



# APRENDIZADO EM MÁQUINAS — 1D

*Machine Learning — 1D*

*Prof. Dr. Waldemar Bonventi Jr.*

*mar2020*



# APLICAÇÃO DO ALGORITMO K-MEANS

## ESTUDO DE CASO

The “famous” dataset IRIS. O conjunto de dados Iris de Fisher é um conjunto de dados multivariado introduzido pelo estatístico e biólogo britânico Ronald Fisher em seu artigo de 1936.



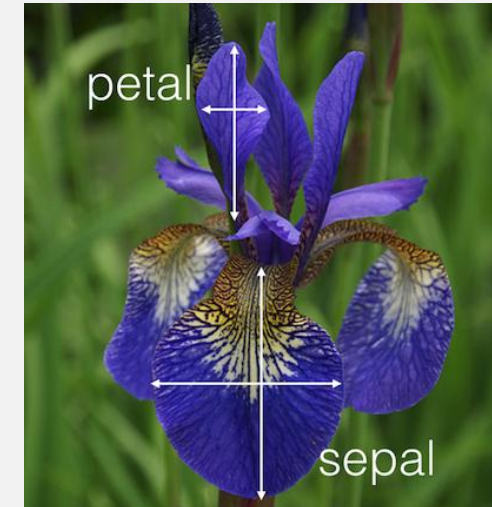
Iris setosa



Iris versicolor



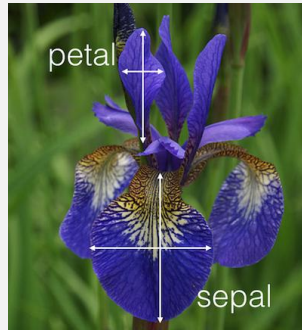
Iris virginica



4 atributos:  
2 comprimentos  
e 2 larguras



# IRIS DATASET



sepal length   sepal width   petal length   petal width   class

5.1,	3.5,	1.4,	0.2,	Iris-setosa
4.9,	3.0,	1.4,	0.2,	Iris-setosa
4.7,	3.2,	1.3,	0.2,	Iris-setosa
4.6,	3.1,	1.5,	0.2,	Iris-setosa
...				
7.0,	3.2,	4.7,	1.4,	Iris-versicolor
6.4,	3.2,	4.5,	1.5,	Iris-versicolor
6.9,	3.1,	4.9,	1.5,	Iris-versicolor
5.5,	2.3,	4.0,	1.3,	Iris-versicolor
...				
6.3,	3.3,	6.0,	2.5,	Iris-virginica
5.8,	2.7,	5.1,	1.9,	Iris-virginica
7.1,	3.0,	5.9,	2.1,	Iris-virginica
6.3,	2.9,	5.6,	1.8,	Iris-virginica

50

50

50

Agrupamento → processo não supervisionado

Verificar se o k-means consegue separar as três classes

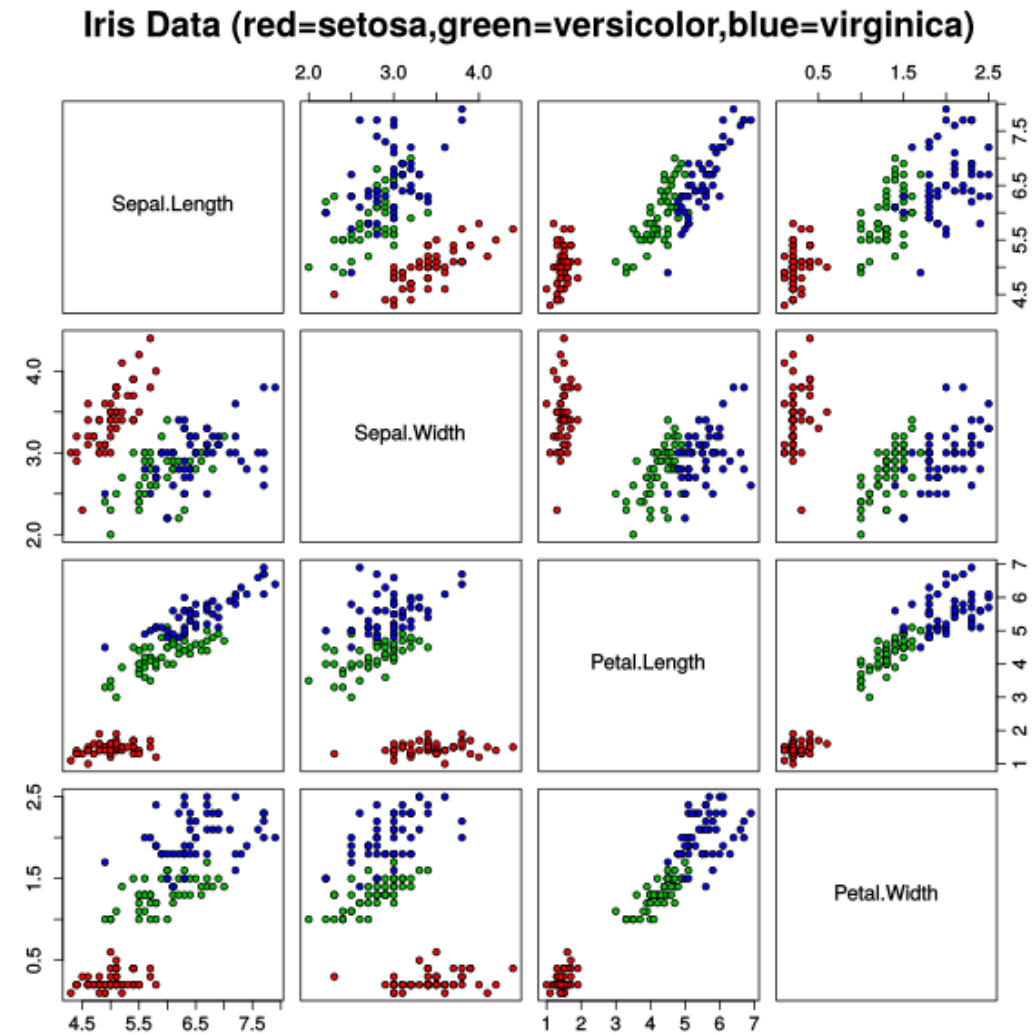
sepal length	sepal width	petal length	petal width
5.1	3.5	1.4	0.2
4.9	3.0	1.4	0.2
4.7	3.2	1.3	0.2
4.6	3.1	1.5	0.2
5.0	3.6	1.4	0.2
5.4	3.9	1.7	0.4
4.6	3.4	1.4	0.3
5.0	3.4	1.5	0.2
4.4	2.9	1.4	0.2
4.9	3.1	1.5	0.1



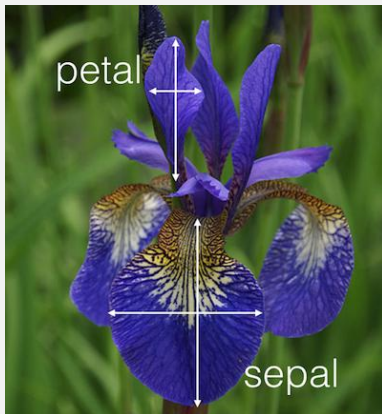
# AGRUPAMENTO

Análise exploratória

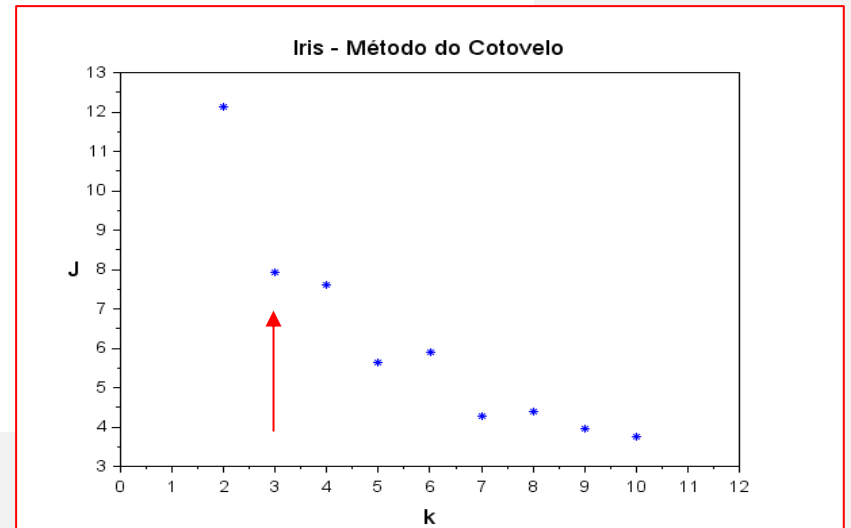
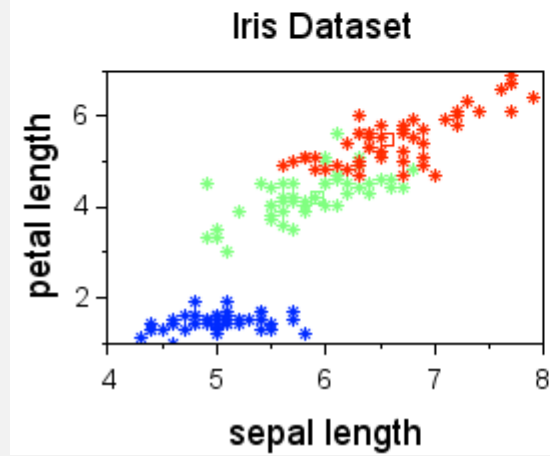
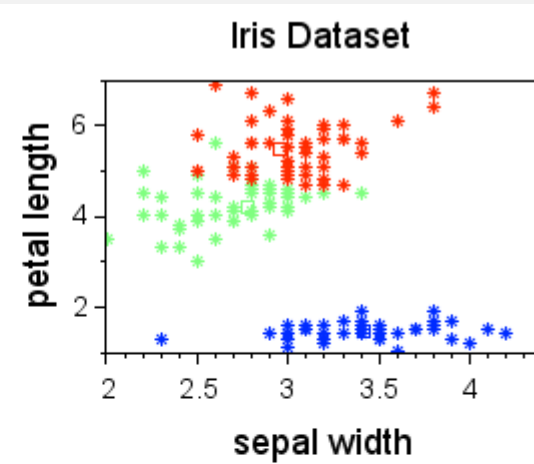
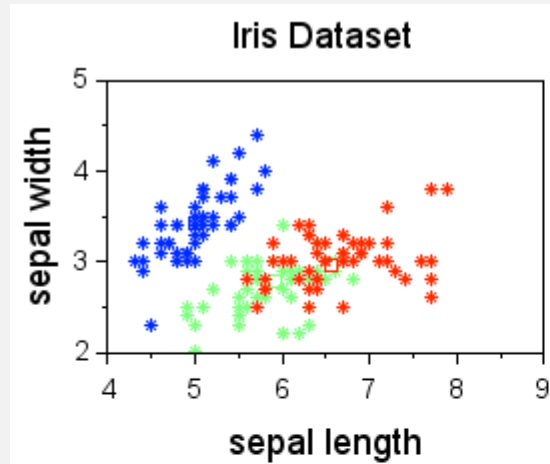
Gráficos. Ex. “pair plot”



# NORMALIZAÇÃO E AGRUPAMENTO



sepal_length	sepal_width	petal_length	petal_width	SLn	SWn	PLn	PWn
5,1	3,5	1,4	0,2	0,222	0,625	0,068	0,042
4,9	3	1,4	0,2	0,167	0,417	0,068	0,042
4,7	3,2	1,3	0,2	0,111	0,500	0,051	0,042
4,6	3,1	1,5	0,2	0,083	0,458	0,085	0,042
5	3,6	1,4	0,2	0,194	0,667	0,068	0,042
5,4	3,9	1,7	0,4	0,306	0,792	0,119	0,125





# CLASSIFICAÇÃO

Matriz de confusão

Grupo	Setosa	Versicolor	Virginica
1	50	0	0
2	0	45	1
3	0	5	49

Acertos (acurácia) =  $(50+45+49)/150 = 96\%$

Setosa  $\rightarrow 50/50 = 100\%$

Versicolor  $\rightarrow 45/50 = 90\%$

Virginica  $\rightarrow 49/50 = 98\%$

sepal_length	sepal_width	petal_length	petal_width	class	grupo
5,1	3,5	1,4	0,2	Iris-setosa	1
4,9	3,0	1,4	0,2	Iris-setosa	1
4,7	3,2	1,3	0,2	Iris-setosa	1
4,6	3,1	1,5	0,2	Iris-setosa	1
5,0	3,6	1,4	0,2	Iris-setosa	1
5,4	3,9	1,7	0,4	Iris-setosa	1
5,8	2,7	4,1	1,0	Iris-versicolor	2
6,2	2,2	4,5	1,5	Iris-versicolor	3
5,6	2,5	3,9	1,1	Iris-versicolor	2
5,9	3,2	4,8	1,8	Iris-versicolor	3
6,1	2,8	4,0	1,3	Iris-versicolor	2
6,3	2,5	4,9	1,5	Iris-versicolor	3
6,6	3,0	4,4	1,4	Iris-versicolor	2
6,8	2,8	4,8	1,4	Iris-versicolor	2
6,7	3,0	5,0	1,7	Iris-versicolor	3
6,0	2,9	4,5	1,5	Iris-versicolor	2
5,7	2,6	3,5	1,0	Iris-versicolor	2
6,0	2,7	5,1	1,6	Iris-versicolor	3
5,4	3,0	4,5	1,5	Iris-versicolor	2
6,0	3,4	4,5	1,6	Iris-versicolor	2
7,2	3,0	5,8	1,6	Iris-virginica	3
7,4	2,8	6,1	1,9	Iris-virginica	3
7,9	3,8	6,4	2,0	Iris-virginica	3
6,4	2,8	5,6	2,2	Iris-virginica	3
6,3	2,8	5,1	1,5	Iris-virginica	2
6,1	2,6	5,6	1,4	Iris-virginica	3
7,7	3,0	6,1	2,3	Iris-virginica	3



# ESTUDO DE CASO 2

## ANÁLISE DO ARRANJO EMPRESARIAL DE UMA COOPERATIVA DE APICULTORES DA REGIÃO DE SOROCABA-SP PELO MÉTODO DE CLUSTERING

SANDRO EDUARDO DE SOUZA; WALDEMAR BONVENTI JR.; DANIEL BERTOLI GONÇALVES (2017)

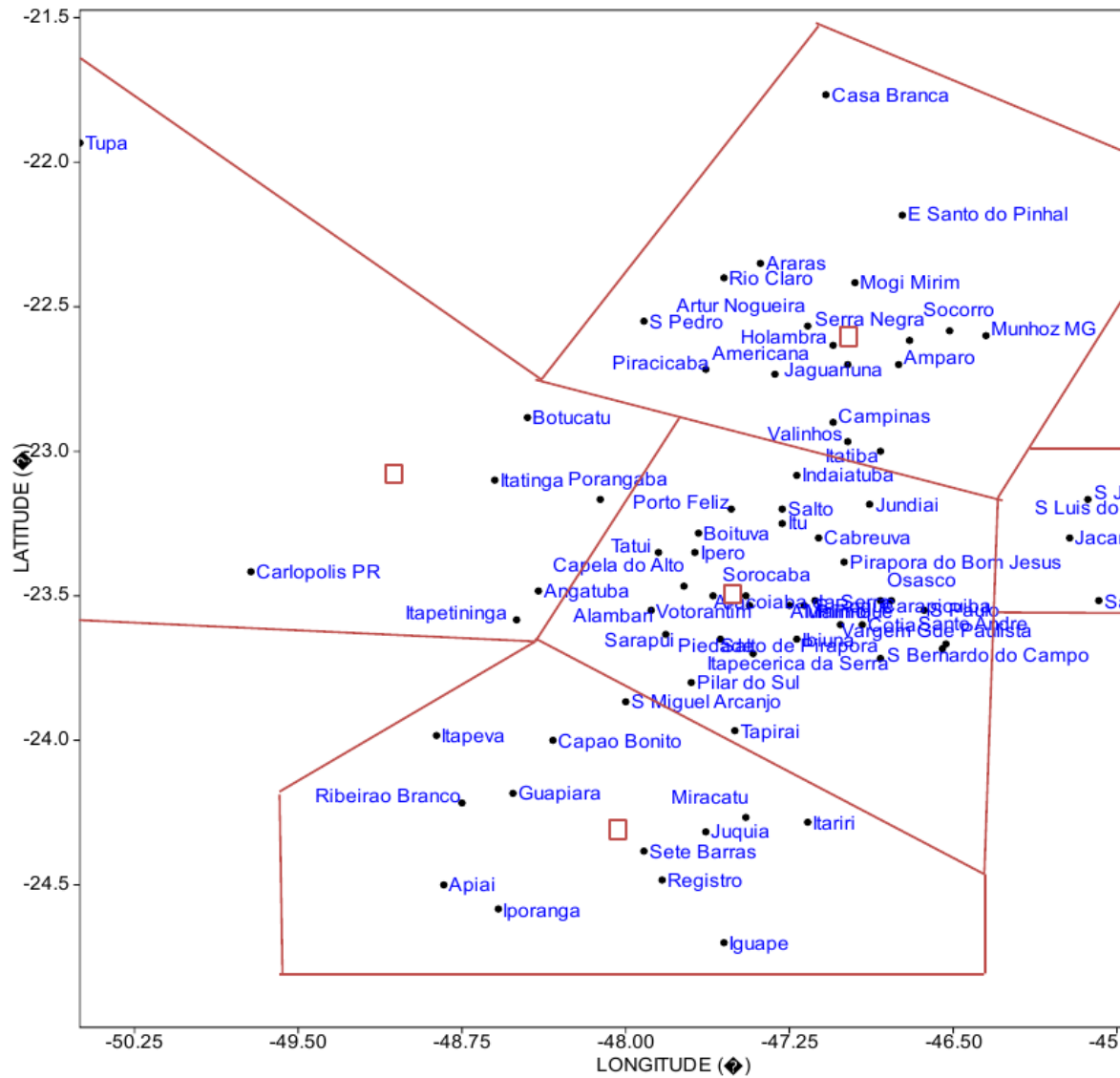
Definir oportunidades de redução de custos na rede de distribuição de mel de uma Cooperativa de Apicultores de Sorocaba-SP, da qual participam 357 cooperados distribuídos em 75 municípios, em um raio de atuação de mais de 200 km.

Dados:

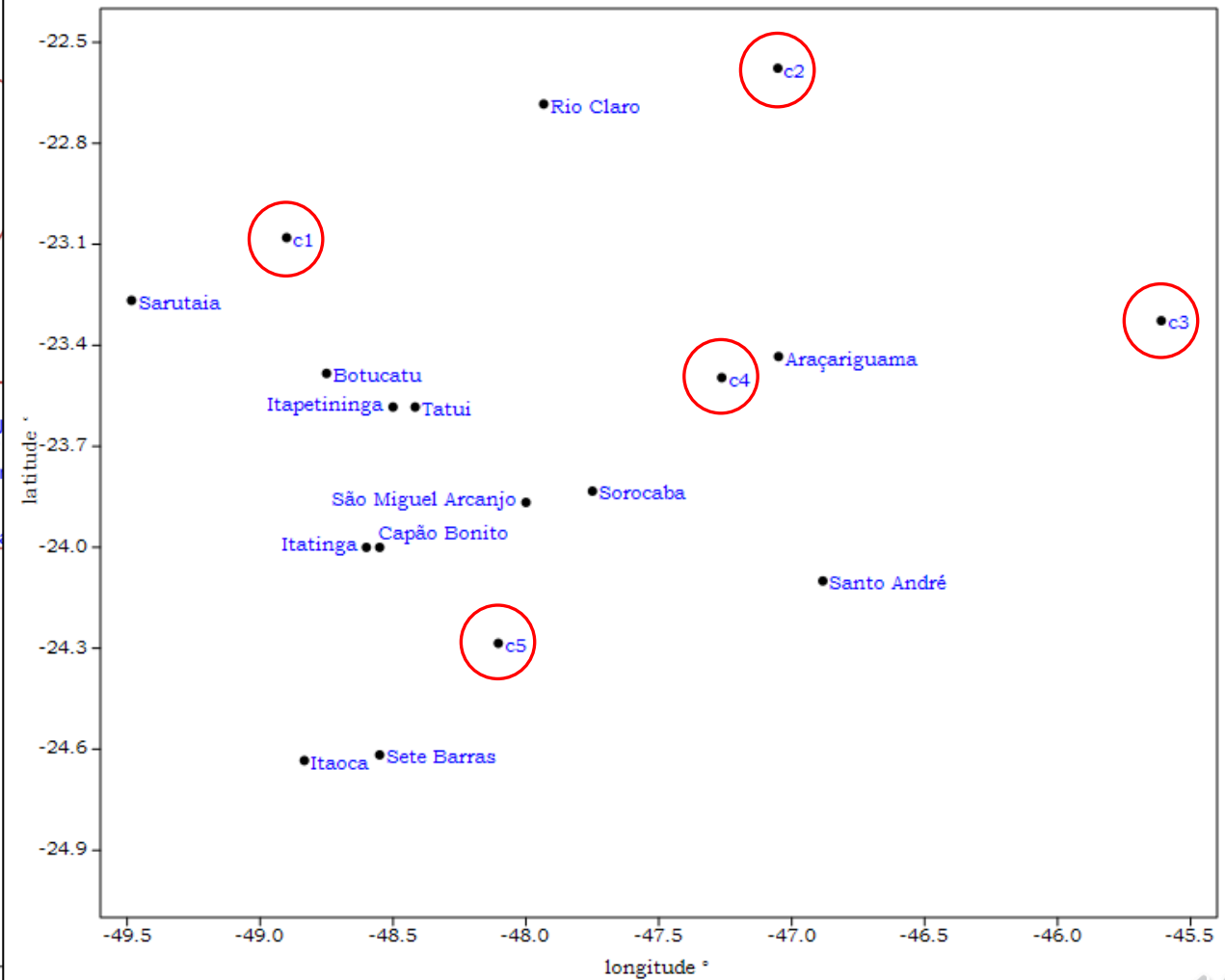
- Latitude do cooperado
- Longitude do cooperado
- Latitude do centro consumidor
- Longitude do centro consumidor
- Demanda do centro consumidor



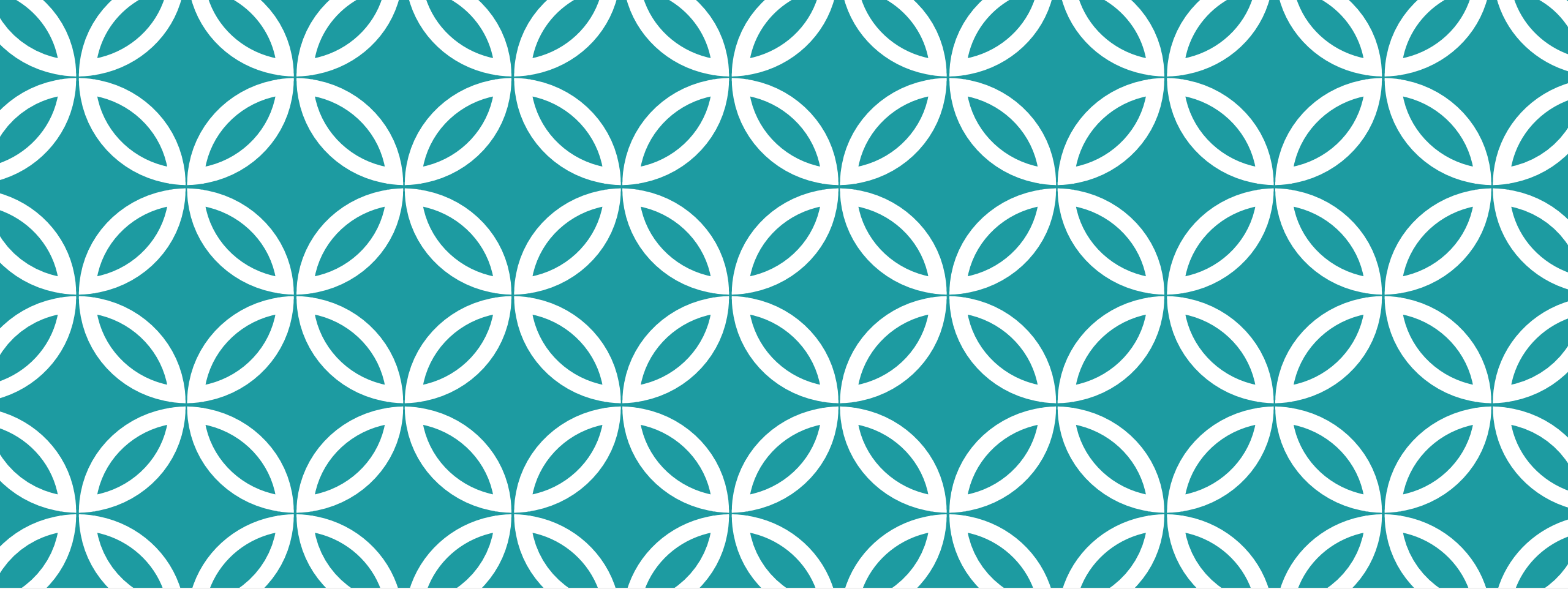
Grupos de cooperados obtidos pela execução do algoritmo k-means



Localização dos clientes em relação aos centros distribuidores propostos







**ATÉ A PRÓXIMA AULA!**

Aprendizado supervisionado

