



# Classificação x Clusterização

## Classificação x Clusterização

Ambas as tarefas trabalham com o agrupamento dos padrões de comportamento, mas com algumas diferenças:

- A classificação reúne os dados em categorias;
- A clusterização reúne os dados por similaridade, sem a definição de categorias.

## Classificação

A classificação trabalha com aprendizado supervisionado, o que significa que ela depende da existência de uma base de dados, com registros e rótulos, que tenha muita representatividade para que seja possível criar um bom modelo para classificação.

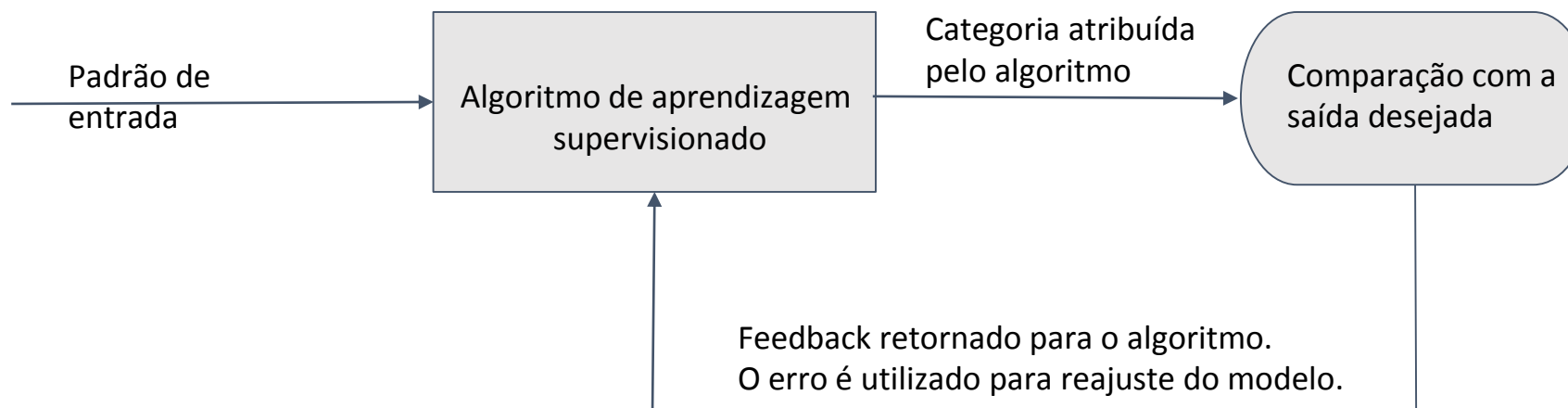
## Base de dados

O que é uma base representativa?

- Uma base representativa é aquela que tem uma boa quantidade de registros de cada uma das categorias do problema.

Isso significa que técnicas para classificação só devem ser aplicadas a problemas que contam com uma boa disponibilidade de dados para treinamento.

## Aprendizado supervisionado



## Clusterização

Na clusterização é utilizado o aprendizado do tipo não supervisionado. O modelo é criado com dados que não contêm rótulo, não sendo possível verificar o aprendizado à medida que ele vai acontecendo.

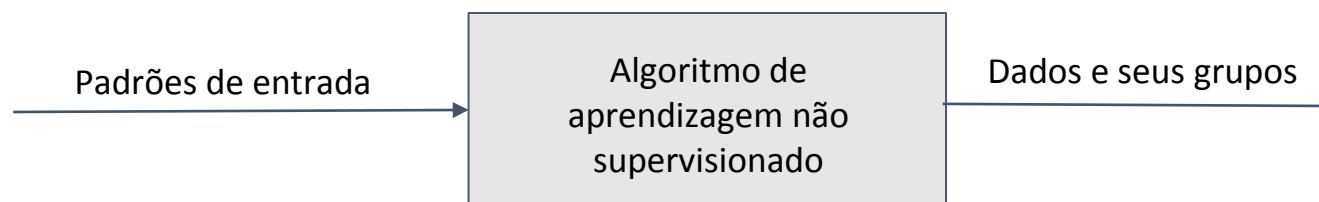
A clusterização então é uma opção mais interessante quando não se tem uma base de dados com rótulos para a realização da etapa de treinamento.

## Abordagens de clusterização

Para a clusterização existem algoritmos que trabalham de formas diferentes:

- Pode haver a necessidade de informar a quantidade de grupos ou não;
- Podem utilizar diferentes métricas para determinar a similaridade;
- As métricas mais comuns são as de densidade ou distância.

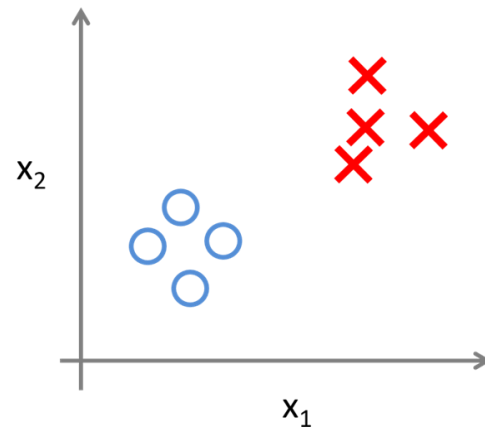
## Aprendizado não supervisionado



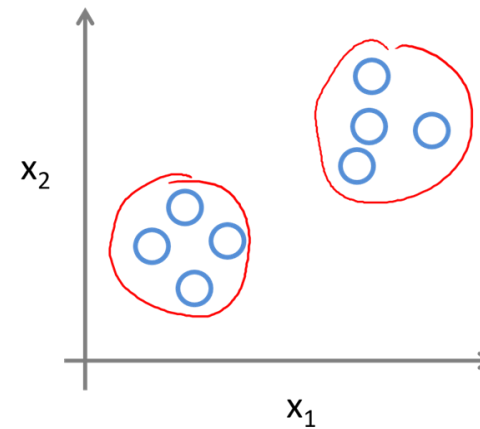


## Aprendizado não supervisionado

Supervised Learning



Unsupervised Learning



Fonte: Barros (2016)

## Referências

BARROS, P.. Aprendizagem de Máquina: Supervisionada ou Não Supervisionada?. 2016. Disponível em: <<https://medium.com/opensanca/aprendizagem-de-maquina-supervisionada-ou-n%C3%A3o-supervisionada-7d01f78cd80a>>. Acesso em: 20. Jan. 2020.



Obrigada!

Ana Laurentino