

# Usando algoritmos genéticos como extrator de características no campo da bioinformática

Calebe Elias Ribeiro Brim

---

# Quem sou eu?

---

Calebe Elias Ribeiro Brim

- Engenheiro de dados - HunterCo (Presente)
- Mestre em Bioinformática - UFPR (2015)
- Bacharel em Sistemas de informação - UNIVILLE (2013)

Áreas de Interesse:

Programação; Inteligência artificial; Arduino; Arquitetura de software.

Aprendendo: Python

Usando: Nodejs

---

# Visão Geral



---

# Problemas - Biologia

---

## Genômica

- Sequenciamento de DNA
- Identificação de Genes
- Identificação de genes correlacionados

## Proteômica

- Identificação de proteínas
- Identificação dos genes que transcrevem aquela proteína
- Relacionamento entre proteínas (Folding)

Outros....

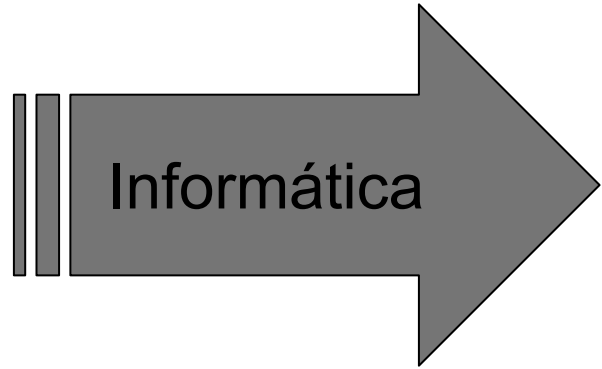
---

# Problemas - Informática

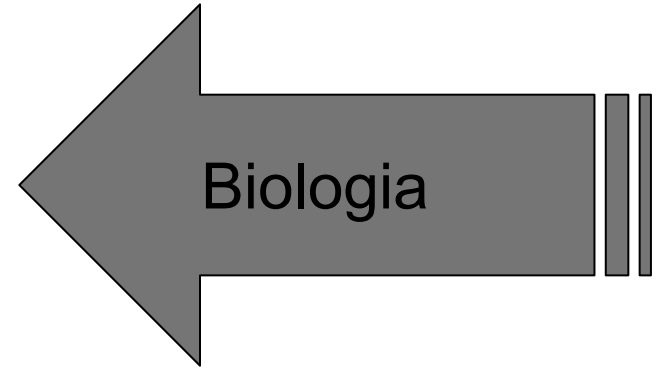
- 
- Velocidade na comunicação através de redes
  - Organização de infraestrutura
  - Engenharia de computadores
  - Arquitetura de software
  - Desenvolvimento de ferramentas
  - Inteligência artificial
  - ...

---

# Bioinformática

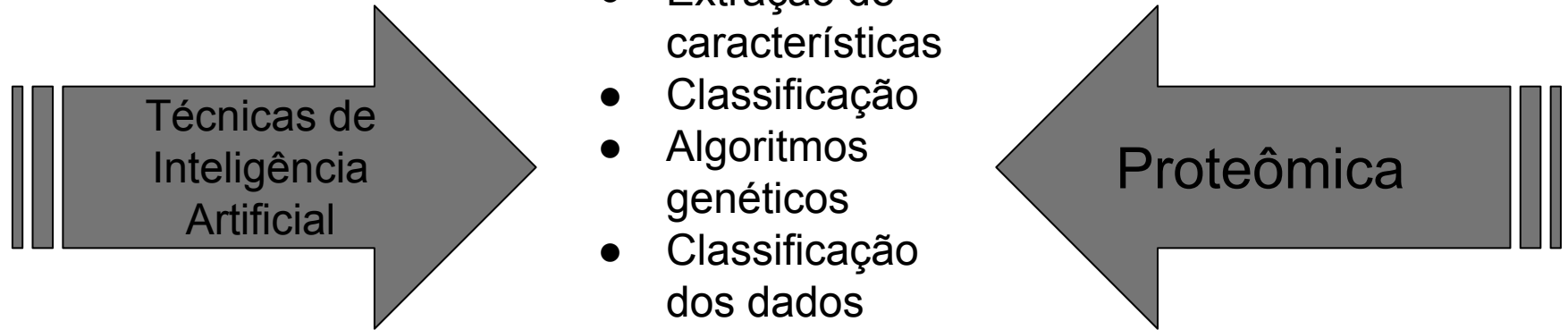


- Ferramentas para facilitação de pesquisa
- Aplicação de algoritmos
- Criação de métodos
- ...



---

# Aplicação - Classificação de Microorganismos



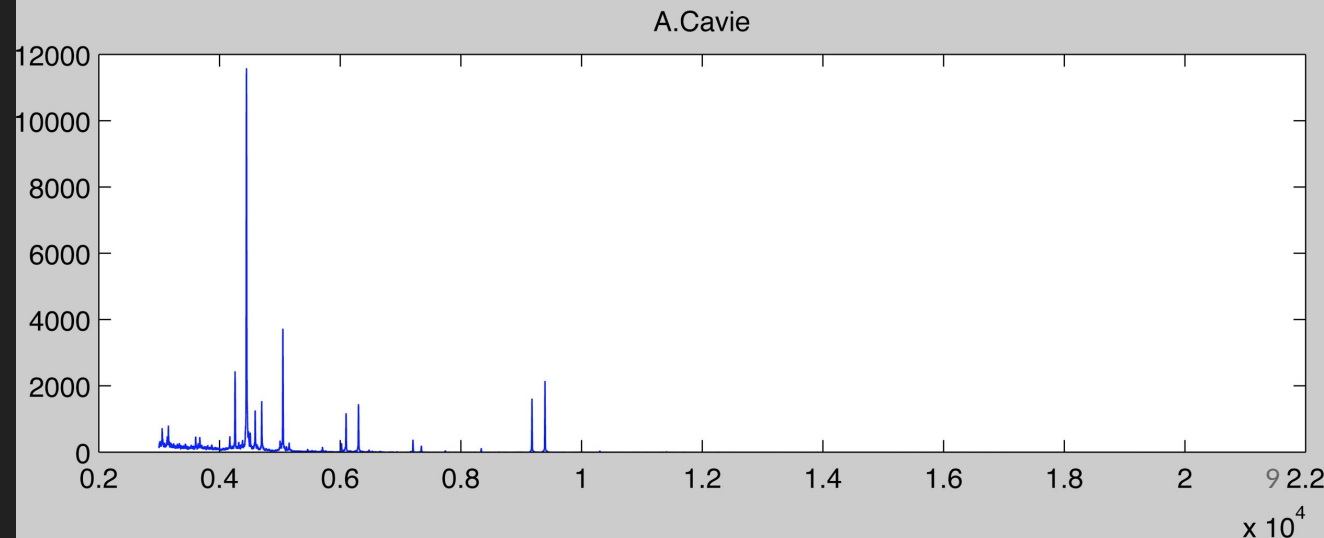
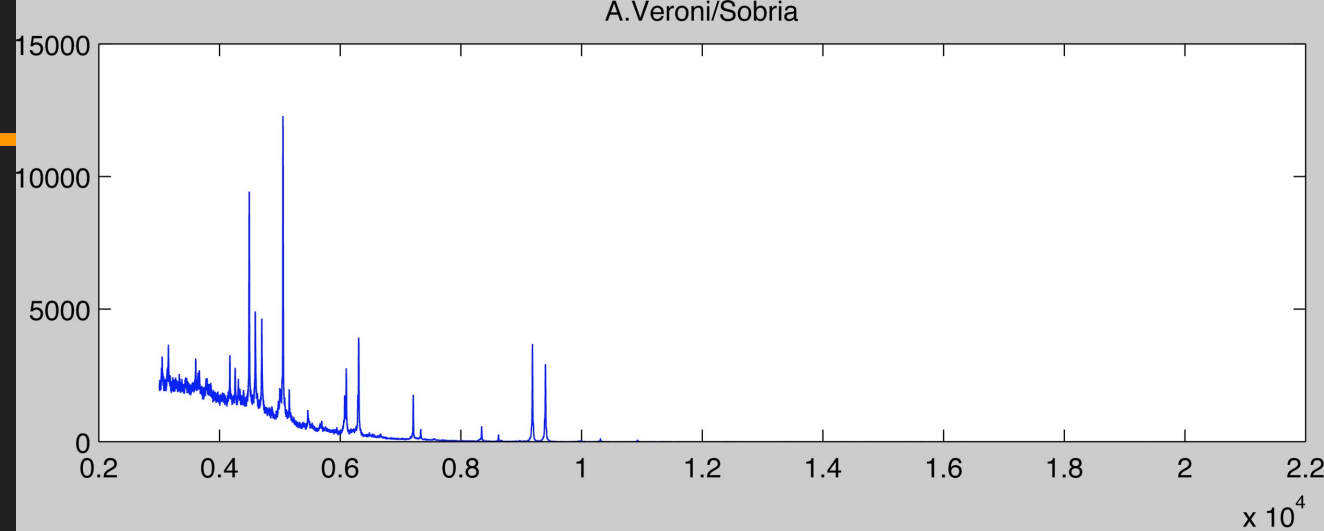
---

# Biologia - Entendendo o Problema

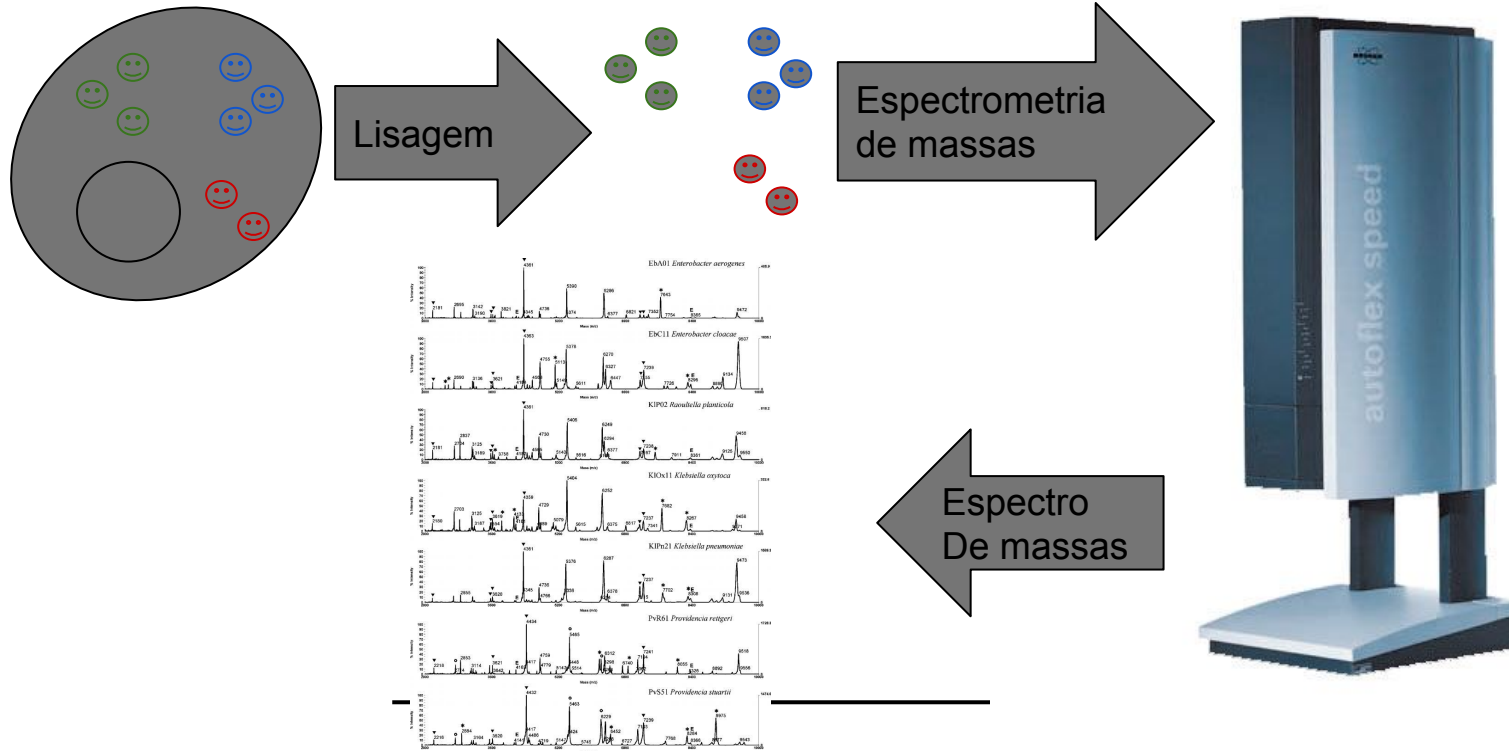
- 
1. O que são microorganismos?
  2. Patogênico?
  3. Qual a importância de estudar a patogenicidade de um microorganismo?
  4. Por que a biologia precisa da informática nessa área?
    - a. Precisão
    - b. Velocidade



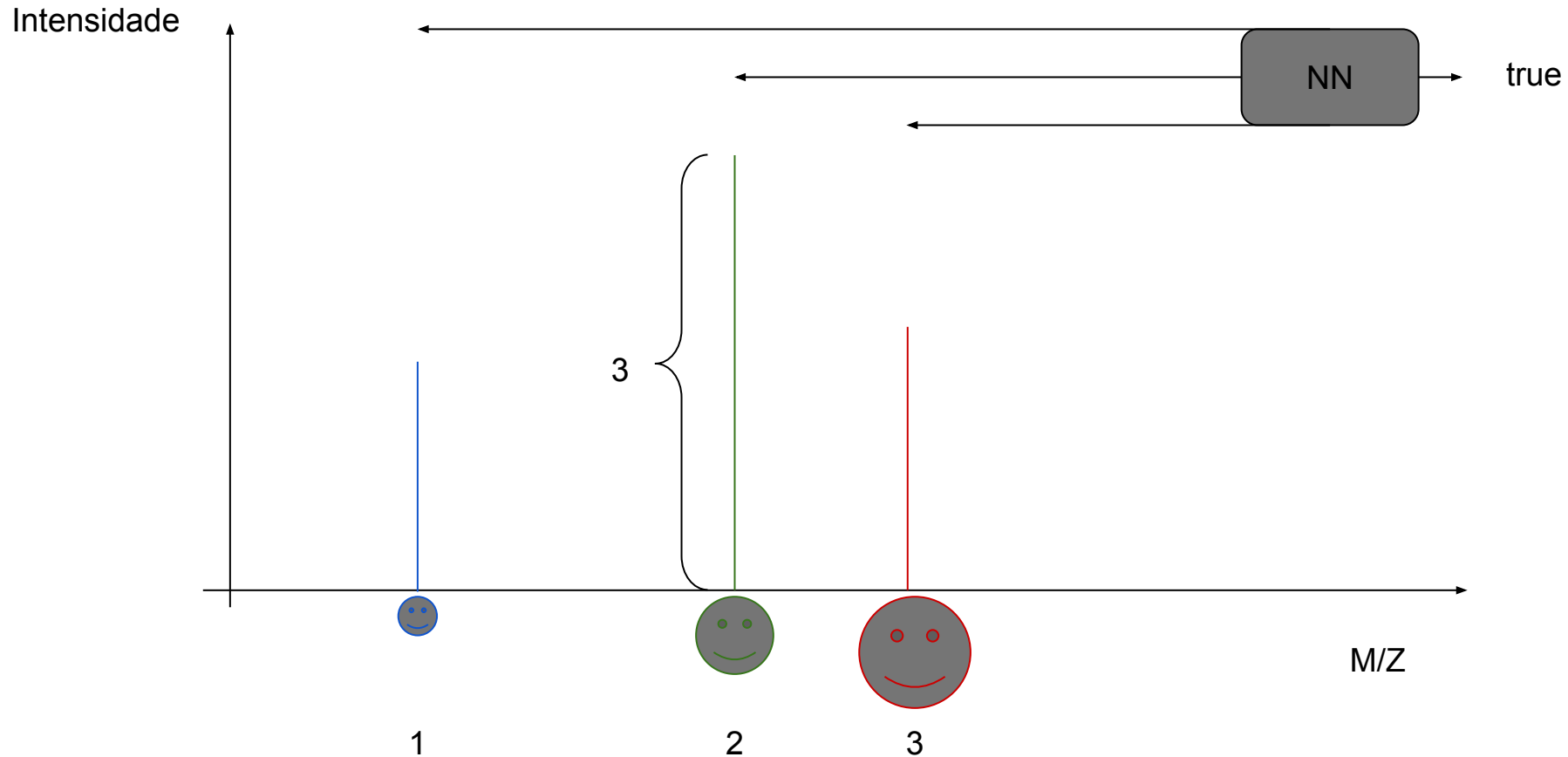
# Biologia - Espectro de massas



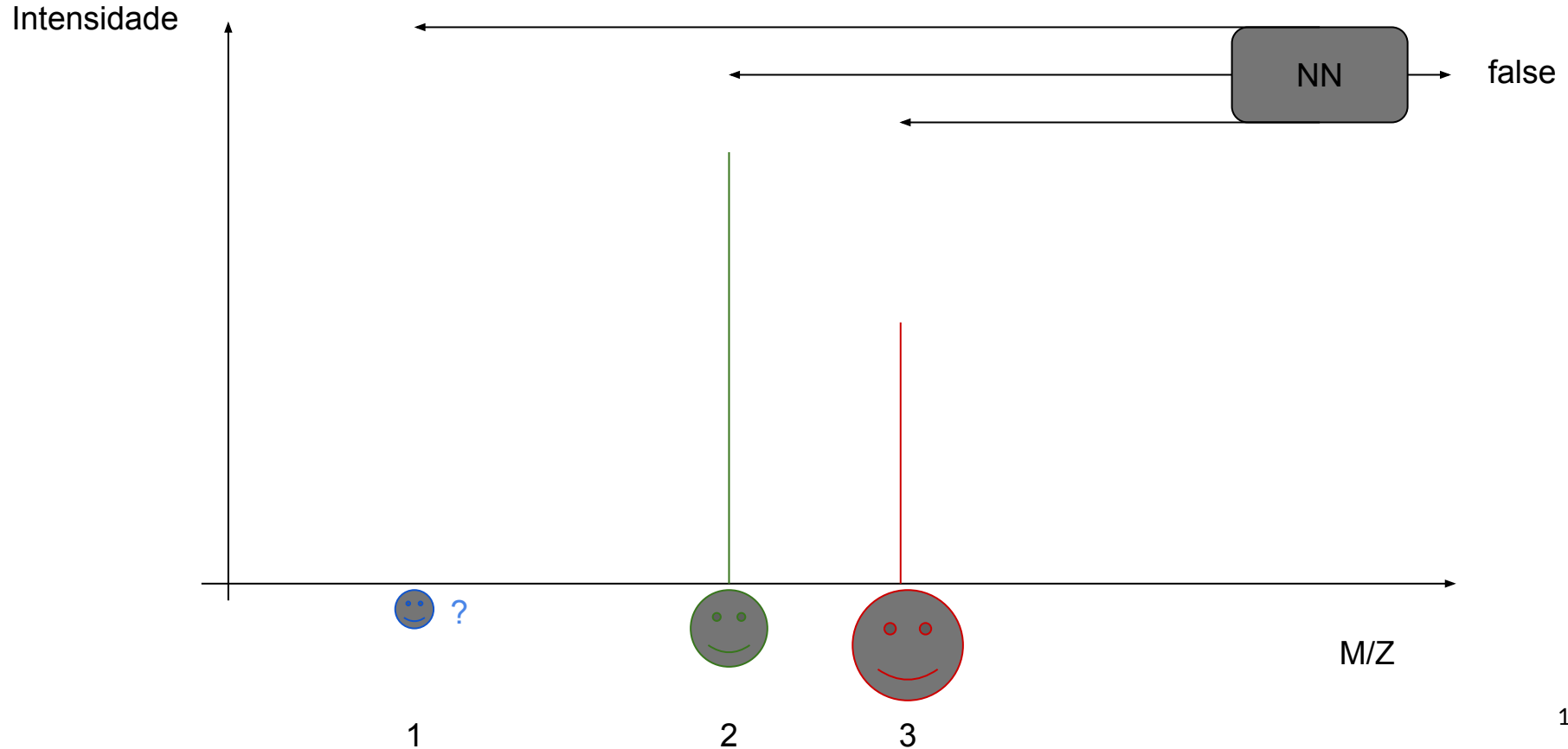
# Biologia - Processo Laboratorial



# Anatomia de um espectro de massas



# Anatomia de um espectro de massas



# Tópicos

## Apresentados

Perguntas ?

1. Bioinformática
2. Bactérias
3. Espectrometria de massas
4. Problemas
5. Precisão x Velocidade
6. Complexidade
7. Qualidade das Amostras



# Artificial Intelligence

---

# Artificial Intelligence

- 
- Feature Selection
  - Feature Extraction
  - Genetic Algorithm
  - Classifier

---

# Genetic Algorithms

---

## Conceitos:

- Algoritmos evolutivos
- Indivíduo
- Gene
- Maximização ou Minimização
- Óptimos Locais
- Multiobjetivo



# Indivíduo

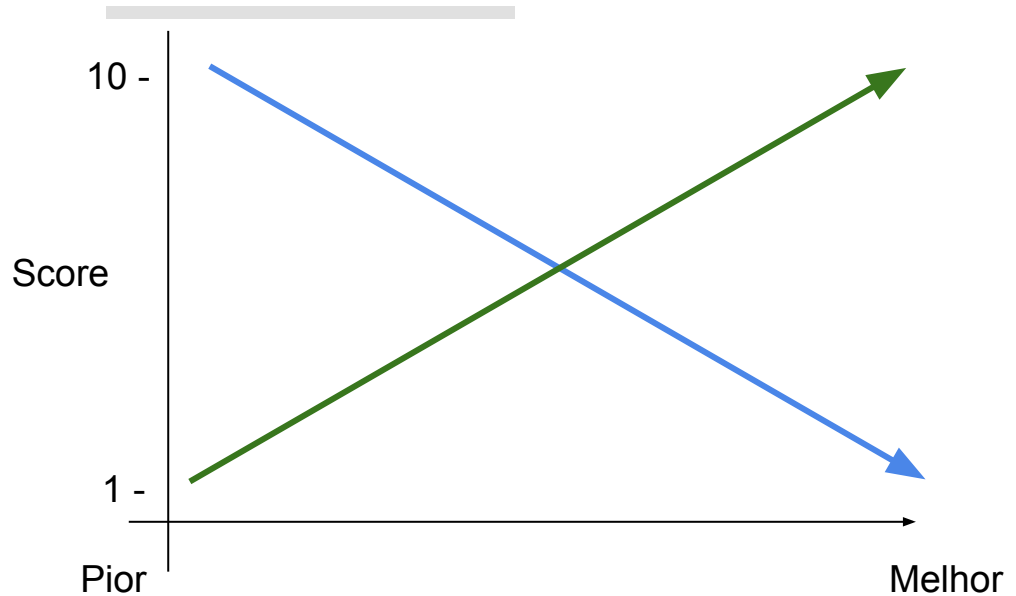
## População

| i   | g1 | g2 | g3 | g5 | g6 | g7 | g8 | g9 |
|-----|----|----|----|----|----|----|----|----|
| 1 - | 0  | 1  | 1  | 0  | 0  | 1  | 0  | 1  |
| 2 - | 1  | 0  | 1  | 0  | 0  | 0  | 1  | 1  |
| 3 - | 0  | 1  | 0  | 0  | 0  | 1  | 0  | 0  |

0 - Não expressa o gene

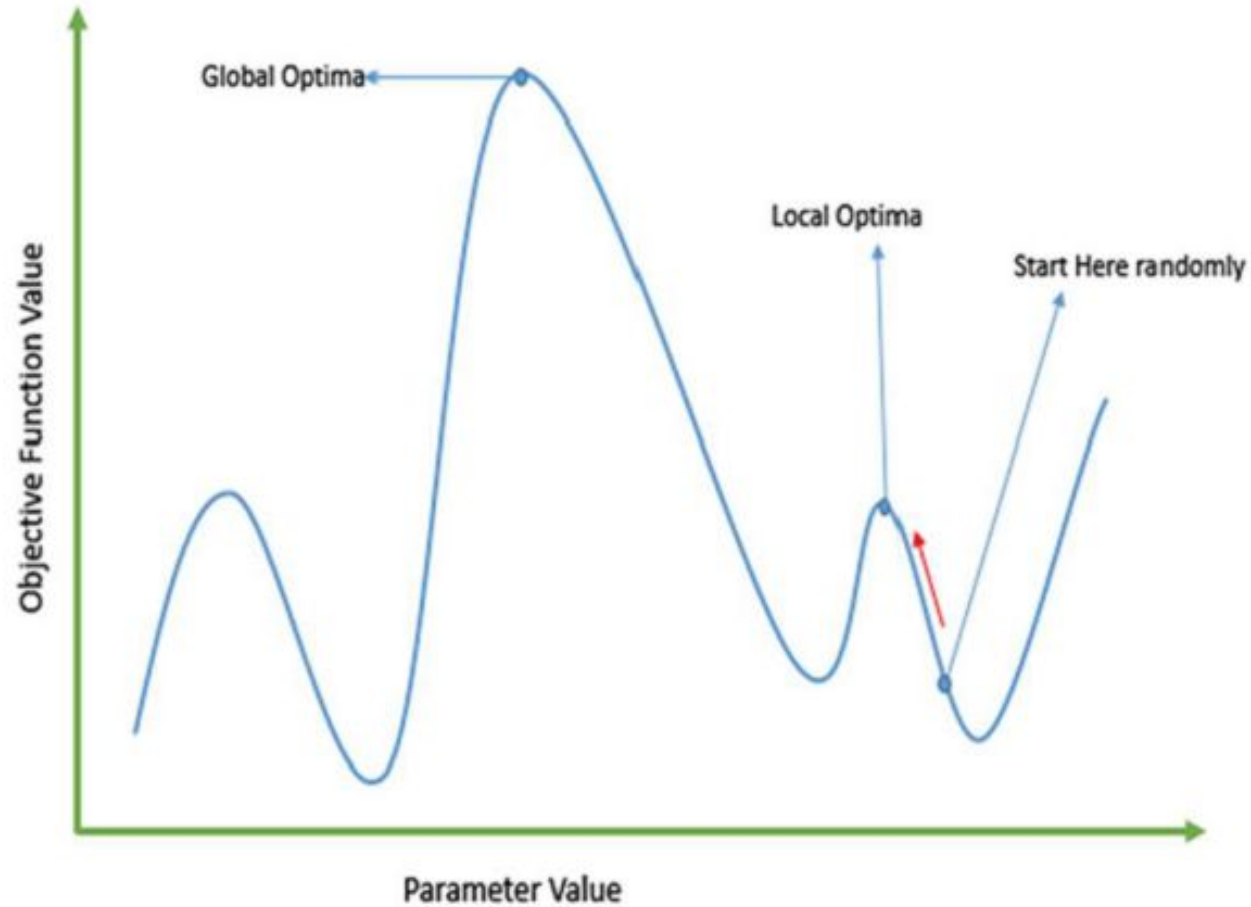
1 - Expressa o gene

# Maximização e Minimização










- Maximização
- Minimização

# Óptimos Locais e Globais



# Algoritmo Genético

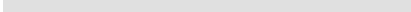
- Geração
- Fitness 
  - Seleção 
  - Cruzamento 
  - Mutação 
  - Seleção 
  - Critério de parada 
- Saída 

# Cruzamento


| i   | g1 | g2 | g3 | g5 | g6 | g7 | g8 | g9 |
|-----|----|----|----|----|----|----|----|----|
| 1 - | 0  | 1  | 1  | 0  | 0  | 1  | 0  | 1  |
| 2 - | 1  | 0  | 1  | 0  | 0  | 0  | 1  | 1  |
| 3 - | 0  | 1  | 1  | 0  | 0  | 0  | 1  | 1  |
| 4 - | 1  | 0  | 1  | 0  | 0  | 1  | 0  | 1  |

Corte

# Mutação



| i   | g1 | g2 | g3 | g5 | g6 | g7 | g8 | g9 |
|-----|----|----|----|----|----|----|----|----|
| 1 - | 0  | 1  | 1  | 1  | 0  | 1  | 0  | 1  |




# Tópicos

---

## Apresentados

Perguntas ?

- 
1. Algoritmo Genético
  2. Indivíduo
  3. Gene
  4. Maximização ou Minimização
  5. Óptimos Locais/Globais
  6. Cruzamento
  7. Mutação
  8. Fitness
  9. Critério de Parada

# Aplicação



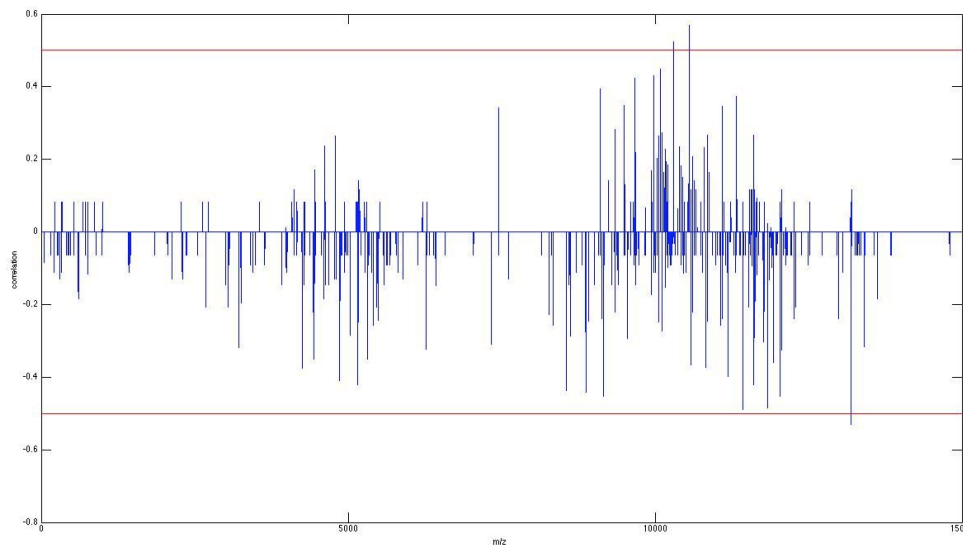
---

# Aplicação

- 
- Correlation Peaks Selection
  - Watch Points
  - Feature Resonance
  - Feature Combination

# Correlation Peaks Selection

- Peaks selection
- Pearson Correlation
  - Positive correlation
  - Negative correlation



# Watch Points

- Fuzzy triangle

For  $L$  odd:

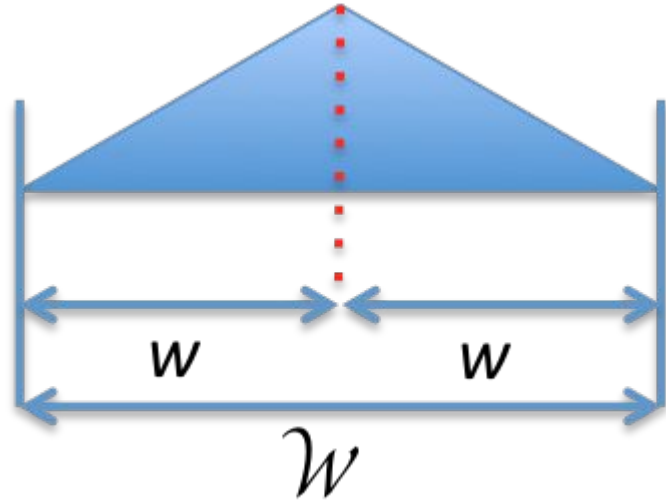
$$w(n) = \begin{cases} \frac{2n}{L+1} & 1 \leq n \leq (L+1)/2 \\ 2 - \frac{2n}{L+1} & (L+1)/2 + 1 \leq n \leq L \end{cases}$$

For  $L$  even:

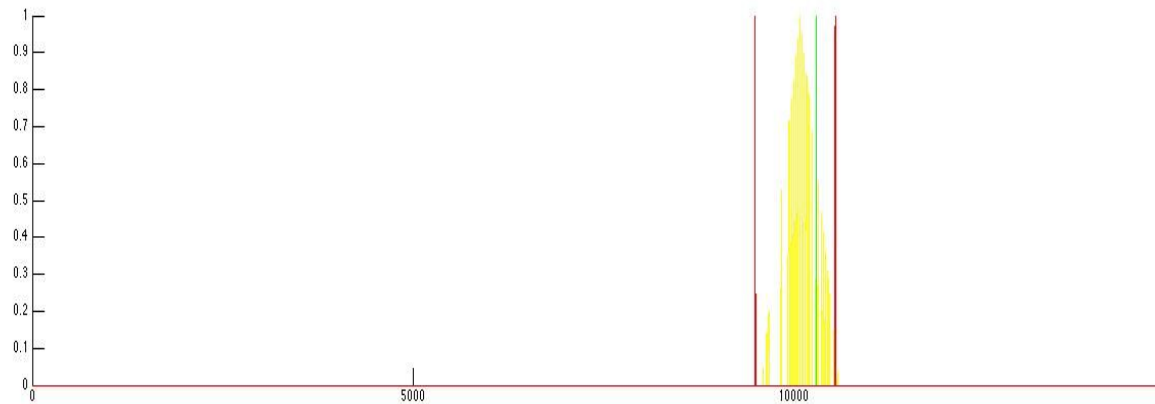
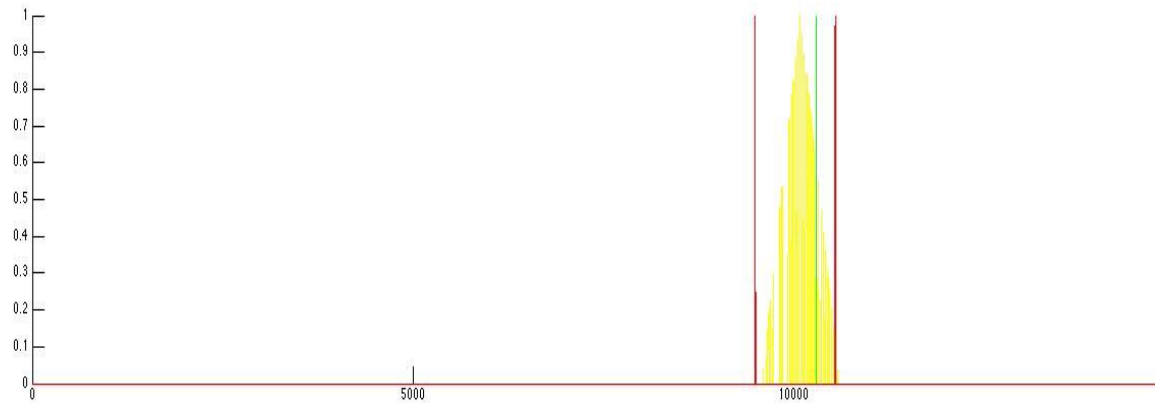
$$w(n) = \begin{cases} \frac{(2n-1)}{L} & 1 \leq n \leq L/2 \\ 2 - \frac{(2n-1)}{L} & L/2 + 1 \leq n \leq L \end{cases}$$

# Watch Points

- Fuzzy triangle



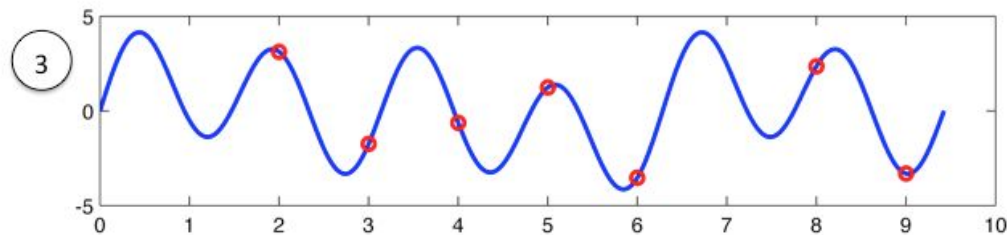
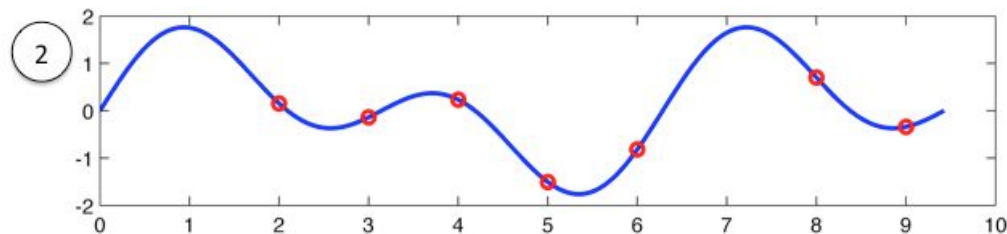
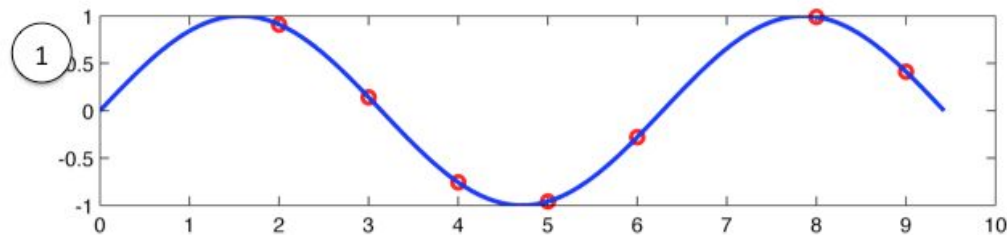
# Watch Points



# Feature Resonance

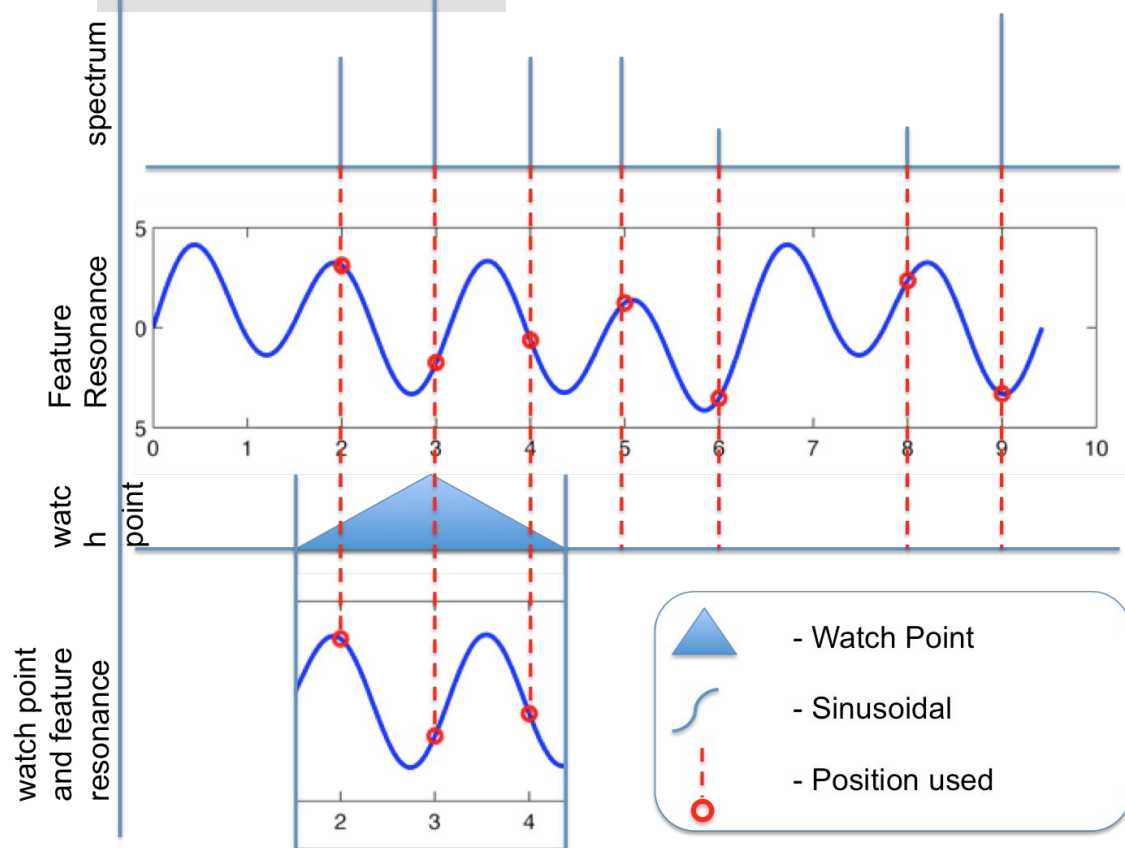
General Sine Formula:

$$y = A \sin(Bx + C) + D$$



# Feature Combination

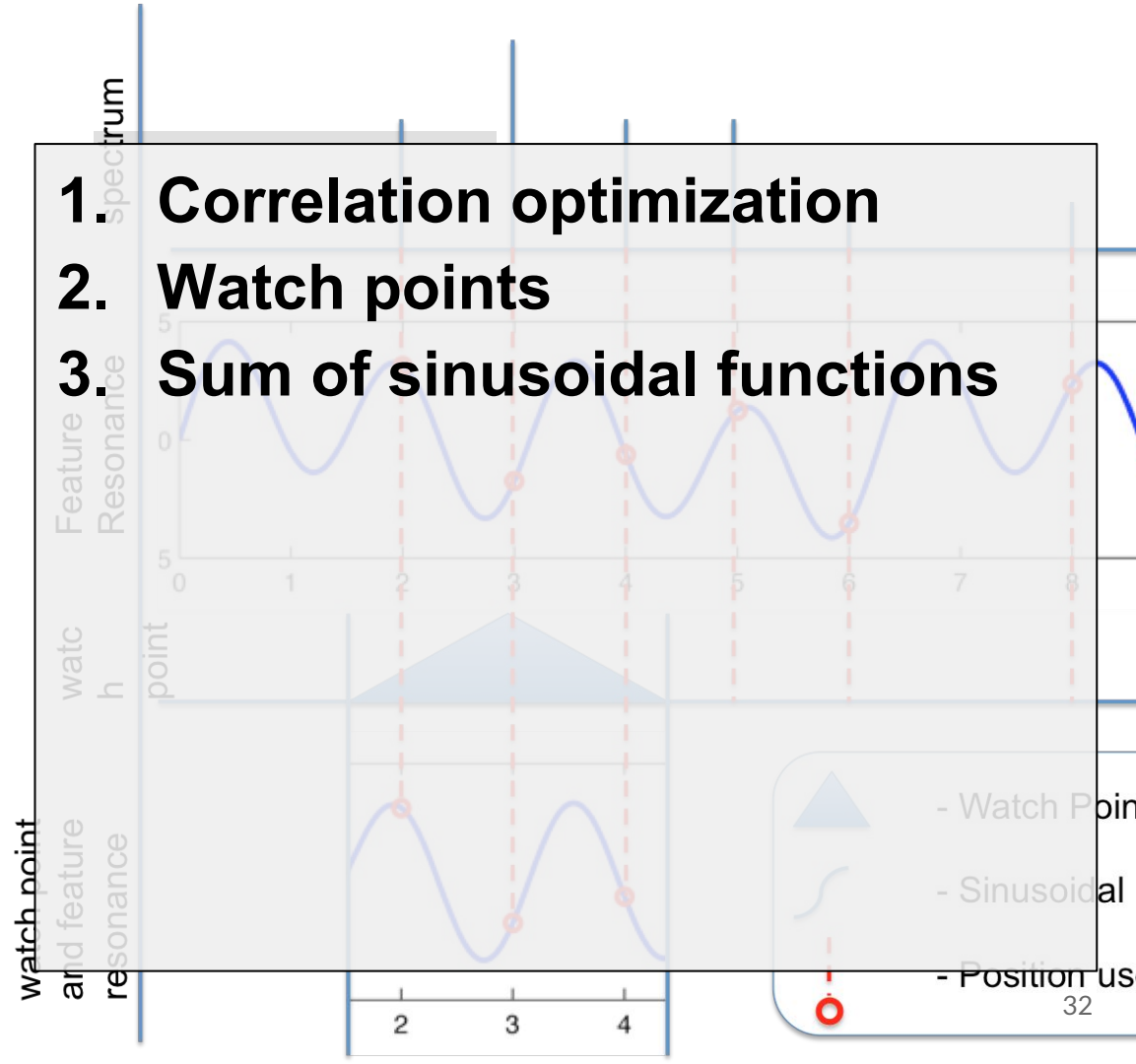
$$\sum_{i=1}^S \sum_{j=1}^X c_i * \sin(a_i + b_i * x_j)$$



# Tópicos

## Apresentados

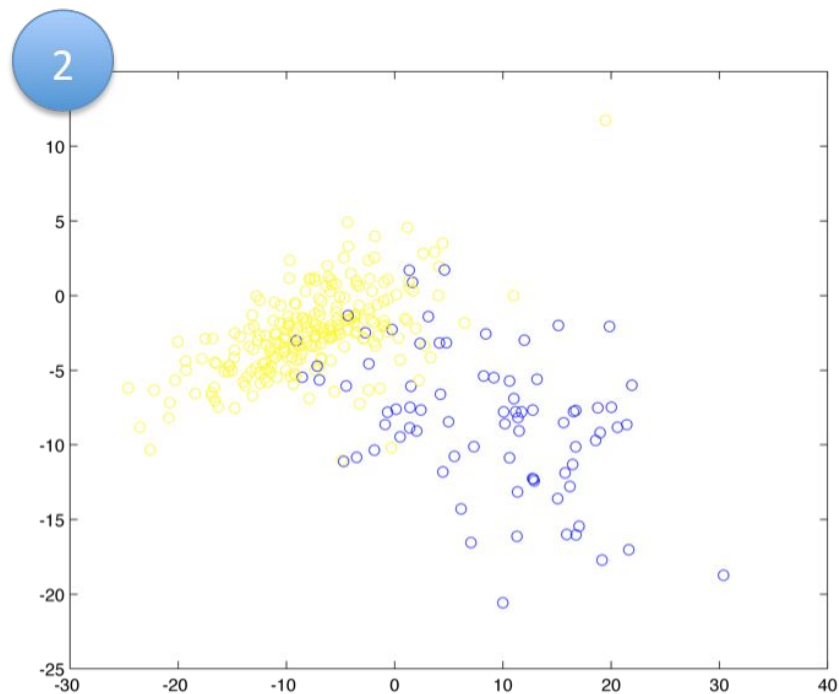
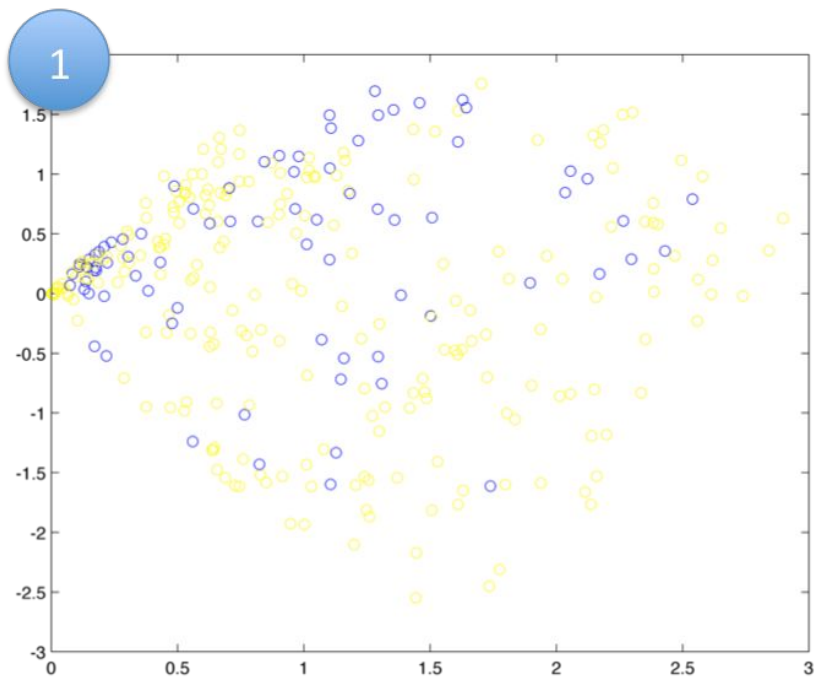
Perguntas ?



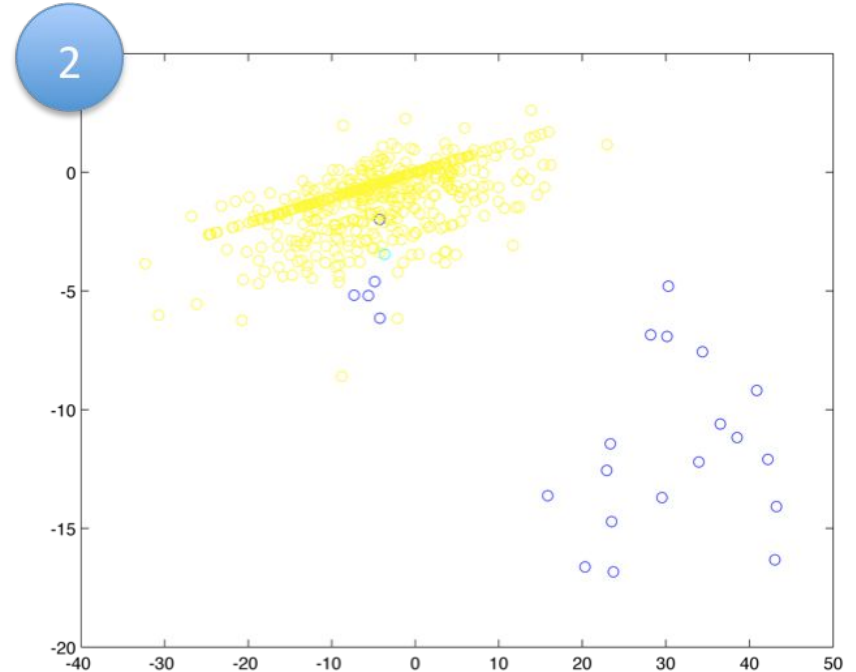
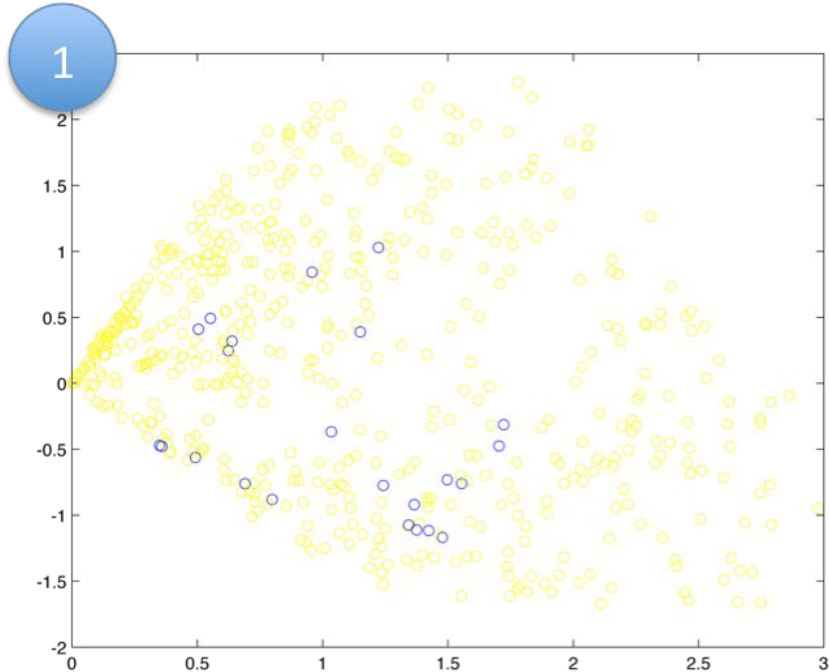


# Results

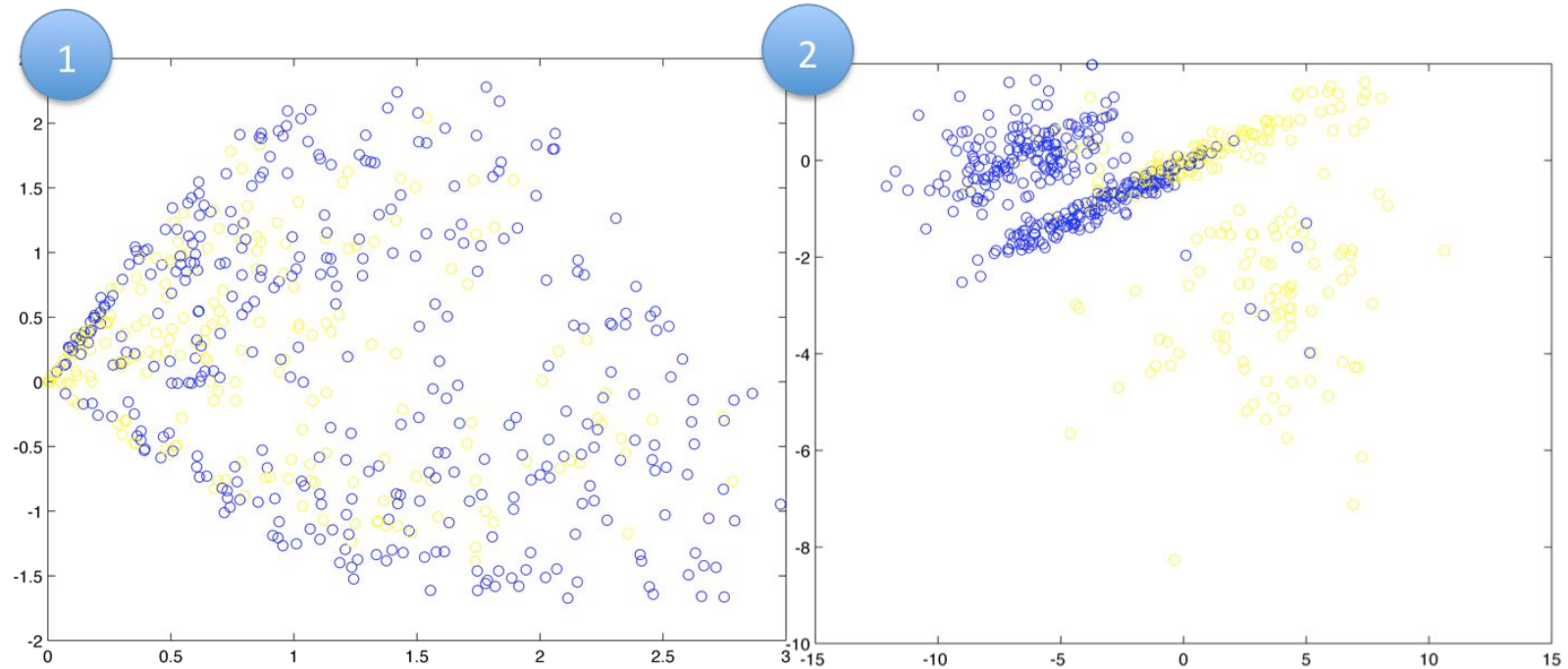
# PCA Veroni Sobria X Todas



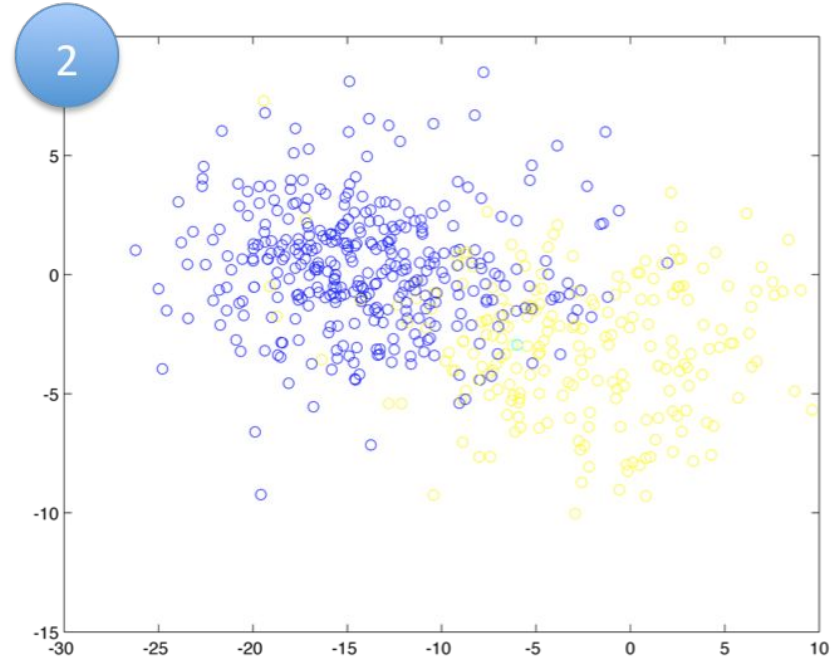
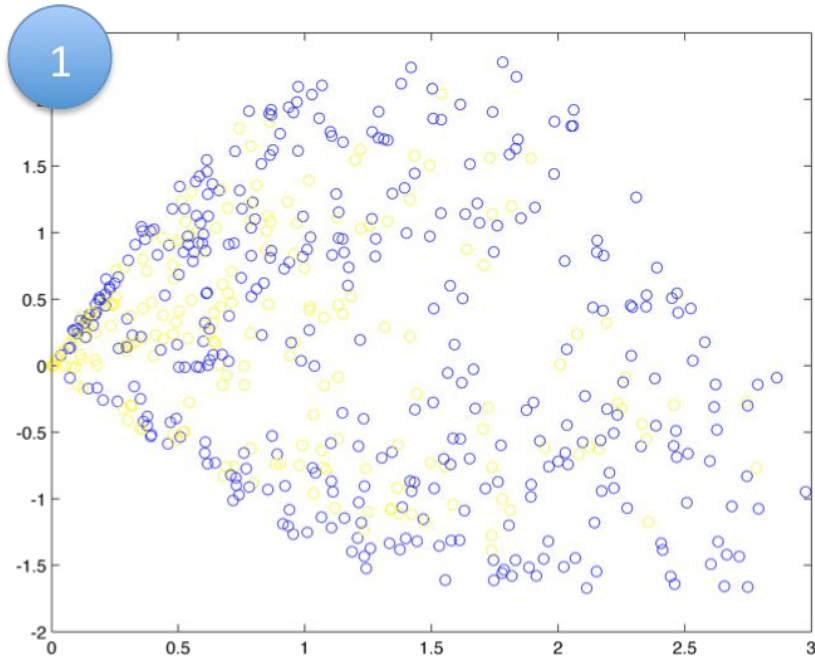
# PCA Trota x Todas



# PCA Hidrophila x Todas



# PCA Cavie x Todas



| Autors      | Experiment(%) | Specificity(%) | Sensibility(%) | PPV(%) | Year |
|-------------|---------------|----------------|----------------|--------|------|
| Petricoin   | 95.00         | 95.00          | 100.00         | 94.00  | 2002 |
| Conrads     | 100.00        | 100.00         | 100.00         | 100.00 | 2003 |
| ML-MS STDIO | 100.00        | 100.00         | 100.00         | 100.00 | 2014 |

Análise comparativa da literatura de um dataset público com dados de Seldi-TOF comprovando a eficácia do extrator de características aplicando as features à uma rede MLP.

**Obrigado!**