

Inteligência Artificial

Aula 02 - Introdução ao Aprendizado de Máquina

Aprendizado de Máquina

- Permite aos computadores
 - Aprender e reconhecer padrões em dados
 - Realizar inferências com base no aprendizado

Imaghe como escrever
um programa que
leia e entendendo esta
escrita.

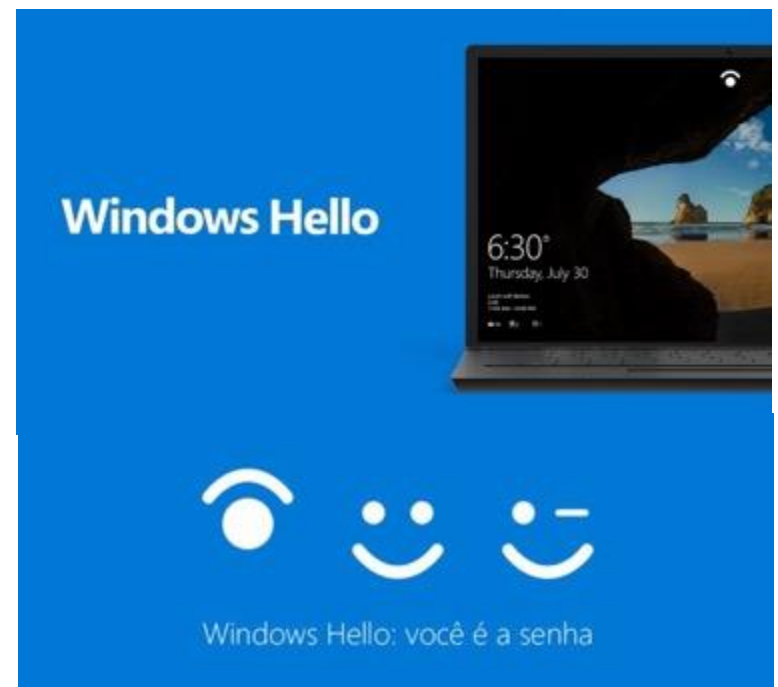
Imagine escrever um programa de computador que:

- Reconheça pessoas pelo rosto
 - Problemas:
 - Diferentes expressões faciais
 - Alterações na face (ex: óculos, bigode)
 - Cortes de cabelo
 - Etc

Que características considerar??



Seres humanos: reconhecimento de padrões,
aprendizado do que deve ser observado após vários exemplos



Imagine escrever um programa de computador que:

- Diagnosticar pacientes a partir de seus sintomas e exames
 - Médico: formação e experiência

Imagine escrever um programa de computador que:

- Responder a questões sobre vendas como
 - Quais produtos são vendidos em **conjunto**?
 - Que produto **recomendar** a um cliente?
 - Como agrupar clientes para melhor **marketing**?

NETFLIX

amazon

LUU, Minh-Duc; LIM, Ee-Peng. Do your friends make you buy this brand?. Data Mining and Knowledge Discovery, vol 32, n 2, p. 287–319, 2018.

Aprendizado de Máquina: **motivações**

- **Automatizar** o processo de aquisição de conhecimento
- **Entender** melhor os mecanismos de aprendizado humano
- Algumas tarefas são melhor definidas e/ou executadas a partir de exemplos
 - Ex: Reconhecer pessoas
- Ser humano não é capaz de explicar (e programar) sua habilidade para executar alguns tipos de tarefas
 - Ex: Dirigir

Aprendizado de Máquina: **motivações**

- **Quantidade de conhecimento** disponível pode ser muito grande para ser descrito (e programado) por humanos
 - Ex: diagnóstico médico
- Algumas tarefas exigem **cálculos complexos**, possíveis apenas com computador
 - Ex: interrelacionar/correlacionar grandes quantidades de dados
- Modelos de AM podem se **adaptar a novas situações**
 - Evita necessidade de reprogramação

O que é aprendizado?

- Essencial para comportamento inteligente
- Algumas **atividades**:
 - Memorizar algo
 - Observar e explorar situações para aprender fatos
 - Melhorar habilidades motoras/cognitivas por prática
 - Organizar conhecimento novo em representações apropriadas

Exemplo

- **Problema:** filtrar mensagens de e-mail
 - **Tarefa:** categorizar mensagens de e-mail como spam ou legítima
 - **Medida de desempenho:** porcentagem de mensagens de spam e legítimas corretamente identificadas
 - **Experiência:** conjunto de e-mails apontados pelo usuário como spams

Google says its machine learning tech now blocks 99.9% of Gmail spam and phishing messages

Posted May 31, 2017 by [Frederic Lardinois \(@fredericl\)](#)



Google today said that its machine learning models can now detect spam and phishing messages with 99.9 percent accuracy.

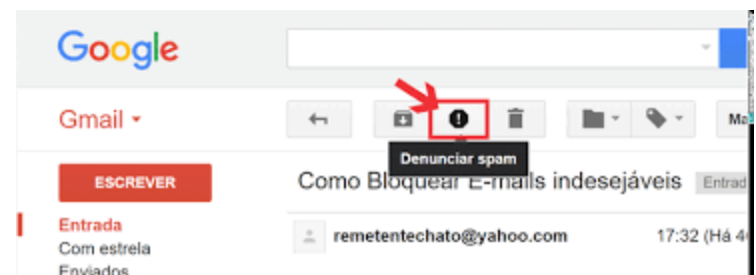


Crunchbase

Gmail

DESCRIPTION
Gmail is Google's free email serv

WEBSITE
<http://gmail.com/>



Exemplo

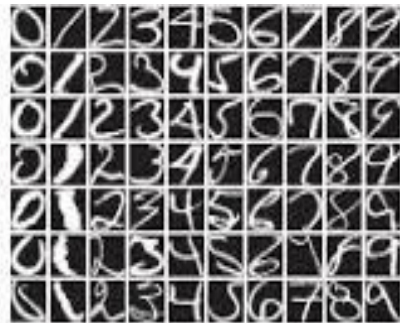
- **Problema:** reconhecer escrita manual
 - **Tarefa:** reconhecer e classificar dígitos manuscritos dentro de imagens
 - **Medida de desempenho:** porcentagem de dígitos corretamente identificados
 - **Experiência:** imagens de dígitos manuscritos por diferentes pessoas

USPS database:

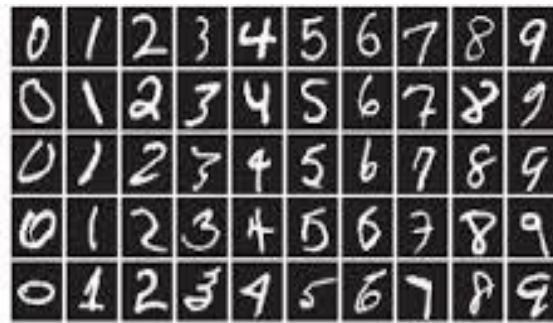
numeric data obtained from the scanning of handwritten digits from envelopes by the U.S. Postal Service

MNIST database:

images of handwritten digits, collected among Census Bureau employees and high-school students. Writers of the training set and test set are disjoint.



(a) USPS



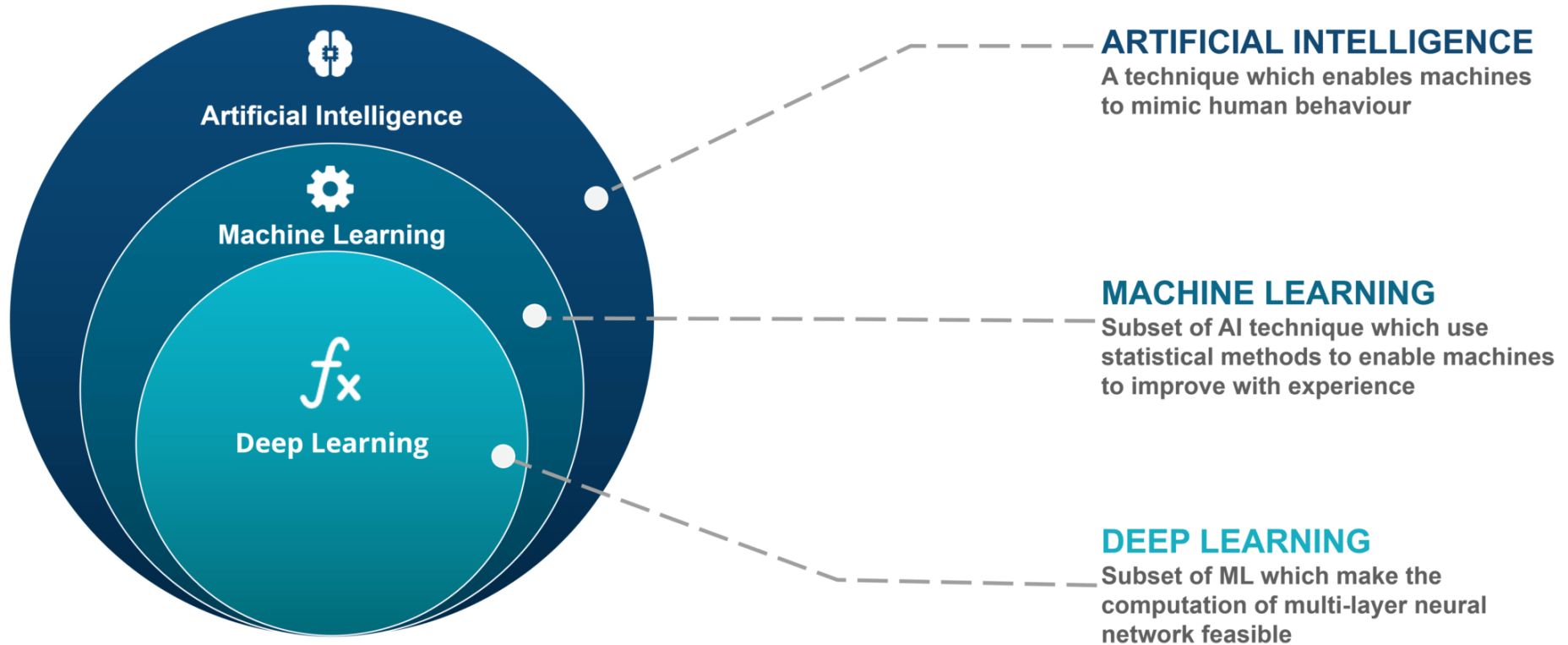
(b) MNIST

Exemplo

- **Problema:** Problema: carro autônomo (aprender a dirigir)
 - **Tarefa:** dirigir em uma rodovia pública usando sensores
 - **Medida de desempenho:** distância média percorrida antes de um erro
 - **Experiência:** sequência de imagens e comandos de direção registrados observando um motorista humano

Exemplo

- **Problema:** detectar bons clientes
 - **Tarefa:** classificar potenciais clientes como bons ou maus pagadores
 - **Medida de desempenho:** porcentagem de clientes classificados corretamente como bons e maus pagadores
 - **Experiência:** uma base de dados histórica com dados de clientes já conhecidos



Inferência Indutiva

- Um dos principais meios para a aquisição de novos conhecimentos
- **Indução:** raciocínio para obter conclusões sobre todos os membros de uma classe pelo exame de alguns membros da classe

Inferência Indutiva

- Se eu noto que:
 - Todos pacientes com déficit de atenção atendidos em 1986 sofriam de ansiedade
 - Todos pacientes com déficit de atenção atendidos em 1987 sofriam de ansiedade
 - ...
- Posso inferir que pacientes que sofrem de déficit de atenção também sofrem de ansiedade
 - Isto pode ser ou não verdade, mas propicia uma boa generalização

Conjunto de Dados

- Experiência pode ser provida por um conjunto de dados (de treinamento)
 - Ex: base de dados de um hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	#	Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2		SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4		MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2		RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8		MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1		PE	Saudável
2301	Ana	22	F	?	Inexistentes	38,0	3		RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6		AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2		GO	Saudável

Conjunto de Dados

- Meta: induzir hipótese para fazer diagnósticos corretos para novos pacientes

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp. #	Int. Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0 2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5 4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0 2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5 8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6 1	PE	Saudável
2301	Ana	22	F	?	Inexistentes	38,0 3	RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0 6	AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4 2	GO	Saudável

Conjunto de Dados

- Cada **linha** (paciente) é um **dado** (amostra, objeto, exemplo, padrão ou registro)

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	#	Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2		SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4		MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2		RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8		MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1		PE	Saudável
2301	Ana	22	F	?	Inexistentes	38,0	3		RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6		AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2		GO	Saudável

Conjunto de Dados

- Cada objeto é uma tupla com valores de **características** (atributos, campos ou variáveis), que descrevem seus principais aspectos

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	#	Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2		SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4		MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2		RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8		MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1		PE	Saudável
2301	Ana	22	F	?	Inexistentes	38,0	3		RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6		AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2		GO	Saudável

Conjunto de Dados

- **Atributo de saída (alvo/meta):** presente em algumas tarefas, seus valores devem ser estimados usando outros atributos (de entrada/preditivos)

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	#	Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP		Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG		Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS		Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG		Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE		Saudável
2301	Ana	22	F	?	Inexistentes	38,0	3	RJ		Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM		Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO		Saudável

Conjunto de Dados

- **Observação:** atributos de identificação e nome não possuem relação com a doença e não são utilizados como entradas

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	#	Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP		Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG		Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS		Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG		Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE		Saudável
2301	Ana	22	F	?	Inexistentes	38,0	3	RJ		Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM		Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO		Saudável

Conjunto de Dados

- **Importante:** lidar com dados imperfeitos (ruídos, ausentes, etc.)



Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	#	Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP		Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG		Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS		Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG		Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE		Saudável
2301	Ana	22	F	?	Inexistentes	38,0	3	RJ		Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM		Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO		Saudável

Tarefas de Aprendizado

- Divisão geral: **Preditivas** x **Descritivas**

Previsão

Encontrar função (modelo ou hipótese) que possa ser utilizada para prever um rótulo ou valor para novos dados

Objetos na forma (entrada, saída)

Descrição

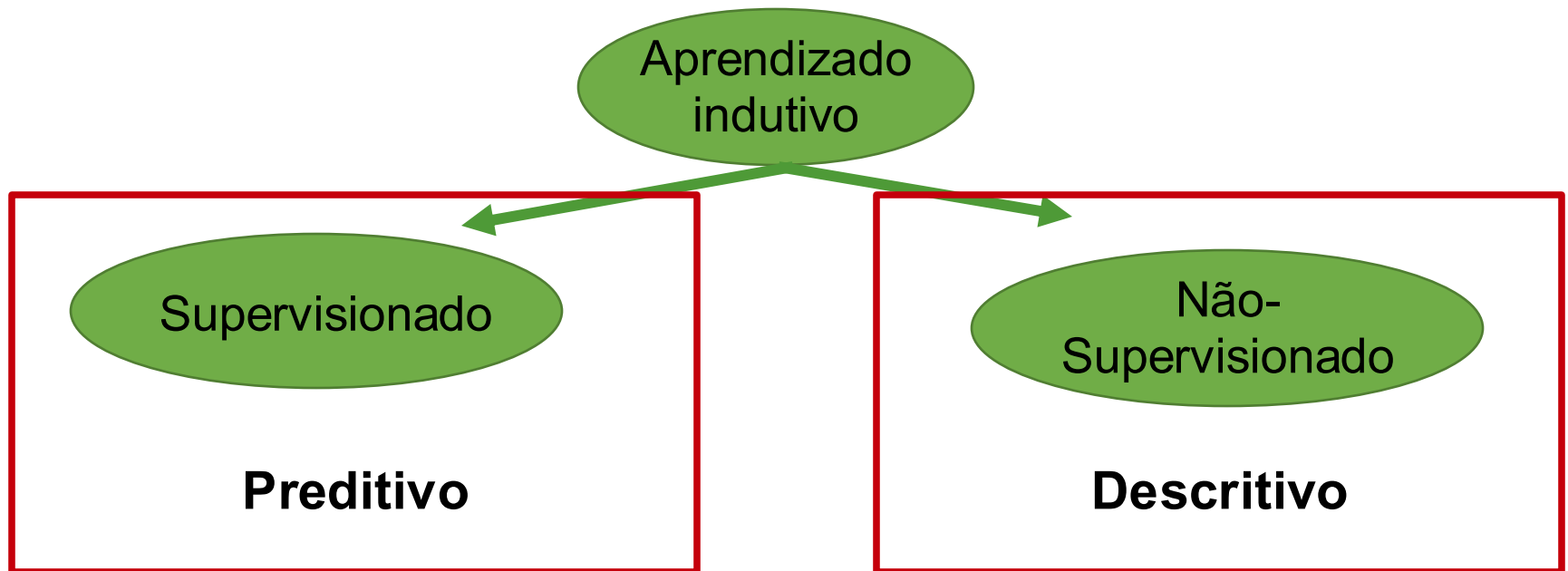
Explorar ou descrever um conjunto de dados

Objetos não possuem saída associada

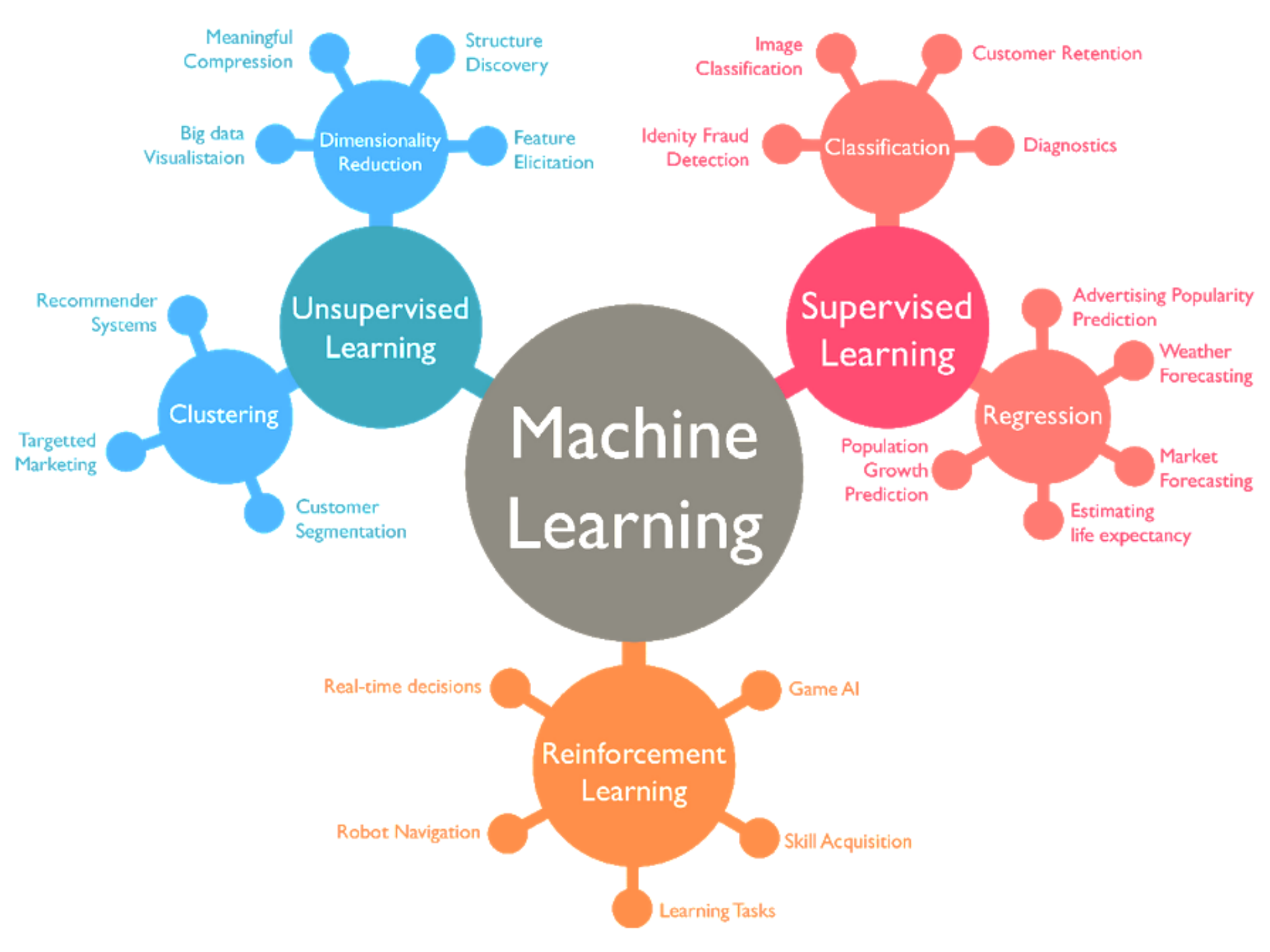
Hierarquia de aprendizado

- Aprendizado supervisionado
- Aprendizado não-supervisionado
- Aprendizado por reforço

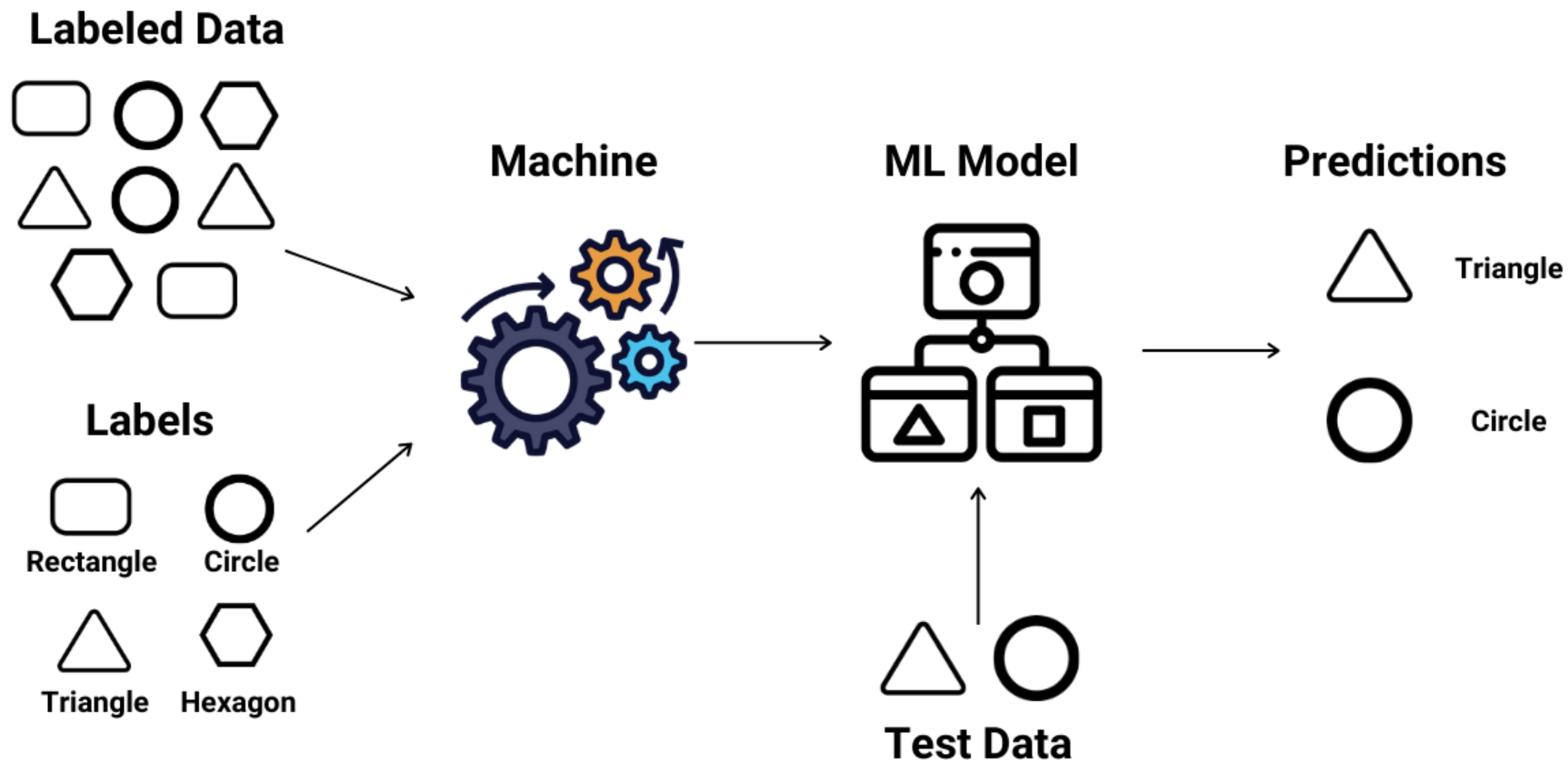
Hierarquia de aprendizado



Importante: divisão não é rígida (modelo preditivo também provê descrição dos dados e modelo descritivo pode prover previsões após validado)
E existem outras formas de realizar o aprendizado preditivo/descritivo

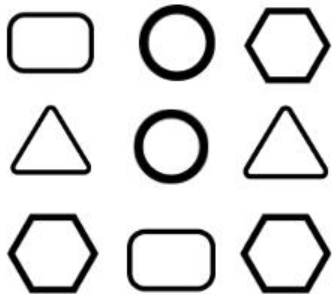


Supervised Learning



Unsupervised Learning

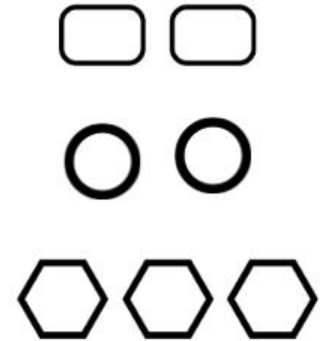
Unlabelled Data



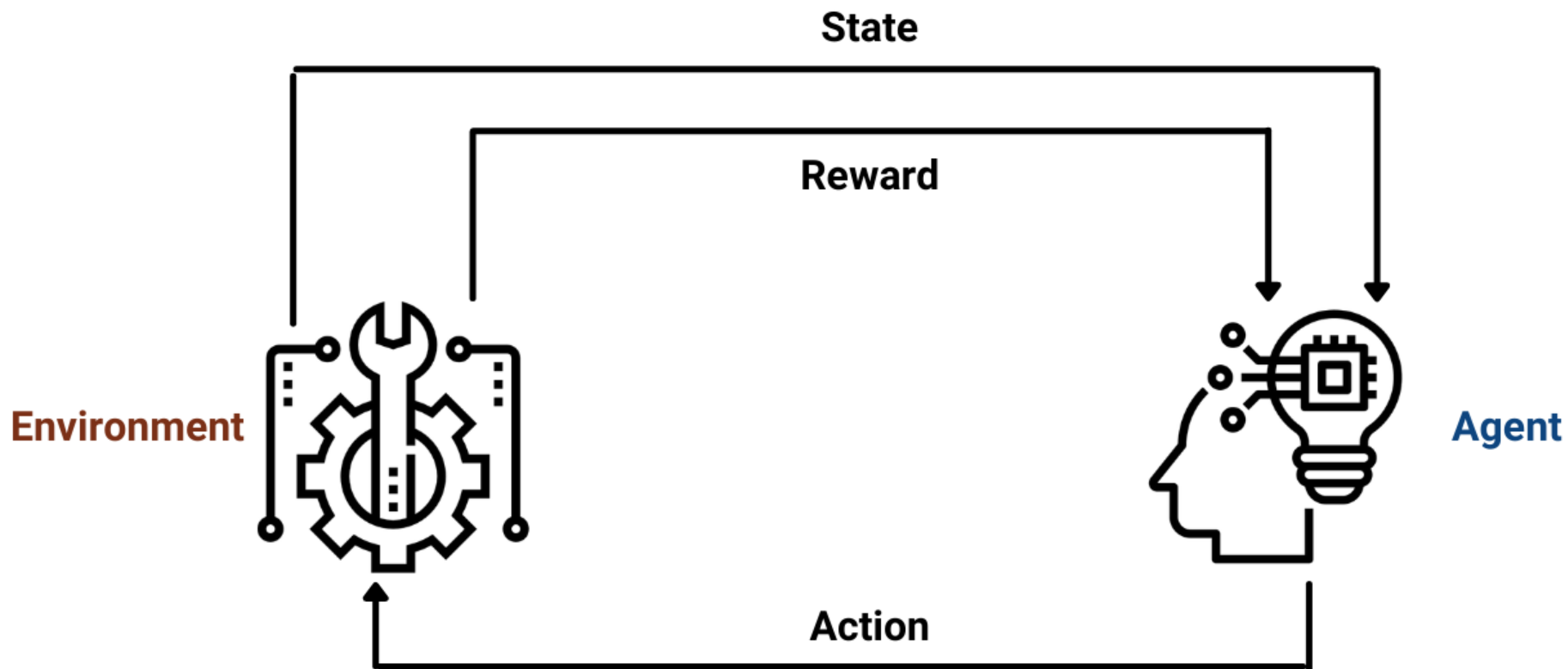
Machine



Results



Reinforcement Learning



Aprendizado Supervisionado

- Conjunto de dados rotulados por um “supervisor”
- Modelo estatístico que aprende regras
- Objetivo é prever a resposta para observações futuras (predição)

Aprendizado Supervisionado

Classificação

Rótulos discretos

Ex.: diagnóstico, bom/mau pagador, etc.

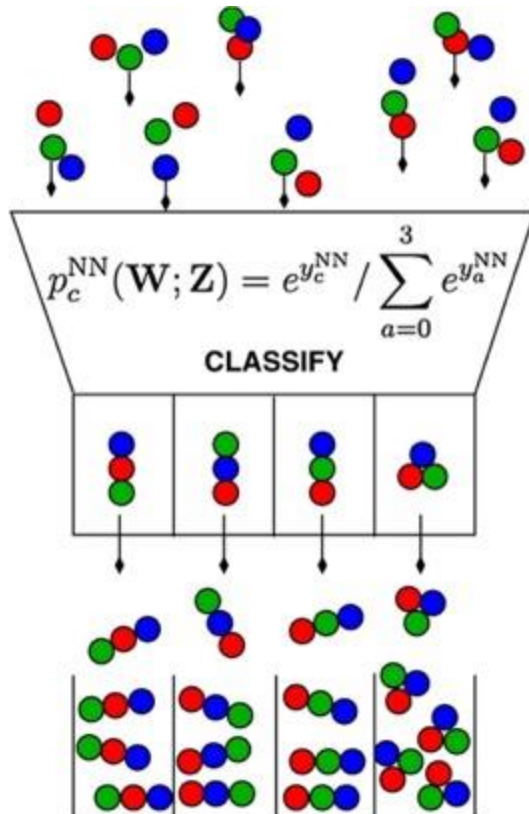
Regressão

Rótulos contínuos

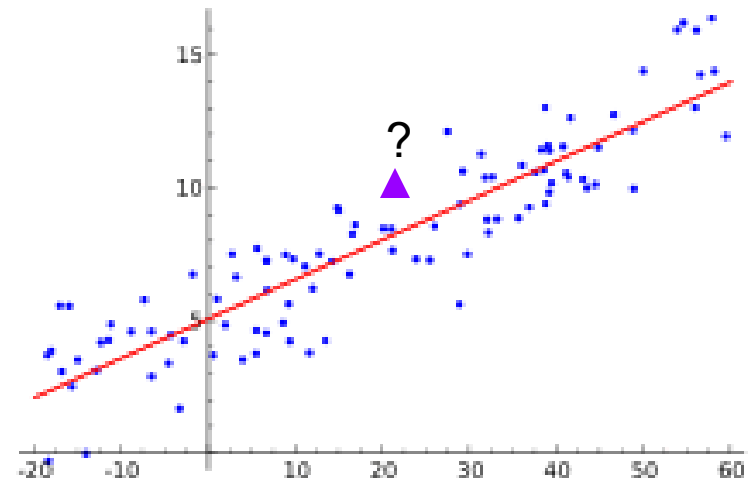
Ex.: peso, altura, etc.

Aprendizado Supervisionado

Classificação



Regressão



Aprendizado Não Supervisionado

- Não há dados rotulados
- Algoritmo recebe os dados de entrada e o objetivo é encontrar regularidades e simetrias nos dados a partir de padrões que possam ser detectados nesses dados

Aprendizado Não Supervisionado

Sumarização

Encontrar descrição compacta para dados

Associação

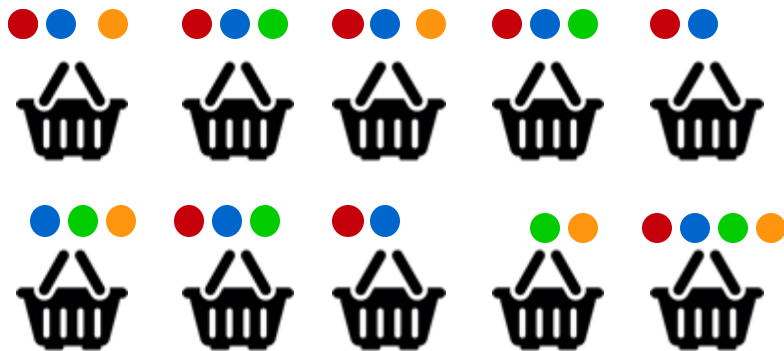
Encontrar padrões frequentes de associações entre atributos

Agrupamento

Dados agrupados de acordo com sua similaridade

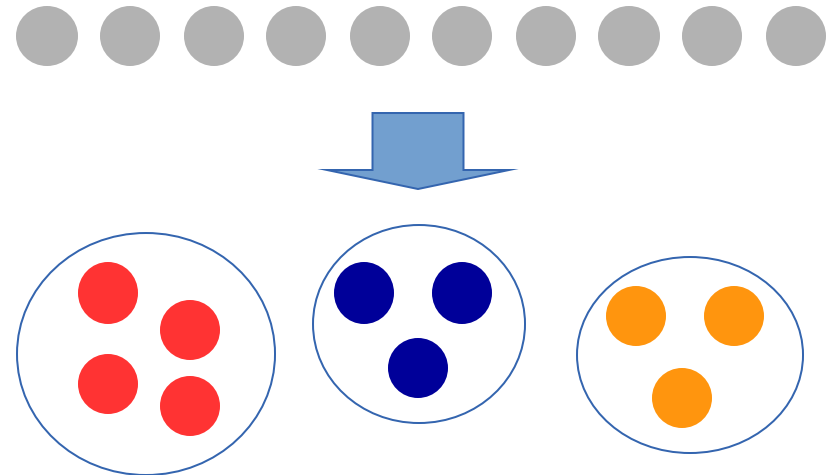
Aprendizado Não Supervisionado

Associação



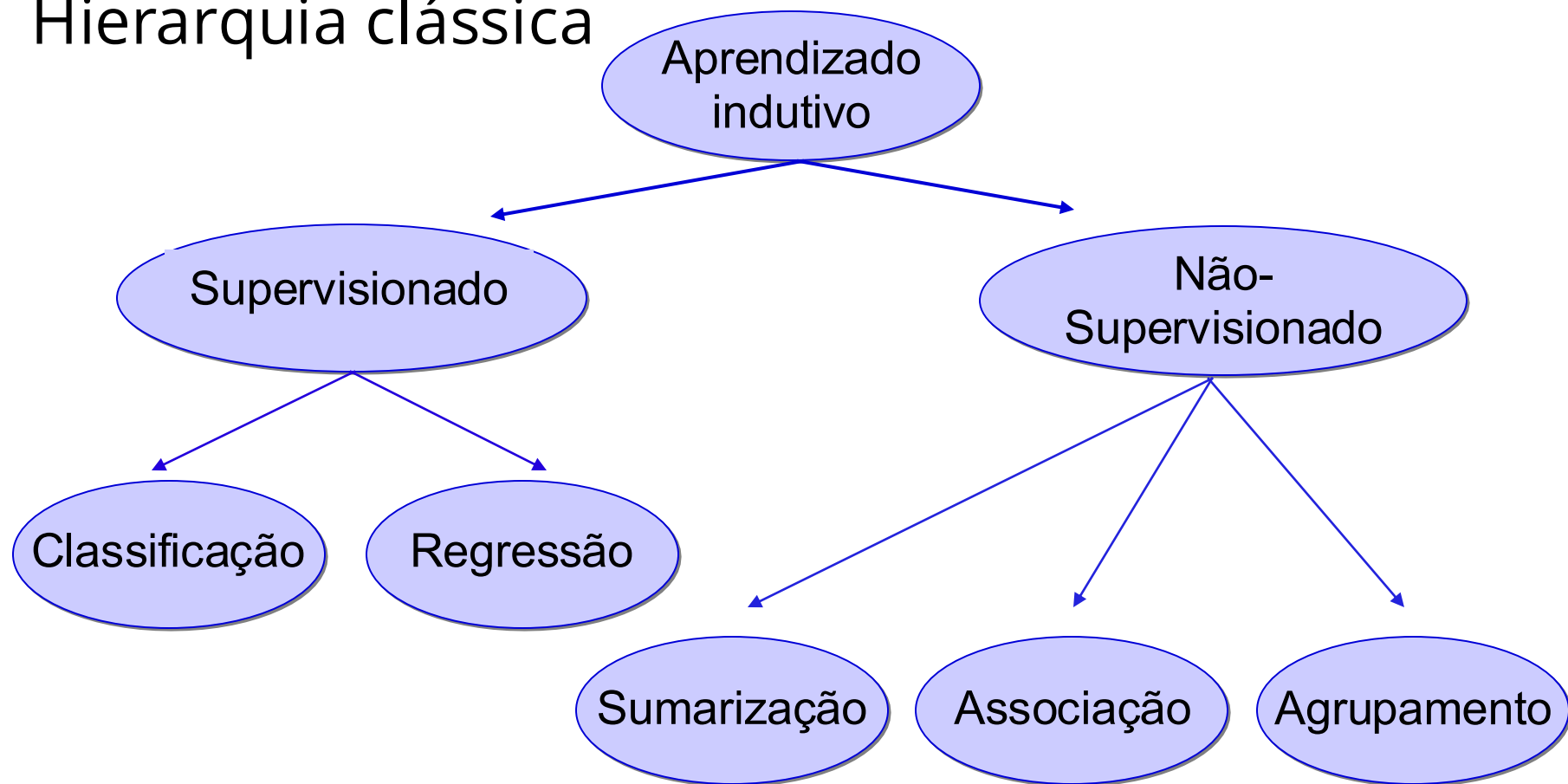
Que itens serão
comprados em
conjunto?

Agrupamento



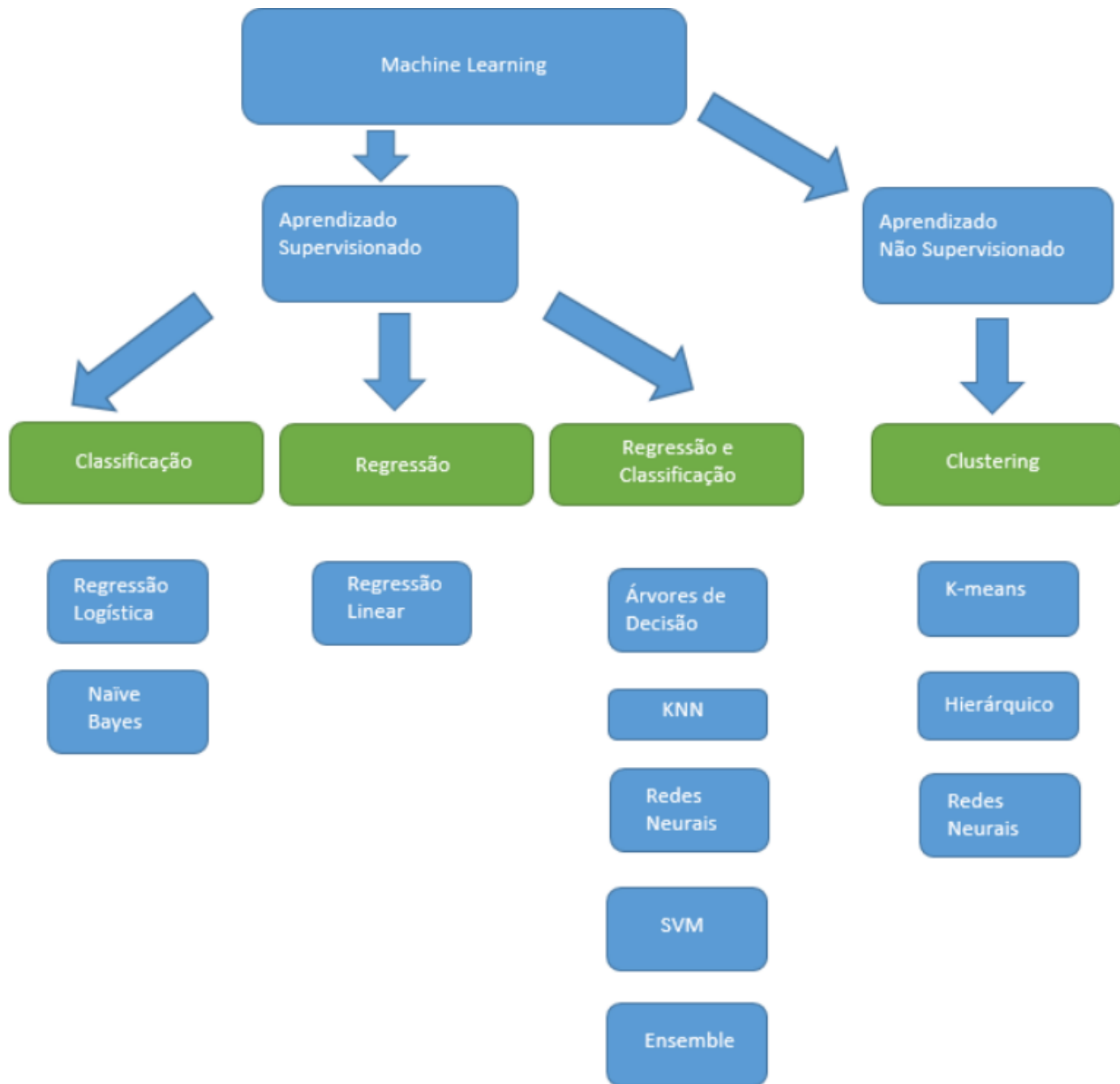
Hierarquia de aprendizado

Hierarquia clássica



Aprendizado por Reforço

- Um agente aprende como atingir um objetivo complexo através da execução de certos passos
- Deve maximizar uma recompensa
- Algoritmo de aprendizado por reforço é treinado para descobrir quais ações tomadas levam à maior recompensa



Exercício 1

- Imagine que você queira criar um sistema que prevê se um cliente clicará em um anúncio online. Defina esse problema usando os quatro componentes discutidos nos exemplos:
 - **Problema:** (Qual é o objetivo geral?)
 - **Tarefa:** (Qual ação específica o sistema deve executar?)
 - **Medida de desempenho:** (Como você saberá se o sistema está indo bem?)
 - **Experiência:** (De que tipo de dados o sistema aprenderia?)

Exercício 2

Dado um conjunto de problemas abaixo, classifique cada um como um problema de **classificação** ou **regressão**:

1. Prever se um e-mail é spam ou não spam.
2. Estimar o preço de um carro com base em suas características.
3. Determinar se um cliente é um bom ou mau pagador.
4. Prever a temperatura de uma cidade para o próximo dia.

Exercício 3

Explique a diferença entre Aprendizado **Supervisionado** e **Não Supervisionado**. Em seguida, identifique a categoria correta para os seguintes problemas:

1. Agrupar clientes com base em seus hábitos de compra.
2. Identificar padrões em imagens médicas para ajudar no diagnóstico de doenças.
3. Descobrir quais produtos são frequentemente comprados juntos em um supermercado.
4. Criar um modelo para prever se um paciente está doente ou saudável com base nos sintomas.

Exercício 4

Imagine que você deseja construir um modelo de aprendizado de máquina para prever se uma pessoa tem uma determinada doença com base em dados médicos. Responda às seguintes perguntas:

1. Quais atributos (variáveis de entrada) você consideraria para treinar o modelo?
2. O problema se encaixa em Aprendizado Supervisionado ou Não Supervisionado? Por quê?
3. Como você poderia medir a qualidade do modelo?