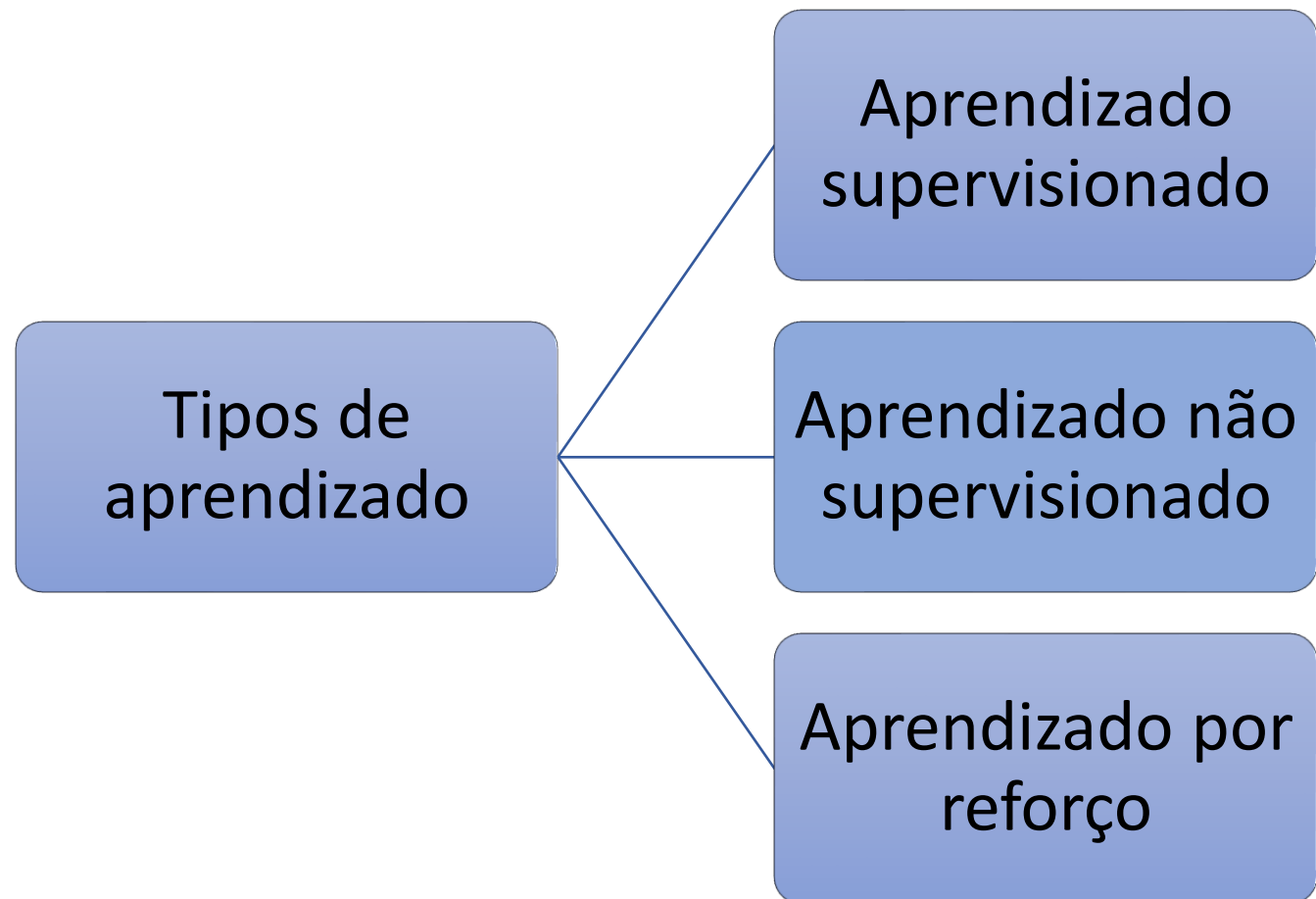


# Aprendizado de Máquina e Deep Learning

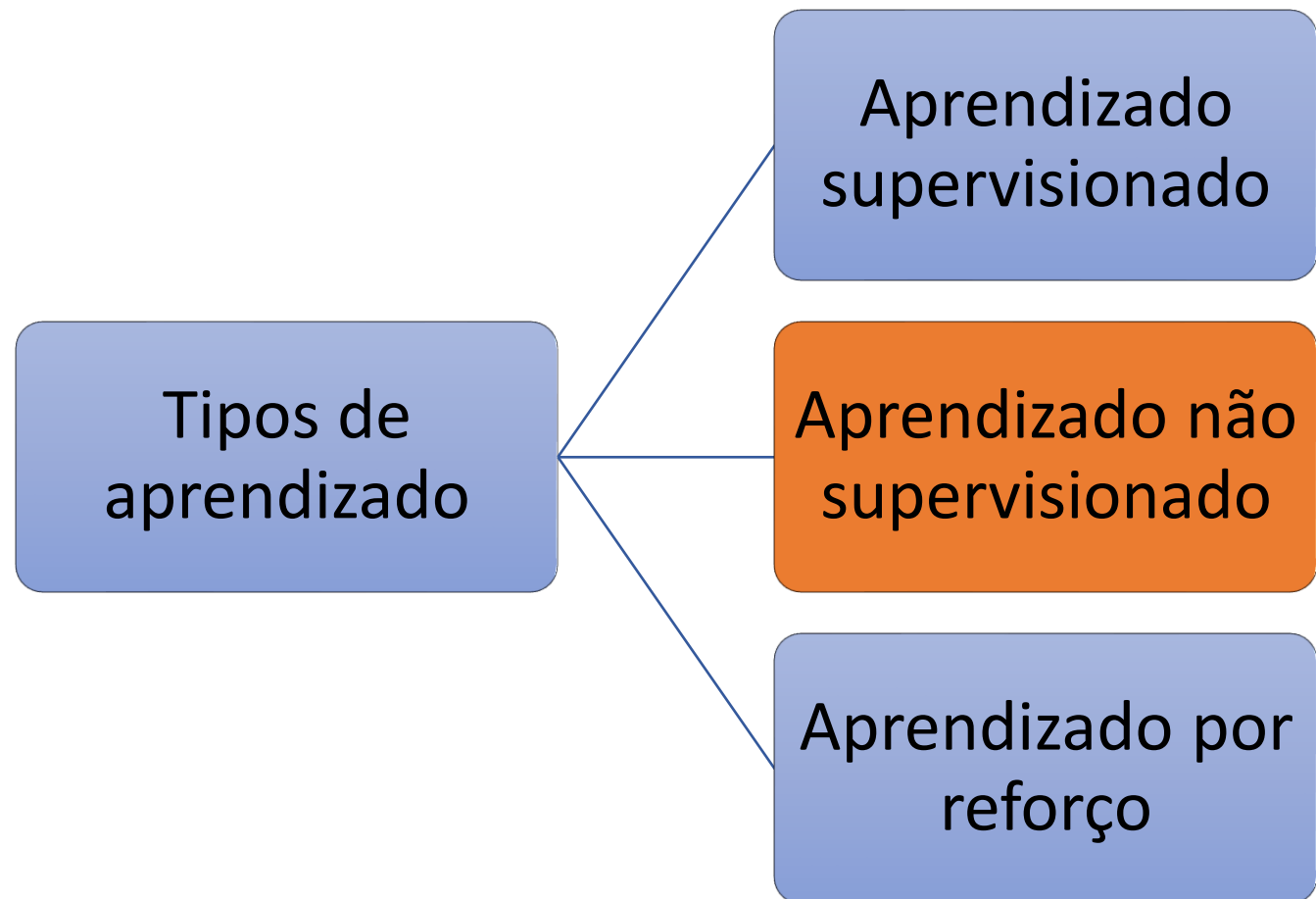
## Aprendizado não supervisionado

Prof. Dr. Thiago Meirelles Ventura

# Tipos de aprendizado



# Tipos de aprendizado



# Aprendizado não supervisionado

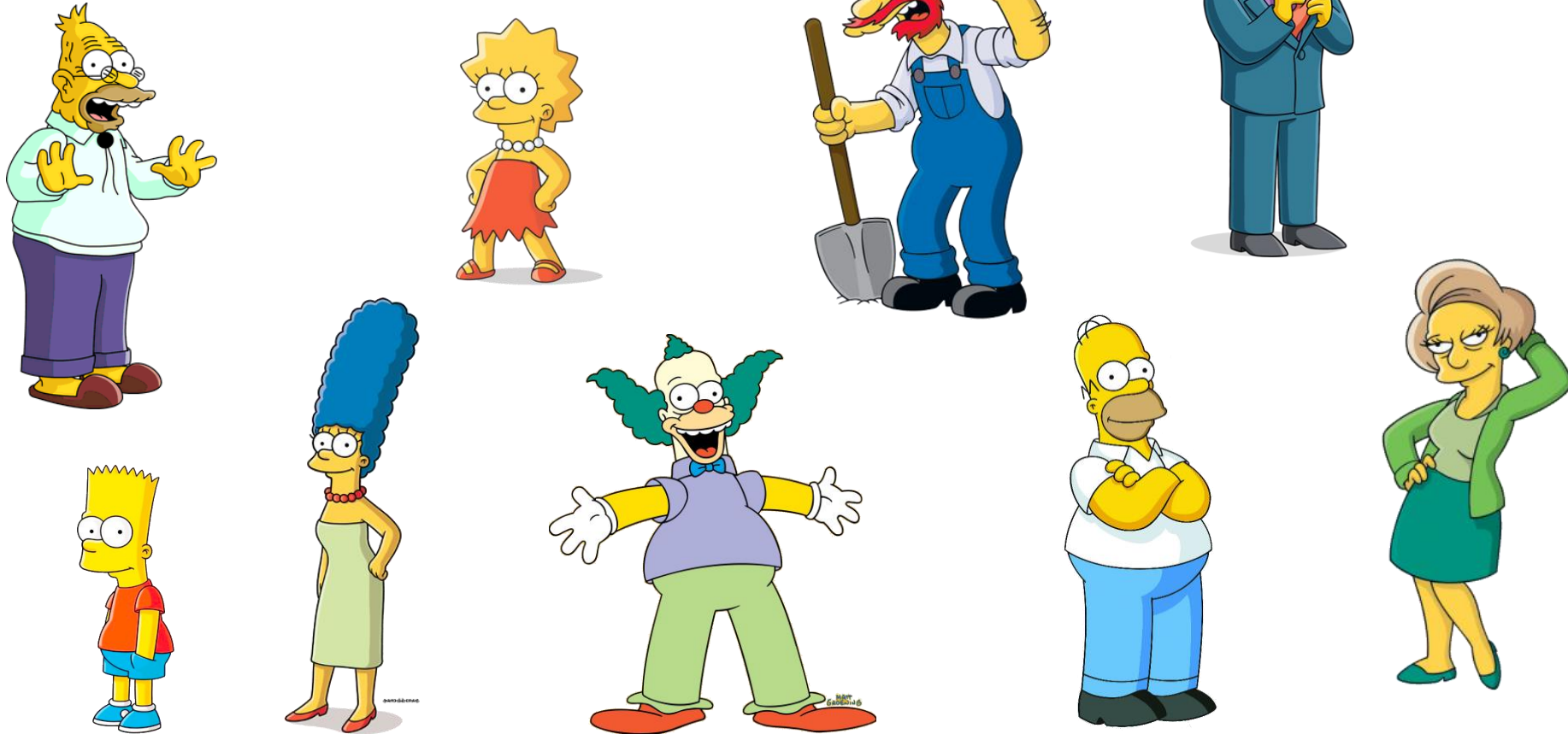
- Ao contrário do aprendizado supervisionado, no aprendizado não-supervisionado não há
  - uma classe (como na classificação) ou
  - valor alvo (como na regressão)
- Os rótulos das classes dos dados de treinamento são desconhecidos

# Aprendizado não supervisionado

- Analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados, formando clusters
- Após os agrupamentos, é necessário uma análise para determinar o que cada grupo significa no contexto do problema que está sendo analisado

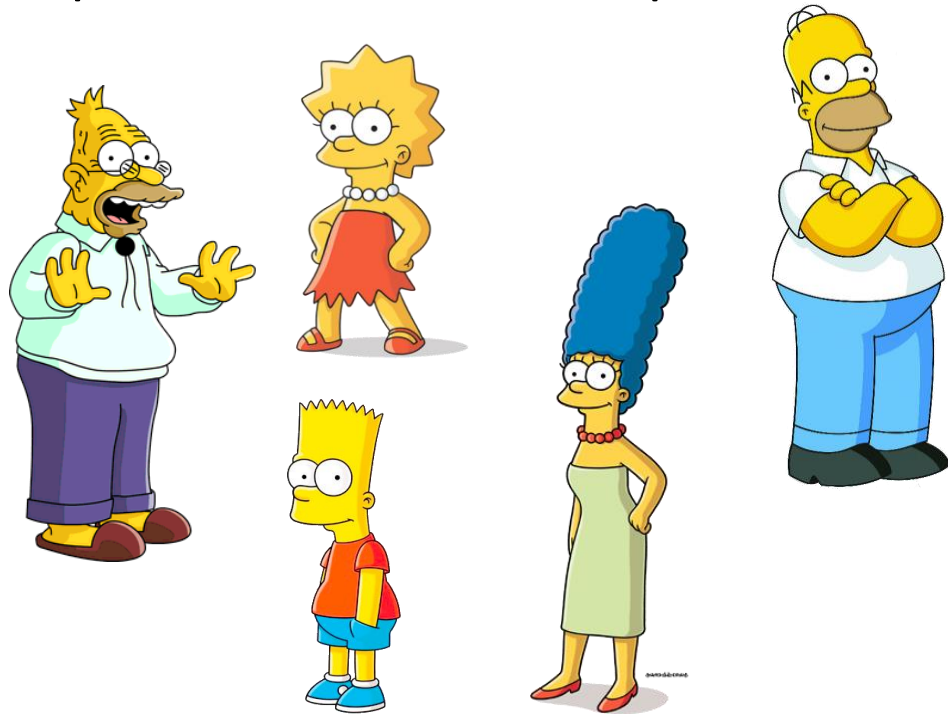
# Agrupamento

- Como agrupar esses exemplos?



# Agrupamento

- Como agrupar esses exemplos?
  - Se pertence a família Simpsom



# Agrupamento

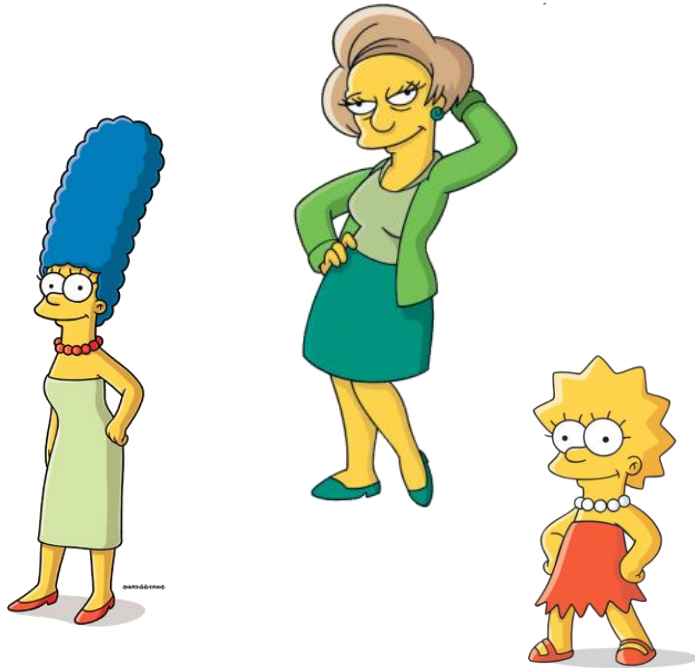
- Como agrupar esses exemplos?
  - Se trabalha na escola





# Agrupamento

- Como agrupar esses exemplos?
  - Por gênero



# Agrupamento

O que é necessário para realizar os agrupamentos?

# Agrupamento

- A definição da característica que estará sendo analisado para realizar a separação
- Pode ser mais de uma característica, fazendo grupos mais específicos
- Deve haver similaridade em algum nível
- A meta é estabelecer a existência de grupos por meio de alguma medida que faça sentido para a resolução do problema proposto

# Similaridade

- É a qualidade, caráter ou condição das coisas similares
- As imagens abaixo são similares?



# Similaridade

- Depende do que você está analisando
  - Se for pelas cores, são similares
  - Se for pelo que representa o conteúdo da imagem, temos uma diferença muito grande
- É necessário ter uma forma para mensurar a similaridade de objetos complexos



# Similaridade

- É fácil comparar objetos que tem uma ordem pré-estabelecida, como números e letras

-5	-4	-3	-2	-1	0	1	2	3	4	5			
g	h	i	j	k	l	m	n	o	p	q	r	s	t

# Similaridade

-5	-4	-3	-2	-1	0	1	2	3	4	5			
g	h	i	j	k	l	m	n	o	p	q	r	s	t

- Sabemos a ordem
- Sabemos o quão distante um objeto está do outro
  - Quanto mais próximo, mais similar

# Exercício 1

- Considere esses dados:
  - Eles representam características de transações de cartão de crédito
- Insira-os em um gráfico 3D:  
<https://technology.cpm.org/general/3dgraph>
- Qual registro pode ser uma fraude?

#	A	B	C
1	9	7	8
2	8	2	4
3	8	9	9
4	4	2	6
5	6	7	5
6	10	7	7
7	10	8	5
8	6	5	8
9	9	3	6
10	9	10	1



# Similaridade

Mas como comparar objetos complexos?

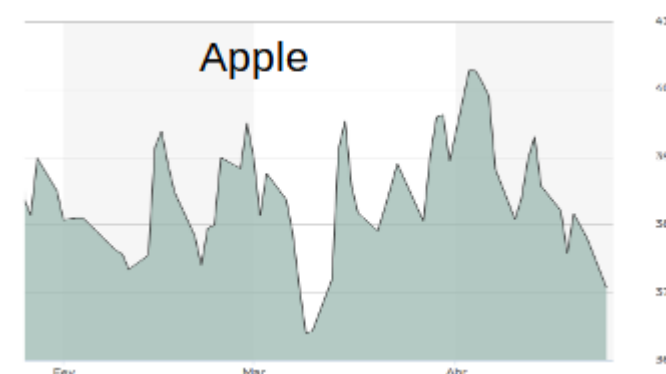
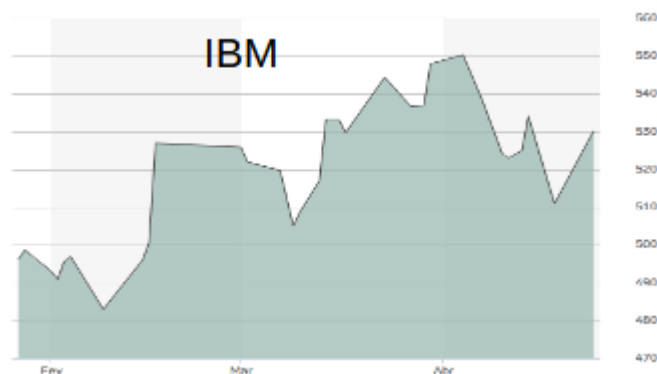
# Comparação de objetos complexos

- Cotação na bolsa de valores da Petrobras, Apple e IBM
- Qual empresa é mais similar à IBM?

Petr	Apple	IBM
9,33	37,07	530,28
9,75	37,81	511
9,48	38,16	534,1
9,62	37,58	525,04
9,24	38,21	523,09
9,69	38,57	524,22
9,16	39,3	538,15
9,51	38,97	550,39
9,03	38,39	547,98
8,39	38,08	537
8,26	38,82	536,75
7,7	39,9	544,45
7,58	40,1	529,74
7,83	40,27	533,29
7,58	40,28	533,29
8,36	38,93	517,05
8,35	39,62	508,7
8,44	39,59	505,1

# Comparação de objetos complexos

- Cotação na bolsa de valores da Petrobras, Apple e IBM
- Qual empresa é mais similar à IBM?



# Comparação de objetos complexos

- E como comparar
  - Imagens
  - Vídeos
  - Sons
  - ...

# Função de distância

- Responsável por medir a similaridade entre dois objetos
- Calculada com base em um conjunto de características que são extraídas a partir do objeto
- Tem seu uso mais efetivo quando ocorre especificamente sobre as características em que ela faz uma melhor discriminação

# Função de distância

- Podemos imaginar a função de distância da seguinte forma:

```
double funcaoDistancia (Object a, Object b) {  
  
    ...  
    return distancia;  
  
}
```

# Função de distância

- Quanto menor o valor de distância, mais similares são os objetos
- Normalmente, os objetos a serem comparados estão representados como vetores de características

# Função de distância

- Vetor de características
  - Estrutura que armazena as características de um objeto complexo
  - Utilizado como parâmetro para o cálculo da função de distância



-5	9	1	0	2	7	-3	4	0	1
----	---	---	---	---	---	----	---	---	---



8	-2	6	3	0	4	2	-1	1	0
---	----	---	---	---	---	---	----	---	---



# Função de distância

- Vetor de características
  - Os valores dependerá das técnicas utilizadas e do domínio dos dados
  - Com o objetivo de classificar gatos, seria mais importante extrair as cores das imagens ou tentar identificar a presença de orelhas, rabo, etc?

# Função de distância

- Vetor de características
  - Os valores devem referenciar os atributos mais importantes dos objetos que estão tentando medir as similaridades
  - Depois que os vetores de características foram extraídos, é possível aplicar alguma função de distância

# Função de distância

- Exemplos

- Euclidiana

$$E(x, y) = \sqrt{\sum_{a=1}^n (x_a - y_a)^2}$$

- Manhattan

$$D(x, y) = \sum_{a=1}^n |x_a - y_a|$$

- Chebychev

$$D(x, y) = \max_{a=1}^n |x_a - y_a|$$

- Camberra

$$D(x, y) = \sum_{a=1}^n \frac{|x_a - y_a|}{|x_a + y_a|}$$

# Exercício 2

- Considere três objetos representados por esses vetores de características
  - Objeto 1: [2, 4, 6]
  - Objeto 2: [3, 3, 3]
  - Objeto 3: [1, 5, 4]
- Utilize a função de distância Manhattan para responder
  - Qual a distância entre os objetos 1 e 2?
  - Qual a distância entre os objetos 1 e 3?
  - Qual objeto é o mais similar ao objeto 1?

$$D(x, y) = \sum_{a=1}^n |x_a - y_a|$$