

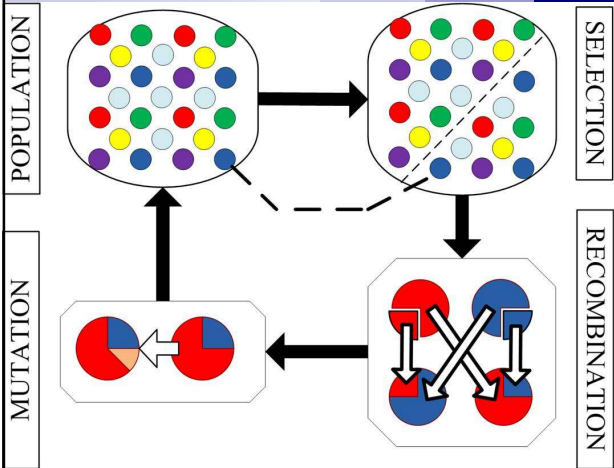


Técnicas de Aprendizado de Máquina EAD0759

Prof. Antonio Geraldo **Vidal**

vidal@usp.br

Sala G175



Introdução ao Algoritmo Genético

GA – *Genetic Algorithm*

Prof. Antonio Geraldo Vidal

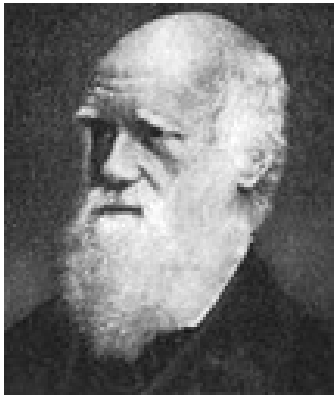


Algoritmo Genético - GA

- Algoritmos Genéticos (GA de *Genetic Algorithm*) aplicam uma abordagem evolutiva (seleção natural das espécies) para a aprendizagem de máquina indutiva.
- O GA tem sido aplicado com sucesso em problemas que são difíceis de resolver através de técnicas convencionais de otimização como:
 - Problemas de programação da produção;
 - Problemas de otimização de portfólio;
 - Problemas de logística (“caixeiro viajante”) ou TSP;
 - Problemas de roteamento em redes;
 - Problemas de classificação em finanças etc.



Inspiração do Algoritmo Genético



Charles
Darwin

*“As espécies evoluem pelo
princípio da seleção
natural e sobrevivência do
mais apto”, 1859*



Seleção Natural + Genética = Otimização

- As técnicas de otimização, geralmente, apresentam:
 - **Espaço de busca:** onde estão todas as possíveis soluções do problema. Nesse espaço de busca há um ponto ou conjunto de pontos que fornece a solução ótima;
 - **Função objetivo:** utilizada para avaliar as soluções produzidas, associando a cada uma delas uma nota (que representa sua aptidão).
- A teoria da evolução começou a partir da conceituação integrada da seleção natural com a genética.
- O GA é um subconjunto de um ramo muito maior da computação, conhecido como **Computação Evolutiva**.



GA - Características Gerais

- Utilizam uma codificação do conjunto de observações (indivíduos) e não os próprios dados (atributos ou variáveis);
- Vasculham várias regiões do espaço de busca de cada vez;
- Utilizam informações diretas de qualidade, em contraste com as derivadas utilizadas nos métodos tradicionais de otimização;
- Utilizam regras de transição probabilísticas e não regras determinísticas;
- Podem ser considerados como métodos que trabalham com buscas paralelas randômicas direcionadas.
- Utilizam apenas informações de custo e recompensa, não requerendo nenhuma outra informação auxiliar (como por exemplo o gradiente).
- Otimizam funções contínuas e discretas e também problemas multiobjetivo.



GA – Características Gerais

- São fáceis de serem implementados em computadores.
- Adaptam-se bem a computadores paralelos.
- São facilmente combinados com outras técnicas.
- Funcionam com parâmetros contínuos ou discretos.
- Consistem em tentar encontrar várias soluções e usar a informação obtida para conseguir soluções cada vez melhores.
- Manipulam uma população de indivíduos que são as possíveis soluções do problema.
- Os indivíduos são combinados (*crossover*) uns com os outros, produzindo gerações descendentes que podem sofrer ou não mutações.
- As populações evoluem através de sucessivas gerações até encontrar a solução ótima (população ideal).



GA - Limitações

- Não são adequados para todos os problemas, especialmente os que são simples e para os quais informações derivadas estão disponíveis.
- O valor de aptidão é calculado repetidamente, o que pode ser custoso do ponto de vista computacional para alguns problemas.
- Sendo estocástico (probabilístico), não há garantia sobre a otimização ou da qualidade da solução.
- Se não for implementado corretamente, o GA pode não convergir para a solução ideal.



GA – Aplicações

- Alocação de Tarefas.
- Configuração de Sistemas Complexos.
- Seleção de Rotas.
- Problemas de Otimização e de Aprendizagem de Máquina.
- Problemas cuja solução seja um estado final e não um caminho.
- São especialmente interessantes em problemas difíceis de otimizar de forma convencional (com técnicas heurísticas).
- Se uma técnica tradicional puder ser empregada, normalmente achará a melhor solução mais rápido, porém:
 - Existem muitos problemas práticos aos quais técnicas determinísticas tradicionais não podem ser aplicadas.
 - Técnicas tradicionais têm natureza serial.
 - Algoritmos Genéticos têm natureza paralela.



GA – Terminologia

■ Indivíduo

- ☐ Um simples membro da população.

■ Cromossomo e Genoma

- ☐ Coleção de genes que caracteriza os indivíduos.
- ☐ Estrutura de dados que codifica a solução de um problema.

■ Genótipo

- ☐ Na biologia, representa a composição genética contida no Genoma.
- ☐ No GA, representa a informação contida no cromossomo ou genoma.

■ Fenótipo

- ☐ Objeto ou estrutura construída a partir das informações do genótipo.
- ☐ É o cromossomo decodificado (apresentando os atributos reais).
- ☐ Se o cromossomo codifica as características de um edifício, então o fenótipo é o edifício construído.

■ Gene

- ☐ Codifica um simples parâmetro do problema.



Algoritmo de Aprendizado

■ Passo 1:

- Inicializar uma população **P** de **n** indivíduos como uma solução potencial.

■ Passo 2:

- Até que uma condição de término especificada for satisfeita:
 - **2a:** Use a função de seleção para avaliar cada indivíduo da solução atual. Se um indivíduo passa pelos critérios de seleção, ele permanece em **P**, em caso contrário não.
 - **2b:** A população agora contém **m** ($m \leq n$) indivíduos. Use operadores genéticos (**cruzamento & mutação**) para criar ($n - m$) novos indivíduos e adicioná-los à população.



Representação Genética Digital

- Uma técnica comum para a representação de conhecimento genético é transformar os atributos dos indivíduos em cadeias binárias.
- Por exemplo, podemos representar a faixa de renda como uma seqüência de dois bits para a atribuição de:
 - "00" a 20-30k,
 - "01" a 30-40k e,
 - "11" a 50-60k.
- Da mesma forma poderiam ser binariamente codificados os outros atributos.



Modelagem do GA

- Indivíduos X Atributos (estados);
- Cada indivíduo possui um código genético (conjunto de atributos);
- Esse código é chamado cromossomo (características do indivíduo);
- Tradicionalmente, um cromossomo é um vetor de bits representando os atributos reais;
- Vetor de bits nem sempre é o ideal. Exemplos:
 - Números reais, (2.345, 4.3454, 5.1, 3.4)
 - Cadeias de bits, (111011011)
 - Números de inteiros, (1,4,2,5,2,8)
 - ou outra estrutura de dados.



Operador Genético – *Selection*

Seleção

- Seleção é o princípio básico para o direcionamento da evolução de uma dada população.
- A seleção serve para substituir indivíduos a serem descartados (**mais fracos**) por indivíduos que passam no teste de seleção com altas pontuações (**mais fortes**).
- Os melhores indivíduos (maior aptidão) são selecionados para gerar descendentes (filhos) através de **cruzamento e mutação**.
- O processo de seleção garante que a qualidade geral da população aumenta a cada época de treinamento (evolução ou otimização).
- Existem vários métodos de seleção.



Principais Métodos de Seleção

- Roleta
- Torneio
- Proporcional à Aptidão
- Amostragem Universal Estocástica

| <i>Indivíduo</i> | <i>Aptidão Absoluta</i> | <i>Aptidão Relativa</i> |
|------------------|-------------------------|-------------------------|
| 1 | 2 | 0,052631579 |
| 2 | 4 | 0,105263158 |
| 3 | 5 | 0,131578947 |
| 4 | 9 | 0,236842105 |
| 5 | 18 | 0,473684211 |
| <i>Total</i> | 38 | 1 |



Método da Roleta

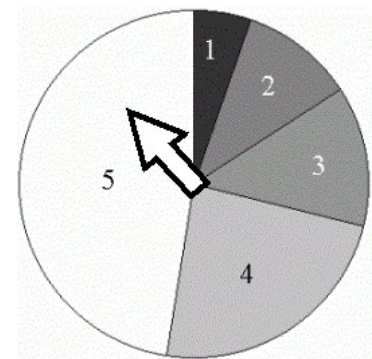
- Coloca-se os indivíduos em uma roleta, dando a cada um uma “fatia” proporcional à sua aptidão relativa.
- Depois roda-se a agulha da roleta. O indivíduo em cuja fatia a agulha parar permanece para a próxima geração.
- Repete-se o sorteio quantas vezes forem necessárias para selecionar a quantidade desejada de indivíduos.

Pop.
inicial

| cromossomos | x | f(x) | Prob. de seleção |
|----------------------------|----|------|------------------|
| A ₁ = 1 1 0 0 1 | 25 | 625 | 54,5% |
| A ₂ = 0 1 1 1 1 | 15 | 225 | 19,6% |
| A ₃ = 0 1 1 1 0 | 14 | 196 | 17,1% |
| A ₄ = 0 1 0 1 0 | 10 | 100 | 8,7% |

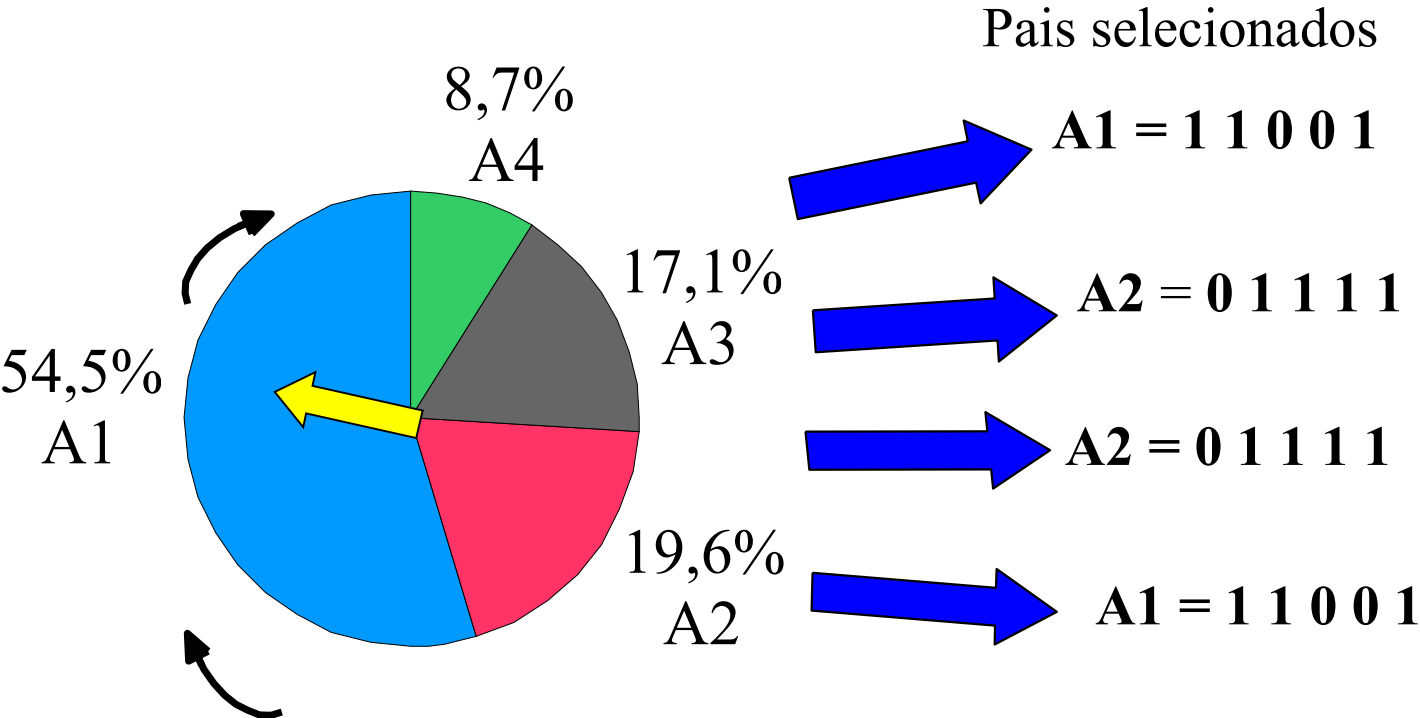
Probabilidade de seleção
proporcional a aptidão

$$p_i = \frac{f(x_i)}{\sum_{k=1}^N f(x_k)}$$





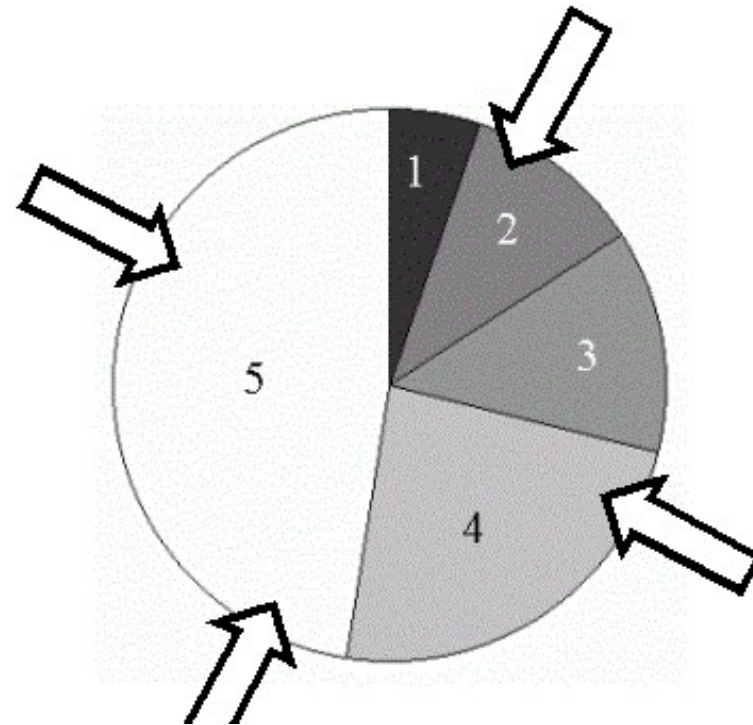
Método da Roleta





Amostragem Universal Estocástica

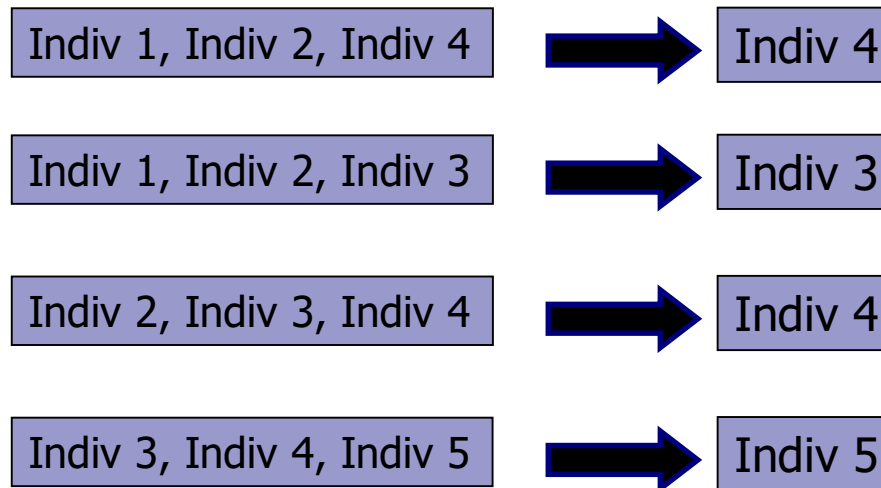
- *SUS – Stochastic Universal Sampling;*
- Semelhante à Roleta, mas para selecionar k indivíduos utiliza k agulhas igualmente espaçadas, girando-as em conjunto uma só vez;
- Apresenta resultados menos variantes que a Roleta.





Método do Torneio

- Utiliza sucessivas *disputas* para realizar a seleção;
- Para selecionar k indivíduos, realiza k disputas, cada disputa envolvendo n indivíduos escolhidos ao acaso;
- O indivíduo de maior aptidão na disputa é selecionado;
- É muito comum utilizar $n = 3$.
- Combinam pais selecionados para produção de filhos.

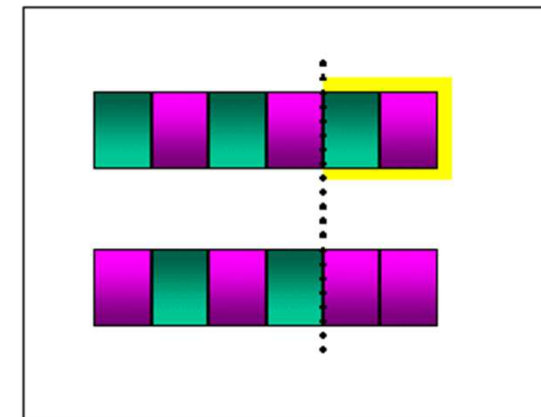




Operador Genético – *Crossover*

Cruzamento

- Os indivíduos mais frequentemente utilizados para o cruzamento são os que são destinados a ser eliminados da população.
- *Crossover* cria (pela reprodução) novos indivíduos para a população através da combinação de partes de dois indivíduos da população atual.
- Combina as informações genéticas de dois indivíduos (*pais*) para gerar novos indivíduos (*filhos*).
- Versões mais comuns criam sempre dois filhos para cada operação.



One-Point Crossover

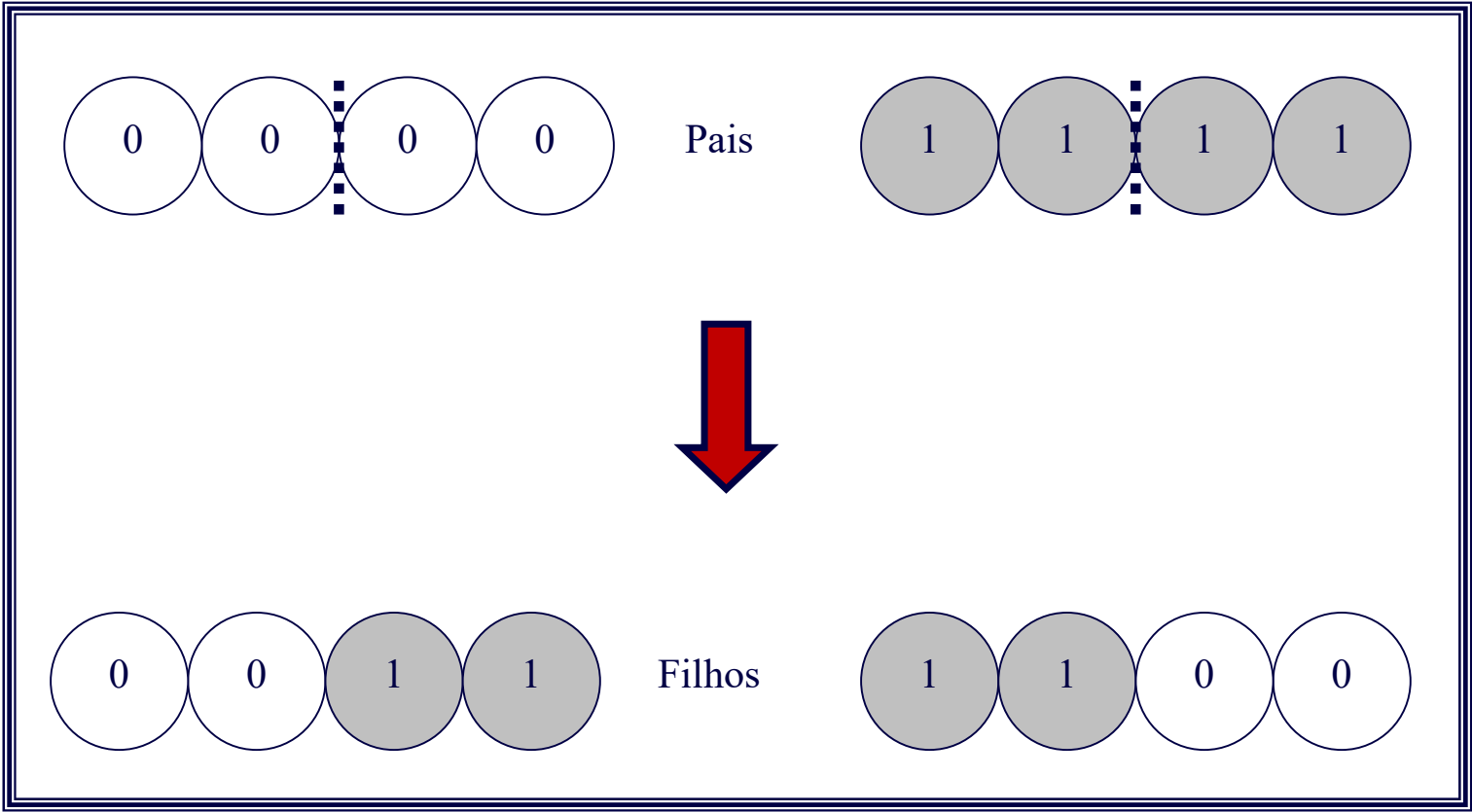


Operador de Cruzamento

- É o operador genético principal;
- Responsável por gerar novos indivíduos *diferentes* (sejam melhores ou piores) a partir de indivíduos existentes;
- Aplicado a cada par de indivíduos com alta probabilidade (normalmente entre 0,6 e 0,99).
- Abordagens para cruzamento:
 - ☐ Cruzamento Um-Ponto
 - ☐ Cruzamento Multi-Pontos
 - ☐ Cruzamento Uniforme

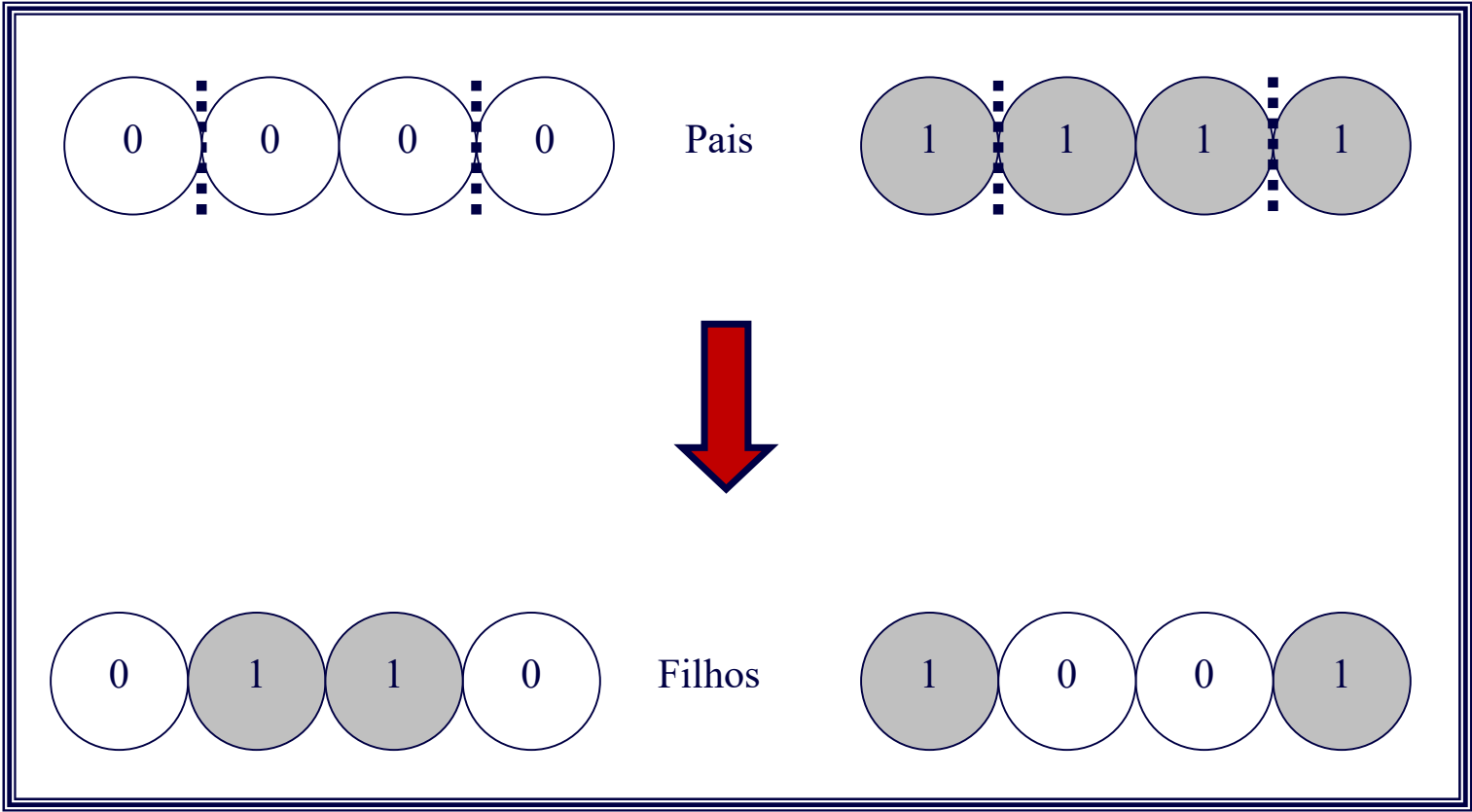


Cruzamento Um-Ponto



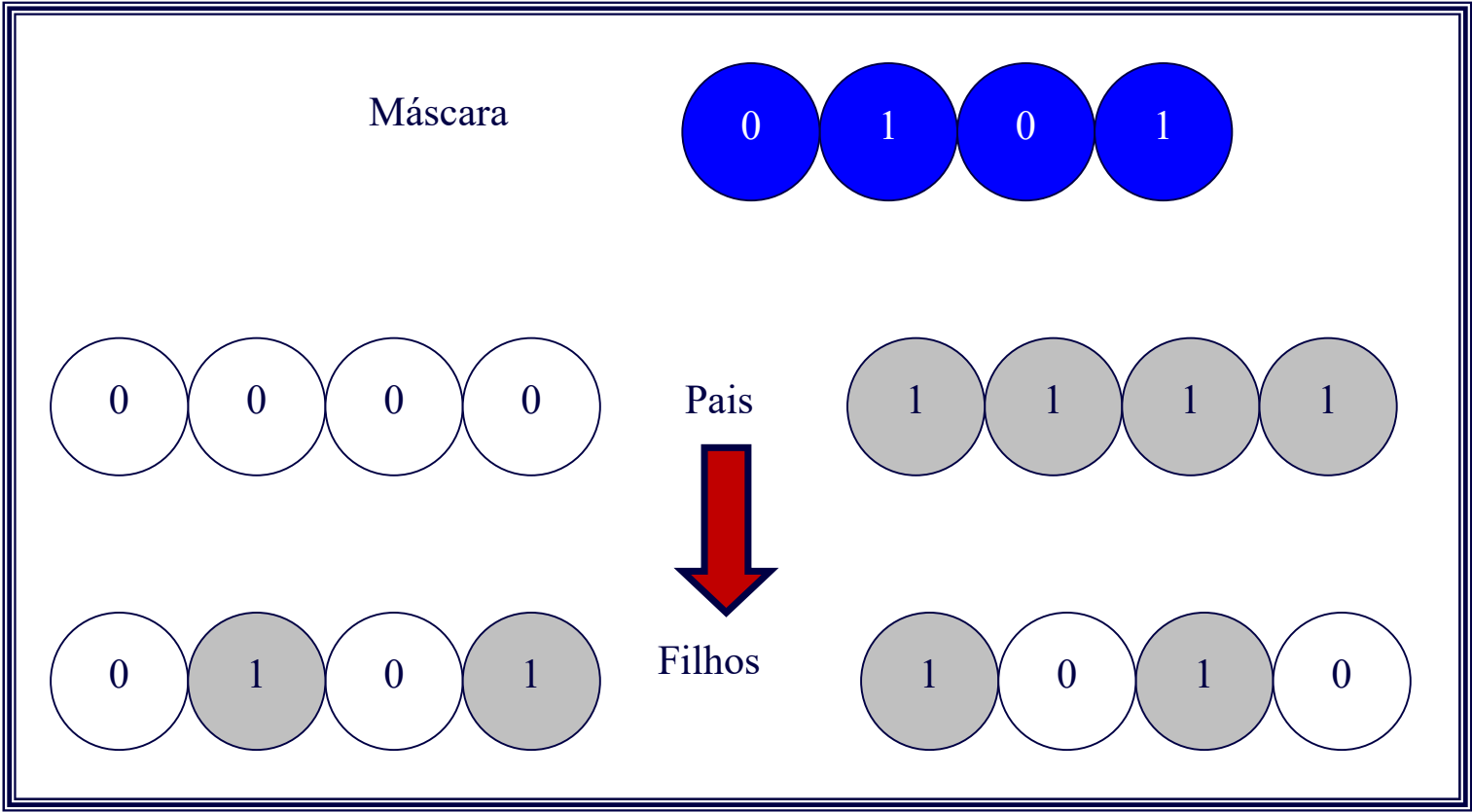


Cruzamento Multi-Ponto





Cruzamento Uniforme





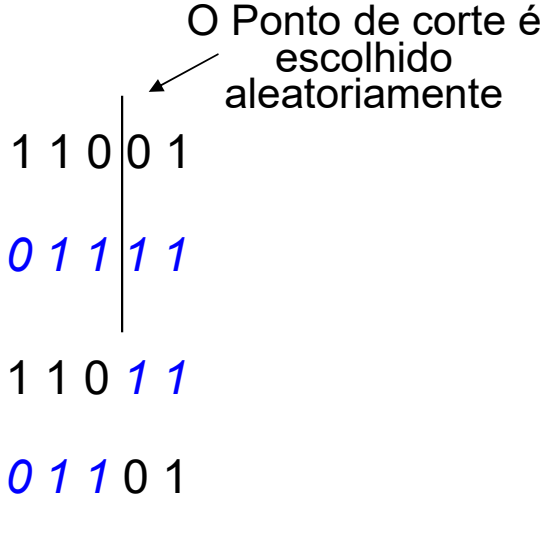
Operador de Cruzamento

O crossover é aplicado com uma dada probabilidade denominada *taxa de crossover* (60% a 90%)

Pais



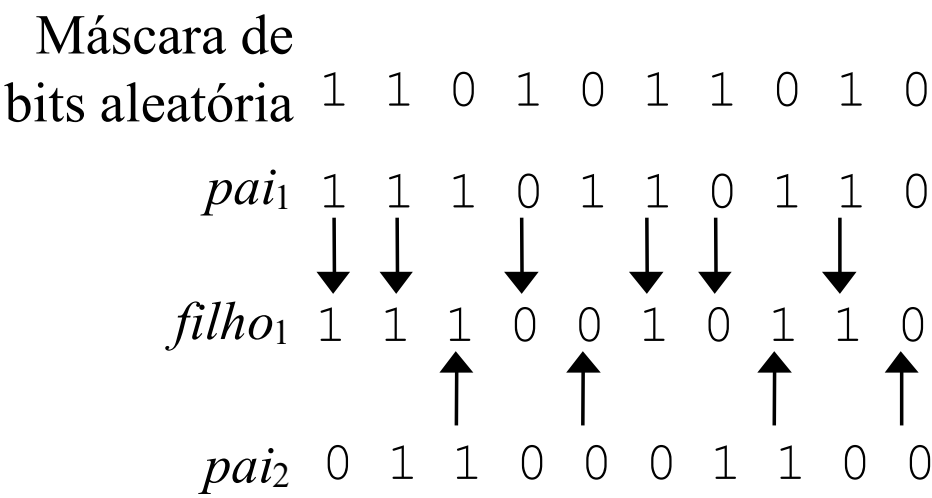
Filhos



Se o crossover for aplicado os pais trocam suas caldas gerando dois filhos, caso contrário os dois filhos serão cópias exatas dos pais.



Crossover Uniforme



- O *filho*₁ têm 50% de chance de adquirir um bit do *pai*₁ e 50% de chance de adquirir um bit do *pai*₂
- O *filho*₂ leva o que sobra do *pai*₁ e *pai*₂



Operador Genético – *Mutation*

Mutação

- Mutação é aplicada com moderação para elementos escolhidos para a eliminação.
- A mutação pode ser aplicada aleatoriamente trocando bits (ou valores de atributo) dentro de um único elemento.
- É um operador genético secundário;
- Operador randômico de manipulação;
- Introduz e mantém a variedade genética da população;
- Garante a possibilidade de se alcançar qualquer ponto do espaço de busca;
- Se seu uso for exagerado, reduz a evolução a uma busca totalmente aleatória;
- Logo um indivíduo sofre mutações com probabilidade baixa (normalmente entre 0,001 e 0,1).



Begin mutation



Operador de Mutação

Mutação inverte os valores dos bits.

A mutação é aplicada com dada probabilidade, denominada *taxa de mutação* (~1%), em cada um dos bits do cromossomo.

**Antes da
mutação** 0 1 1 0 1

Depois 0 0 1 0 1

Aqui, apenas o 2º bit
passou no teste de
probabilidade.

A taxa de mutação não deve ser nem muito alta nem muito baixa, mas o suficiente para assegurar a diversidade de cromossomos na população.



Parâmetros Genéticos do Algoritmo

- Tamanho da população
- Função de seleção ou objetivo
- Taxa de cruzamento
- Taxa de mutação
- Intervalo de geração
- Critério de parada



Critérios de Parada

- Número de gerações.
- Encontrou a solução (quando esta é conhecida).
- Perda de diversidade.
- Convergência
 - Nas últimas k gerações não houve melhora significativa da aptidão:
 - Média
 - Máxima




Algoritmo de Aprendizado

■ Passo 1:

- Inicializar uma população P de n indivíduos como uma solução potencial.

■ Passo 2:

- Até que uma condição de término especificada for satisfeita:
 - **2a:** Use a função de seleção para avaliar cada indivíduo da solução atual. Se um indivíduo passa pelos critérios de seleção, ele permanece em P , em caso contrário não.
 - **2b:** A população agora contém m ($m \leq n$) indivíduos. Use operadores genéticos (cruzamento & mutação) para criar $(n - m)$ novos indivíduos e adicioná-los na população.



Primeira Geração

A1 = 1 1 0 | 0 1

A2 = 0 1 1 | 1 1

crossover

→

1 1 0 1 1

0 1 1 0 1

mutação

→

1 1 0 1 1

0 0 1 0 1

A2 = 0 1 1 1 | 1

A1 = 1 1 0 0 | 1

crossover

→

0 1 1 1 1

1 1 0 0 1

mutação

→

1 0 1 1 1

1 1 0 0 1

Nova pop.

| cromossomos | x | $f(x)$ | prob. de seleção |
|-------------|-----|--------|------------------|
| 1 1 1 0 1 1 | 27 | 729 | 29,1% |
| 1 1 0 0 1 | 25 | 625 | 24,9% |
| 1 1 0 0 1 | 25 | 625 | 24,9% |
| 1 0 1 1 1 | 23 | 529 | 21,1% |

32

Prof. Antonio Geraldo da Rocha Vidal

32



Segunda e Terceira Gerações

Segunda Geração

| | | x | $f(x)$ |
|---|-----------|-----|--------|
| 1 | 1 1 0 1 1 | 27 | 729 |
| 2 | 1 1 0 0 0 | 24 | 576 |
| 3 | 1 0 1 1 1 | 23 | 529 |
| 4 | 1 0 1 0 1 | 21 | 441 |

Terceira Geração

| | | x | $f(x)$ |
|---|-----------|-----|--------|
| 1 | 1 1 0 1 1 | 27 | 729 |
| 2 | 1 0 1 1 1 | 23 | 529 |
| 3 | 0 1 1 1 1 | 15 | 225 |
| 4 | 0 0 1 1 1 | 7 | 49 |



Quarta e Quinta Gerações

Quarta Geração

| | | x | $f(x)$ |
|---|-----------|-----|--------|
| 1 | 1 1 1 1 1 | 31 | 961 |
| 2 | 1 1 0 1 1 | 27 | 729 |
| 3 | 1 0 1 1 1 | 23 | 529 |
| 4 | 1 0 1 1 1 | 23 | 529 |

Quinta Geração

| | | x | $f(x)$ |
|---|-----------|-----|--------|
| 1 | 1 1 1 1 1 | 31 | 961 |
| 2 | 1 1 1 1 1 | 31 | 961 |
| 3 | 1 1 1 1 1 | 31 | 961 |
| 4 | 1 0 1 1 1 | 23 | 529 |



Passo 1 do Aprendizado Genético

- Este passo inicializa uma população P de n indivíduos.
- O algoritmo modifica os indivíduos da população até que uma condição de terminação seja satisfeita.
- A condição pode exigir, por exemplo, que todos os indivíduos da população satisfaçam alguns critérios mínimos.
- Uma outra alternativa para finalização é definir um número fixo de iterações do processo de aprendizagem.



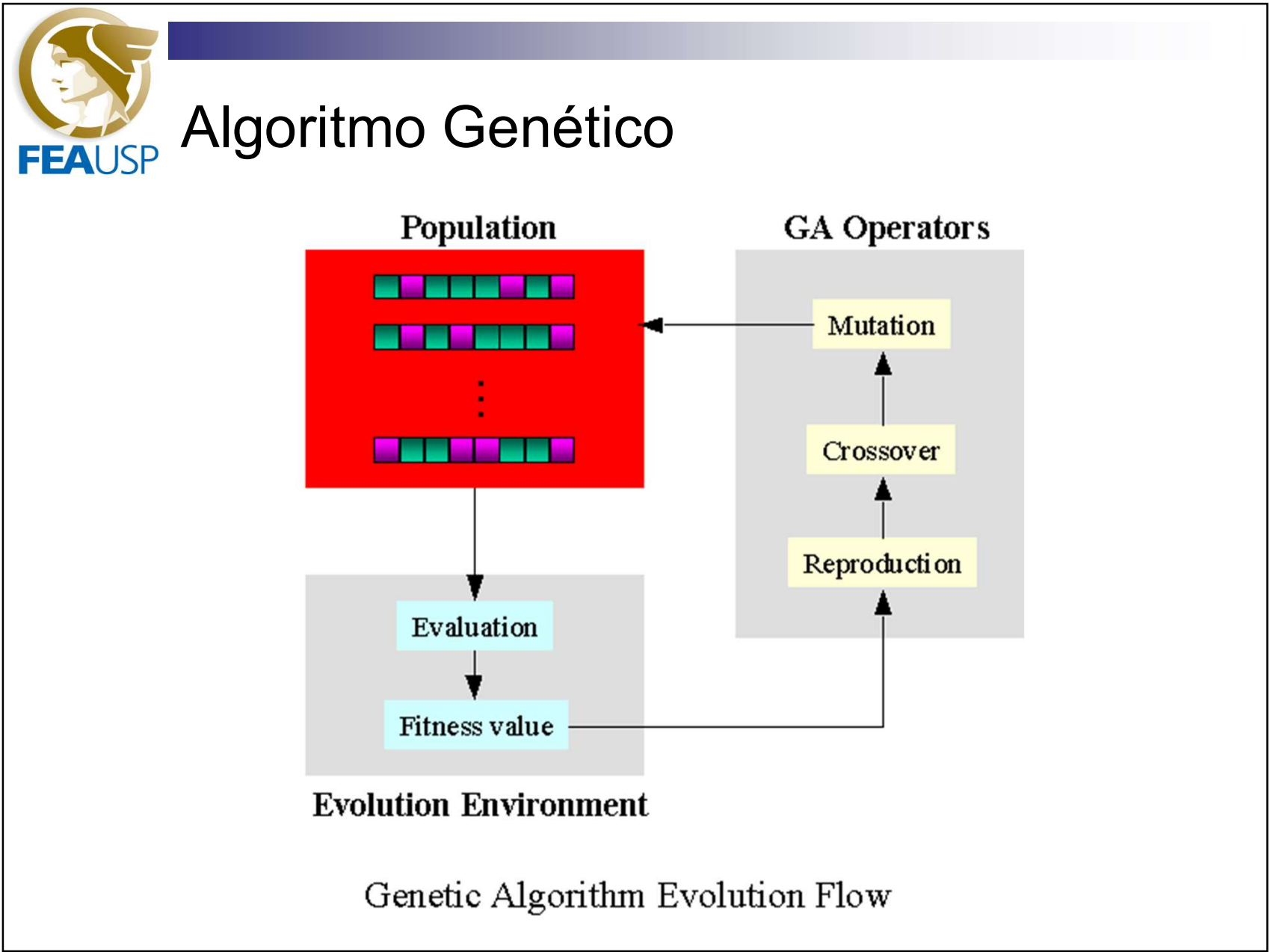
Passo 2A do Aprendizado Genético

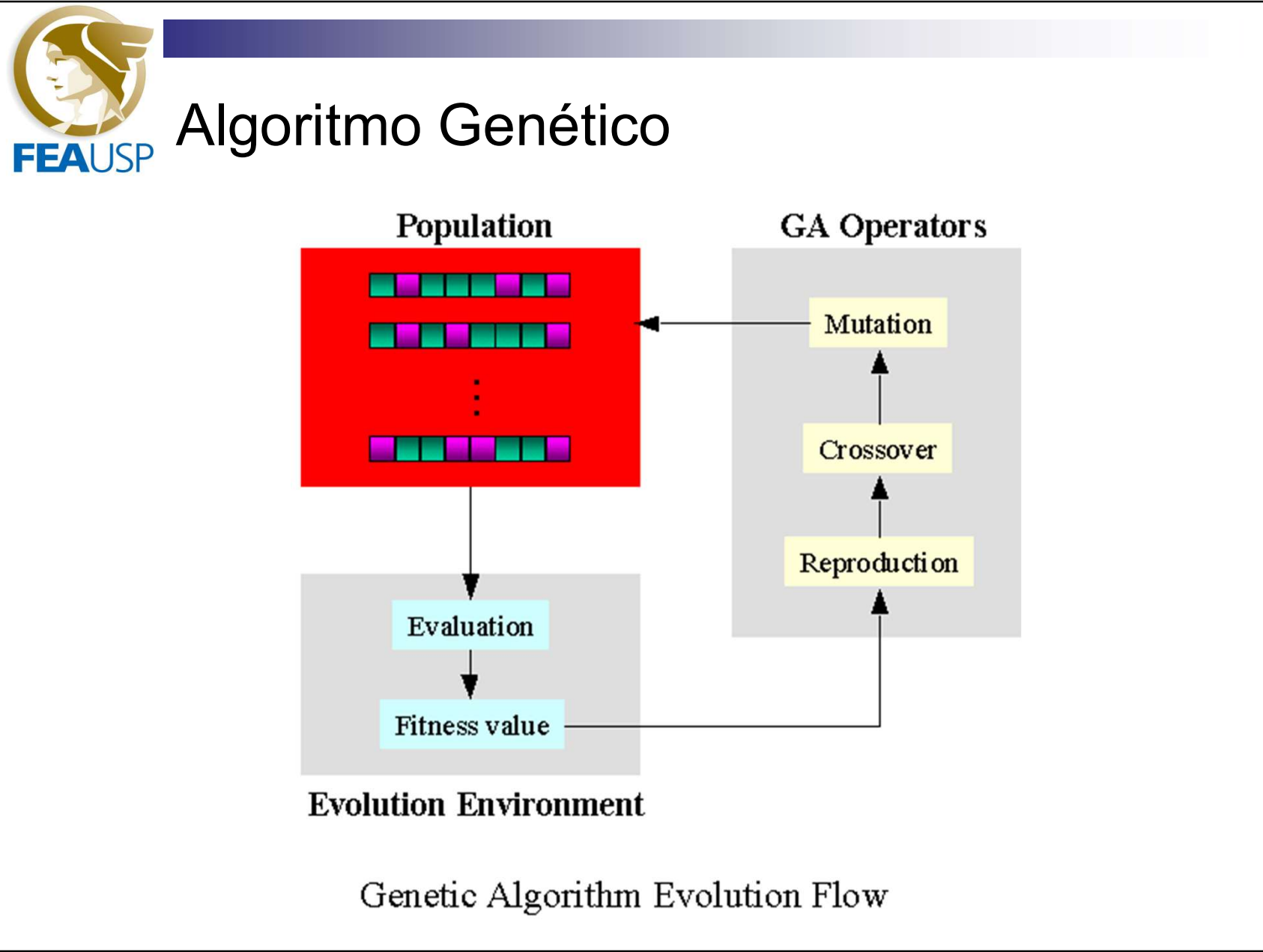
- Inicialmente é aplicada uma função de seleção, para avaliar cada indivíduo atualmente na população.
- A cada iteração, os indivíduos que não satisfaçam os critérios de seleção são eliminados da população.
- O resultado final de uma sessão de aprendizagem genética supervisionada é, de acordo com a função de seleção, um conjunto de indivíduos da população que melhor representa os dados de treinamento.




Passo 2B do Aprendizado Genético

- São então acrescentados novos indivíduos à população para substituir quaisquer indivíduos eliminados na etapa 2A.
- Novos indivíduos são formados a partir de indivíduos excluídos anteriormente pela aplicação de operações de:
 - ☐ Cruzamento e
 - ☐ Mutação.







Exemplo – População Inicial

| Elemento População | Faixa de Renda | Seguro de Vida | Seguro Cartão de Crédito | Sexo | Idade |
|--------------------|----------------|----------------|--------------------------|-----------|-------|
| 1 | 20-30k | Não | Sim | Masculino | 30-39 |
| 2 | 30k-40k | Sim | Não | Feminino | 50-59 |
| 3 | ? | Não | Não | Masculino | 40-49 |
| 4 | 30k-40k | Sim | Sim | Masculino | 40-49 |

Um ponto de interrogação (?) na população significa que é uma condição que "não faz diferença", o que implica que o atributo não é importante para o processo de aprendizagem.



Dados de Treinamento para o Aprendizado Genético

| Elemento de Treinamento | Faixa de Renda | Seguro de Vida | Seguro Cartão de Crédito | Sexo | Idade |
|-------------------------|----------------|----------------|--------------------------|-----------|-------|
| 1 | 30-40k | Sim | Sim | Masculino | 30-39 |
| 2 | 30-40k | Sim | Não | Feminino | 40-49 |
| 3 | 50-60k | Sim | Não | Feminino | 30-39 |
| 4 | 20-30k | Não | Não | Feminino | 50-59 |
| 5 | 20-30k | Não | Não | Masculino | 20-29 |
| 6 | 30-40k | Não | Não | Masculino | 40-49 |



Objetivo e Condição

- Nosso objetivo é criar um modelo capaz de diferenciar os indivíduos que possuem seguro de vida daqueles que não possuem.
- Exigimos que após cada iteração do algoritmo, exatamente dois elementos de cada classe (seguro de vida = sim) e (seguro de vida = não) permanecem na população.



Função de Seleção

1. Seja **N** o número de casamentos dos valores de atributos de entrada **E** com instâncias de treinamento de sua própria classe.
2. Seja **M** o número de casamentos dos valores de atributos de entrada para todas as instâncias de treinamento das classes concorrentes.
3. Adicione 1 para M.
4. Divida N por M.

Nota: quanto maior a pontuação de seleção, menor será a taxa de erro para a solução.



Função de Seleção para o Elemento 1 com Seguro de Vida = Não


| | | | | | |
|---|--------|-----|-----|-----------|-------|
| 1 | 20-30k | Não | Sim | Masculino | 30-39 |
|---|--------|-----|-----|-----------|-------|

- 1. Faixa de renda = 20-30k casa com instâncias de treinamento 4 e 5.
- 2. Sem casamento para Seguro de cartão de crédito = Sim
- 3. Sexo = Masculino casa com instâncias de treinamento 5 e 6.
- 4. Não há casamentos para Idade = 30-39.
- 5. $\therefore N = 4$



Dados de Treinamento para o Aprendizado Genético

| Elemento de Treinamento | Faixa de Renda | Seguro de Vida | Seguro Cartão de Crédito | Sexo | Idade |
|-------------------------|----------------|----------------|--------------------------|-----------|-------|
| 1 | 30-40k | Sim | Sim | Masculino | 30-39 |
| 2 | 30-40k | Sim | Não | Feminino | 40-49 |
| 3 | 50-60k | Sim | Não | Feminino | 30-39 |
| 4 | 20-30k | Não | Não | Feminino | 50-59 |
| 5 | 20-30k | Não | Não | Masculino | 20-29 |
| 6 | 30-40k | Não | Não | Masculino | 40-49 |



Função de Seleção para o Elemento 1 com Seguro de Vida = Sim

| | | | | | |
|---|--------|-----|-----|-----------|-------|
| 1 | 20-30k | Não | Sim | Masculino | 30-39 |
|---|--------|-----|-----|-----------|-------|

1.

Sem casamento para Faixa de Renda = 20-30k

2.

Seguro de cartão de crédito = Sim casa com uma instância de treinamento.

3.

Sexo = Masculino casa com instância de treinamento 1.

4.

Idade = 30-39 casa com instâncias de treinamento 1 e 3.

5.

$\therefore M = 4$

6.

Adicionar 1 a M $\rightarrow M+1 = 5$

7.


$\therefore F(1) = 4/5 = 0,8$

8.

Similarmente $F(2) = 0,86$, $F(3) = 1,2$, $F(4) = 1,0$

9.

Manter elementos 3 e 4 que possuem maior pontuação!



Cruzamento dos Elementos 1 & 2

| Elemento | Faixa de Renda | Seguro de Vida |
|----------|----------------|----------------|
| 1 | 20-30K | Não |

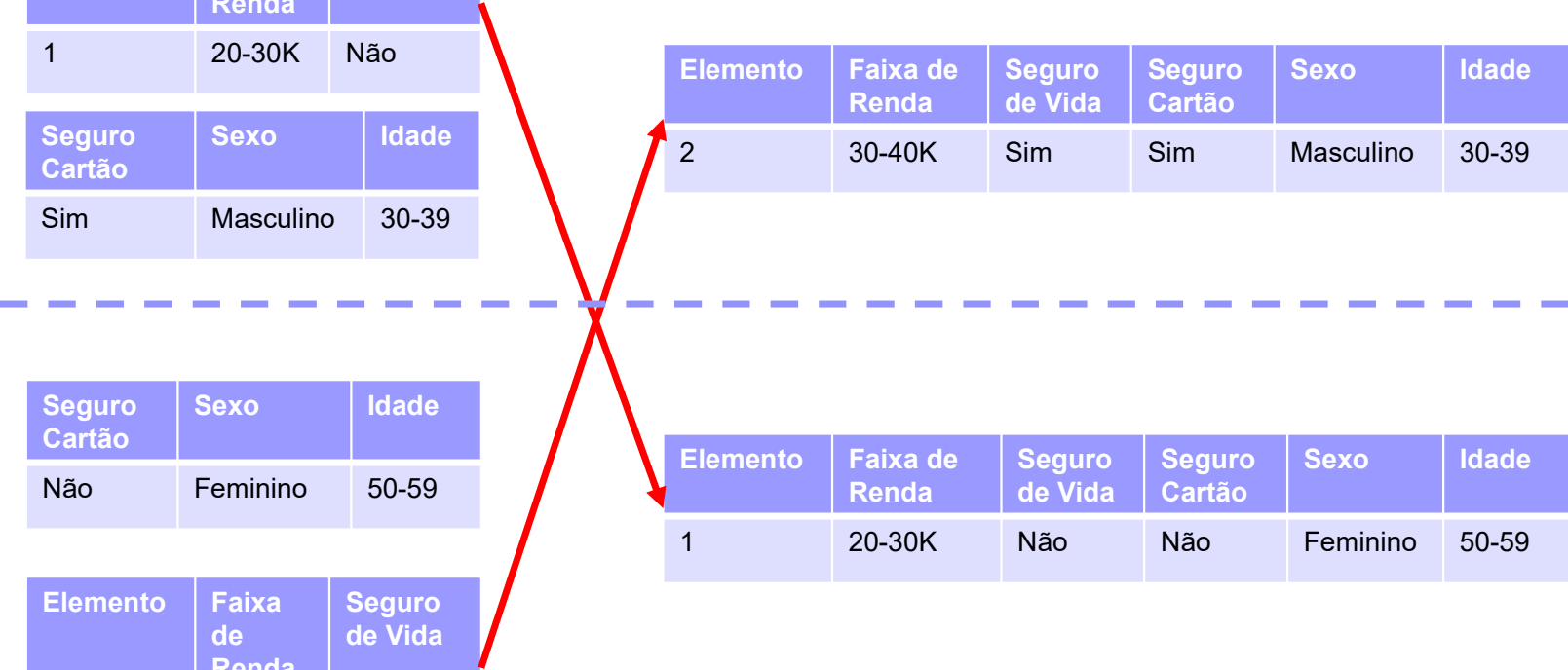
| Seguro Cartão | Sexo | Idade |
|---------------|-----------|-------|
| Sim | Masculino | 30-39 |

| Elemento | Faixa de Renda | Seguro de Vida | Seguro Cartão | Sexo | Idade |
|----------|----------------|----------------|---------------|-----------|-------|
| 2 | 30-40K | Sim | Sim | Masculino | 30-39 |

| Seguro Cartão | Sexo | Idade |
|---------------|----------|-------|
| Não | Feminino | 50-59 |

| Elemento | Faixa de Renda | Seguro de Vida |
|----------|----------------|----------------|
| 2 | 30-40K | Sim |

| Elemento | Faixa de Renda | Seguro de Vida | Seguro Cartão | Sexo | Idade |
|----------|----------------|----------------|---------------|----------|-------|
| 1 | 20-30K | Não | Não | Feminino | 50-59 |





Segunda Geração da População

| Elemento População | Faixa de Renda | Seguro de Vida | Seguro Cartão de Crédito | Sexo | Idade |
|--------------------|----------------|----------------|--------------------------|-----------|-------|
| 1 | 20-30k | Não | Não | Feminino | 50-59 |
| 2 | 30k-40k | Sim | Sim | Masculino | 30-39 |
| 3 | ? | Não | Não | Masculino | 40-49 |
| 4 | 30k-40k | Sim | Sim | Masculino | 40-49 |



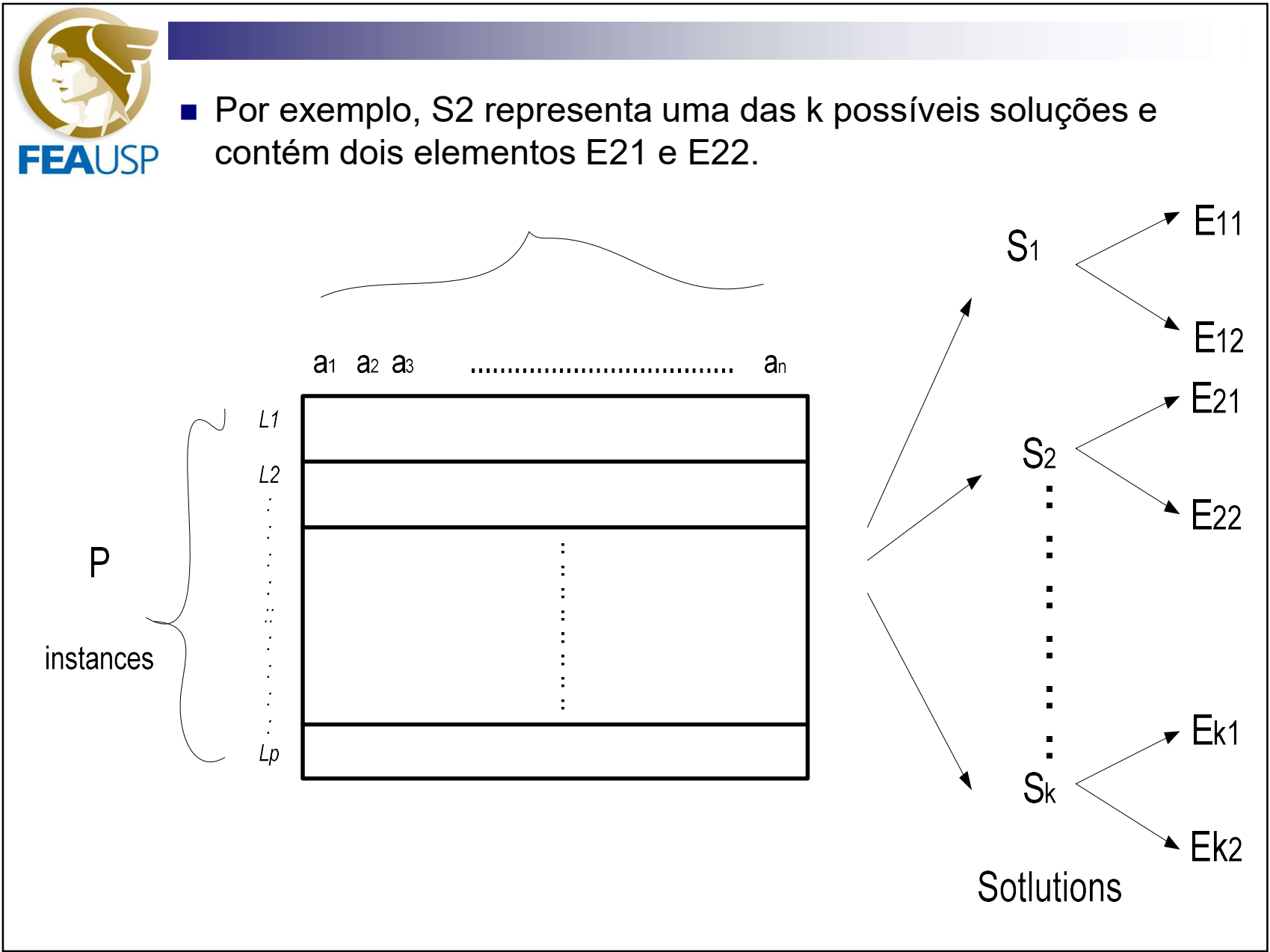
Aplicação do Modelo (Fase de Teste)

- Para utilizar o modelo, pode-se comparar um novo indivíduo desconhecido (dados de teste) com os indivíduos da população final.
- Uma técnica simples é dar ao indivíduo desconhecido a mesma classificação do indivíduo da população para o qual ele é muito semelhante.
- O algoritmo, em seguida, escolhe um dos m indivíduos e dá ao indivíduo desconhecido a classificação do indivíduo escolhido aleatoriamente.



Algoritmo Genético & Cluster Não Supervisionado

- Suponha que há objetos de dados P dentro do espaço onde cada objeto de dados consiste em valores de atributos n .
- Suponha m aglomerados são desejados.
- O modelo irá gerar k possíveis soluções.
- Uma solução específica contém m pontos n -dimensionais, onde cada ponto é o melhor elemento representante de um dos grupos m .





Operação de Cruzamento

- A operação de cruzamento é realizada por elementos móveis (pontos n-dimensionais) de solução para solução S_i S_j .
- Existem várias possibilidades para a execução de operações de mutação.
- Uma maneira de transformar a solução S_i é trocar uma ou mais coordenadas do ponto dos elementos dentro de S_i .



Função de Seleção

- Uma função de seleção aplicável para solução S_j é a distância euclidiana média dos objetos P no espaço n -dimensional a partir de seu objeto mais próximo dentro de S_j .
- Tomamos cada objeto l em P e calculamos a distância euclidiana de l a cada um dos objetos m em S_j .
- Os valores mais baixos representam melhores pontuações de seleção.
- Uma vez que a aprendizagem genética termina, a melhor das soluções possíveis k é selecionada como a solução final.
- Cada objeto, no espaço n -dimensional é atribuído ao agrupamento associado com o objeto mais próximo na solução final.



Conjunto de Entrada

| Objeto | X | Y |
|--------|-----|-----|
| 1 | 1,0 | 1,5 |
| 2 | 1,0 | 4,5 |
| 3 | 2,0 | 1,5 |
| 4 | 2,0 | 3,5 |
| 5 | 3,0 | 2,5 |
| 6 | 5,0 | 6,0 |



Função de Seleção para GA Não Supervisionado

- Aplicamos a função de seleção para os dados de treinamento.
- Instruímos o algoritmo para começar com um conjunto de soluções que consiste em três soluções plausíveis ($k = 3$).
- Com $m = 2$, $P = 6$, e $k = 3$, o algoritmo gera o conjunto inicial de soluções.
- Um elemento no espaço de solução contém um único ponto de dados representativos para cada cluster.
- Por exemplo, os pontos de dados para a solução são S1 (1,0, 1,0) e (5.0,5.0).




Distância Euclidiana

$$d(i, j) = \sqrt{(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2)}$$

Pontuação de seleção de $d(1.0, 1.0)$ e $d(5.0, 5.0)$

$$\begin{aligned} &= \min (\text{Raiz}(|1.0 - 1.0|^2 + |1.0 - 1.5|^2), \text{Raiz}(|5.0 - 1.0|^2 + |5.0 - 1.5|^2) + \\ &\quad \min (\text{Raiz}(|1.0 - 1.0|^2 + |1.0 - 4.5|^2), \text{Raiz}(|5.0 - 1.0|^2 + |5.0 - 4.5|^2) + \\ &\quad \min (\text{Raiz}(|1.0 - 2.0|^2 + |1.0 - 1.5|^2), \text{Raiz}(|5.0 - 2.0|^2 + |5.0 - 1.5|^2) + \\ &\quad \min (\text{Raiz}(|1.0 - 2.0|^2 + |1.0 - 3.5|^2), \text{Raiz}(|5.0 - 2.0|^2 + |5.0 - 3.5|^2) + \\ &\quad \min (\text{Raiz}(|1.0 - 3.0|^2 + |1.0 - 2.5|^2), \text{Raiz}(|5.0 - 3.0|^2 + |5.0 - 2.5|^2) + \\ &\quad \min (\text{Raiz}(|1.0 - 5.0|^2 + |1.0 - 6.0|^2), \text{Raiz}(|5.0 - 5.0|^2 + |5.0 - 6.0|^2) \\ &= 0.5 + 3.5 + 1.11 + 2.69 + 2.5 + 1 \\ &= \mathbf{11.3} \end{aligned}$$



Solução da População para Agrupamento não Supervisionado

| | S1 | S2 | S3 |
|----------------------------------------|------------------------|------------------------|------------------------|
| Objetos da Solução (População Inicial) | (1.0,1.0) (5.0,5.0) | (3.0,2.0) (3.0,5.0) | (4.0,3.0) (5.0,1.0) |
| Pontuação | 11.31 | 9.78 | 15.55 |
| ----- | | | |
| Objetos da Solução (Segunda Geração) | (5.0,1.0) (5.0,5.0) | (3.0,2.0) (3.0,5.0) | (4.0,3.0) (1.0,1.0) |
| Pontuação | 17.96 | 9.78 | 11.34 |
| ----- | | | |
| Objetos da Solução (Terceira Geração) | (5.0,5.0) (1.0,5.0) | (3.0,2.0) (3.0,5.0) | (4.0,3.0) (1.0,1.0) |
| Pontuação | 13.64 | 9.78 | 11.34 |
| ----- | | | |

Usualmente, soluções relacionadas a agrupamento envolvem cálculo proximidade, qualquer que seja a técnica.



Solução de Primeira Geração

- Para calcular a pontuação de seleção de 11,31 para solução S1 a distância euclidiana entre cada objeto e seu ponto mais próximo de dados em S1 é somado.
- Para ilustrar isso, considere um exemplo de dados de treinamento.
- A distância euclidiana entre (1.0,1.0) e (1.0,1.5) é calculada como 0,50.
- A distância entre (5.0,5.0) e (1.0,1.5) é 5,32.
- O menor valor de 0,50 é representado na pontuação de seleção total para o S1 solução.
- S2 é a melhor solução de primeira geração.



Solução de Segunda Geração

- A segunda geração é obtida através da realização de um cruzamento (crossover) entre as soluções S1 e S3.
- O objeto da solução S1 (1.0,1.0) troca de lugar com objeto da solução S3 (5.0,1.0).
- O resultado da operação de cruzamento melhora (diminui) a pontuação de seleção para o S3, enquanto a pontuação para S1 aumenta.



Solução de Terceira Geração

- A terceira geração é obtida através da mutação S1.
- A mutação transporta a coordenada y do primeiro objeto em S1 com a coordenada x do segundo objeto.
- A mutação resulta em uma melhor pontuação de seleção para S1.
- Mutações e cruzamentos continuam até que uma condição de término seja satisfeita.
- Se a terceira geração é a última, então, a solução final é S2.



Solução para Agrupamento

- Se S2 (3.0, 2.0) e (3.0, 5.0) são a solução final, então calculam-se as distâncias entre S2 e os pontos seguintes:
 - Objetos 1, 3 e 5, formando um primeiro cluster e objetos 2 e 6 formando segundo cluster e objeto 4 pode estar em qualquer um dos clusters.

| Cluster 1 centro (3.0, 2.0) | | | Cluster 2 centro (3.0, 5.0) | | |
|-----------------------------|-----|-----|-----------------------------|-----|-----|
| Objeto | X | Y | Objeto | X | Y |
| 1 | 1.0 | 1.5 | 2 | 1.0 | 4.5 |
| 3 | 2.0 | 1.5 | 4 | 2.0 | 3.5 |
| 5 | 3.0 | 2.5 | 6 | 5.0 | 6.0 |



Considerações Gerais sobre o Algoritmo Genético

- GA é útil para encontrar soluções globalmente otimizadas.
- A função de seleção determina a complexidade computacional de um Algoritmo Genético.
- GA explica seus resultados na medida em que a função de seleção é compreensível.
- A transformação dos dados para um formato adequado para um Algoritmo Genético pode ser um desafio.



Machine Learning

Prof. Antonio Geraldo da Rocha Vidal