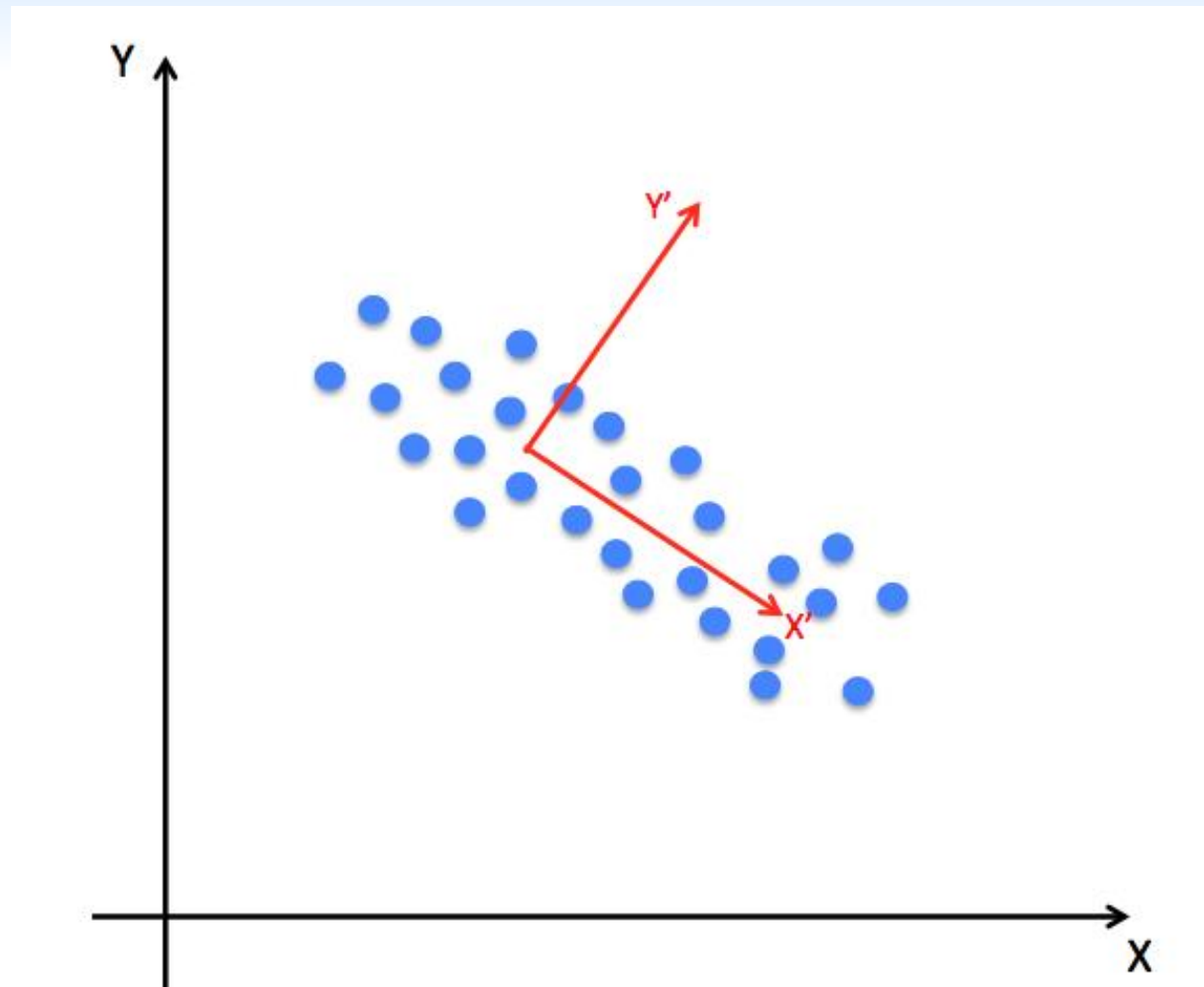
pattern 1 < HC (warm)
pattern 1 > HC (cold)

pattern 2 < HC

pattern 1 < pattern 2 (warm)
pattern 1 > pattern 2 (cold)

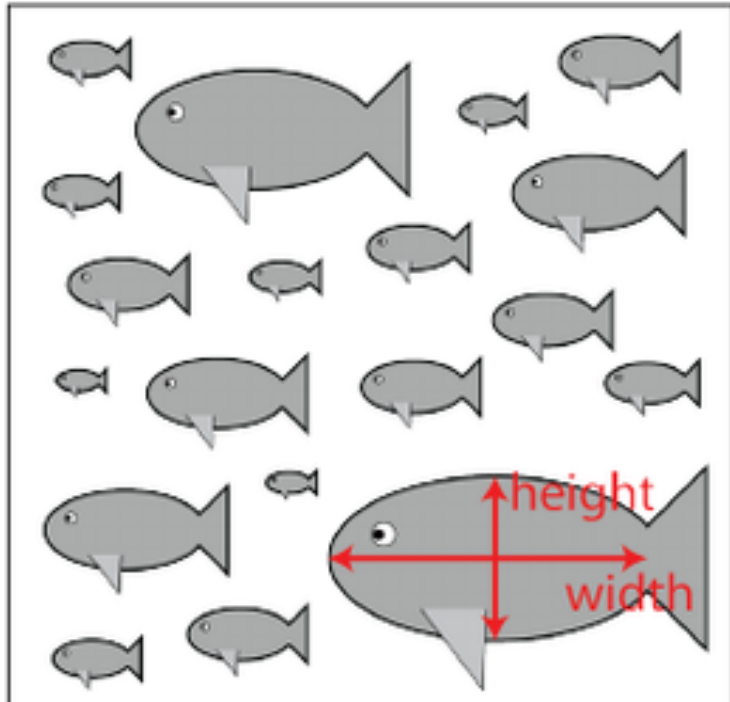
<.05  <.0001 corrected p <.05  <.0001

Redução de dimensão

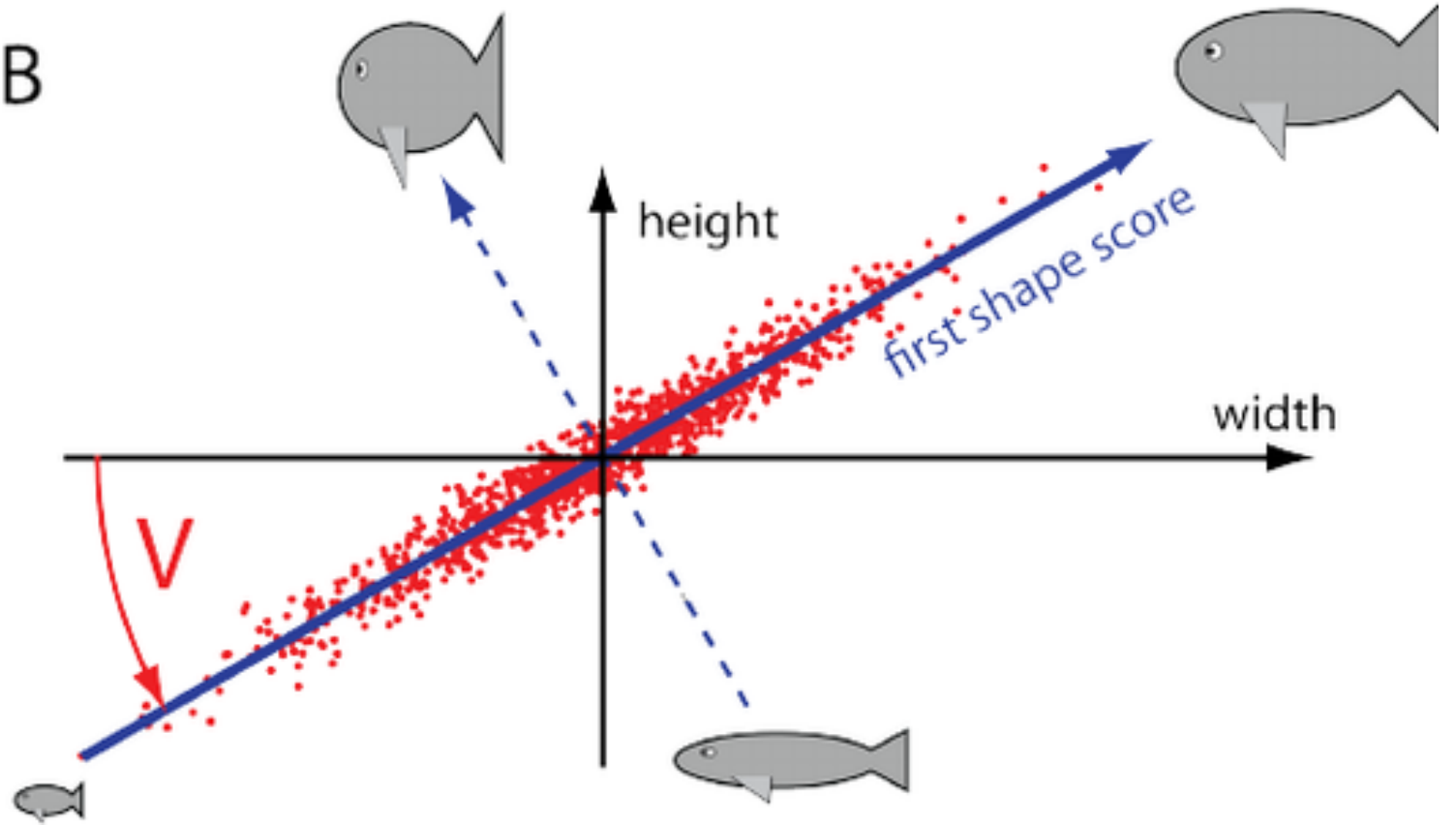


Análise de componentes principais

A



B





Migraine features, associated symptoms and triggers: A principal component analysis in the Women's Health Study

Cephalalgia

31(7) 861–869

© International Headache Society 2011

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

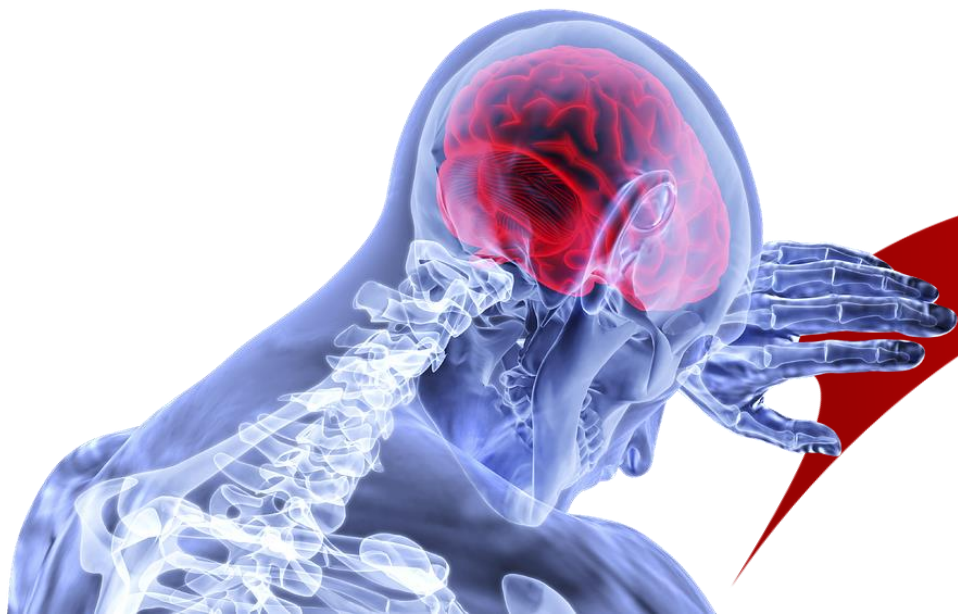
DOI: 10.1177/0333102411401635

cep.sagepub.com



Markus Schürks^{1,2}, Julie E Buring¹ and Tobias Kurth^{1,3,4}

<https://doi.org/10.1177/0333102411401635>



n = 9427



Sintomas

	Component 1: CNS sensitization	Component 2: attack frequency and pain location	Component 3: aura and visual phenomena
Eigenvalue	2.69	1.10	1.04
Variance explained (%)	26.9	11.0	10.4
Migraine feature			
Inhibition of daily activities	0.68	0.09	0.01
Sensitivity to sound	0.67	0.02	0.17
Aggravation by routine physical activity	0.62	0.03	-0.10
Sensitivity to light	0.61	0.08	0.41
Nausea/Vomiting	0.56	0.16	-0.05
Pulsating pain quality	0.40	0.36	-0.28
Unilateral pain location	0.12	0.64	0.01
Attack duration of 4–72 hours	0.47	0.47	-0.19
Migraine attack frequency	0.05	-0.73	-0.18
Aura	0.03	0.08	0.86



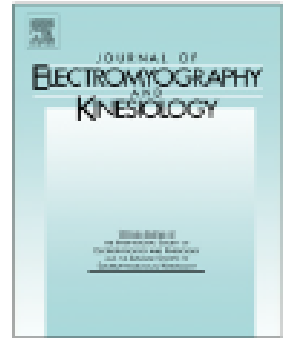




Contents lists available at ScienceDirect

Journal of Electromyography and Kinesiology

journal homepage: www.elsevier.com/locate/jelekin



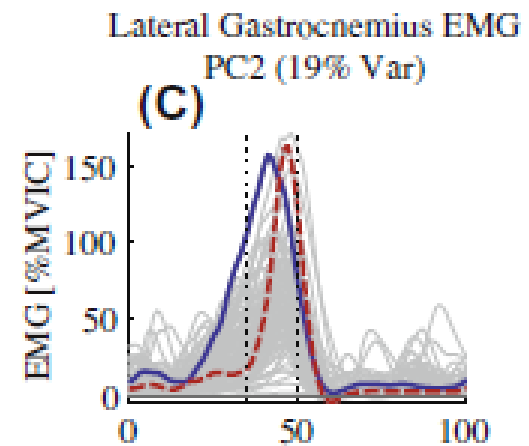
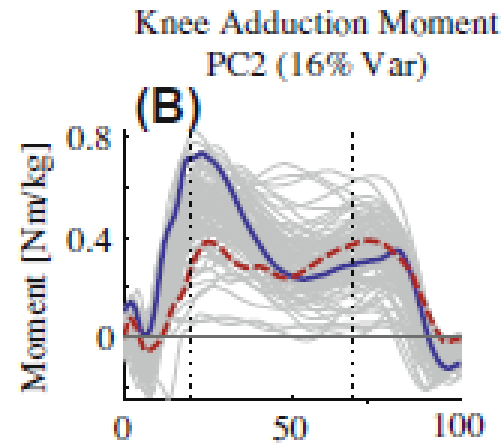
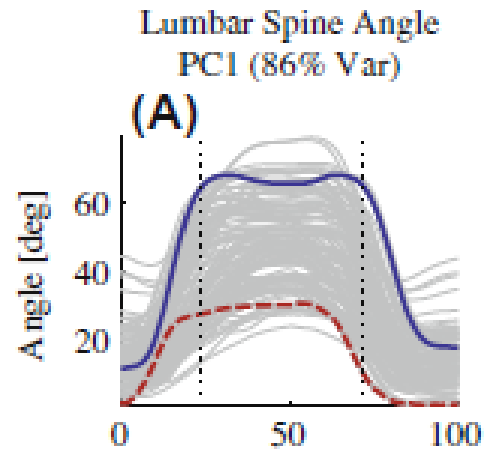
Interpreting principal components in biomechanics: Representative extremes and single component reconstruction



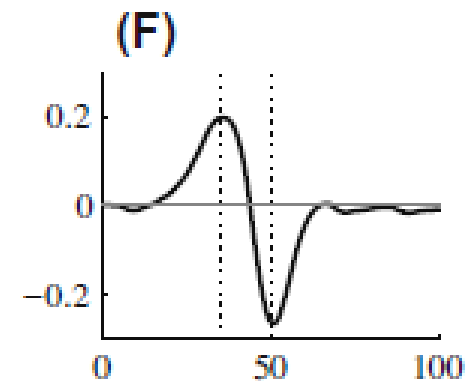
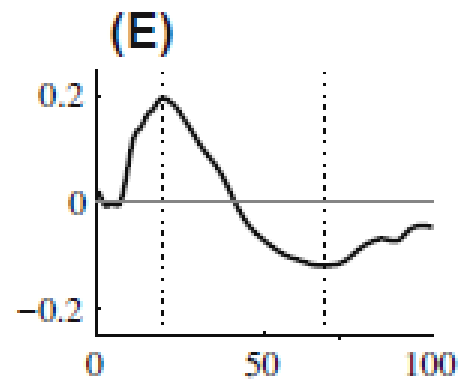
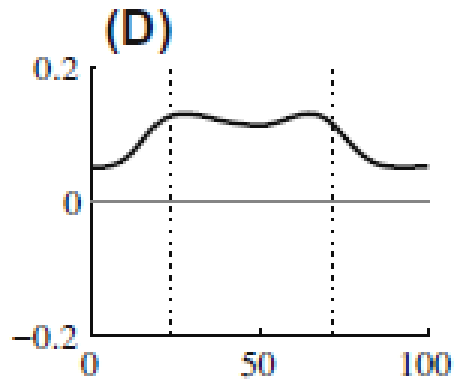
Scott C.E. Brandon^{a,b,*}, Ryan B. Graham^c, Sivan Almosnino^{b,d}, Erin M. Sadler^d, Joan M. Stevenson^{b,d},
Kevin J. Deluzio^{a,b}

<https://doi.org/10.1016/j.jelekin.2013.09.010>

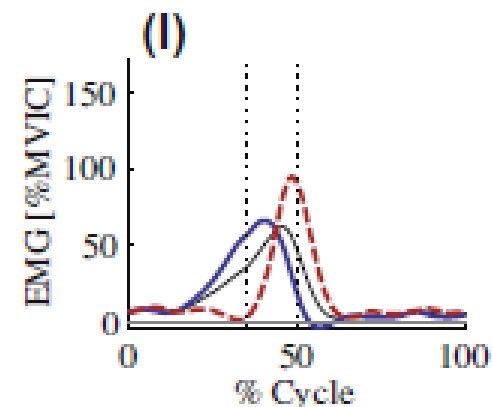
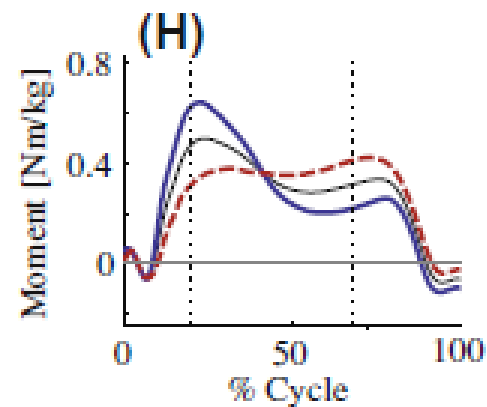
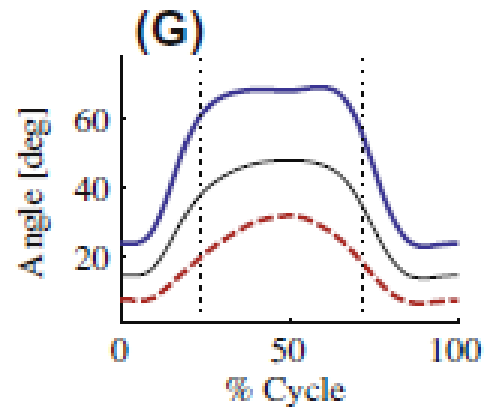
Raw Waveforms



PC Loading Vectors



Single PC Reconstructions





Outros métodos

- Árvores de decisões / Random forest
- Naive Bayes classification
- Independent component analysis
- Redes neurais
- Deep learning



Limitações

- Necessidade de grandes número de dados
- Qualidade dos dados
- Capacidade computacional
- Interpretabilidade