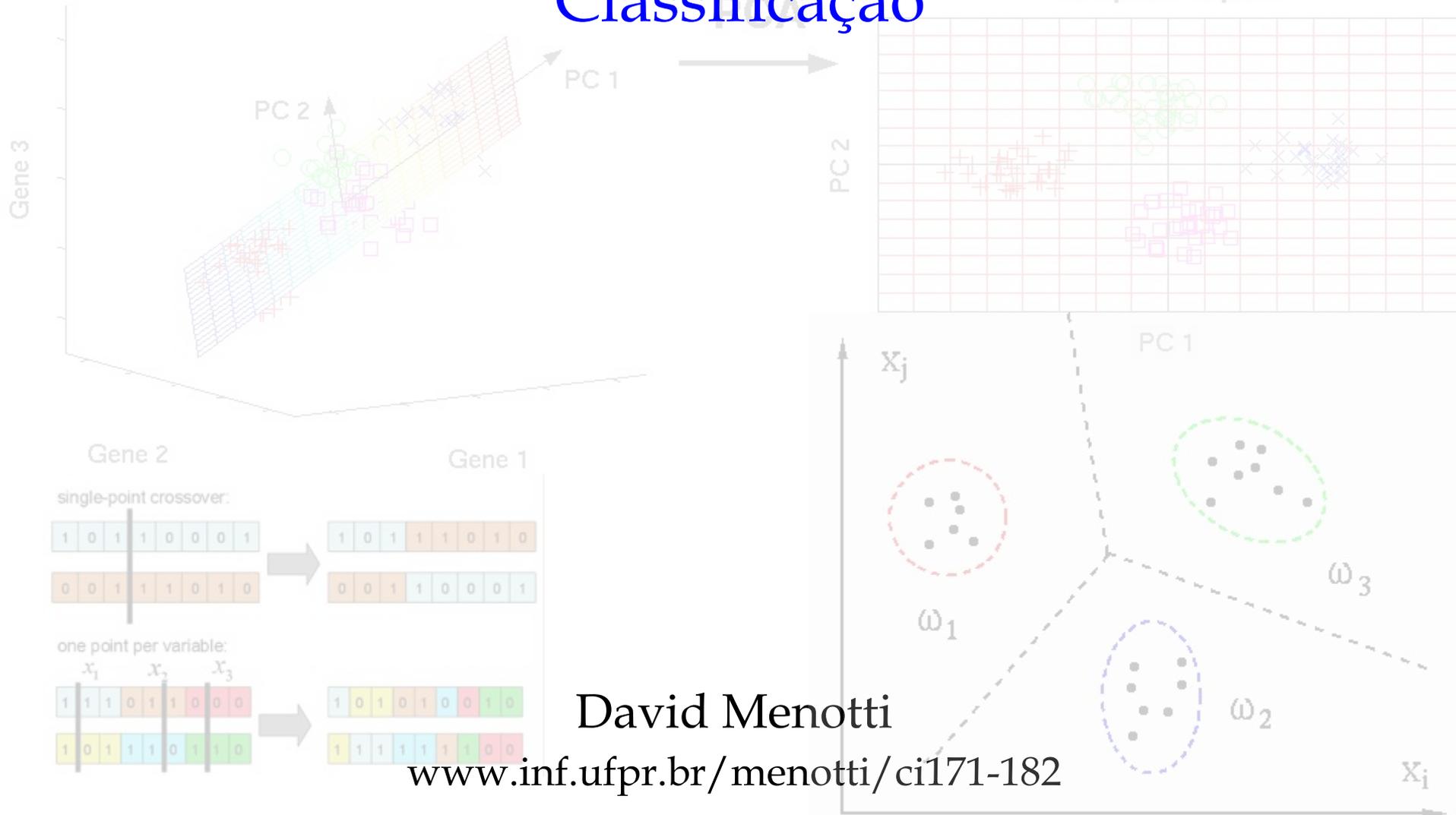


# Classificação

original data space

component space



David Menotti

[www.inf.ufpr.br/menotti/ci171-182](http://www.inf.ufpr.br/menotti/ci171-182)

# Hoje

- Seleção de Características

# Seleção de Características

# Introdução

- Um dos principais aspectos na construção de um bom classificador é a utilização de características discriminantes.
- Não é difícil encontrar situações nas quais centenas de características são utilizadas para alimentar um classificador.
- A adição de uma nova característica não significa necessariamente um bom classificador.
  - Depois de um certo ponto, adicionar novas características pode piorar o desempenho do classificador.

# Introdução

- Aspectos diretamente relacionados com a escolha das características:
  - Desempenho
  - Tempo de aprendizagem
  - Tamanho da base de dados.
- Seleção de características
  - Tarefa de identificar e selecionar um subconjunto de características relevantes para um determinado problema, a partir de um conjunto inicial
    - Características relevantes, correlacionadas, ou mesmo irrelevantes.

# Introdução

- Não é um problema trivial
  - Em problemas reais, características discriminantes não são conhecidas *a priori*.
  - Características raramente são totalmente independentes.
  - Duas características irrelevantes, quando unidas podem formar uma nova característica relevante e com bom poder de discriminação.

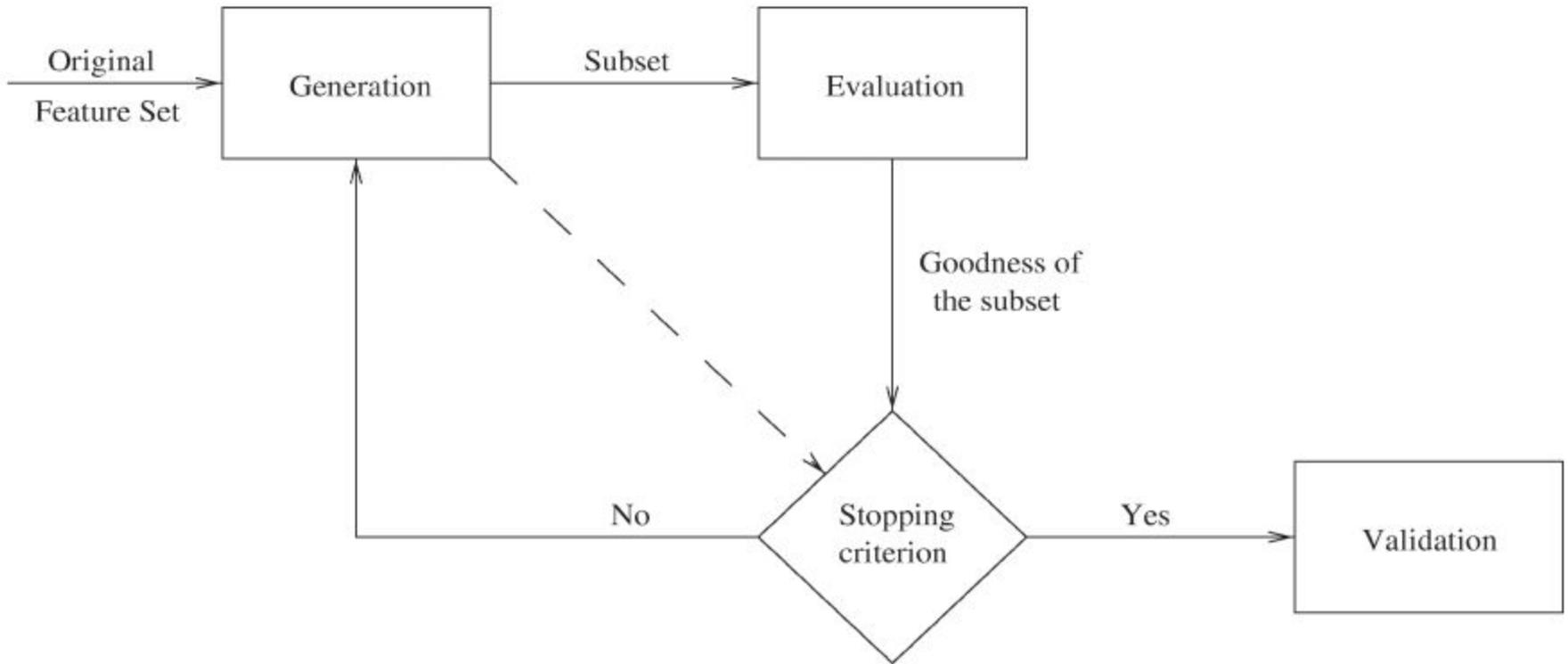
# Objetivos

- O objetivo é encontrar um subconjunto que pode ser
  - Ideal
    - O menor subconjunto necessário e suficiente para resolver um dado problema.
  - Clássico
    - Selecionar um subconjunto de  $M$  características a partir de  $N$  características, na qual  $M < N$ , de maneira a minimizar uma dada função objetivo.
  - Melhor desempenho
    - Buscar um subconjunto que melhore o desempenho de um dado classificador.

# Visão Geral

- Um método de seleção de características deve utilizar um método de busca para encontrar um subconjunto  $M$  a partir de  $N$  características
  - Espaço de busca é  $2^N$
- Para cada solução encontrada nessa busca, uma avaliação se faz necessária.
- Critério de parada
- Validação

# Visão Geral



Dash, Liu, 1997.

# Gerando Subconjuntos

- Existem diferentes abordagens que podem ser usadas para gerar os subconjuntos
  - Exaustiva
    - Exponencial
  - Heurística
    - Aproximada

# Busca Exaustiva

- Explora todas as possíveis combinações do espaço de busca ( $2^N$ )
- Garante que o subconjunto ótimo será encontrado.
- Muito caro computacionalmente
  - Inviável quando o espaço de busca é grande.

# Busca Heurística

- Com o objetivo de tornar o processo de busca mais rápido, vários algoritmos de busca foram propostos
  - *Hill climbing*
    - *Forward selection*
    - *Backward elimination*
    - Busca Flutuante
  - Computação Evolutiva
    - Algoritmos genéticos
    - PSO

# Funções de Avaliação

- Para julgar se um dado subconjunto é ótimo, temos que avaliar o mesmo.
- As funções de avaliação podem ser divididas em *filter* e *wrapper*.
  - *Filter*
    - Independentes do algoritmo de aprendizagem.
  - *Wrapper*
    - Dependente do algoritmo de aprendizagem.

# *Filter vs Wrapper*

- *Wrapper* geralmente produz os melhores resultados
  - Entretanto, os resultados podem não ser iguais se trocarmos o algoritmo de aprendizagem em questão
  - O tempo é uma questão crucial para métodos *wrapper*.
  - *Filter* não produz resultados tão bons, porém é a solução para grandes bases de dados.

# Medidas de Correlação

- Para abordagens *filter*, precisamos de medidas que forneçam a correlação entre as características
  - Entropia
  - Coeficiente de Correlação
  - *F-Score*

# PCA

- Uma ferramenta que pode ser utilizada tanto para extração como para seleção de características é a Análise de Componentes Principais (PCA)
  - A ideia é aplicar PCA na base de aprendizagem e encontrar os principais autovetores da base.
  - Abordagem *filter*, visto que o algoritmo de aprendizagem não é utilizado.

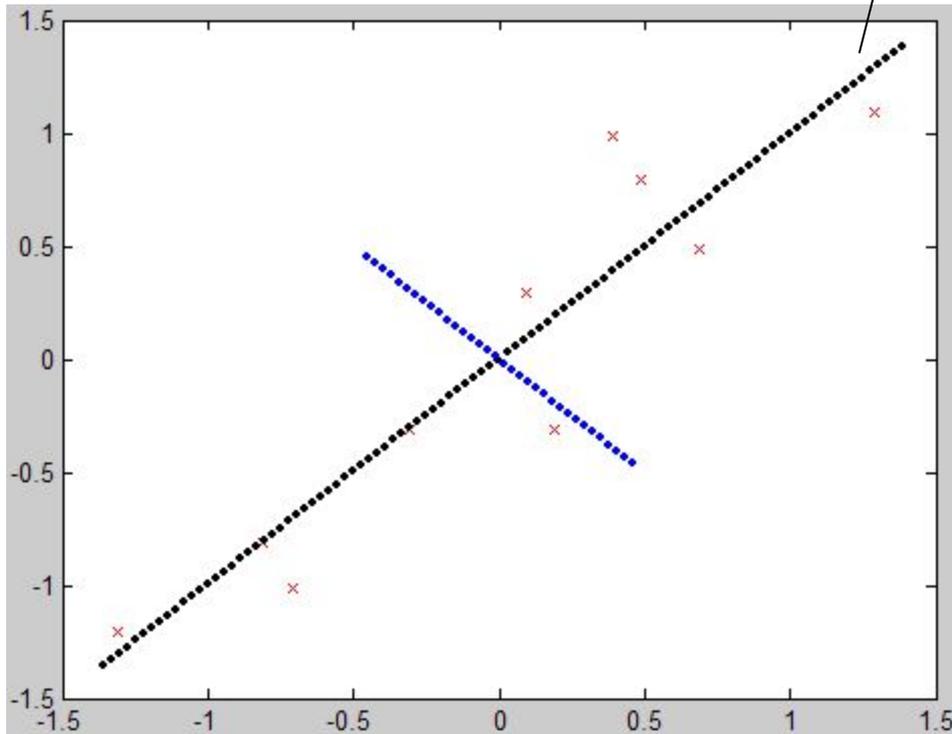
# PCA

- Note que após o PCA, os dados se encontram em um novo espaço de representação.
- Apesar de uma possível redução, todas as características devem continuar sendo extraídas / usadas.
- O custo da extração de características não é alterado
  - Somente o custo do algoritmo de aprendizagem.

# PCA

- Exemplo

$$y = \frac{-W_1 \times x - b}{W_2}$$



AUTOVETORES

-0.6779 -0.7352

-0.7352 0.6779

AUTOVALORES

1.2840

0.0491

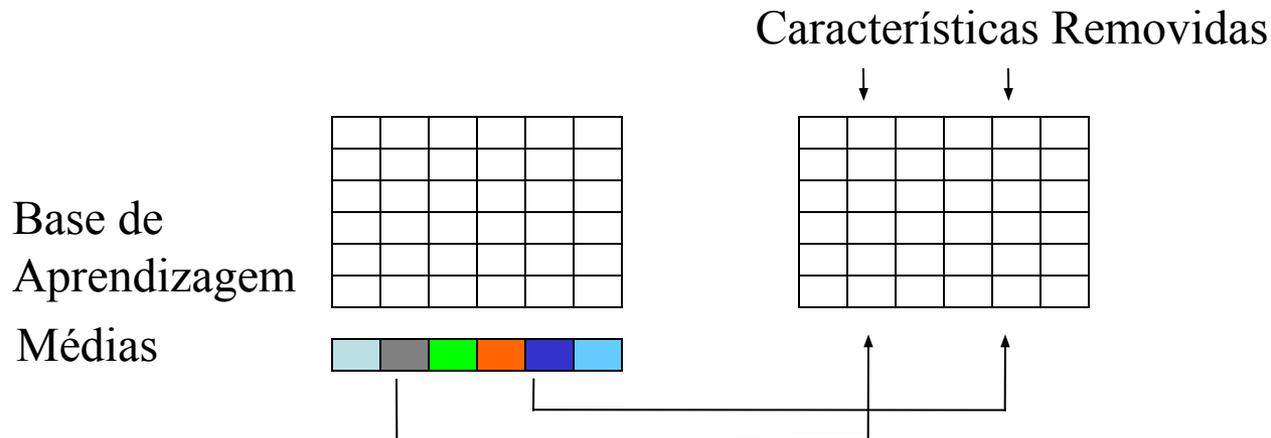
VARIÂNCIA EXPLICADA

96.3181

3.6819

# Análise de Sensibilidade

- Uma estratégia que pode ser utilizada para reduzir o custo computacional da abordagem *wrapper* é a Análise de Sensibilidade.
  - A idéia é trocar as características não selecionadas pelo algoritmo de seleção de características pela média calculada na base de aprendizagem.



# Seleção e Aprendizagem

- É importante ter em mente que o processo de seleção de características deve ser visto como um processo de aprendizagem
- Sendo assim, é importante utilizar uma base de validação para evitar *over-fitting*.
- Quando possível utilize uma base diferente de todas para calcular a **função de avaliação**

# Algoritmos Genéticos para Seleção de Características

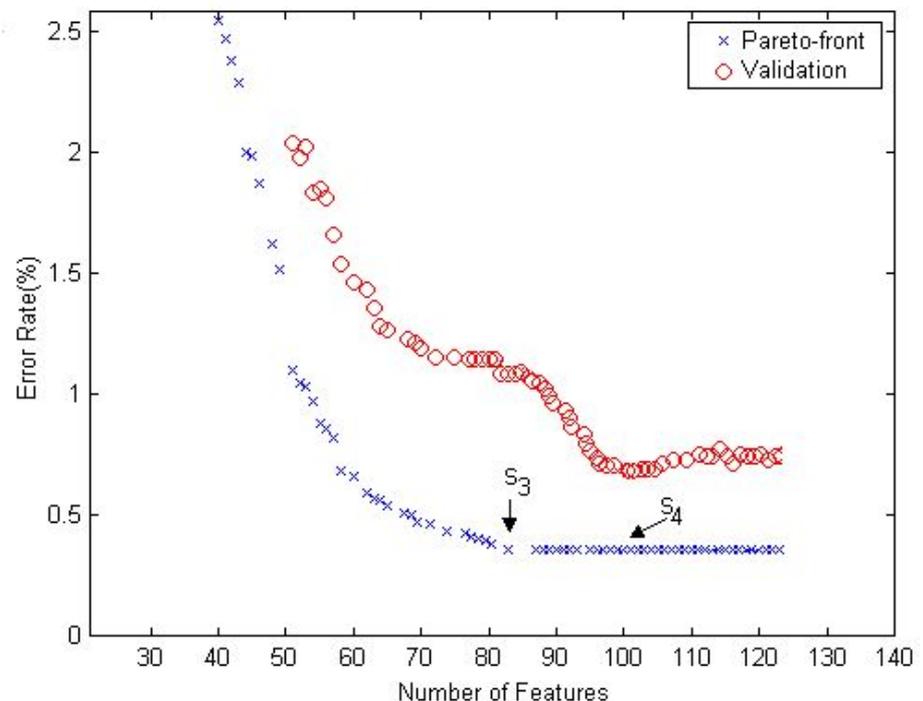
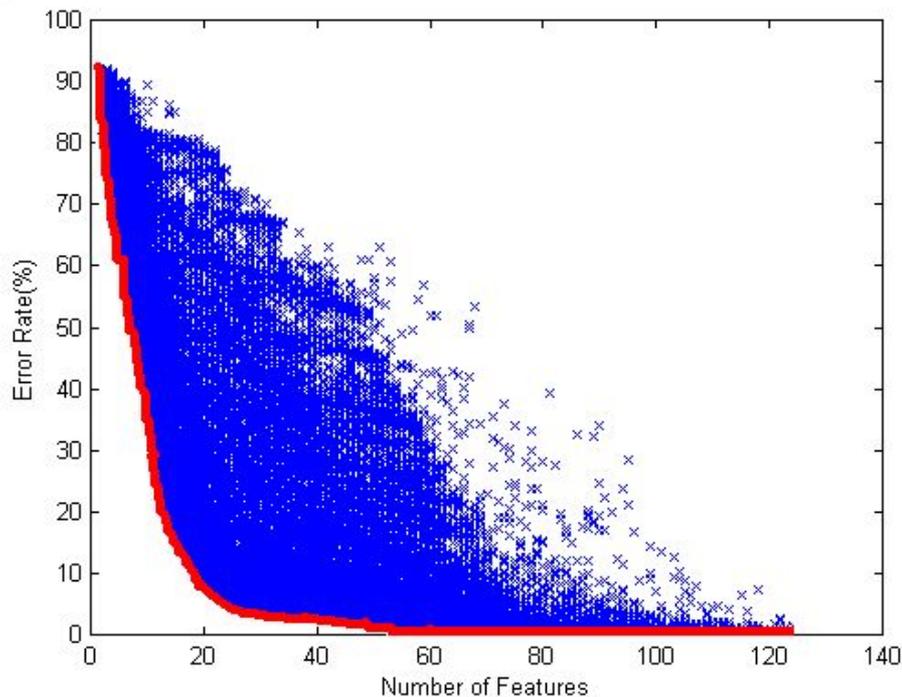
- Devido ao poder de explorar grandes espaços de busca, algoritmos genéticos tem sido largamente utilizados em problemas de seleção de características
  - Um objetivo  
(desempenho ou um índice qualquer)
  - Múltiplos objetivos  
(qtde de características, desempenho, etc..)

# Algoritmos Genéticos para Seleção de Características

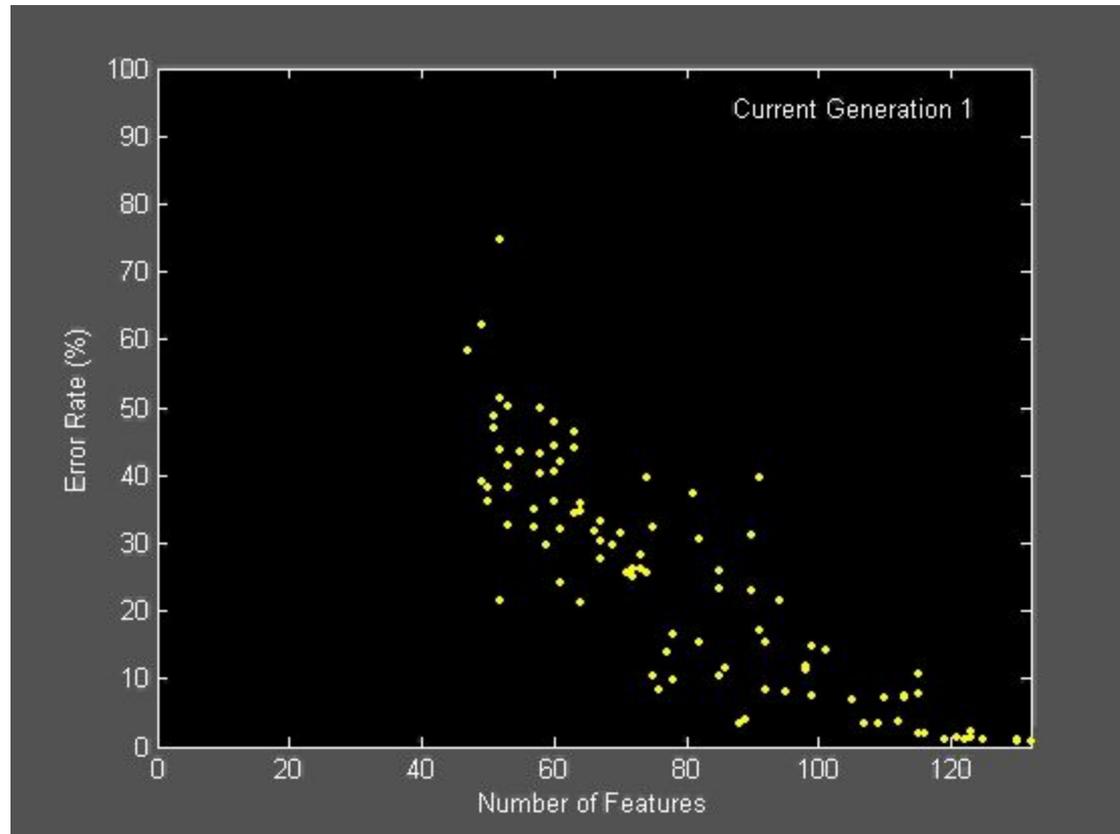
- Metodologia comumente utilizada
  - Bit representation,
    - cruzamento de um ponto,
    - mutação *bit-flip* e elitismo.
  - *Fitness*
    - Desempenho
    - Quantidade de características selecionadas

# Algoritmos Genéticos para Seleção de Características

- Abordagem baseada em Pareto
  - Converge para o Pareto
  - Uso de uma base de validação é importante.



# Algoritmos Genéticos para Seleção de Características





# Referências