

# SCC0221

# Introdução à Ciência da Computação I

## Trabalho 2

Ano/Semestre: 2020/02

Prof. Ricardo M. Marcacini  
ricardo.marcacini@icmc.usp.br



# Sobre o trabalho

- Você recebeu um arquivo de texto com as seguintes informações:
  - Número de linhas: 150
  - Número de colunas: 3
- As linhas representam 50 amostras de três espécies de *Flores Iris* (*Iris setosa*, *Iris virginica*, *Iris versicolor*)
- As colunas representam:
  - Largura da pétala
  - Altura da pétala
  - Espécie da flor



Setosa



Versicolor



Virginica

# Sobre o trabalho

- O objetivo geral é escrever um programa que leia o arquivo de dados para organizar as flores em  $k$  grupos, em que flores similares (altura e largura da pétala) devem ser alocadas no mesmo grupo.
  - A terceira coluna (espécie) não deve ser utilizada para a geração dos grupos. É uma informação que usaremos apenas para verificar a “qualidade” dos grupos obtidos.



Setosa



Versicolor



Virginica

# Arquivo de dados

- O arquivo de dados fornecido é um arquivo texto, em que as colunas são separadas por vírgulas.

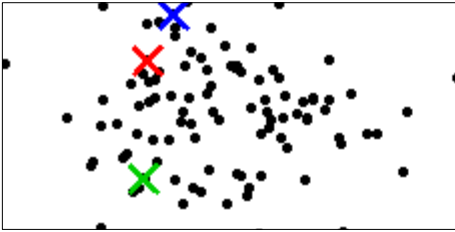
Vamos ver o arquivo  
iris\_petalas.csv

# Método para gerar os grupos

- Implementar o seguinte pseudocódigo para gerar os grupos:
  1. Ler o arquivo de dados
  2. Ler o valor de  $k$  do usuário, com  $k > 1$  (número de grupos)
  3. Escolher aleatoriamente  $k$  flores do conjunto de dados para serem representantes iniciais de cada um dos  $k$  grupos
  4. Para cada flor do conjunto de dados, alocar ao grupo mais próximo
    - 4.1. A proximidade de uma flor a um grupo é dada pela distância euclidiana da flor ao representante do grupo
  5. Atualizar o representante de cada grupo.
    - 5.1. Agora o representante de um grupo é formado pelo ponto médio das flores daquele grupo
  6. Repetir os passos 4 e 5 uma quantidade  $x$  de vezes.
  7. Gerar um novo arquivo de dados adicionando uma quarta coluna. Esta coluna indica o número do grupo em que a flor foi alocada.

# Método para gerar os grupos

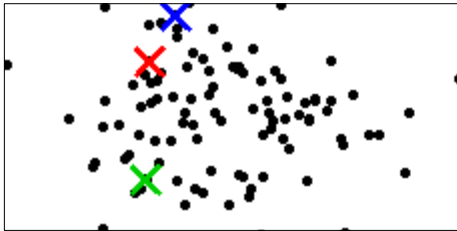
## ■ Explicação visual do método:



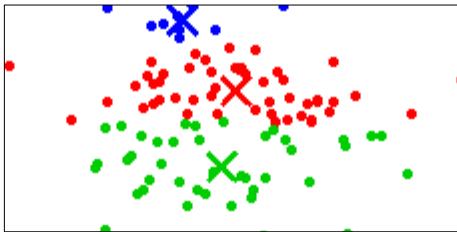
Usuário informa  $k=3$  grupos.  
Escolher 3 representantes iniciais.  
OBS: cada cor indica um grupo.  
(Linhas 1 a 3 do pseudocódigo)

# Método para gerar os grupos

## ■ Explicação visual do método:



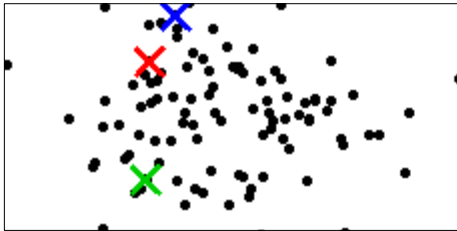
Usuário informa  $k=3$  grupos.  
Escolher 3 representantes iniciais.  
OBS: cada cor indica um grupo.  
(Linhas 1 a 3 do pseudocódigo)



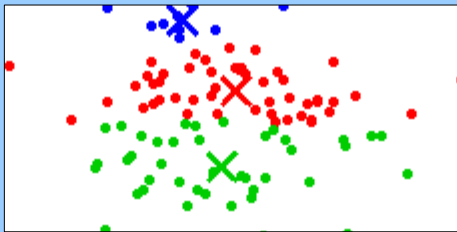
Cada ponto (flor) é associado ao cluster via representante mais próximo (distância euclidiana).  
Em seguida, os representantes são atualizados para um ponto médio do cluster.  
(Linhas 4 e 5 do pseudocódigo)

# Método para gerar os grupos

## ■ Explicação visual do método:



Usuário informa  $k=3$  grupos.  
Escolher 3 representantes iniciais.  
OBS: cada cor indica um grupo.



Cada ponto (flor) é associado ao cluster via representante mais próximo (distância euclidiana).  
Em seguida, os representantes são atualizados para um ponto médio do cluster.  
(Linhas 4 e 5 do pseudocódigo)

Essa etapa é repetida um número  $\underline{x}$  de vezes, em que  $\underline{x}$  é definido pelo usuário.



# Requisitos do trabalho

- O código deve ser organizado de forma modular, usando funções. Exemplo de funções:
  - Ler conjunto de dados
  - Inicializar representantes
  - Alocar ao grupo mais próximo
  - Atualizar representantes
  - Escrever resultados
- Todos os vetores/matrizes utilizados no trabalho devem ser alocados dinamicamente!
- Trabalho individual ou em duplas.
  - Em caso de dupla, ambos devem saber explicar o código durante a entrevista.