# Classificação e Pesquisa de Dados

**Cristiano Santos** 

cristiano.santos@amf.edu.br

#### **OBJETIVO**

- Capacitar o aluno na análise e seleção de algoritmos de classificação de dados e de pesquisa de dados
- Específicos:
  - Conhecer os principais conceitos e técnicas para implementação de algoritmos de classificação e pesquisa de dados existentes e comparar a sua eficácia em diferentes conjuntos de dados;
  - Compreender os aspectos introdutórios do processo de análise de complexidade de algoritmos com base nos algoritmos estudados na disciplina, sabendo determinar os que apresentam o melhor e o pior custo em termos de tempo e espaço;
  - <u>Implementar algoritmos de manipulação</u> de dados e analisar sua complexidade para solucionar problemas específicos.

#### O que é?

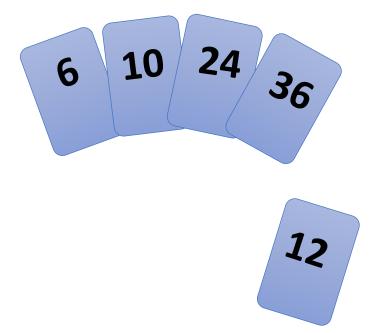
- Analise de algoritmos
  - Em ciência da computação, é a área de pesquisa cujo o foco são os algoritmos.
  - Busca responder:
    - Podemos fazer algoritmos mais eficientes?

### Ordenação (Sort)

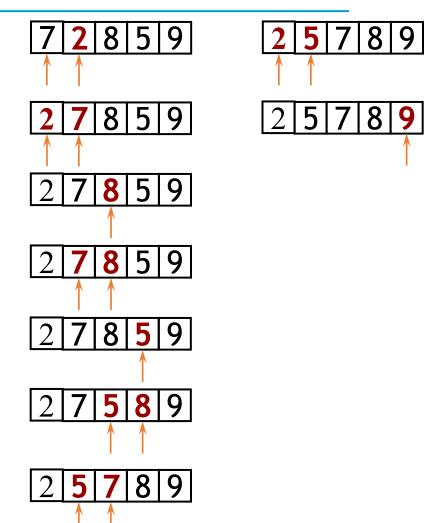
- Dado um conjunto de dados armazenados em ordem aleatória, reordena-los com base em uma regra de precedência
  - Qual a regra de precedência para:
    - Números
    - Nomes
    - Alunos

### Método de Ordenação

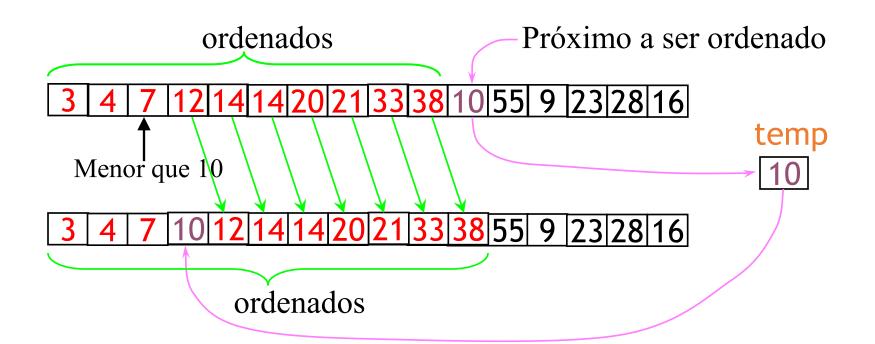
• Método do jogador de cartas



# Ordenação (Exemplo)



#### Um passo de ordenação



7

#### Algoritmos e Complexidade

- Eficiência de um algoritmo
  - Complexidade de tempo: quanto "tempo" é necessário para computar o resultado para uma instância do problema de tamanho n
    - **Pior caso**: Considera-se a instância que faz o algoritmo funcionar mais lentamente
    - Caso médio: Considera-se todas as possíveis instâncias e mede-se o tempo médio
- Eficiência de uma estrutura de dados
  - Complexidade de espaço: quanto "espaço de memória/disco" é preciso para armazenar a estrutura (pior caso e caso médio)
- Complexidade de espaço e tempo estão frequentemente relacionadas

```
algoritmoSort (int data[], int n) {
        int tmp,i,j;
        for (j=1; j<n; j++) {
                i =j - 1;
                tmp = data[j];
                while ( (i>=0) && (tmp < data[i]) ) {
                        data[i+1] = data[i];
                         i--;
                }//while
                data[i+1] = tmp;
        }//for
```

```
algoritmoSort (int data[], int n) {
        int tmp,i,j;
                                                           5
                                                    4
                                              1
        for (j=1; j<n; j++) {
                i =j - 1;
                tmp = data[j];
                while ( (i > = 0) && (tmp < data[i]) ) {
                         data[i+1] = data[i];
                         i--;
                }//while
                data[i+1] = tmp;
        }//for
```

```
algoritmoSort (int data[], int n) {
        int tmp,i,j;
                                                          5
                                        2
                                              1
                                                    4
        for (j=1; j<n; j++) {
                i = j - 1;
                tmp = data[j];
                while ( (i>=0) && (tmp < data[i]) ) {
                        data[i+1] = data[i];
                         i--;
                }//while
                data[i+1] = tmp;
        }//for
```

```
algoritmoSort (int data[], int n) {
        int tmp,i,j;
                                                         5
                                             1
                                                    4
        for (j=1; j<n; j++) {
                i =j - 1;
                tmp = data[j];
                while ( (i>=0) && (tmp < data[i]) ) {
                        data[i+1] = data[i];
                         i--;
                }//while
                data[i+1] = tmp;
        }//for
```

```
algoritmoSort (int data[], int n) {
        int tmp,i,j;
                                                         5
                                       2
                                             1
                                                   4
       for (j=1; j<n; j++) {
                i = j - 1;
                tmp = data[j];
                while ((i>=0) && (tmp < data[i])) {
                        data[i+1] = data[i];
                         i--;
                }//while
                data[i+1] = tmp;
        }//for
```

```
algoritmoSort (int data[], int n) {
        int tmp,i,j;
                                                          5
                                       2
                                              1
                                                    4
       for (j=1; j<n; j++) {
                i =j - 1;
                tmp = data[j];
                while ( (i>=0) && (tmp < data[i])) {
                        data[i+1] = data[i];
                         i--;
                }//while
                data[i+1] = tmp;
        }//for
```

```
algoritmoSort (int data[], int n) {
        int tmp,i,j;
                                                         5
                                             1
                                                   4
       for (j=1; j<n; j++) {
                i = j - 1;
                tmp = data[j];
                while ((i \ge 0) & (tmp < data[i]))
                        data[i+1] = data[i];
                         i--;
                }//while
                data[i+1] = tmp;
        }//for
```

```
algoritmoSort (int data[], int n) {
        int tmp,i,j;
                                                           5
                                               1
                                                     4
        for (j=1; j<n; j++) {
                i = j - 1;
                tmp = data[j];
                while ( (i > = 0) \&\& (tmp < data[i]) ) {
                         data[i+1] = data[i];
                         i--;
                }//while
                data[i+1] = tmp;
        }//for
```

```
algoritmoSort (int data[], int n) {
        int tmp,i,j;
                                                           5
                                         3
                                               1
                                                     4
        for (j=1; j<n; j++) {
                i = j - 1;
                tmp = data[j];
                while ( (i > = 0) \&\& (tmp < data[i]) ) {
                         data[i+1] = data[i];
                          i--;
                 }//while
                 data[i+1] = tmp;
        }//for
```

```
algoritmoSort (int data[], int n) {
        int tmp,i,j;
                                                         5
                                       3
                                             1
                                                   4
       for (j=1; j<n; j++) {
                i = j - 1;
                tmp = data[j];
                while(((i>=0))&& (tmp < data[i])) {
                        data[i+1] = data[i];
                         i--;
                }//while
                data[i+1] = tmp;
        }//for
```

```
algoritmoSort (int data[], int n) {
        int tmp,i,j;
                                                         5
                                       3
                                             1
                                                   4
       for (j=1; j<n; j++) {
                i =j - 1;
                tmp = data[j];
                while ((i > = 0) & (tmp < data[i]))
                        data[i+1] = data[i];
                         i--;
                }//while
                data[i+1] = tmp; // -1 +1 = 0
        }//for
```

#### Estudo de caso

- Qual o melhor caso?
  - Os dados já estarem ordenados

1	2	3	4	5	6	7	8	9	
---	---	---	---	---	---	---	---	---	--

- Qual o pior?
  - Os dados estarem na ordem inversa

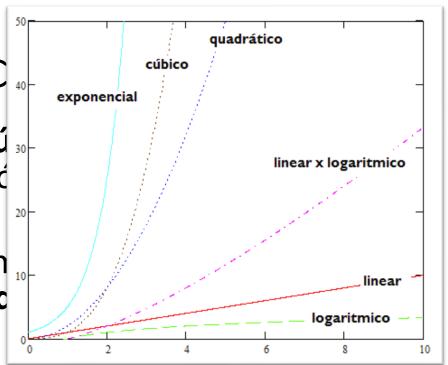
9	8	7	6	5	4	3	2	1
---	---	---	---	---	---	---	---	---

#### Algoritmos e Complexidade

- Eficiência medida objetivamente depende de:
  - Como o programador implementou o algoritmo/ED
  - Características do computador usado para fazer experimentos:
    - Velocidade da CPU
    - Capacidade e velocidade de acesso à memória primária / secundária
    - Etc
  - Linguagem / Compilador / Sistema Operacional / etc
- Portanto, a medição formal de complexidade tem que ser subjetiva, porém matematicamente consistente
  - ⇒ Complexidade assintótica

# Complexidade Assintótic

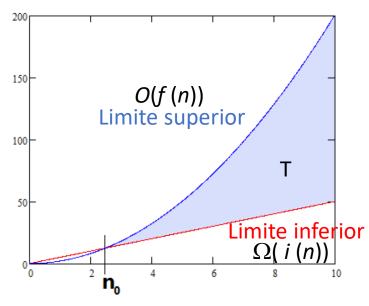
- Tempo / espaço medidos em nú 30algoritmo / "palavras" de memó bytes
- Análise do algoritmo / E.D. pern depende do tamanho da entrac armazenados (n).
  - Ex.:  $T(n) = 13n^3 + 2n^2 + 6n \log n$
- Percebe-se que à medida que *n* aumenta, o termo **cúbico** começa a **dominar**
- A constante que multiplica o termo cúbico tem relativamente a mesma importância que a velocidade da CPU / memória
- Diz-se que  $T(n) \in O(n^3)$



#### Limites

#### • Definição:

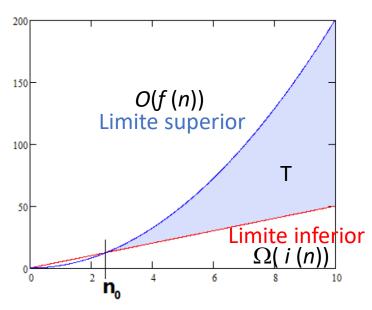
 $T(n) \in O(f(n))$  se existem constantes  $c \in n_0$  tais que  $T(n) \le c f(n)$  para todo  $n \ge n_0$ 



**Figura 2.** Algoritmo com limites assintóticos superior e inferior quadrático e linear, respectivamente. Estes dois limites correspondem ao melhor caso e ao pior caso, mas a área em azul compreende todos os outros possíveis casos.

#### Limites Justos

- Observe que se  $T(n) \in O(n^3)$  então:
  - $T(n) \in O(n^4), T(n) \in O(n^5), \text{ etc}$
- Analogamente, se  $T(n) \in \Omega(n^3)$  então:
  - $T(n) \in \Omega(n^2)$ ,  $T(n) \in \Omega(n)$ , etc



**Figura 2.** Algoritmo com limites assintóticos superior e inferior quadrático e linear, respectivamente. Estes dois limites correspondem ao melhor caso e ao pior caso, mas a área em azul compreende todos os outros possíveis casos.

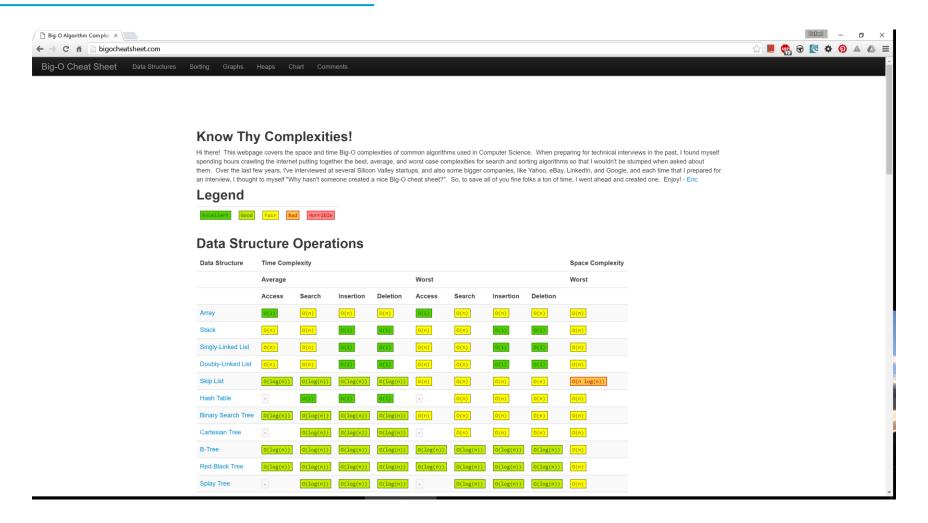
- Se uma função T(n) tem como limites superior e inferior a mesma função f(n), então:
  - f(n) é um **limite justo** de T(n), ou  $T(n) \in \Theta(f(n))$

#### Inventário de funções de complexidade

- $T(n) \in O(1)$ : constante mais rápido, impossível
- $T(n) \in O(\log \log n)$ : super-rápido
- $T(n) \in O(\log n)$ : logarítmico muito bom
- T(n) ∈ O(n): linear é o melhor que se pode esperar se algo não pode ser determinado sem examinar toda a entrada
- T (n) ∈ O (n log n): limite de muitos problemas práticos, ex.: ordenar uma coleção de números
- $T(n) \in O(n^2)$ : quadrático
- $T(n) \in O(n^k)$ : polinomial ok para n pequeno
- $T(n) \in O(k^n)$ , O(n!),  $O(n^n)$ : exponencial evite!

Qual a complexidade do insertion sort?

## http://bigocheatsheet.com/



#### Complexidade por contagem

**}//Insertionsort** 

```
Insertionsort (int data[], int n) {
       int tmp,i,j;
                                                                   N+1
       for (j=1; j<n; j++) {
                                               E agora?
                                                                   (1)N+1
               i = j - 1;
                                                   E agora?
                                                                   (1+1)N+1
               tmp = data[j];
                                                                   (T+1+1)N+1
               while ( (i>=0) && (tmp < data[i]) ) {
                                                                   ((1)T+1+1)N+1
                       data[i+1] = data[i];
                                                                   ((1+1)T+1+1)N+1
                       i--;
               }//while
                                                                   ((1+1)T+1+1+1)N+1
               data[i+1] = tmp;
                                                                    TN
       }//for
```

Essa é a complexidade?

#### complexidade do algoritmo de cada caso

Qual o melhor caso?

Qual o T em ambos os casos?

•	Os d	ados	já	estarem	ord	lenad	os
---	------	------	----	---------	-----	-------	----

1	2	3	4	5	6	7	8	9
---	---	---	---	---	---	---	---	---

$$T = 1 \longrightarrow 1N \longrightarrow \Omega(N)$$

- Qual o pior?
  - Os dados estarem na ordem inversa

$$T = N \longrightarrow NN \longrightarrow O(N^2)$$

# Revisão:

#### Definição: Estruturas de dados

• Uma estrutura de dados é um meio para armazenar e organizar dados com o objetivo de facilitar o acesso e as modificações.

• Nenhuma estrutura de dados única funciona bem para todos os propósitos, e assim é importante conhecer os pontos fortes e as limitações de várias delas

#### Ordenação

• Vamos ordenar um lista de números



### Somente para números?

- Essas técnicas só funcionam para números?
  - Se definirmos uma função que diga a partir de dois elementos qual o maior, podemos aplicar esses algoritmos para diversos tipos de dados
  - Exemplos?
    - Alfabeto
    - Cerveja
    - •

#### Tarefa:

- Identificar qual algoritmo de ordenação apresentado em cada SLIDE e justificar
- 2. Descreva/pesquise o que é notação assintótica?
- 3. Identificar qual o melhor e o pior caso (pesquisar/encontrar qual o melhor e o pior caso de cada algoritmo estudado)
- 4. Qual dos algoritmos estudados está(ão) faltando nestes slides?

# Qual é o Algoritmo que segue nos próximos?

```
???
         (int data[], int n) {
   int tmp,i,j;
   for (j=1; j<n; j++) {
           i =j - 1;
           tmp = data[j];
           while ( (i>=0) && (tmp < data[i]) ) {
                   data[i+1] = data[i];
                    i--;
           }//while
           data[i+1] = tmp;
   }//for
```

```
???
         (int data[], int n) {
   int tmp,i,j;
                                                     5
                                         1
                                               4
   for (j=1; j<n; j++) {
           i =j - 1;
           tmp = data[j];
           while ( (i>=0) && (tmp < data[i]) ) {
                    data[i+1] = data[i];
                    i--;
           }//while
           data[i+1] = tmp;
   }//for
```

```
(int data[], int n) {
???
   int tmp,i,j;
                                                     5
                                   2
                                         1
                                               4
   for (j=1; j<n; j++) {
           i = j - 1;
           tmp = data[j];
           while ( (i>=0) && (tmp < data[i]) ) {
                    data[i+1] = data[i];
                    i--;
           }//while
           data[i+1] = tmp;
   }//for
```

```
(int data[], int n) {
???
   int tmp,i,j;
                                                     5
                                   2
                                         1
                                               4
   for (j=1; j<n; j++) {
           i =j - 1;
           tmp = data[j];
           while ( (i>=0) && (tmp < data[i]) ) {
                   data[i+1] = data[i];
                    i--;
           }//while
           data[i+1] = tmp;
   }//for
```

```
(int data[], int n) {
???
   int tmp,i,j;
                                                     5
                                   2
                                         1
                                               4
   for (j=1; j<n; j++) {
           i = j - 1;
           tmp = data[j];
           while ((i>=0) && (tmp < data[i])) {
                   data[i+1] = data[i];
                    i--;
           }//while
           data[i+1] = tmp;
   }//for
```

```
(int data[], int n) {
???
   int tmp,i,j;
                                                     5
                                   2
                                         1
                                               4
   for (j=1; j<n; j++) {
           i = j - 1;
           tmp = data[j];
           while ( (i>=0) && (tmp < data[i])) {
                   data[i+1] = data[i];
                    i--;
           }//while
           data[i+1] = tmp;
   }//for
```

```
(int data[], int n) {
???
   int tmp,i,j;
                                                     5
                                   2
                                         1
                                               4
   for (j=1; j<n; j++) {
           i = j - 1;
           tmp = data[j];
           while ((i \ge 0) & (tmp < data[i]))
                   data[i+1] = data[i];
                    i--;
           }//while
           data[i+1] = tmp;
   }//for
```

```
(int data[], int n) {
???
   int tmp,i,j;
                                                       5
                                    3
                                          1
                                                 4
   for (j=1; j<n; j++) {
           i = j - 1;
           tmp = data[j];
           while ( (i > = 0) \&\& (tmp < data[i]) ) {
                    data[i+1] = data[i];
                     i--;
           }//while
            data[i+1] = tmp;
   }//for
```

```
(int data[], int n) {
???
   int tmp,i,j;
                                                      5
                                    3
                                          1
                                                4
   for (j=1; j<n; j++) {
           i = j - 1;
           tmp = data[j];
           while ( (i > = 0) && (tmp < data[i]) ) {
                    data[i+1] = data[i];
                    i--;
           }//while
           data[i+1] = tmp;
   }//for
```

```
(int data[], int n) {
???
   int tmp,i,j;
                                                     5
                                   3
                                         1
                                               4
   for (j=1; j<n; j++) {
           i = j - 1;
           tmp = data[j];
           while(((i>=0))&& (tmp < data[i])) {
                   data[i+1] = data[i];
                    i--;
           }//while
           data[i+1] = tmp;
   }//for
```

```
(int data[], int n) {
???
   int tmp,i,j;
                                                     5
                                   3
                                         1
                                               4
   for (j=1; j<n; j++) {
           i =j - 1;
           tmp = data[j];
           while ((i > = 0) & (tmp < data[i]))
                   data[i+1] = data[i];
                    i--;
           }//while
           data[i+1] = tmp; // -1 +1 = 0
   }//for
```

```
(int data[], int n) {
???
   int tmp,i,j;
                                                     5
                                   3
                                         1
                                               4
   for (j=1; j<n; j++) {
           i =j - 1;
           tmp = data[j];
           while ((i > = 0) & (tmp < data[i]))
                   data[i+1] = data[i];
                    i--;
           }//while
           data[i+1] = tmp; // -1 +1 = 0
   }//for
```

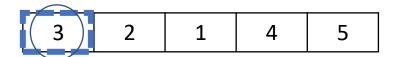
## Qual é o Algoritmo que segue nos próximos?

```
(int data[],int n) {
 555
int menor, troca, i, j, menor id;
for (i=0; i<n-1; i++) {
  menor = data[i];
  for (j=i+1; j<n; j++)
    if (data[j] < menor)</pre>
      menor = data[j];
      menor id = j;
  troca = data[i];
  data[i] = data[menor id];
  data[menor id] = troca;
```

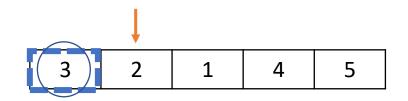
```
(int data[],int n) {
 555
int menor, troca, i, j, menor id;
for (i=0; i<n-1; i++) {
  menor = data[i];
  for (j=i+1; j<n; j++)
    if (data[j] < menor)</pre>
      menor = data[j];
      menor id = j;
  troca = data[i];
  data[i] = data[menor id];
  data[menor id] = troca;
```

	$\sim$					
(	3	)	2	1	4	5

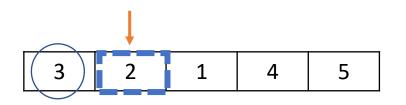
```
(int data[],int n) {
 555
int menor, troca, i, j, menor id;
for (i=0; i<n-1; i++) {
  menor = data[i];
  for (j=i+1; j<n; j++)
    if (data[j] < menor)</pre>
      menor = data[j];
      menor id = j;
  troca = data[i];
  data[i] = data[menor id];
  data[menor id] = troca;
```



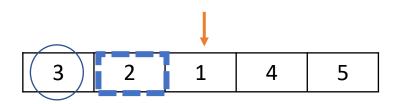
```
(int data[],int n) {
 333
int menor, troca, i, j, menor id;
for (i=0; i<n-1; i++) {
  menor = data[i];
  for (j=i+1; j<n; j++)
    (data[j] < menor)</pre>
      menor = data[j];
      menor id = j;
  troca = data[i];
  data[i] = data[menor_id];
  data[menor id] = troca;
```



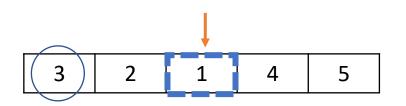
```
(int data[],int n) {
 333
int menor, troca, i, j, menor id;
for (i=0; i<n-1; i++) {
  menor = data[i];
  for (j=i+1; j<n; j++)
    (data[j] < menor)</pre>
      menor = data[j];
      menor id = j;
  troca = data[i];
  data[i] = data[menor_id];
  data[menor id] = troca;
```



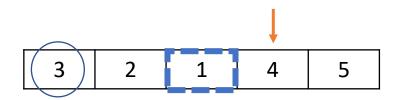
```
(int data[],int n) {
 555
int menor, troca, i, j, menor id;
for (i=0; i<n-1; i++) {
  menor = data[i];
  for (j=i+1; j<n; j++)
    (data[j] < menor)</pre>
      menor = data[j];
      menor id = j;
  troca = data[i];
  data[i] = data[menor id];
  data[menor id] = troca;
```



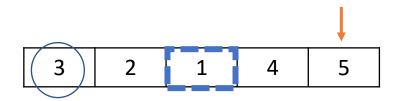
```
(int data[],int n) {
 555
int menor, troca, i, j, menor id;
for (i=0; i<n-1; i++) {
  menor = data[i];
  for (j=i+1; j<n; j++)
    (data[j] < menor)</pre>
      menor = data[j];
      menor id = j;
  troca = data[i];
  data[i] = data[menor id];
  data[menor id] = troca;
```



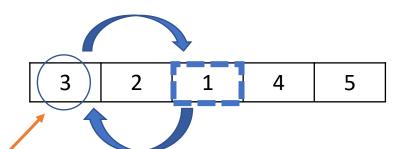
```
(int data[],int n) {
 555
int menor, troca, i, j, menor id;
for (i=0; i<n-1; i++) {
  menor = data[i];
  for (j=i+1; j<n; j++)
    (data[j] < menor)</pre>
      menor = data[j];
      menor id = j;
  troca = data[i];
  data[i] = data[menor id];
  data[menor id] = troca;
```



```
(int data[],int n) {
 555
int menor, troca, i, j, menor id;
for (i=0; i<n-1; i++) {
  menor = data[i];
  for (j=i+1; j<n; j++)
    (data[j] < menor)</pre>
      menor = data[j];
      menor id = j;
  troca = data[i];
  data[i] = data[menor id];
  data[menor id] = troca;
```



```
(int data[],int n) {
 555
int menor, troca, i, j, menor id;
for (i=0; i<n-1; i++) {
  menor = data[i];
  for (j=i+1; j<n; j++)
    if (data[j] < menor)</pre>
      menor = data[j];
      menor id = j;
  troca = data[i];
  data[i] = data[menor id];
  data[menor id] = troca;
```



```
(int data[],int n) {
 555
int menor, troca, i, j, menor id;
for (i=0; i<n-1; i++) {
  menor = data[i];
  for (j=i+1; j<n; j++)
    if (data[j] < menor)</pre>
      menor = data[j];
      menor id = j;
  troca = data[i];
  data[i] = data[menor id];
  data[menor id] = troca;
```



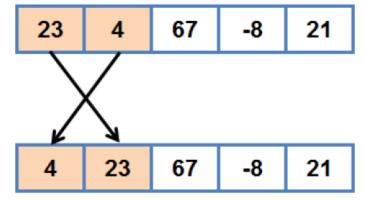
# Qual é o Algoritmo que segue nos próximos?

```
(int* v, int n) {
void
     ???
  int i, fim, aux;
    for (fim = n-1; fim > 0; --fim) {
      for (i = 0; i < fim; ++i) {
                                          6 5 3 1 8 7 2 4
        if (v[i] > v[i+1]) {
          aux = v[i];
          v[i] = v[i+1];
          v[i+1] = aux;
```

# Qual é o Algoritmo que segue nos próximos?

```
???
               (int *∇ , int N) {
void
    int i, continua, aux, fim = N;
    do{
        continua = 0;
        for (i = 0; i < fim-1; i++) {
            if (V[i] > V[i+1]){
                aux = V[i];
                V[i] = V[i+1];
                V[i+1] = aux;
                continua = i;
        fim--;
    }while(continua != 0);
```

Troca dois valores consecutivos no vetor

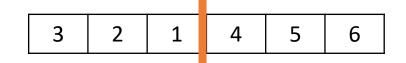


# Qual é o Algoritmo que segue nos próximos?

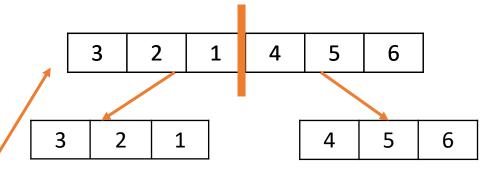
```
void
                  (int *v, int inicio, int fim){
           ???
       int meio;
       if (inicio < fim){</pre>
               meio = floor((inicio + fim)/2);
                            (v, inicio, meio);
                    ???
                            (v, meio+1, fim);
                    ???
                      (v, inicio, meio, fim)
                 ???
```

```
(int *v, int inicio, int fim){
void
          ???
       int meio;
       if (inicio < fim){</pre>
              meio = floor((inicio + fim)/2);
                           (v, inicio, meio);
                   ???
                                                     Divisão
                           (v, meio+1, fim);
                   ???
                      (v, inicio, meio, fim)
                 ???
                                                    Conquista
```

```
void
                  (int *v, int inicio, int fim){
           ???
       int meio;
       if (inicio < fim){</pre>
               meio = floor((inicio + fim)/2);
                            (v, inicio, meio);
                    ???
                            (v, meio+1, fim);
                    ???
                       (v, inicio, meio, fim)
                 ???
```



```
void
                  (int *v, int inicio, int fim){
           ???
       int meio;
       if (inicio < fim){</pre>
               meio = floor((inicio + fim)/2);
                            (v, inicio, meio);
                    ???
                            (v, meio+1, fim);
                    ???
                       (v, inicio, meio, fim)
                 ???
```



```
Igoritmo 5 (int *v, int ini, int meio, int fim) {
void
             int *temp, p1, p2, tam, i, j, k;
             int fim1 = 0, fim2 = 0;
             tam = fim - ini +1;
              p1 = ini;
              p2 = meio + 1;
             temp = (int *) malloc(tam * sizeof(int));
             if(temp != NULL){
                           for(i = 0; i < tam; i++){
                                         if(!fim1 && !fim2){
                                                      if(v[p1] < v[p2])
                                                                    temp[i] = v[p1++];
                                                      else
                                                                    temp[i] = v[p2++];
                                                      if(p1>meio) fim1 = 1;
                                                      if(p2>fim) fim2 = 1;
                                         }else{
                                                      if (!fim1)
                                                                    temp[i] = v[p1++];
                                                      else
                                                                    temp[i] = v[p2++];
                           for(j = 0, k = ini; j < tam; j++, k++)
                                        v[k] = temp[j];
```

free(temp);

## Qual é o Algoritmo que segue nos próximos?

```
(int *V, int inicio, int final ) {
   ∃int
19
20
          int esq, dir, pivo, aux;
21
          esq = inicio;
22
          dir = final;
23
          pivo = V[inicio];
24
          while(esq < dir){</pre>
25
              while(esq <= final && V[esq] <= pivo)</pre>
26
                  esq++;
27
28
              while(dir >= 0 && V[dir] > pivo)
29
                  dir--;
30
31
              if(esq < dir){</pre>
32
                  aux = V[esq];
33
                  V[esq] = V[dir];
34
                  V[dir] = aux;
35
36
37
          V[inicio] = V[dir];
38
          V[dir] = pivo;
39
          return dir;
40
```

## Classificação e Pesquisa de Dados

**Cristiano Santos** 

cristiano.santos@amf.edu.br