



ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO

## PQI 3403 Análise de Processos da Indústria Química

Ardson dos Santos Vianna Júnior - ASVJ  
e-mail: [ardson@usp.br](mailto:ardson@usp.br)





ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO

# Aula Machine learning

PQI 3403 Análise de Processos da Indústria Química



# Sumário

1. Introdução
2. Algoritmos de agrupamento
3. Conclusão

# Algoritmos de agrupamento

- Algoritmos hierárquicos
  - Algoritmos baseados nas conectividades
- Algoritmos baseados em centroides
- Algoritmos baseados em distribuição
- Algoritmos baseados em densidade

# Algoritmos de agrupamento

- Algoritmos hierárquicos (a ver com árvores! dendogramas)
  - Agrupamento baseado em conectividade diferem pela forma como são calculadas as distâncias
  - Single-linkage clustering
  - complete-linkage clustering
  - UGPMA ou WPGMA (unweighted or weighted pair group with arithmetic mean)

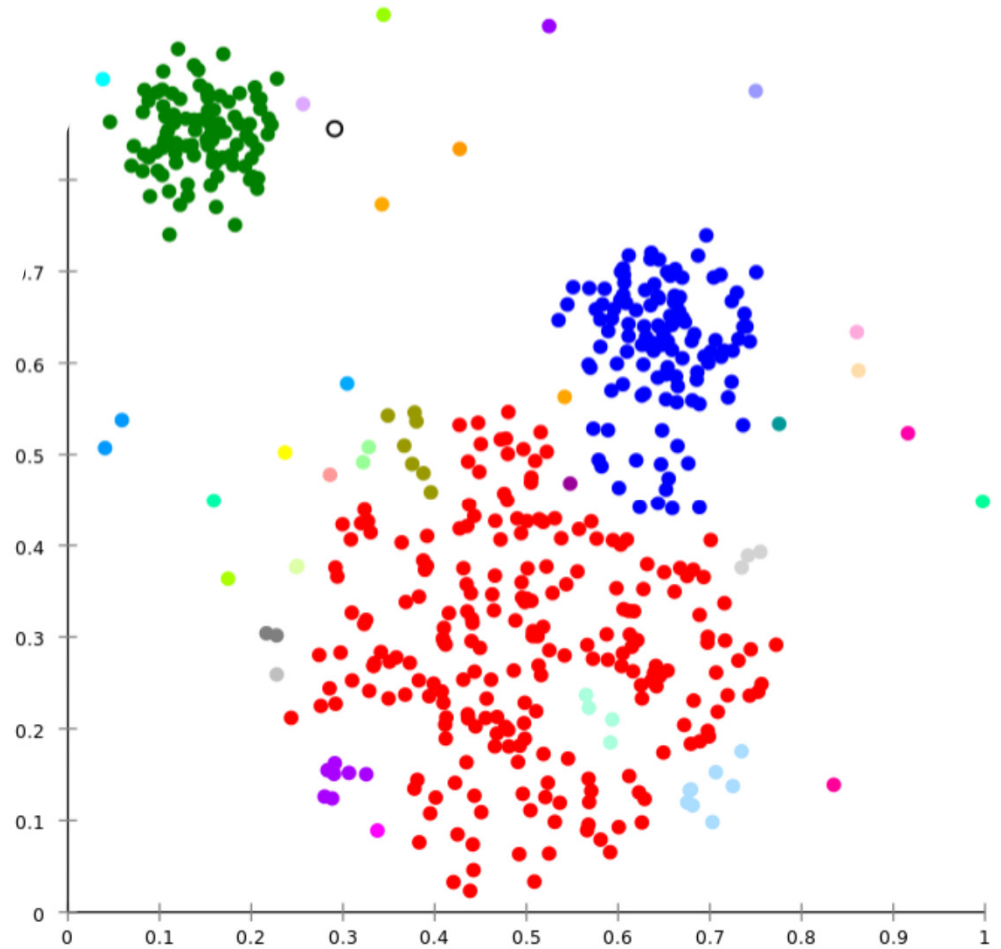
# Algoritmos de agrupamento

- Algoritmos hierárquicos - Agrupamento baseado em conectividade
  - Determinístico
  - Flexível com relação ao critério de similaridades
  - Lento
  - Algoritmo naife  $n^3$ .
  - Algoritmos  $n^2$  a partir de algum critério de ligação

# Algoritmos de agrupamento

---

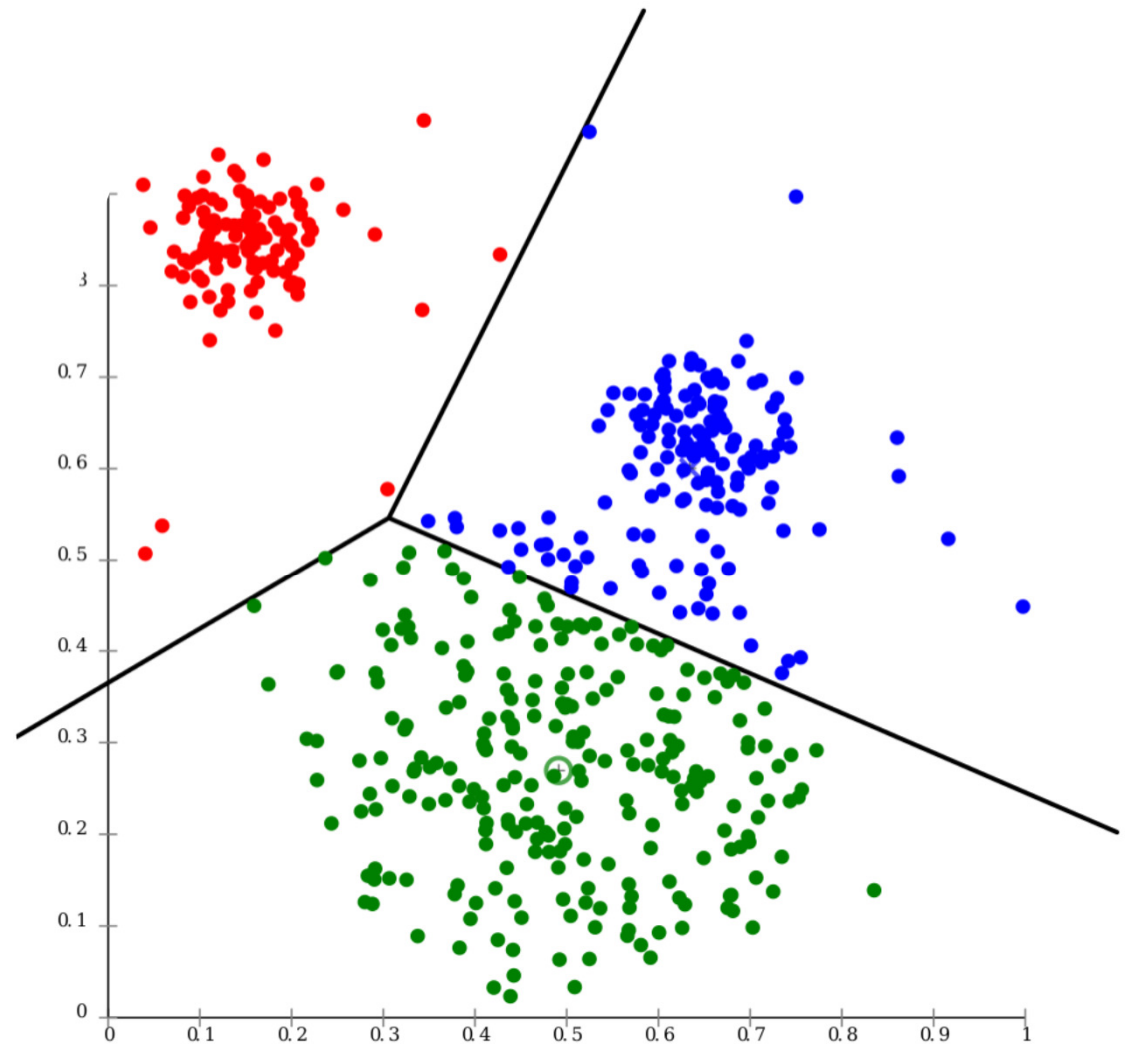
- Algoritmos hierárquicos (a ver com árvores!)
  - Agrupamento baseado em conectividade diferem pela forma como são calculadas as distâncias
  - Single-linkage clustering -  $O(n^3)$



# Algoritmos de agrupamento

---

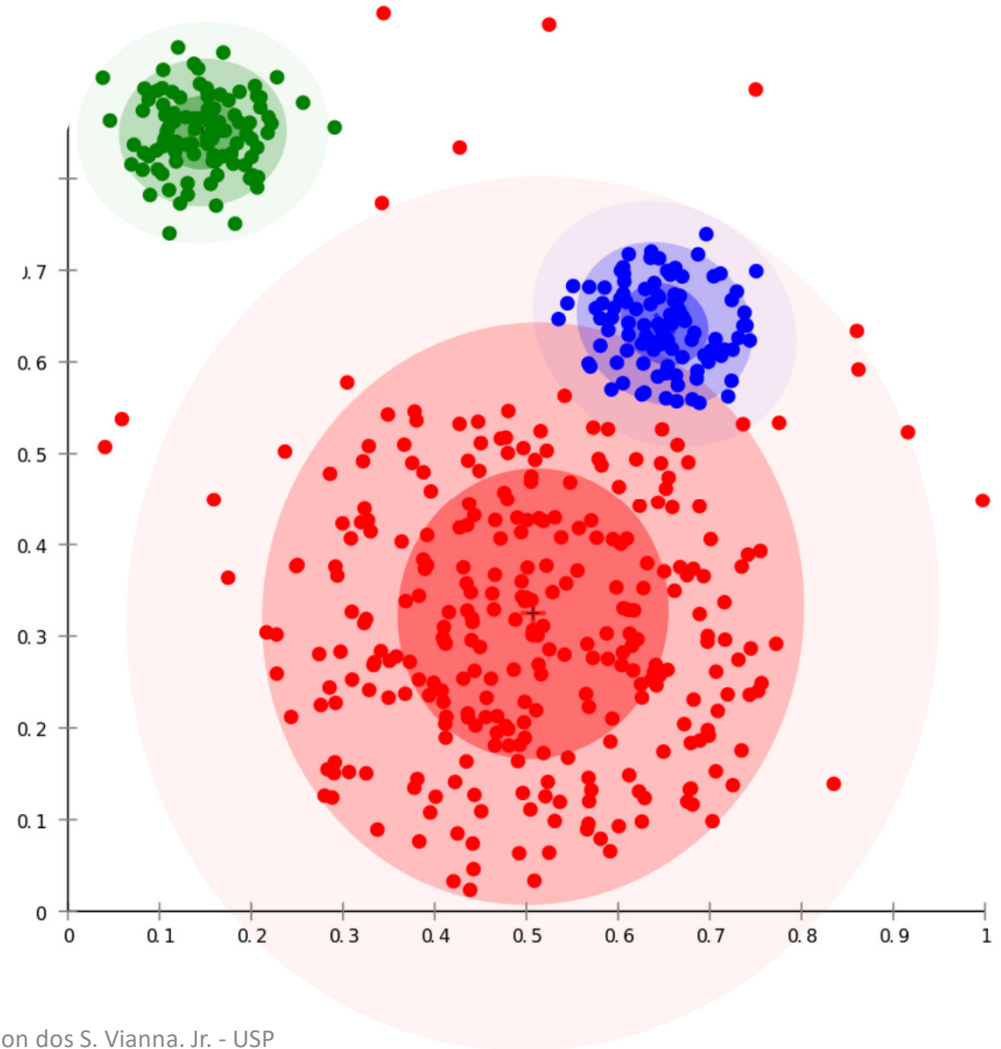
- Algoritmos baseados em centróide
- K médio: muito mais rápido, algoritmo guloso (*greedy*)
  - Mais útil quando se sabe quantos clusters são necessários





# Algoritmos de agrupamento

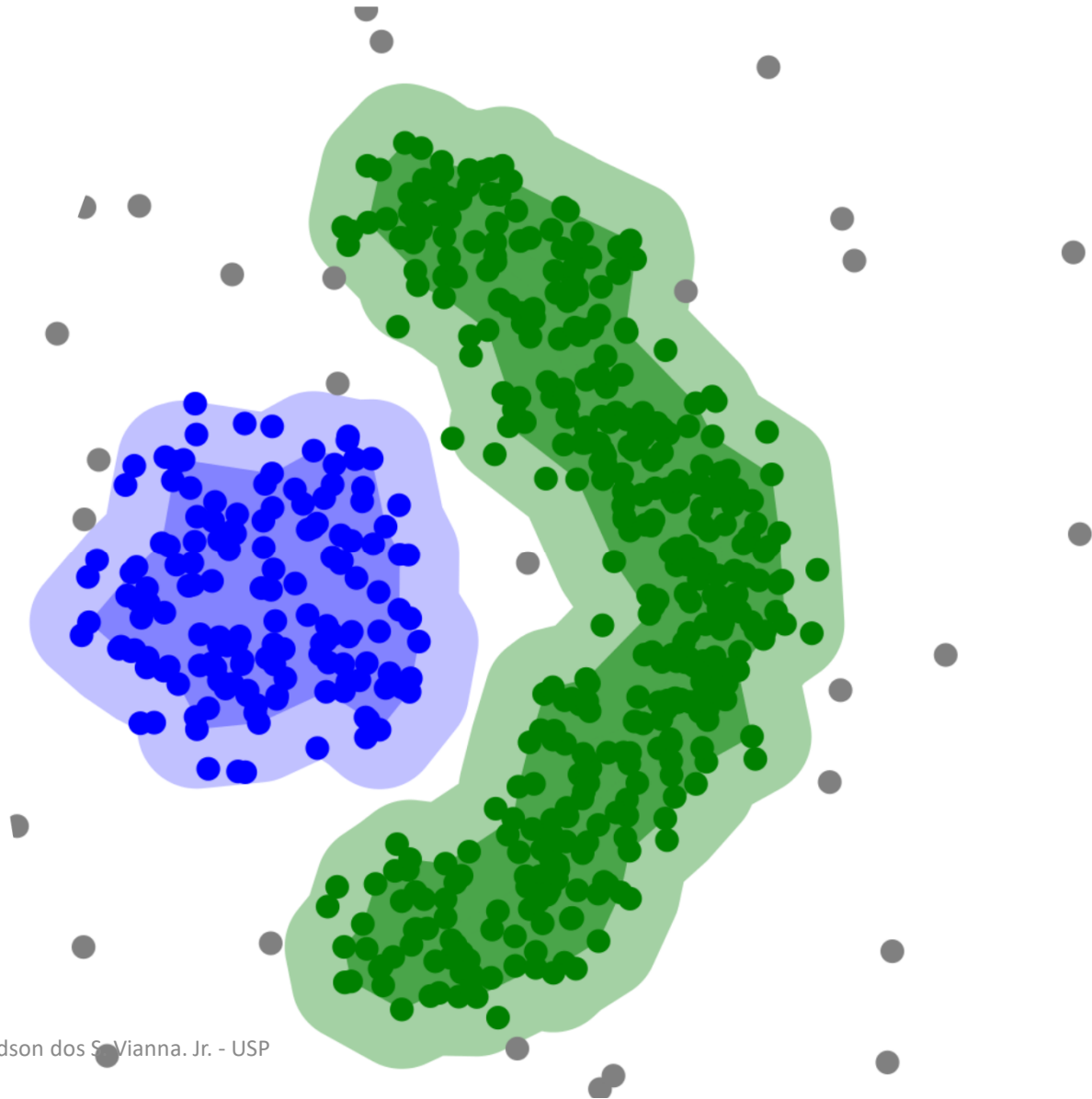
- Algoritmos baseados em distribuição
- objetos pertencem a uma mesma distribuição
- Misturas de distribuição de Gauss



# Algoritmos de agrupamento

---

- Algoritmos baseados em densidades
- áreas esparsas são consideradas como ruídos
- Método DBSCAN (Density-based spatial clustering of applications with noise)



# Conclusões

- Não tem uma solução específica,
- Encontrar o algoritmo faz parte da tarefa (Wikipedia)
- Processo iterativo onde se descobre o conhecimento.

# Bibliografia

- [https://en.wikipedia.org/w/index.php?title=Cluster\\_analysis&oldid=1007902291](https://en.wikipedia.org/w/index.php?title=Cluster_analysis&oldid=1007902291)