

# INF721

2023/2



# Aprendizado em Redes Neurais Profundas

## A2: Aprendizado de Máquina

# Logística

## Avisos

- ▶ Aula A1 – Introdução publicada no site [slides, vídeo]

## Última aula

- ▶ Organização da disciplina
- ▶ Visão geral de aprendizado de máquina redes neurais

# Plano de Aula

- ▶ Aprendizado de Máquina
- ▶ Tipos de Aprendizado
- ▶ Tipos de Dados
- ▶ Espaço de hipóteses
- ▶ Generalização

# Computação Clássica x Aprendizado de Máquina



**Computação  
Clássica**

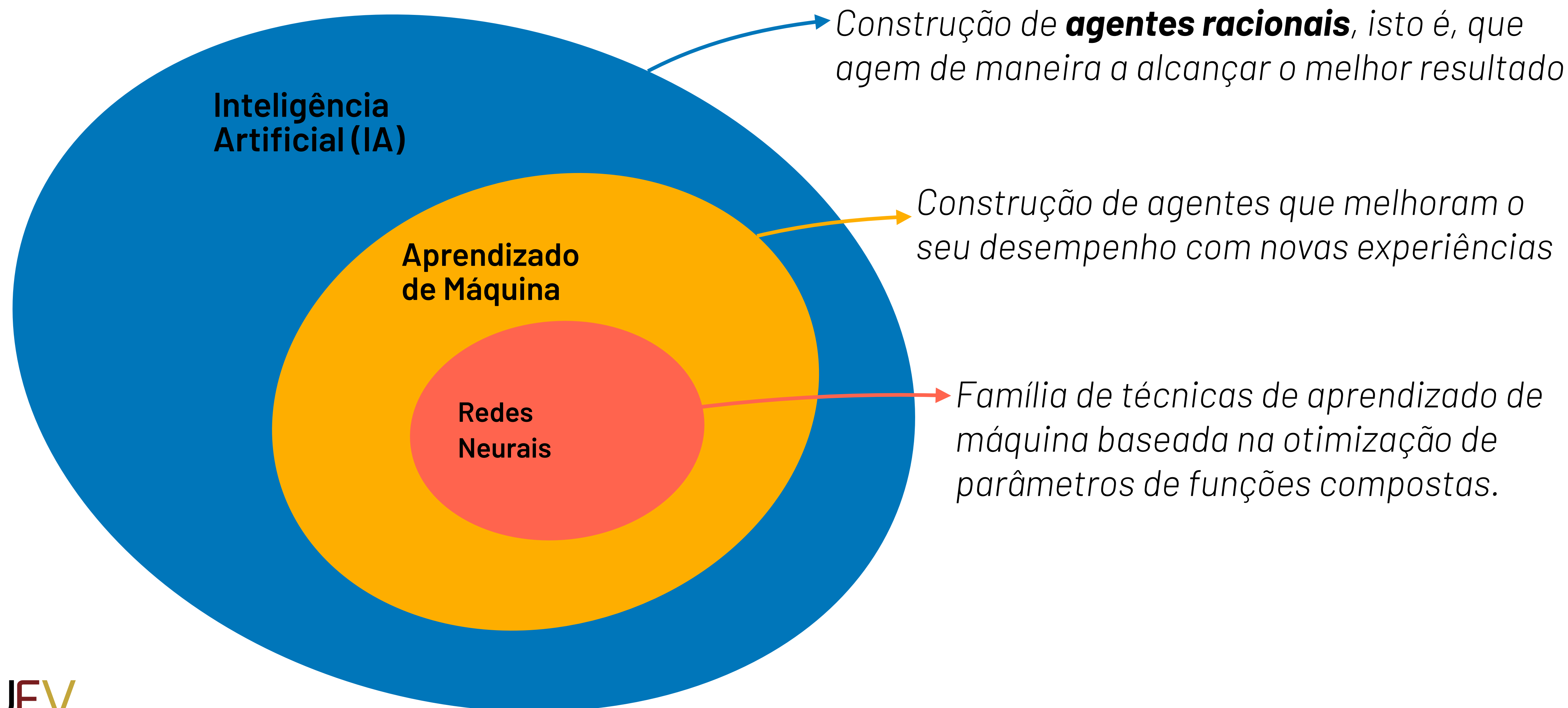
*Funções programadas  
explicitamente*



**Aprendizado  
de Máquina**

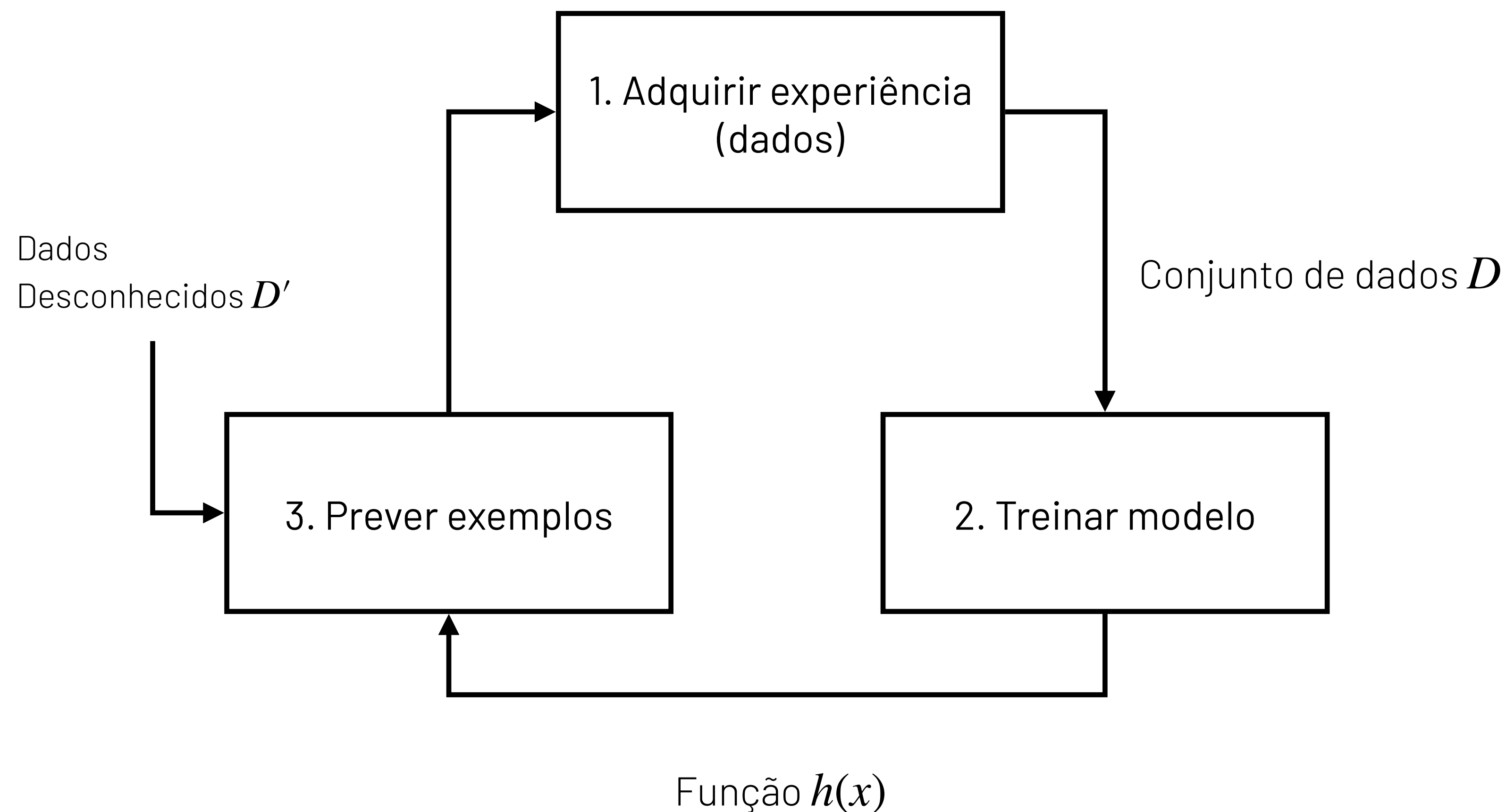
*Funções encontradas  
a partir de dados*

# Inteligência Artificial x Aprendizado de Máquina



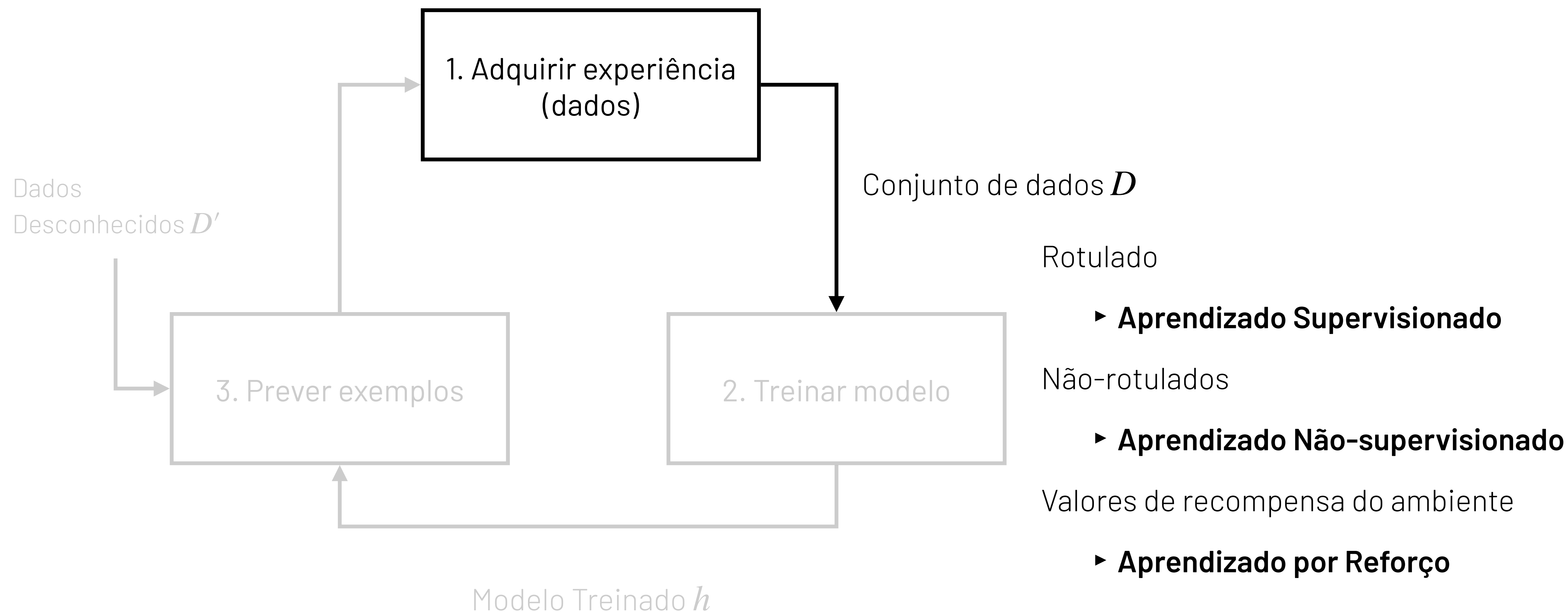
# Aprendizado de Máquina

Construção de agentes que melhoram o seu desempenho com novas experiências



# Tipos de Aprendizado

O tipo de aprendizado é definido pelo tipo de experiência (i.e., dados) observada pelo agente.



# Aprendizado Supervisionado

Quando todos os *exemplos* do conjunto de dados são pares  $(x_i, y_i)$ , chamamos o problema de **Aprendizado Supervisionado**.

Formalmente:

$D = \{(x_1, y_1), \dots, (x_N, y_N)\} \subseteq \mathbb{R}^d \times C$ , onde:


- ▶  $x_i$  é o vetor de características do  $i$ –ésimo exemplo
- ▶  $y_i$  é o rótulo do  $i$ –ésimo exemplo
- ▶  $\mathbb{R}^d$  é o espaço de características
- ▶  $C$  é o espaço de classes





# Exemplos de Aprendizado Supervisionado


## Classificação de Imagens de Gatos e Cachorros

$$D = \{$$

$(x_1 =$  $, y_1 = 1),$

$(x_2 =$  $, y_2 = 1),$

$(x_3 =$  $, y_3 = 0),$

$(x_4 =$  $, y_4 = 0)\}$

►  $x_i$ : vetor com os pixels da imagem achatada

►  $y_i$ : gato (1) ou cachorro (0)

►  $d \sim 100.000 - 10M$

►  $C = \{0, 1\}$

# Exemplos de Aprendizado Supervisionado

## Classificação de Imagens de Dígitos Escritos Manualmente

$D = \{$

$(x_1 =$    $, y_1 = 0),$

▶  $x_i$ : vetor com os pixels da imagem achatada

$(x_2 =$    $, y_2 = 1),$

▶  $y_i$ : o valor do dígito da imagem

▶  $d \sim 100.000 - 10M$

$(x_3 =$    $, y_3 = 5),$

▶  $C = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

$(x_4 =$    $, y_4 = 8)\}$

# Exemplos de Aprendizado Supervisionado

## Previsão de Preços de Imóveis

$$D = \{$$

$$(x_1 = [72, \text{Centro}, 2] , y_1 = 252,000),$$

$$(x_2 = [54, \text{Centro}, 1] , y_2 = 349,999),$$

$$(x_3 = [72, \text{Clélia}, 3] , y_3 = 380,250),$$

$$(x_4 = [182, \text{Ramos}, 4] , y_4 = 640,900)\}$$

►  $x_i$ : [tamanho (m<sup>2</sup>), bairro, número de quartos]

►  $y_i$ : preço do imóvel

►  $d = 3$

►  $C = \mathbb{R}$

# Aprendizado Supervisionado

## Classificação

Quando o espaço de classes  $C$  é conjunto finito de rótulos, chamamos o problema de **Classificação**.

### Classificação Binária

- ▶ Apenas dois rótulos possíveis:  $C = \{0, 1\}$
- ▶ Exemplo: Classificação de Imagens de Gatos e Cachorros

### Classificação Multiclasse

- ▶  $N > 2$  rótulos possíveis:  $C = \{0, 1, 2, \dots, N\}$
- ▶ Exemplo: Classificação de Imagens de Dígitos Escritos Manualmente

# Aprendizado Supervisionado

## Regressão

Quando o espaço de classes  $C$  é um conjunto infinito de números, chamamos o problema de **Regressão**.

Outros exemplos:

- ▶ Previsão de temperatura
- ▶ Previsão da nota de INF110 baseado no ENEM
- ▶ Regressão de caixa delimitadora

# Aprendizado Não-supervisionado

Quando todos os exemplos do conjunto de dados são apenas vetores  $x_i$ , **sem rótulos**, chamamos o problema de **Aprendizado Não-supervisionado**.

Formalmente:

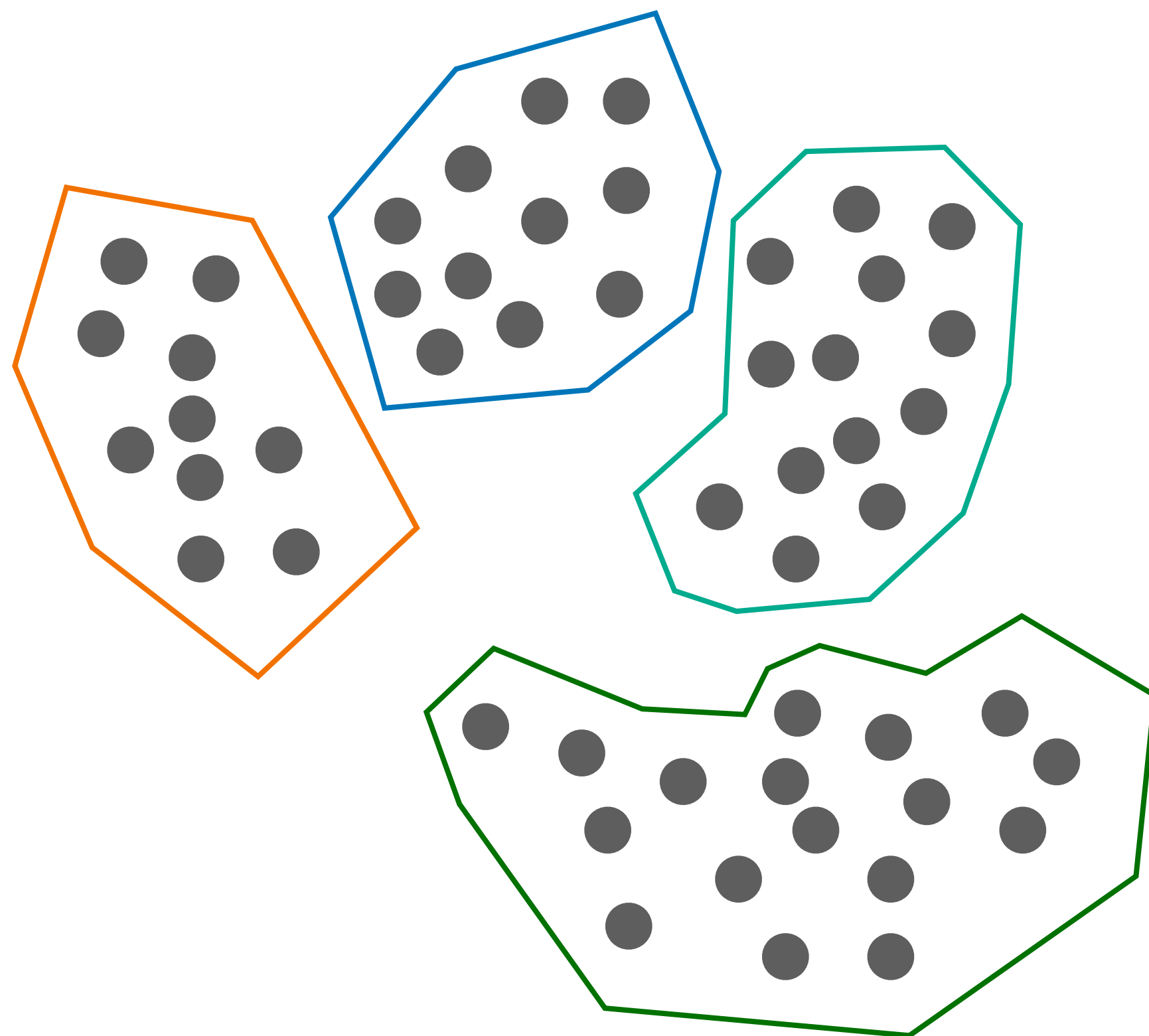
$D = \{x_1, x_2, \dots, x_N\} \subseteq \mathbb{R}^d$ , onde:

- ▶  $x_i$  é o vetor de características do  $i$ –ésimo exemplo
- ▶  $\mathbb{R}^d$  é o espaço de características

# Exemplos de Aprendizado Não-supervisionado

## Agrupamento

Agrupar os exemplos do conjunto de dados baseado em similaridade





# Exemplos de Aprendizado Não-supervisionado

## Redução de Dimensionalidade

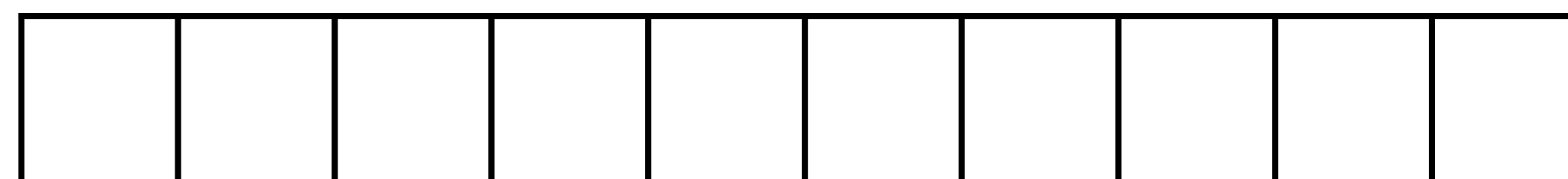
Reduzir a dimensionalidade  $d$  dos exemplos do conjunto de dados



Áudio

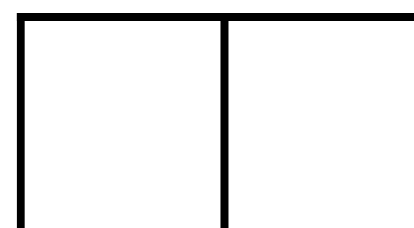


$x_i$



1 2 3 4 5 6 7 8 9 10

$x'_i$



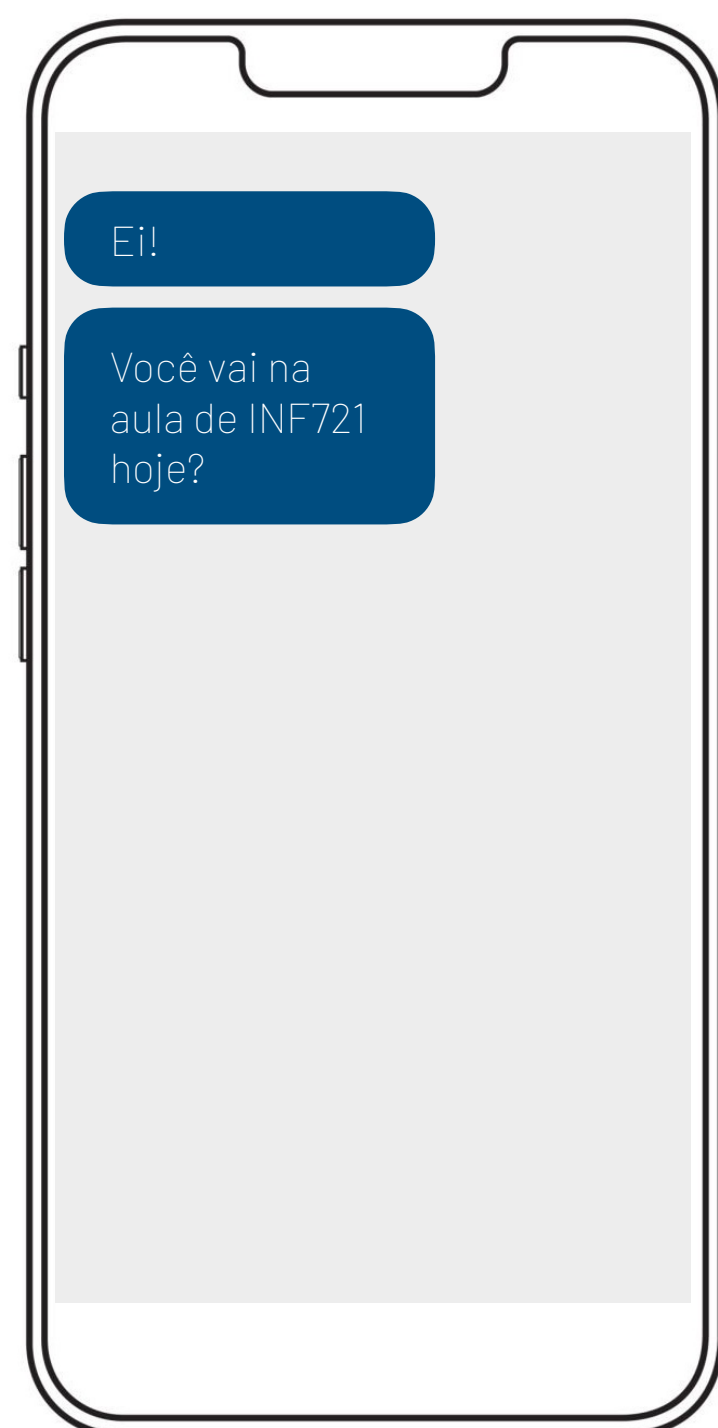
1 2



# Exemplos de Aprendizado Não-supervisionado

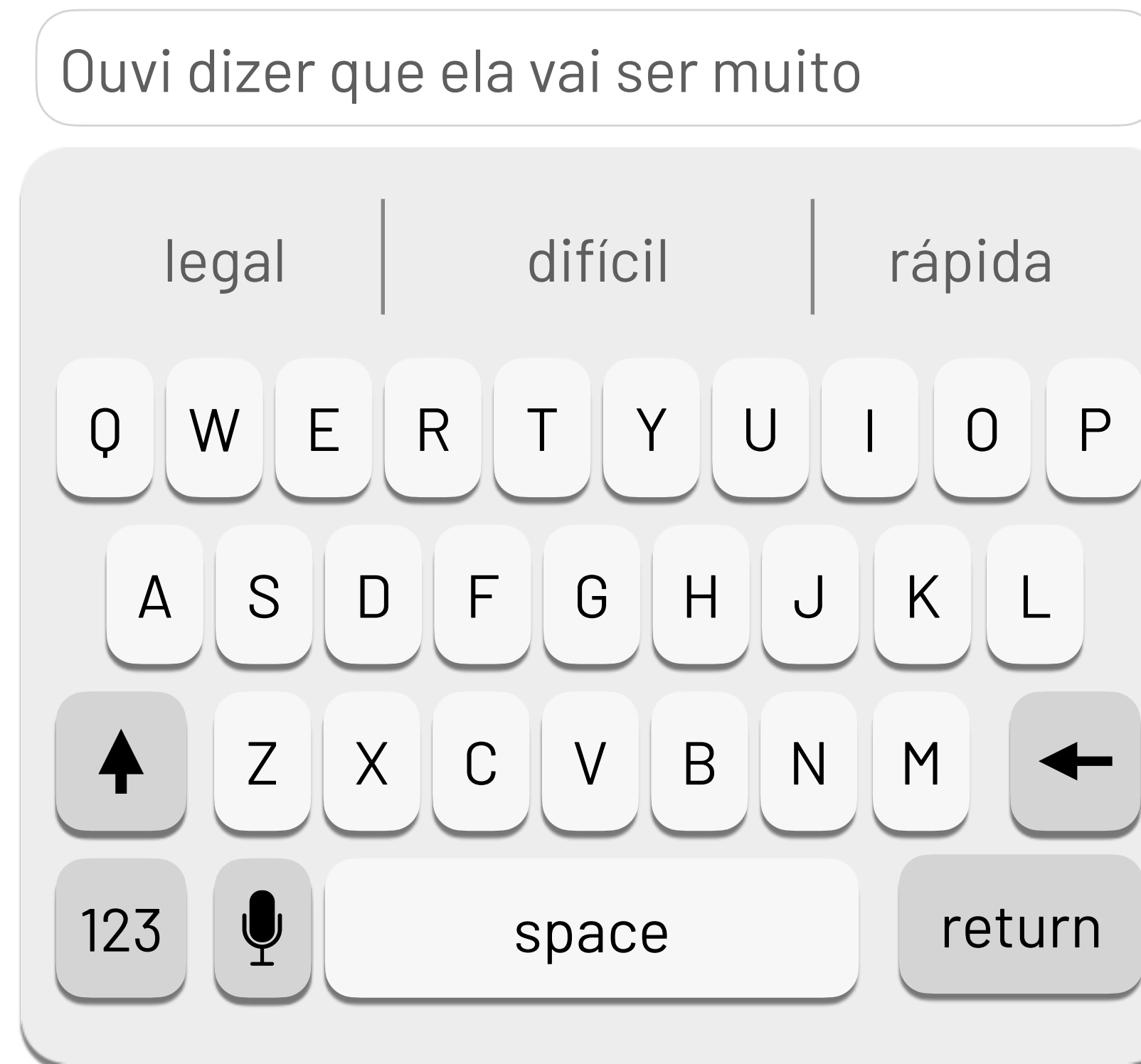
## Geração de Dados

Inferir a distribuição que gerou os dados do conjunto de dados



$$P(x_n | x_{n-1}, x_{n-2}, \dots, x_1)$$

Modelo de linguagem



# Aprendizado por Reforço

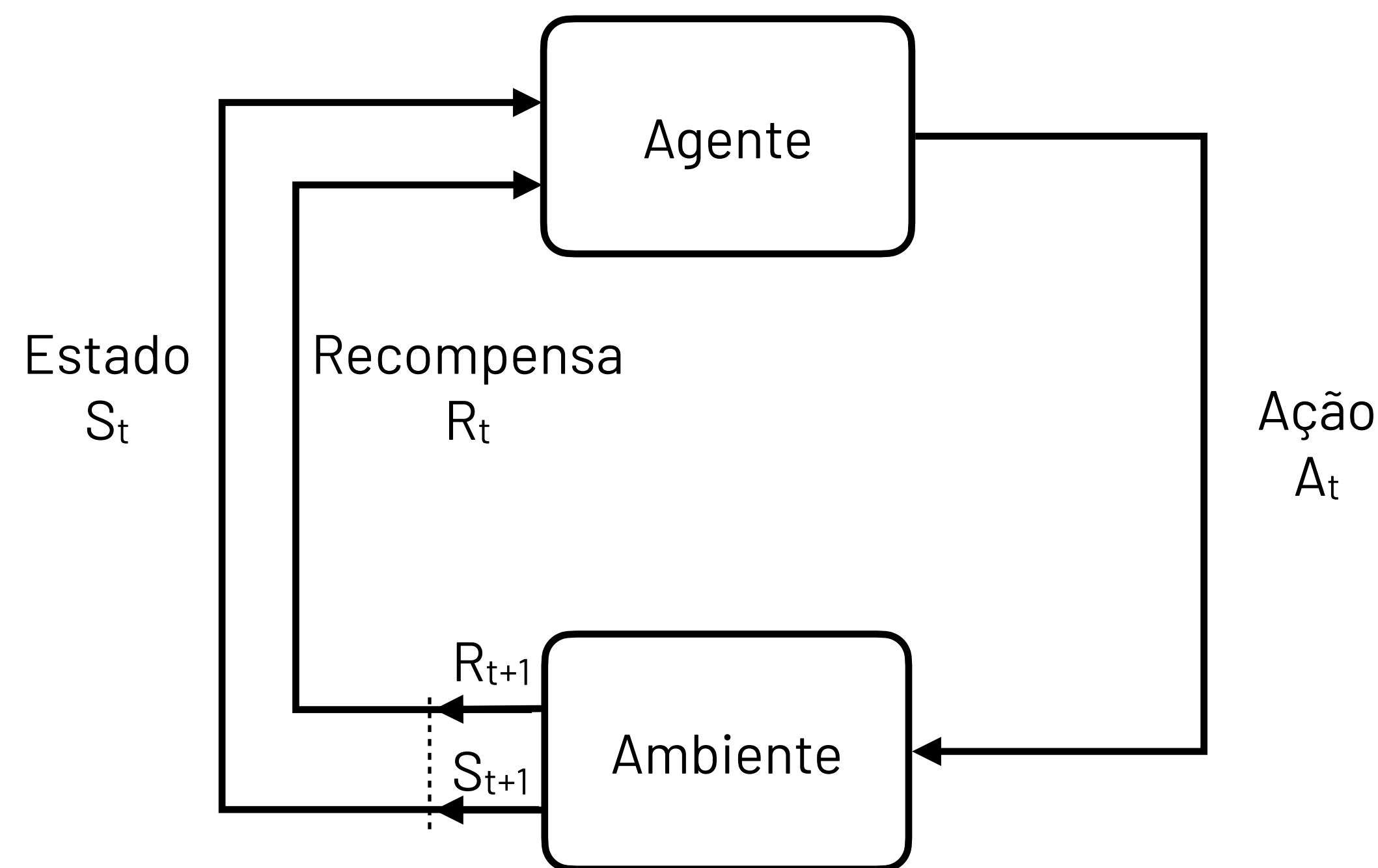
Aprender uma função  $\pi(s) = a$  que prevê a ação  $a$  que um agente deve tomar no estado  $s$ , maximizando as recompensas recebidas pelo ambiente.

## Agente

- ▶ Observa um estado  $s_t$  no tempo  $t$
- ▶ Produz uma ação  $a_t$  no tempo  $t$

## Ambiente

- ▶ Retorna uma recompensa  $r_{t+1}$
- ▶ Gera o próximo estado  $s_{t+1}$



# Tipos de Dados

## Estruturados (tabulares)

Tamanho	Bairro	# de quartos	...	Preço
72	Centro	2		
54	Centro	1		
...	...	...		...
72	Clélia	3		

Idade	Estado	Ad Id	...	Click
72	MG	93242		1
54	SP	93287		0
...	...	...		...
72	RJ	71244		1

## Não-estruturados (não-tabulares)



Imagens

Você vai na aula de INF721 hoje?

Texto

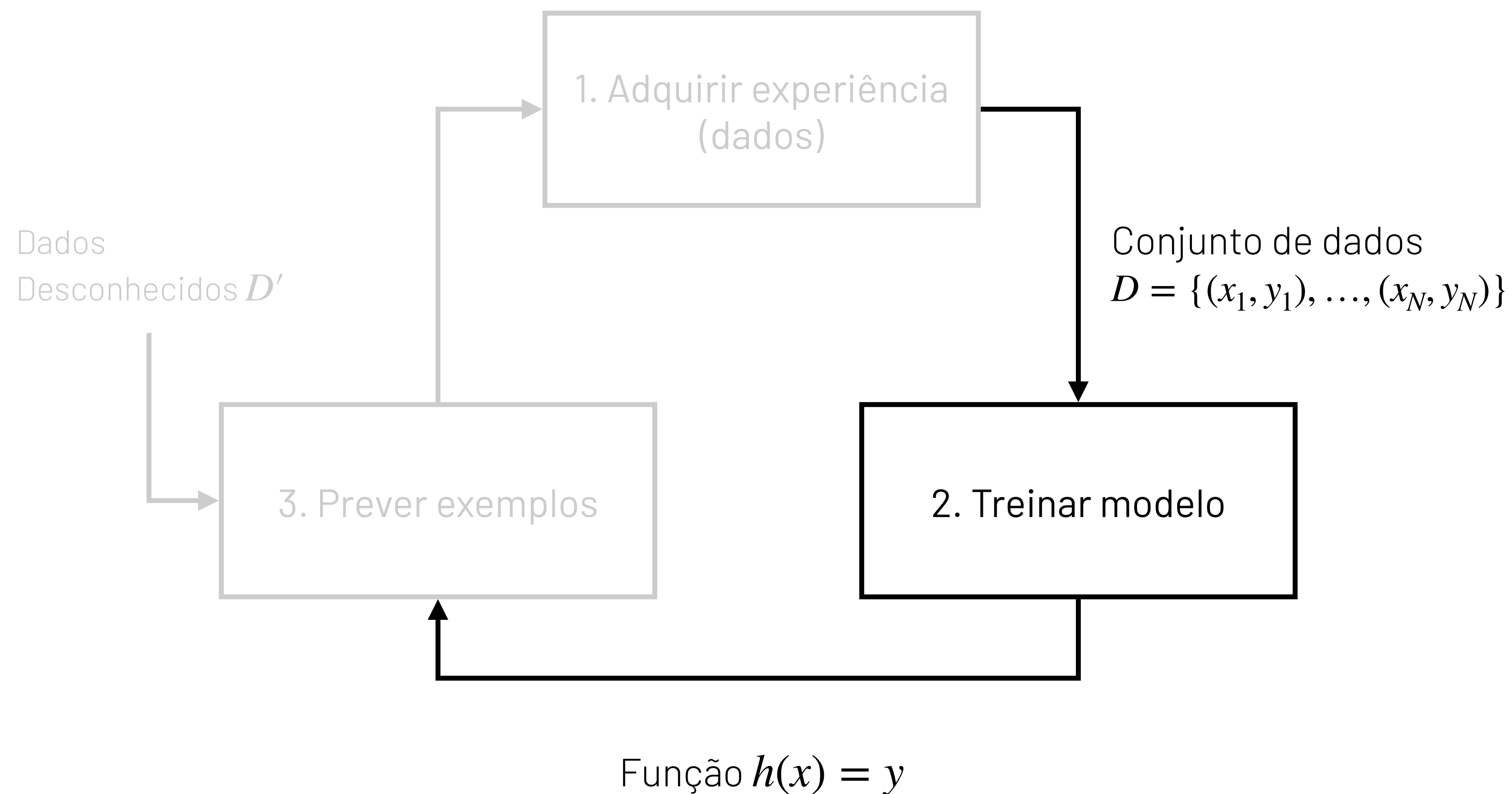


Áudio

# Aprendizado Supervisionado

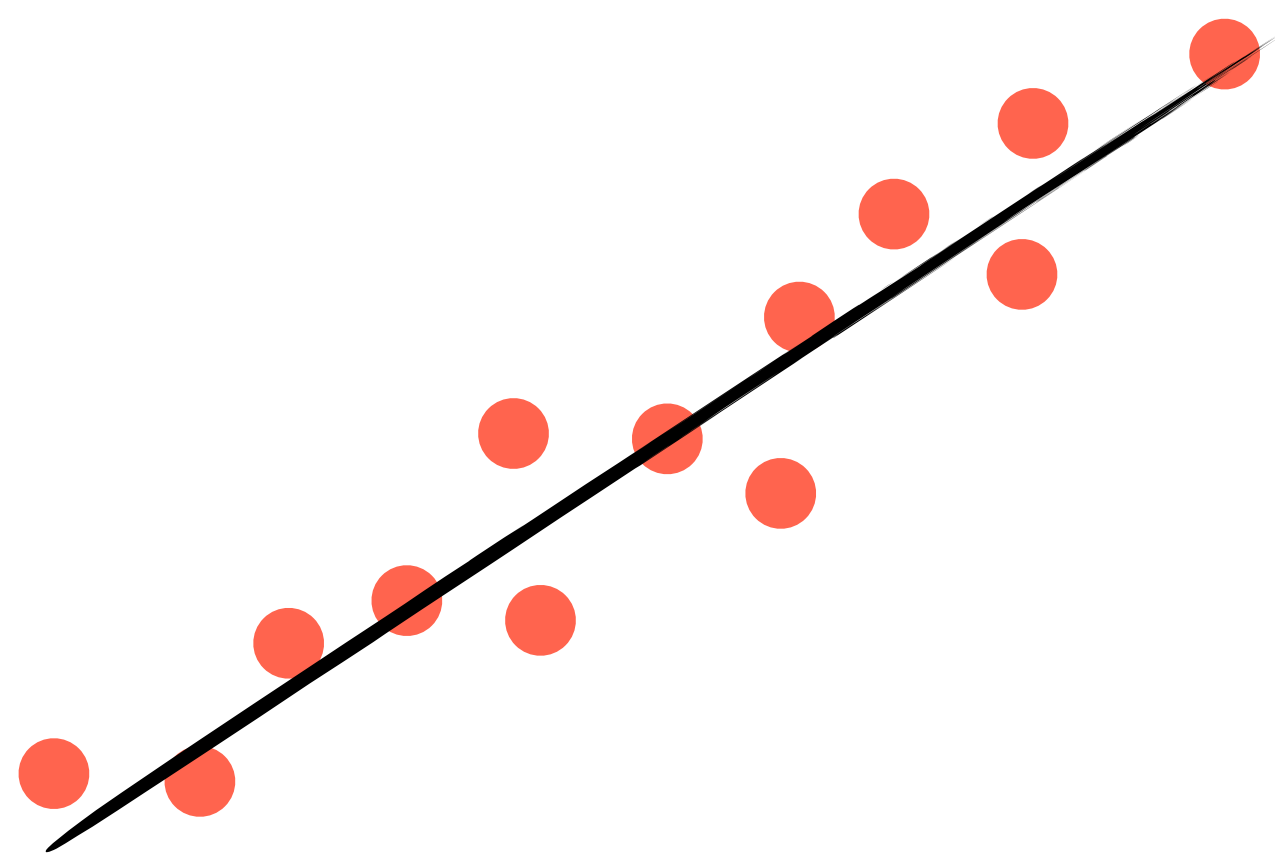
# Aprendizado Supervisionado

O objetivo de aprendizado supervisionado é aprender uma função  $h(x) = y$  a partir de dados



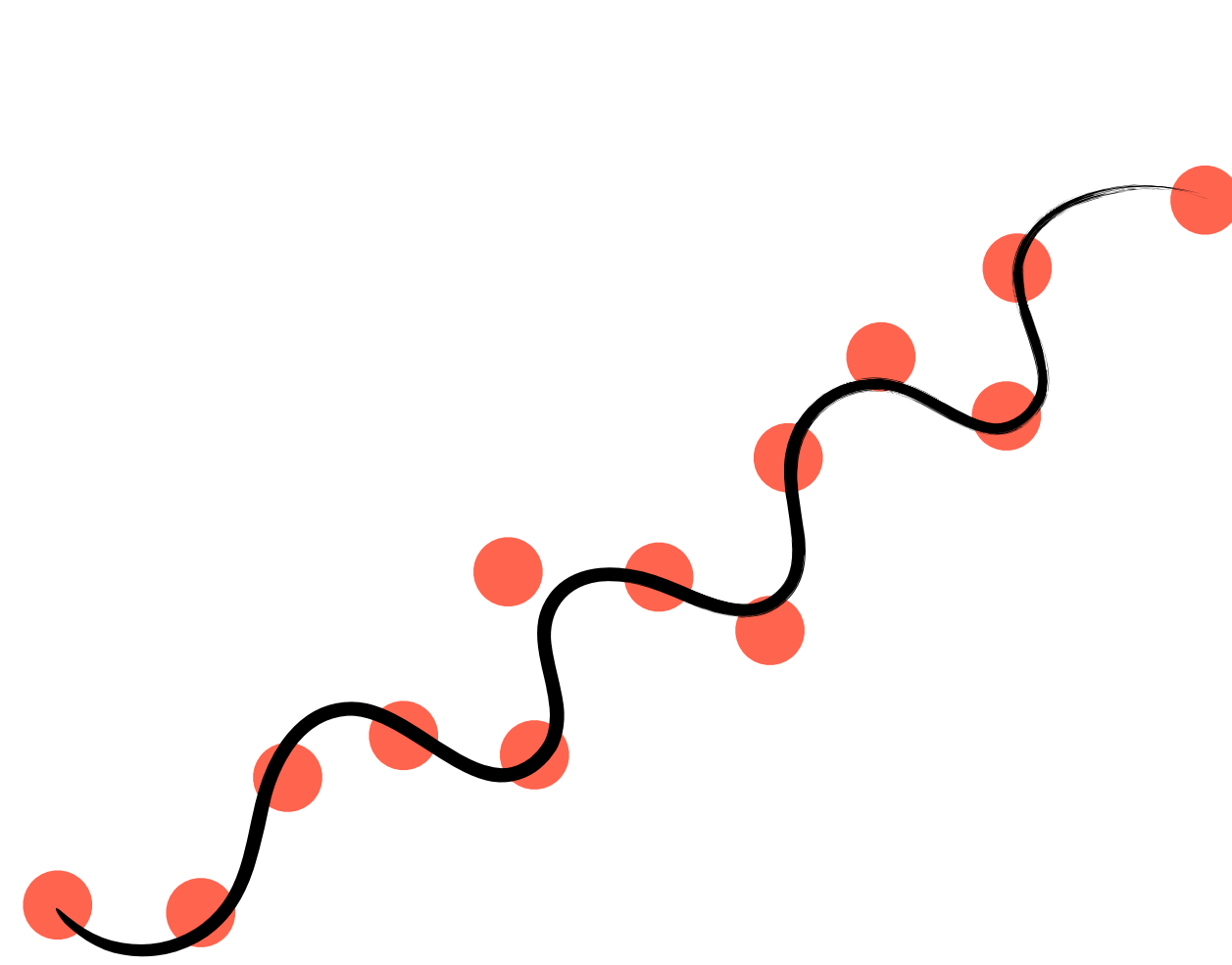
# Espaço de Hipóteses

Para encontrar uma função, um algoritmo de aprendizado supervisionado precisa assumir uma *hipótese* sobre os dados para definir um espaço de funções  $H$ .



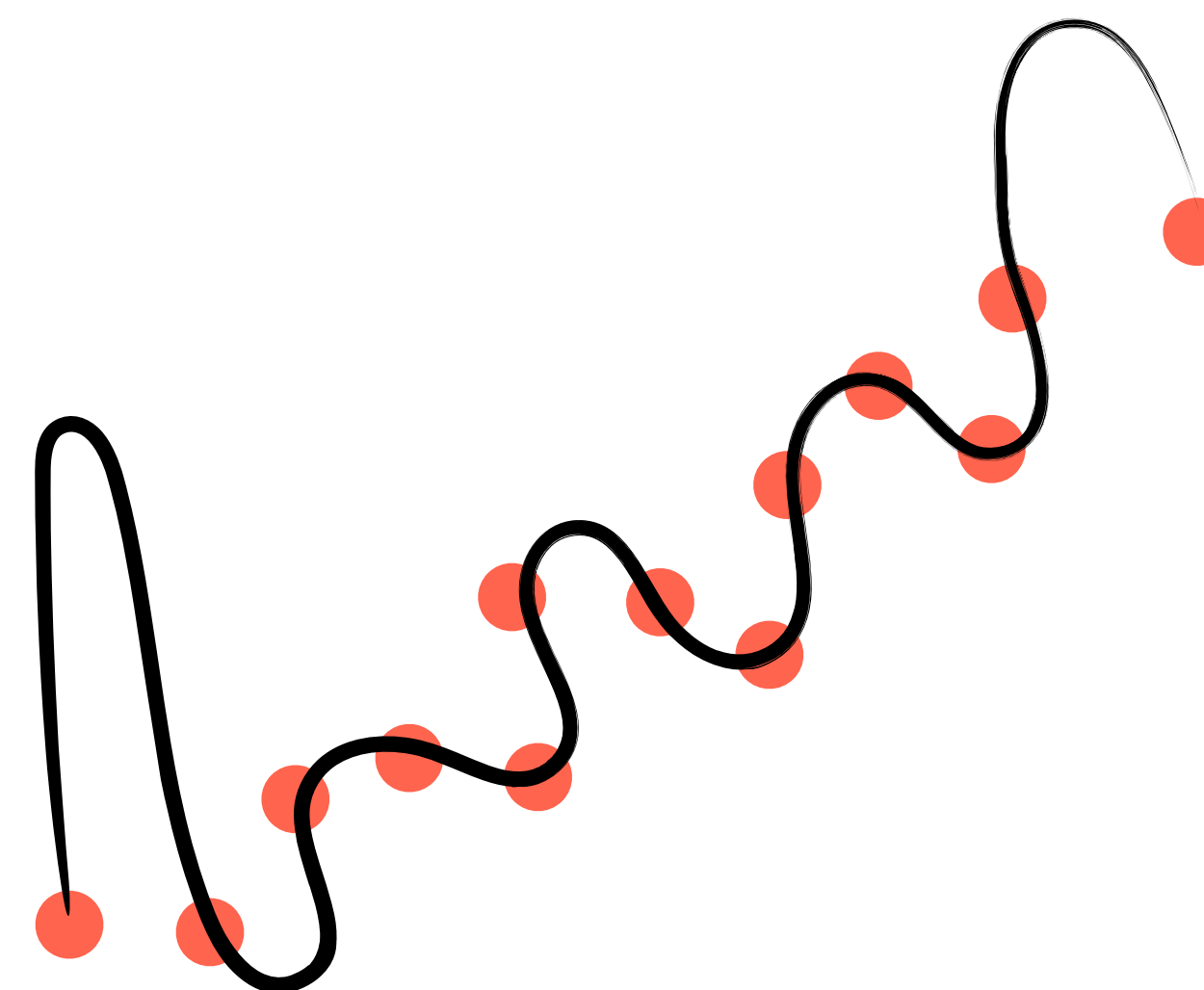
Reta

$$h(x) = w_1x + w_0$$



Senoide

$$h(x) = w_1x + \sin(w_0x)$$

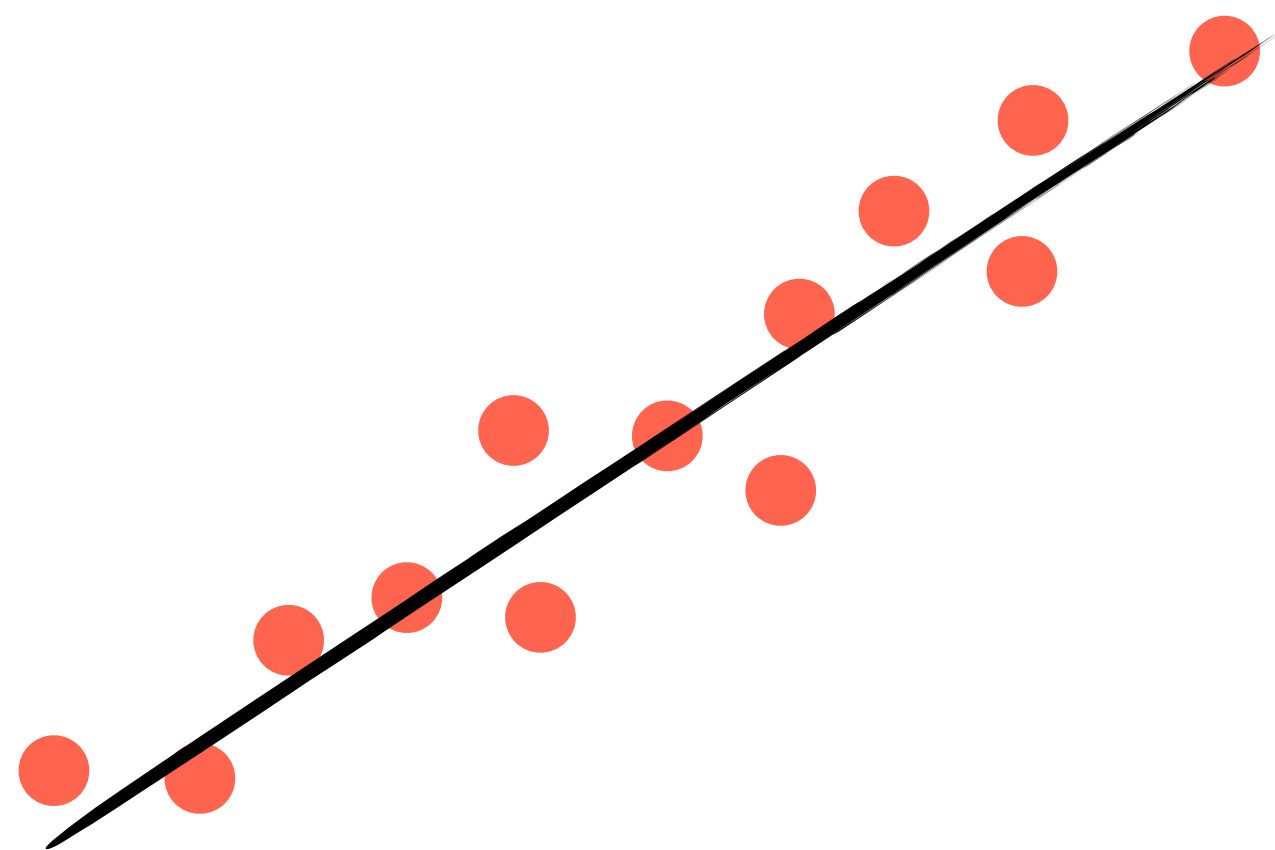


Polinômio de grau 12

$$h(x) = \sum_{i=0}^{12} w_i x^i$$

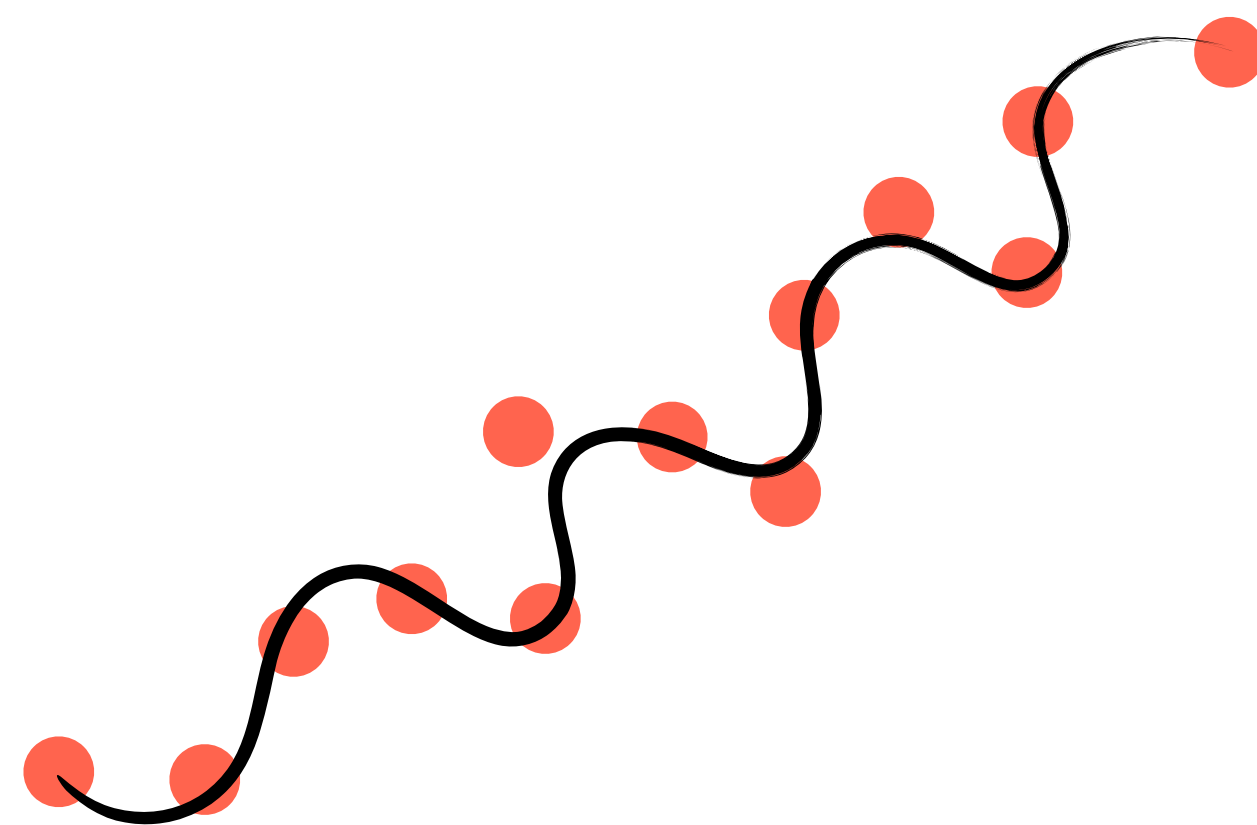
# Espaço de Hipóteses

Qual é a melhor hipótese?



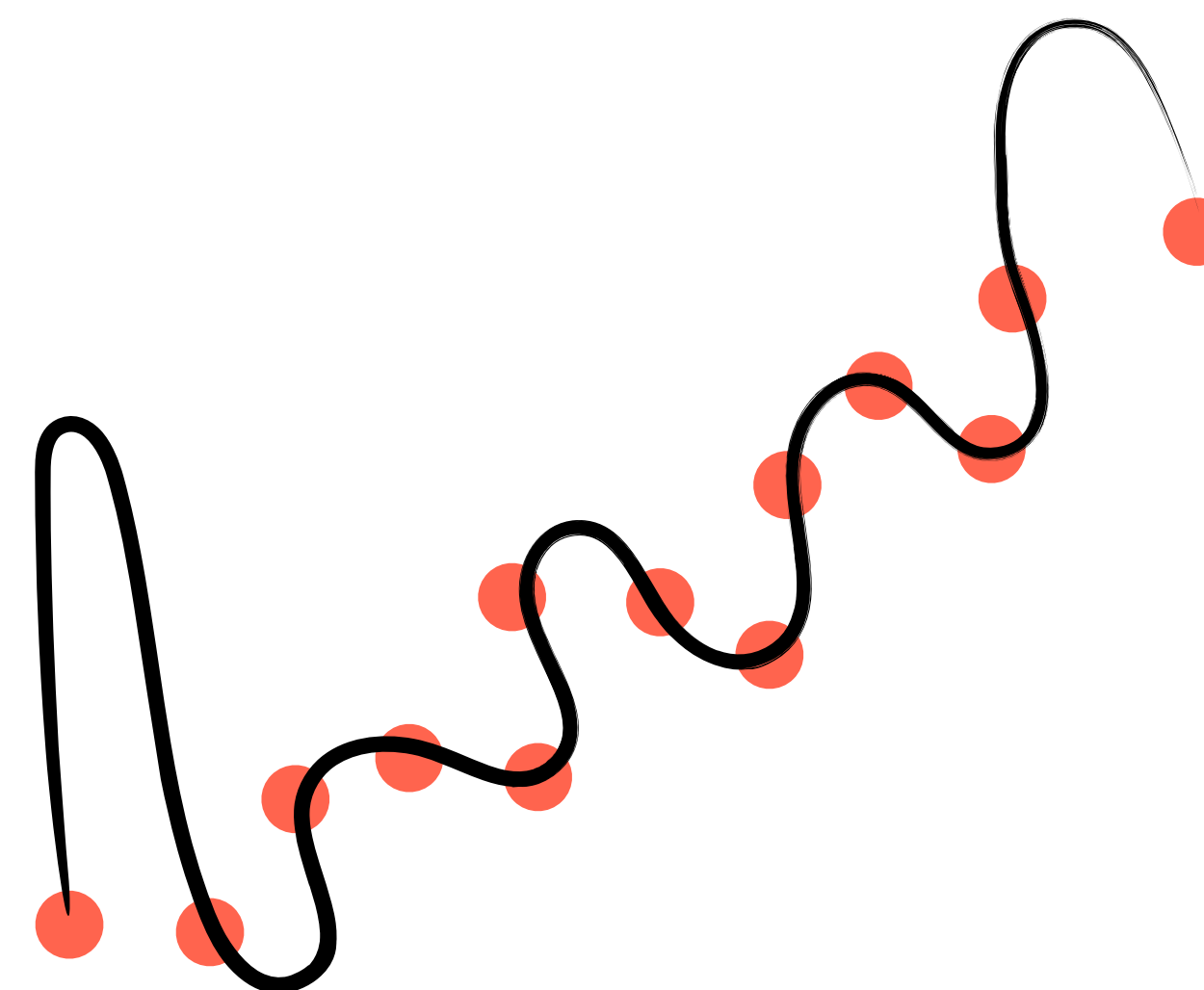
Reta

$$h(x) = w_1x + w_0$$



Senoide

$$h(x) = w_1x + \sin(w_0x)$$



Polinômio de grau 12

$$h(x) = \sum_{i=0}^{12} w_i x^i$$

# Generalização

Para medir a qualidade de uma hipótese, temos que avaliar o seu desempenho em dados desconhecidos, ou seja, que não vimos durante o treinamento. Assim, dividimos o conjunto de dados em 3 conjuntos disjuntos:

## 1. Conjunto de Treinamento

- ▶ Utilizado para ajustar uma hipótese aos dados
- ▶ Exemplo: os pesos  $w$  de uma rede neural

## 2. Conjunto de Validação

- ▶ Utilizado para definir os hiper-parâmetros da hipótese
- ▶ Exemplo: o número de camadas de uma rede neural

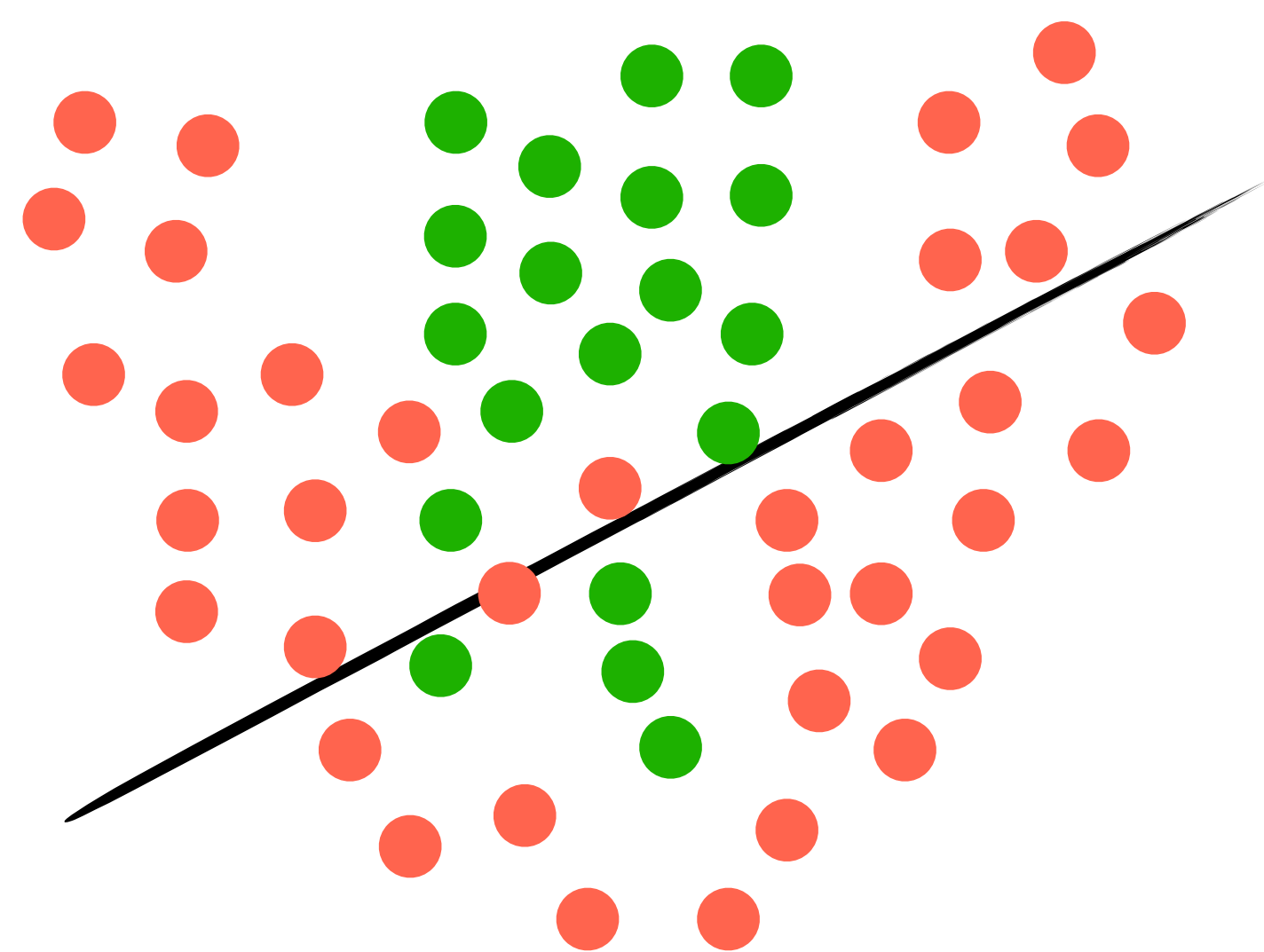
## 3. Conjunto de Teste

- ▶ Utilizado para avaliar o desempenho de um modelo
- ▶ Exemplo: a acurácia de uma rede neural de classificação

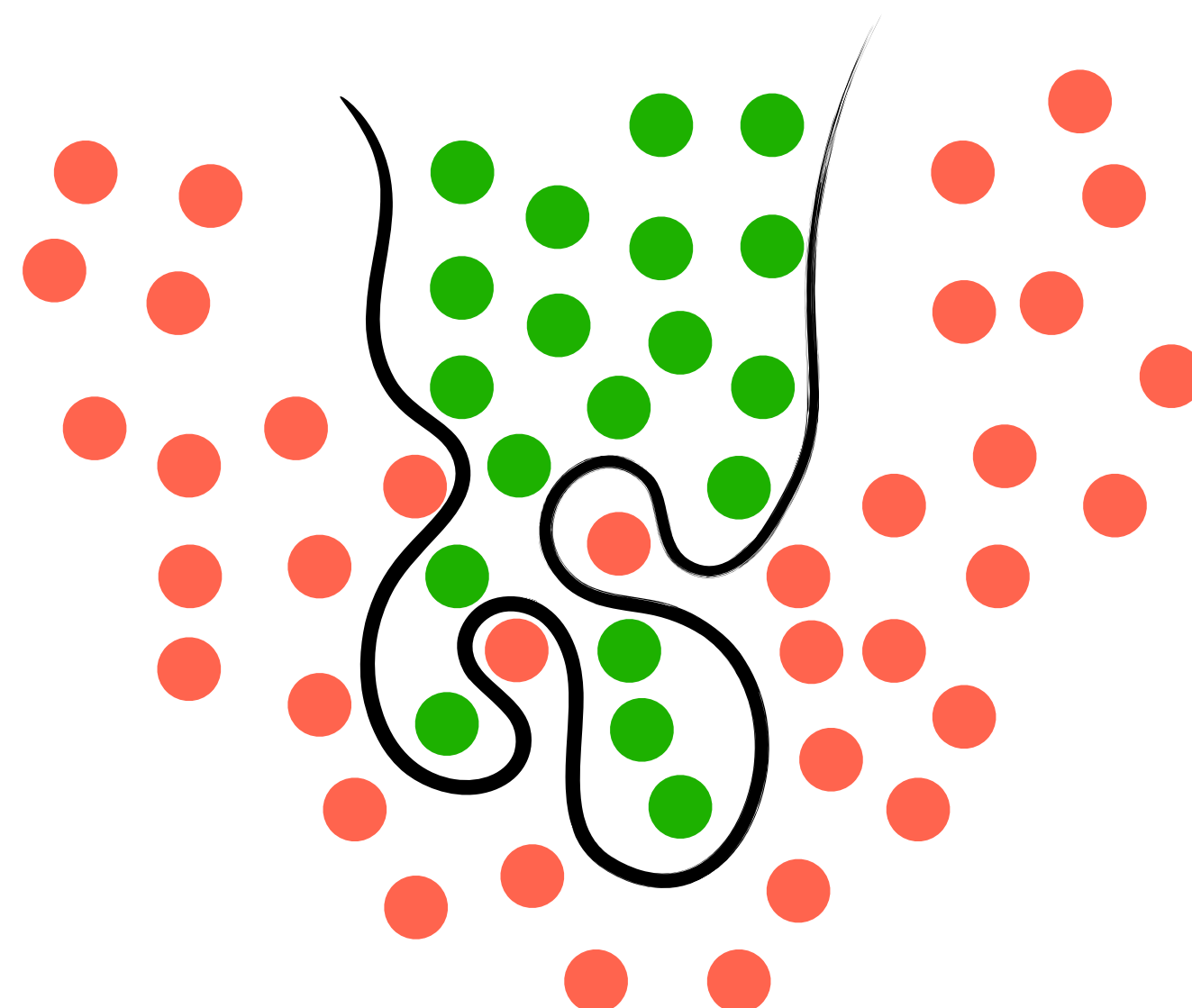


# Generalização

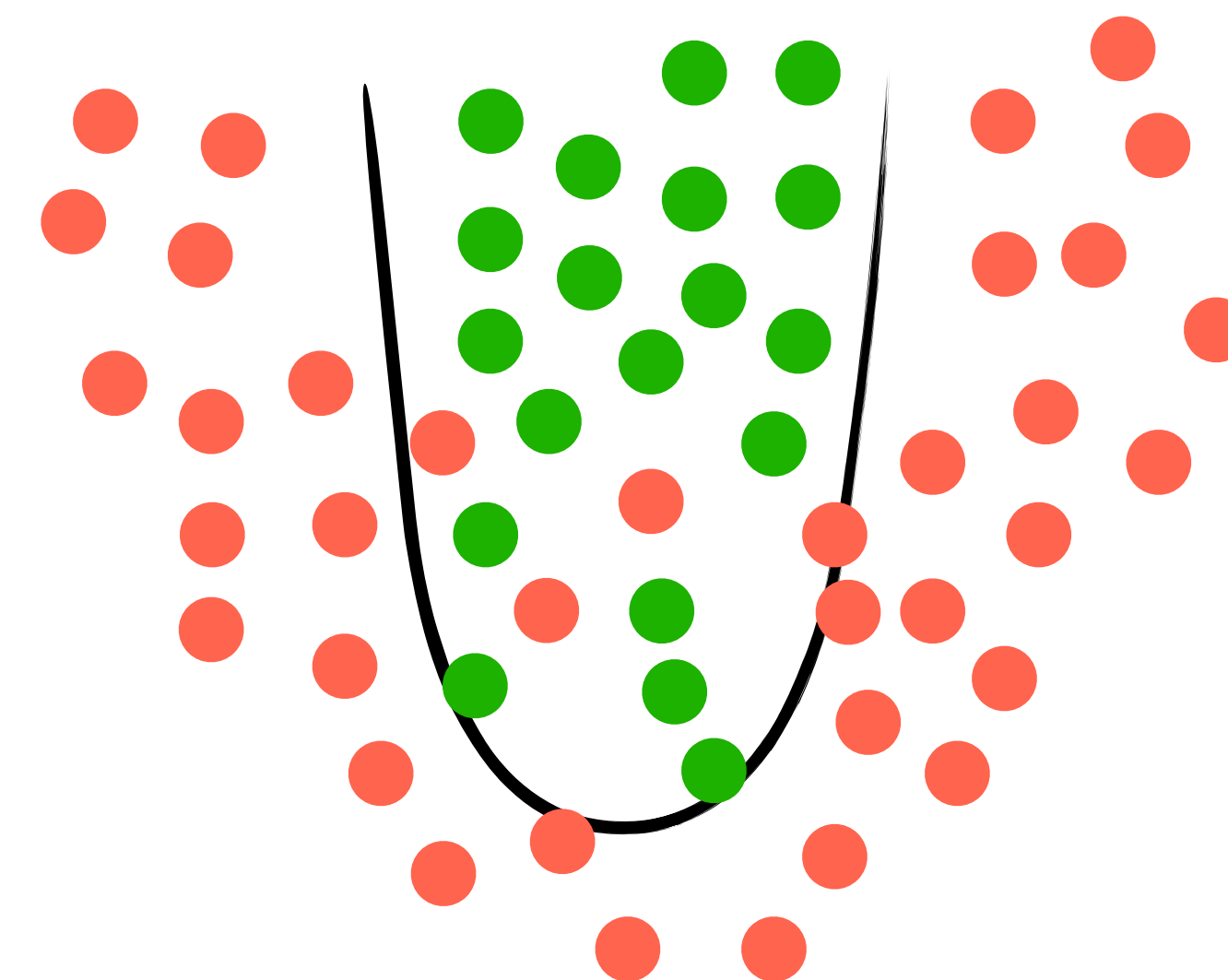
A melhor hipótese é aquela que melhor **generaliza** aos dados de teste, evitando tanto subajuste (*underfitting*), quanto sobreajuste (*overfitting*)



Subajuste

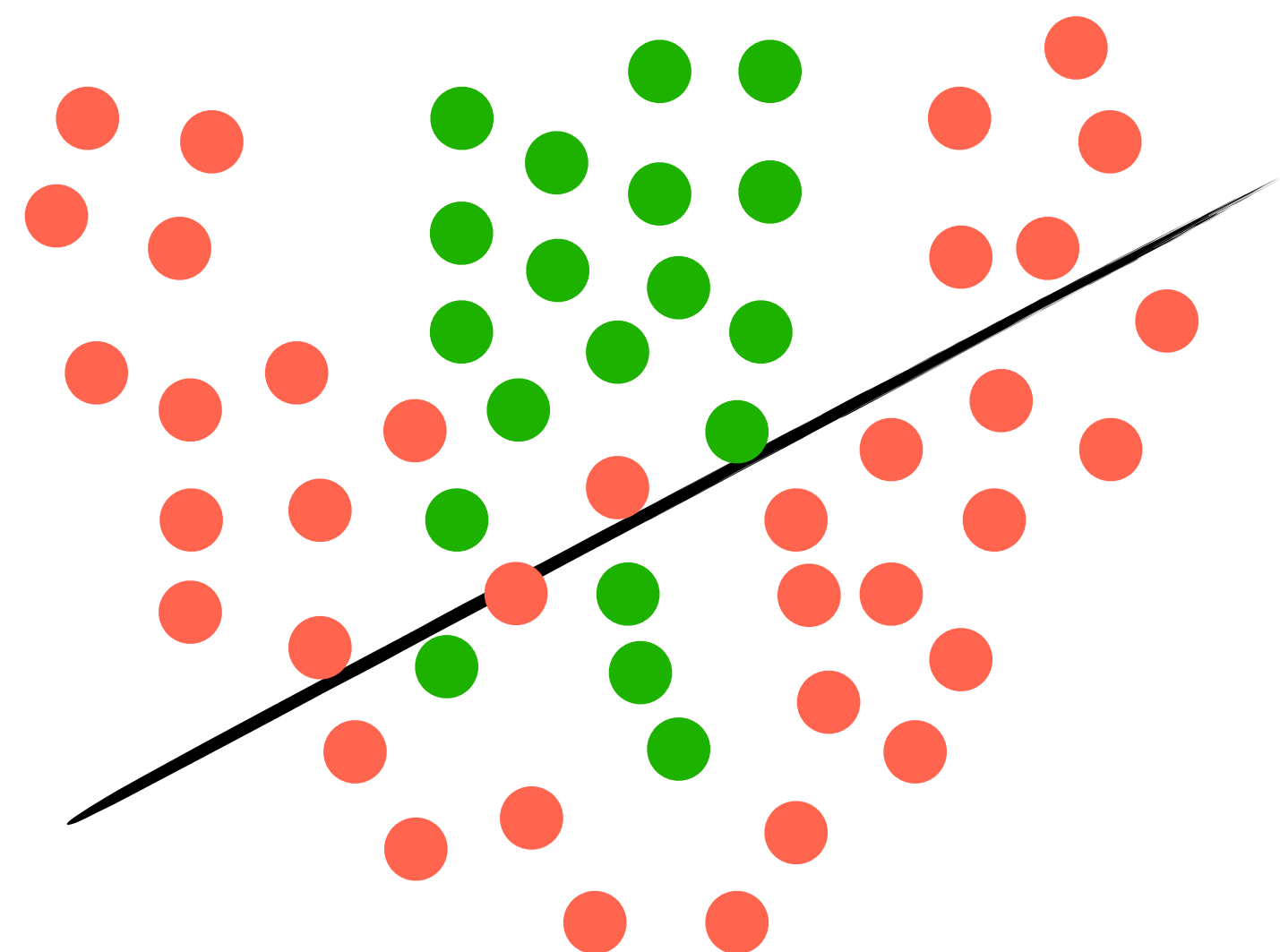


Sobreajuste



Ajuste Adequado

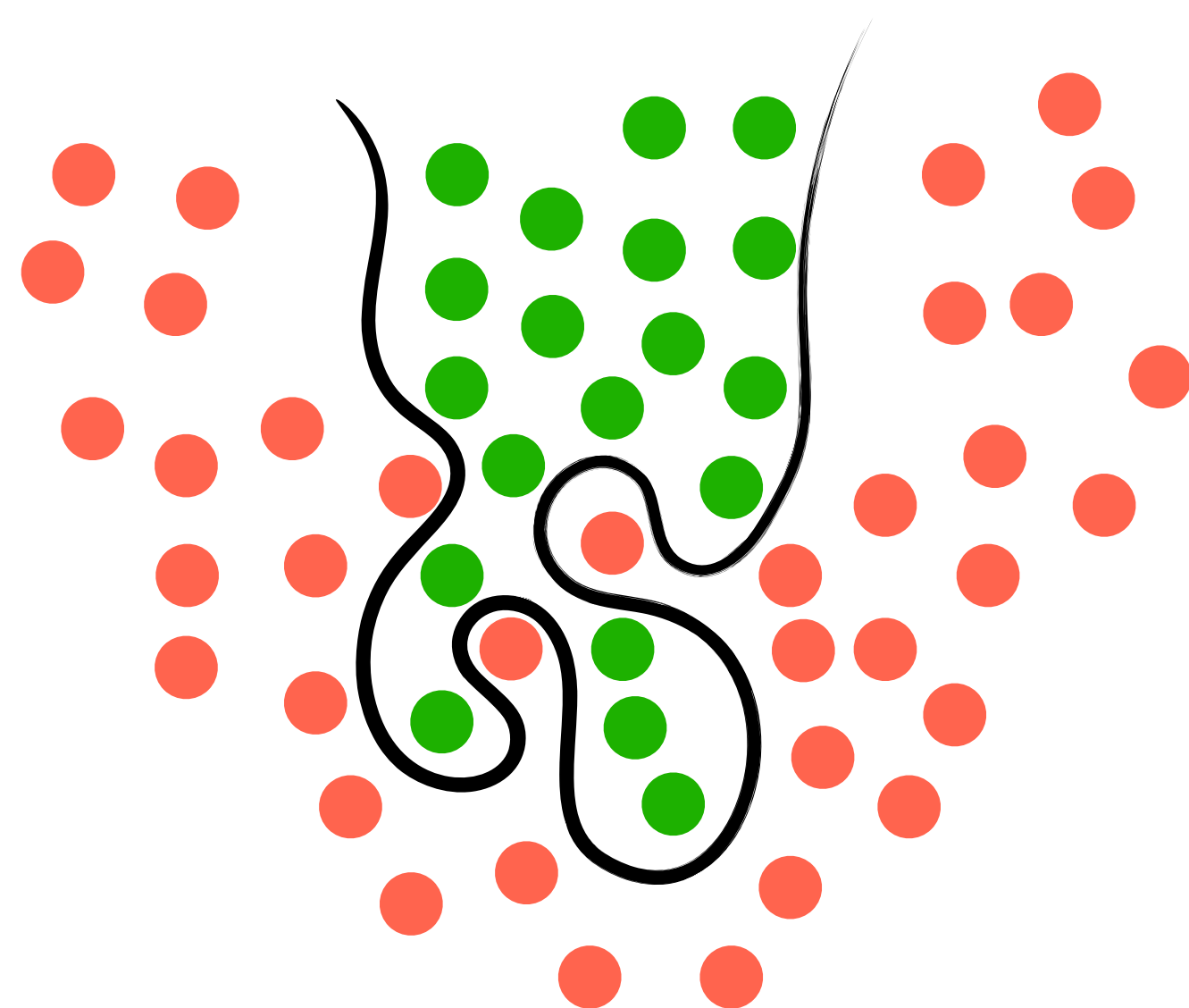
# Generalização



## Subajuste (underfitting)

Quando a hipótese se ajusta pouco aos dados de treinamento, apresentando baixo desempenho de previsão tanto no conjunto de treinamento quanto no de teste.

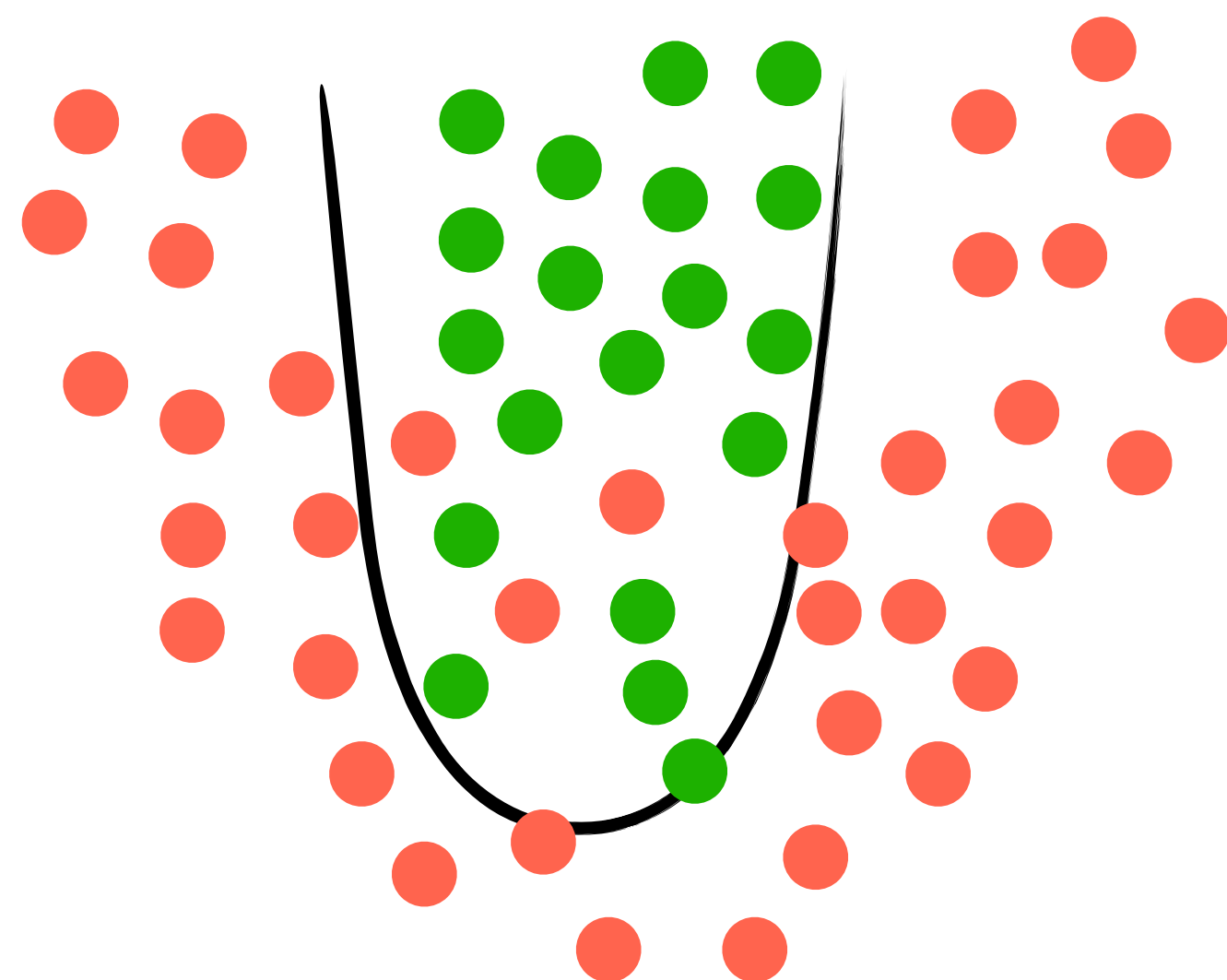
# Generalização



## Sobreajuste (overfitting)

Quando a hipótese se ajusta muito aos dados de treinamento, apresentando alto desempenho de previsão no conjunto de treinamento, mas baixo no conjunto de teste.

# Generalização



## Ajuste Adequado

Quando a hipótese se ajusta bem aos dados de treinamento, apresentando alto desempenho de previsão tanto no conjunto de treinamento quanto no de teste.

# Algoritmos de Aprendizado Supervisionado

Cada algoritmo de aprendizado supervisionado assume uma *hipótese* diferente sobre os dados para definir um espaço de funções  $H$ .

- ▶ Regressão Linear
- ▶ Regressão Logística
- ▶ Árvores de Decisão
- ▶ K-Nearest Neighbors (KNN)
- ▶ Naive Bayes
- ▶ Suport Vector Machines (SVMs)
- ▶ Redes Neurais

# Próxima aula

## **A3:** Regressão Logística

Regressão Logística como uma rede neural para problemas linearmente separáveis.