Instituto Iracema
PESQUISA E INOVAÇÃO

Capacitação em Inteligência Artificial e Aplicações

Universidade Estadual do Ceará

LUMEN AD VIAM

Classificadores Supervisionados: Avaliação da Classificação

- Prof. Gerson Vieira Albuquerque Neto
- Prof. Rodrigo Carvalho Souza Costa
- Prof. Yves Augusto Romero

MCTI FUTURO

Softex

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO

GOVERNO FEDERAL
BRASIL
UNIÃO E RECONSTRUÇÃO

IA

Planejamento da Disciplina

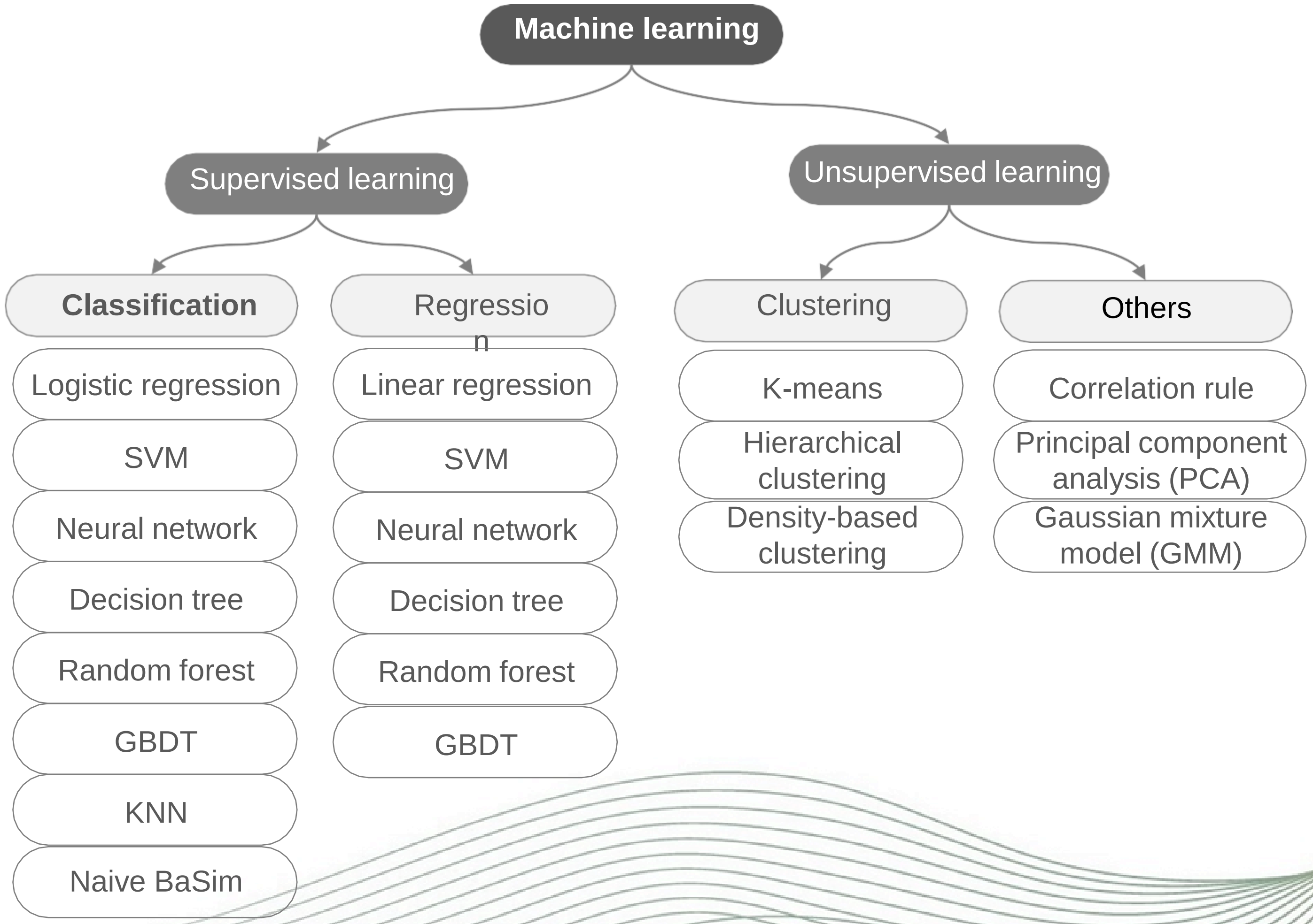
D	S	T	Q	Q	S	S
26	27 Introdução ao curso	28 Áreas e aplicações de IA	29 Tipos e definições de Inteligência artificial	30 Revisão de álgebra e probabilidade	31 Laboratório Python 1	1
2	3 Introdução aos classificadores supervisionados	4 Aula teórica Naive BaSim	5 Aula prática Naive BaSim	6 Feriado Semana Santa	7 Feriado Semana Santa	8
9	10 KNN + Métricas de Avaliação	11 Regressão Linear e e Introdução à árvores de decisão	12 Prática Regressão Lienar + Árvores de Decisão	13 Introdução à Clusterização + KMédias	14 Introdução ao PCA / prática com classificadores já implementados	15
16	17 Introdução ao Perceptron Simples – Prática	18 Teoria MLP / Aplicação scilearn	19 Introdução ao DeepLearning	20 Uso de biblioteca DeepLearning	21 Feriado Tiradentes	28
23	24 Introdução ao TensorFlow / Keras	25 Introdução ao Pytorch	26 Tensorflow for android	27	28	29

Objetivos da Aula

- Após a conclusão deste módulo, você será capaz de:
  - Compreensão de algoritmos de regressão linear
  - Compreensão do algoritmo de árvore de decisão

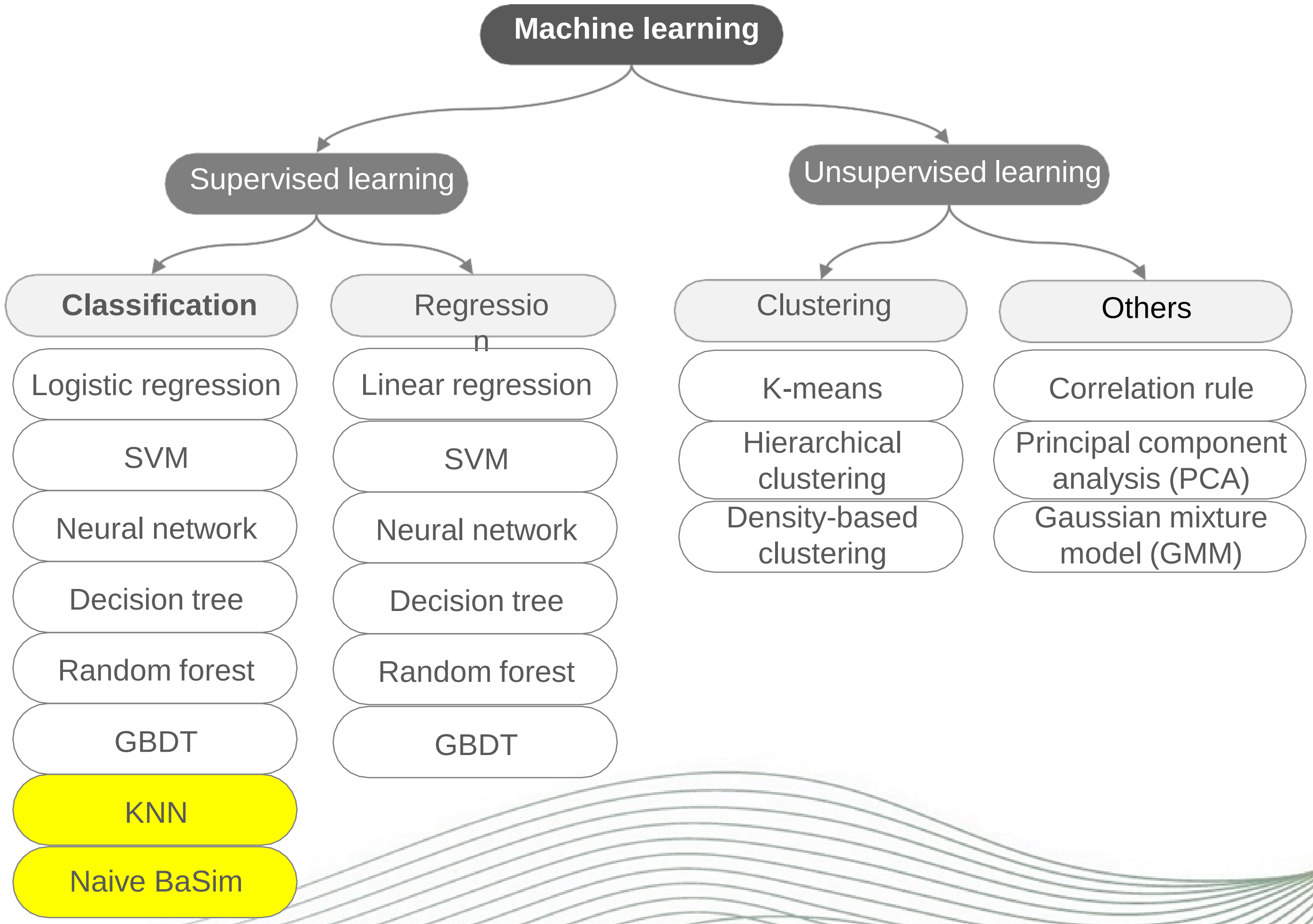
IA

Revisão: Algoritmos de Machine Learning



IA

Revisão: Algoritmos de Machine Learning

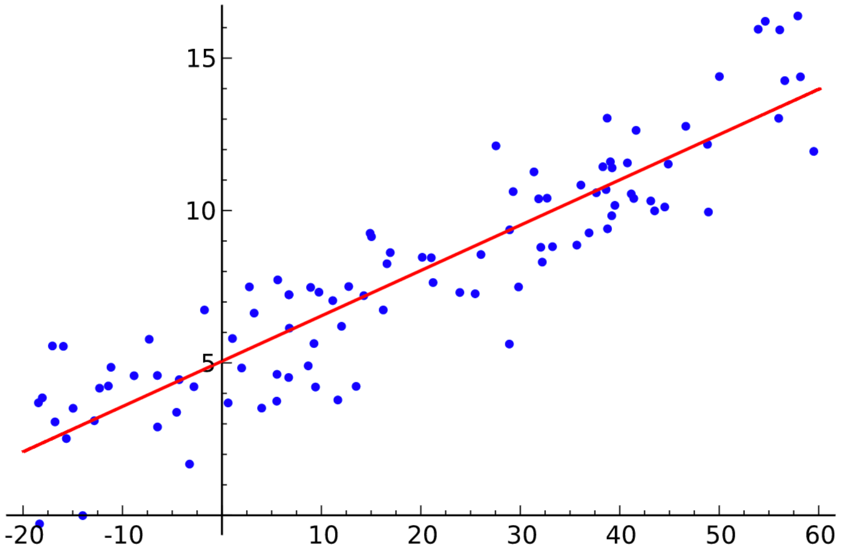


Classificadores Supervisionados:

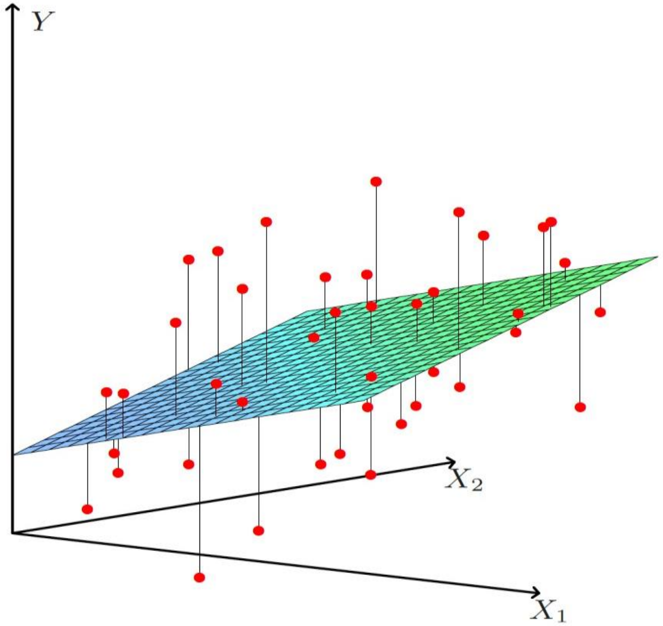
- Princípios sobre regressão
- Exemplo de regressão de uma função linear
- Utilizando a regressão como classificação
- Árvore de decisão

Regressão linear

- **Regressão linear**: um método de análise estatística para determinar as relações quantitativas entre duas ou mais variáveis através da análise de regressão em estatística matemática.
- A regressão linear é um tipo de aprendizagem supervisionada.



Regressão linear unidimensional



Regressão linear multidimensional

IA

Regressão Linear

- A função do modelo de regressão linear é a seguinte, onde **w** indica o parâmetro de peso, **b** indica o viés e **x** indica o vetor de características da amostra **i**.

$$h_w(x_i) = w^T x_i + b$$

- A relação entre o valor previsto pelo modelo e o valor real é a seguinte, onde **y** indica o valor real e **ε** indica o erro.

$$y = w^T x + b + \varepsilon$$

- O erro **ε** é influenciado por muitos fatores de forma independente. De acordo com o teorema do limite central, o erro **ε** segue a distribuição normal. De acordo com a função de distribuição Normal e a estimativa de máxima verossimilhança, a função de erro $J(w)$ da regressão linear é a seguinte:

$$J(w) = \frac{1}{2m} \sum (h_w(x) - y)^2$$

- Para tornar o valor previsto próximo do valor real, precisamos minimizar o valor de perda. Podemos usar o **método do gradiente descendente** para calcular o parâmetro de peso **w** quando a função de perda atinge o mínimo e, em seguida, completar a construção do modelo.

IA

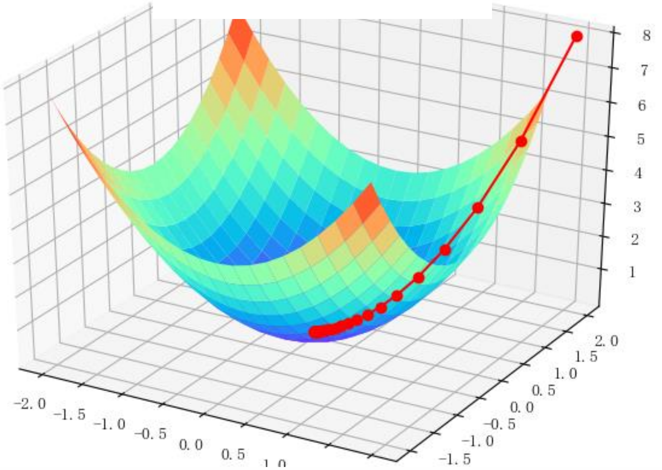
Gradiente Descendente

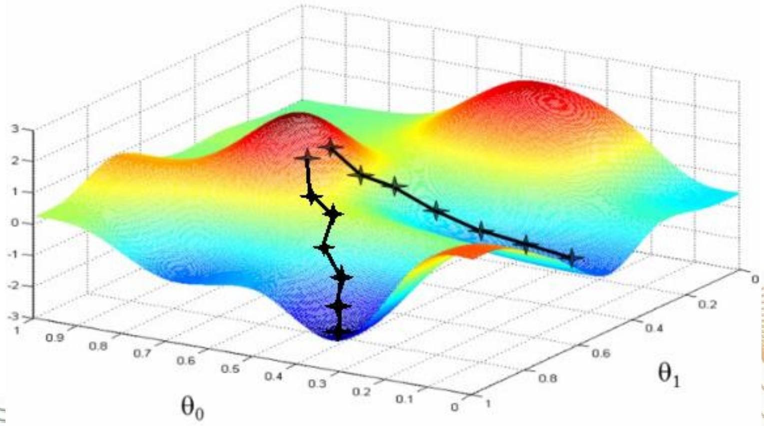
- O método do gradiente descendente usa a direção de gradiente negativo da posição atual como a direção de pesquisa, que é a direção mais íngreme. A fórmula é a seguinte:

$$w_{k+1} = w_k - \eta \nabla f_{w_k}(x^i)$$

- Na fórmula, **η** indica a taxa de aprendizagem e **i** indica o número de registro de dados **i**. O parâmetro de peso **w** indica a alteração em cada iteração.
- **Convergência**: O valor da função objetivo muda muito pouco, ou o número máximo de iterações é atingido.

Superficie de custo





Gradiente da função de erro da regressão linear

- Exemplo para uma função linear $\mathbf{y} = w_1 x + w_2$

$$J(w) = \frac{1}{2m}\sum_{i=1}^{m}(h_w(x_i) - y)^2 = \frac{1}{2m}\sum_{k=1}^{m}(w_1 x_k + w_2 - y)^2$$

$$\frac{dJ(w)}{dw_k} = \frac{1}{2m}\sum \frac{d(w_1 x + w_2 - y)^2}{dw_k}$$

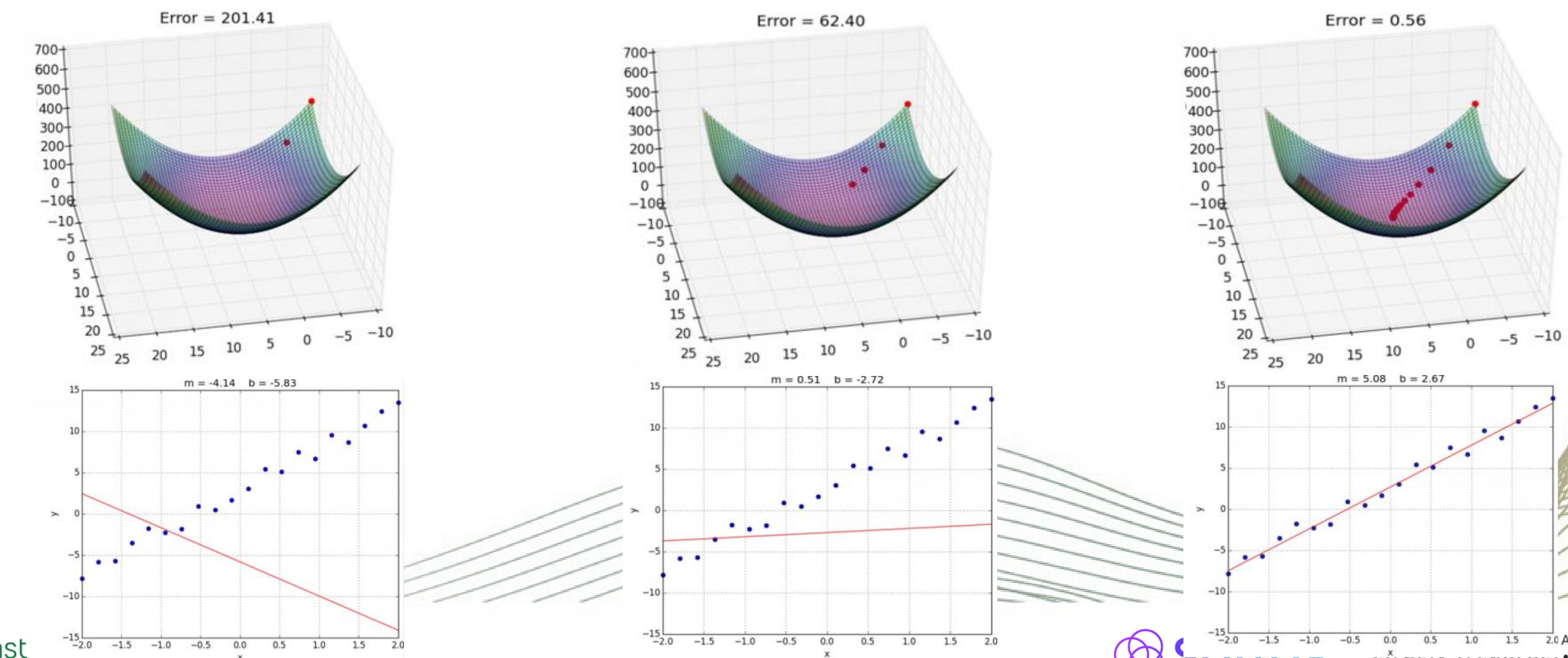
$$\frac{dJ(w)}{dw_1} = \frac{1}{m}\sum_{k=1}^{m} x_k \cdot (w_1 x + w_2 - y)$$

$$\frac{dJ(w)}{dw_2} = \frac{1}{m}\sum_{k=1}^{m} 1 \cdot (w_1 x + w_2 - y)$$

IA

Gradiente Descendente

- A seguir está um exemplo de uma iteração de descida de gradiente. Podemos ver que, à medida que os pontos vermelhos na superfície da função de perda se aproximam gradualmente de um ponto mais baixo, o ajuste da linha vermelha de regressão linear com os dados torna-se cada vez melhor.
- Neste momento, podemos obter os melhores parâmetros.



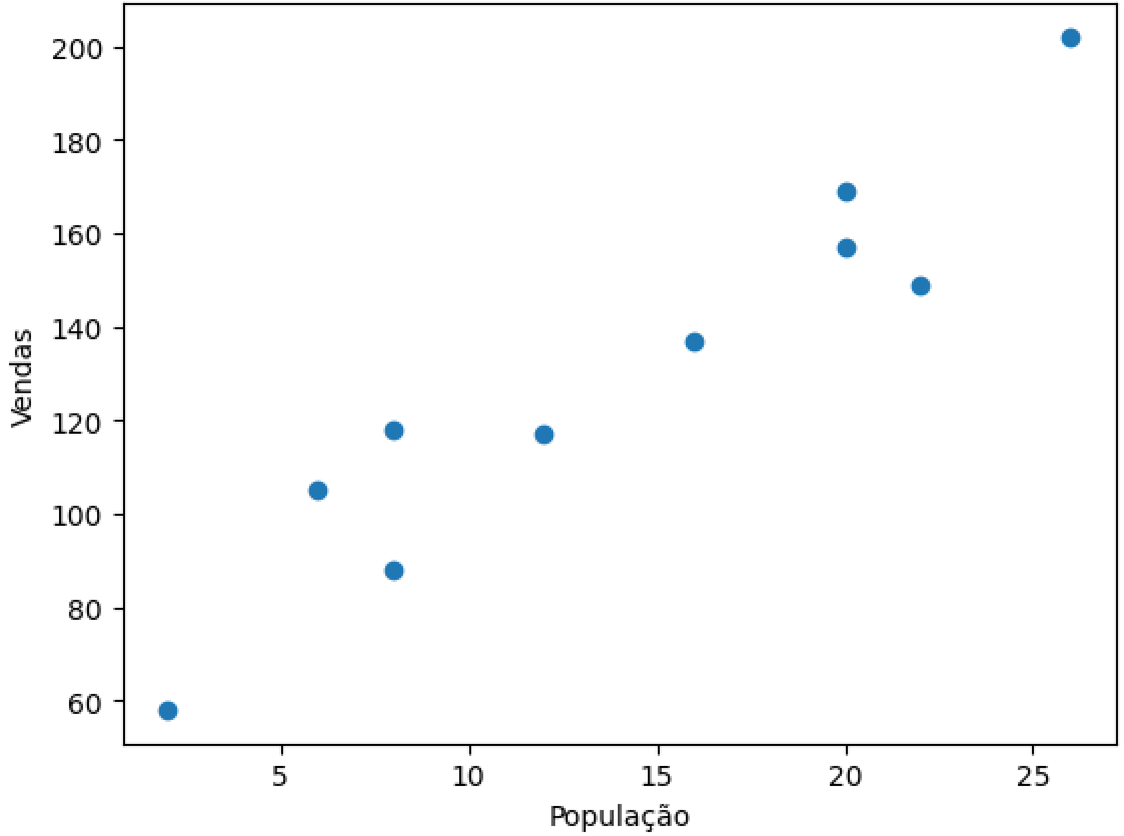
Classificadores Supervisionados:

- Princípios sobre regressão
- **Exemplo de regressão de uma função linear**
- Regressão como classificador
- Árvore de decisão

Exemplo

Restaurante	População
1	2
2	6
3	8
4	8
5	12
6	16
7	20
8	20
9	22
10	26

$$y = w_1 x + w_2$$
$$w_1 = 4$$
$$w_2 = 30$$



IA

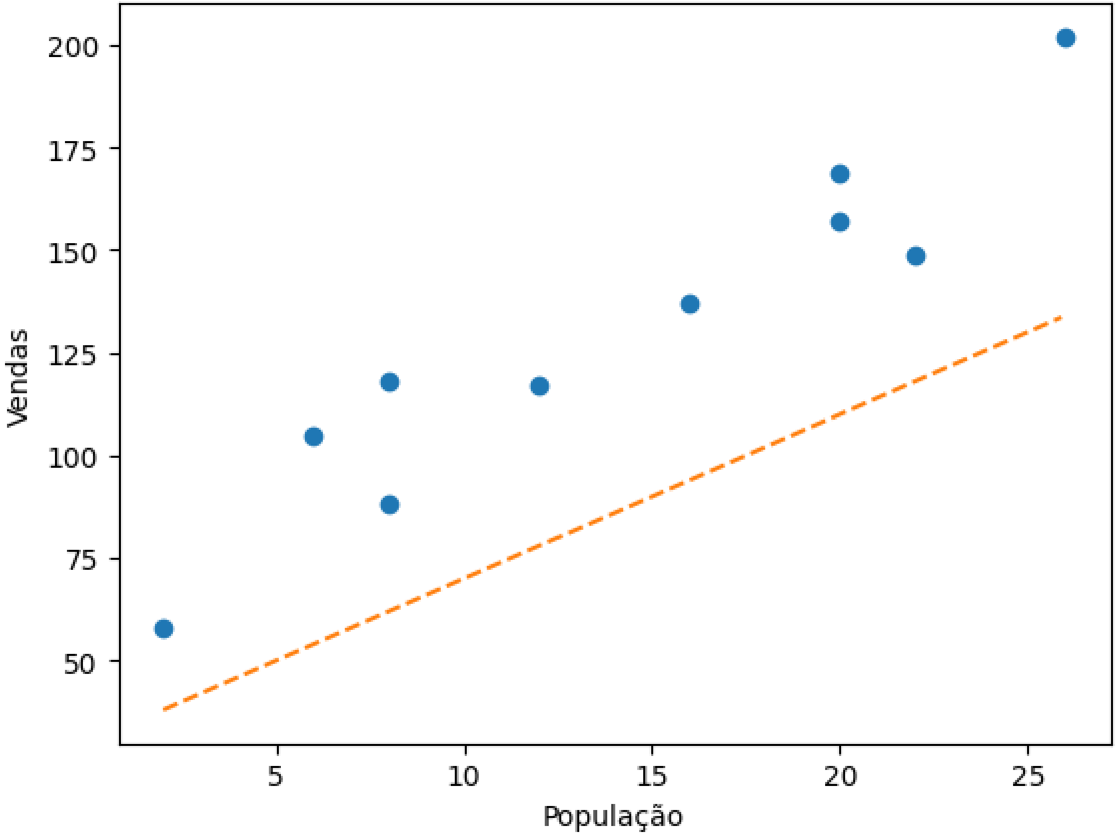
Exemplo

Restaurante	População
1	2
2	6
3	8
4	8
5	12
6	16
7	20
8	20
9	22
10	26

$$y = w_1 x + w_2$$
$$w_1 = 4$$
$$w_2 = 30$$

= 4

1ª Iteração



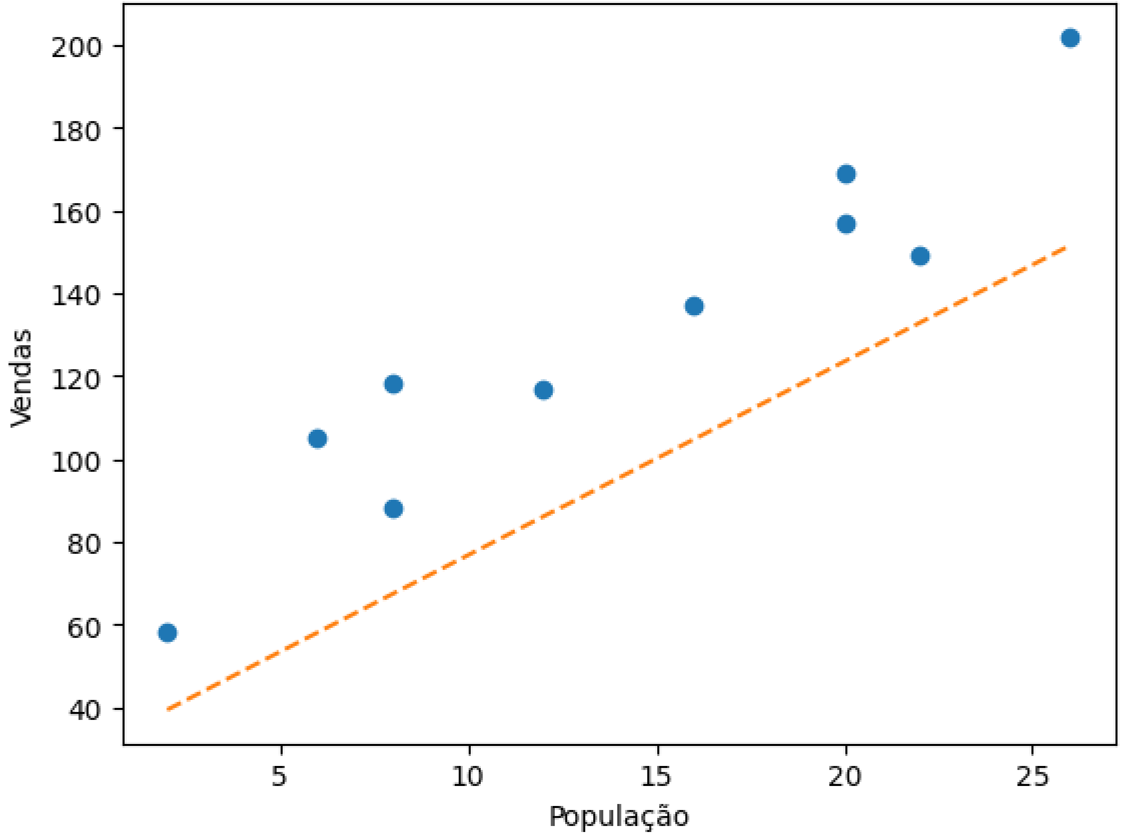
IA

Exemplo

2ª Iteração

Restaurante	População
1	2
2	6
3	8
4	8
5	12
6	16
7	20
8	20
9	22
10	26

$$y = w_1 x + w_2$$
$$w_1 = 4.6728$$
$$w_2 = 30.044$$

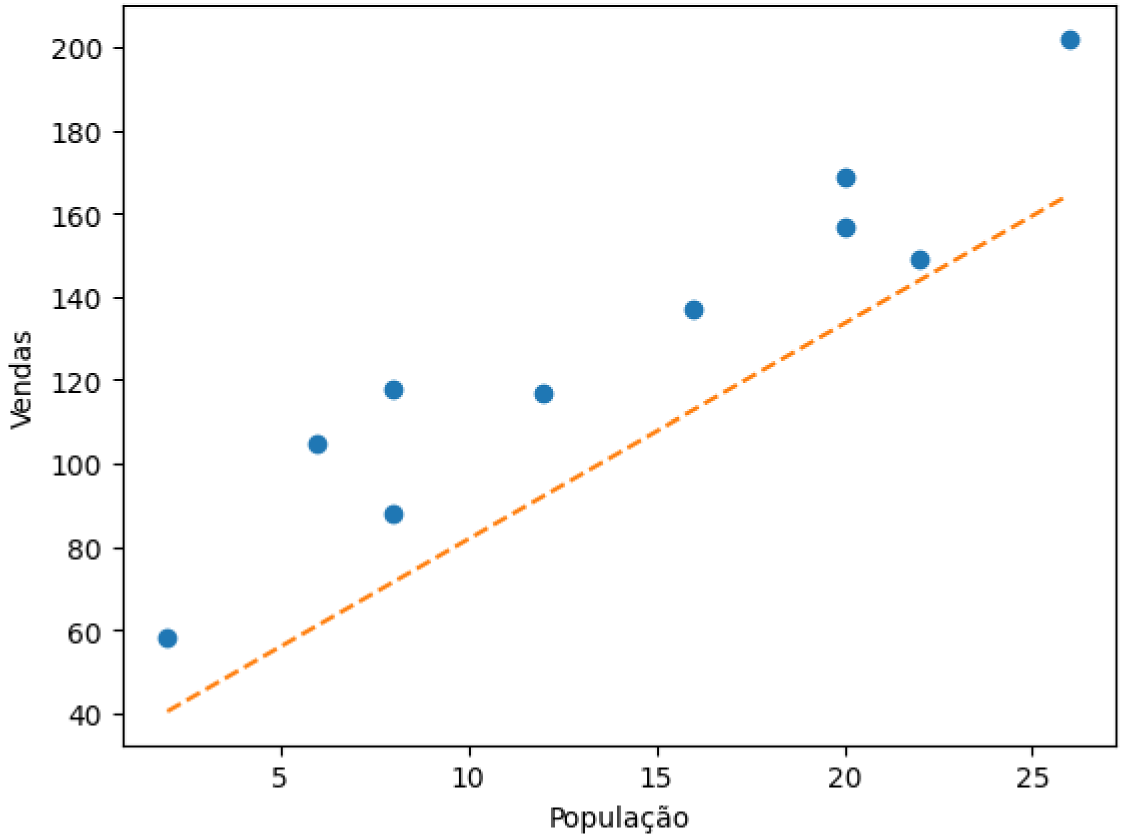


Exemplo

3ª Iteração

Restaurante	População
1	2
2	6
3	8
4	8
5	12
6	16
7	20
8	20
9	22
10	26

$$y = w_1 x + w_2$$
$$w_1 = 5.1749$$
$$w_2 = 30.0785$$

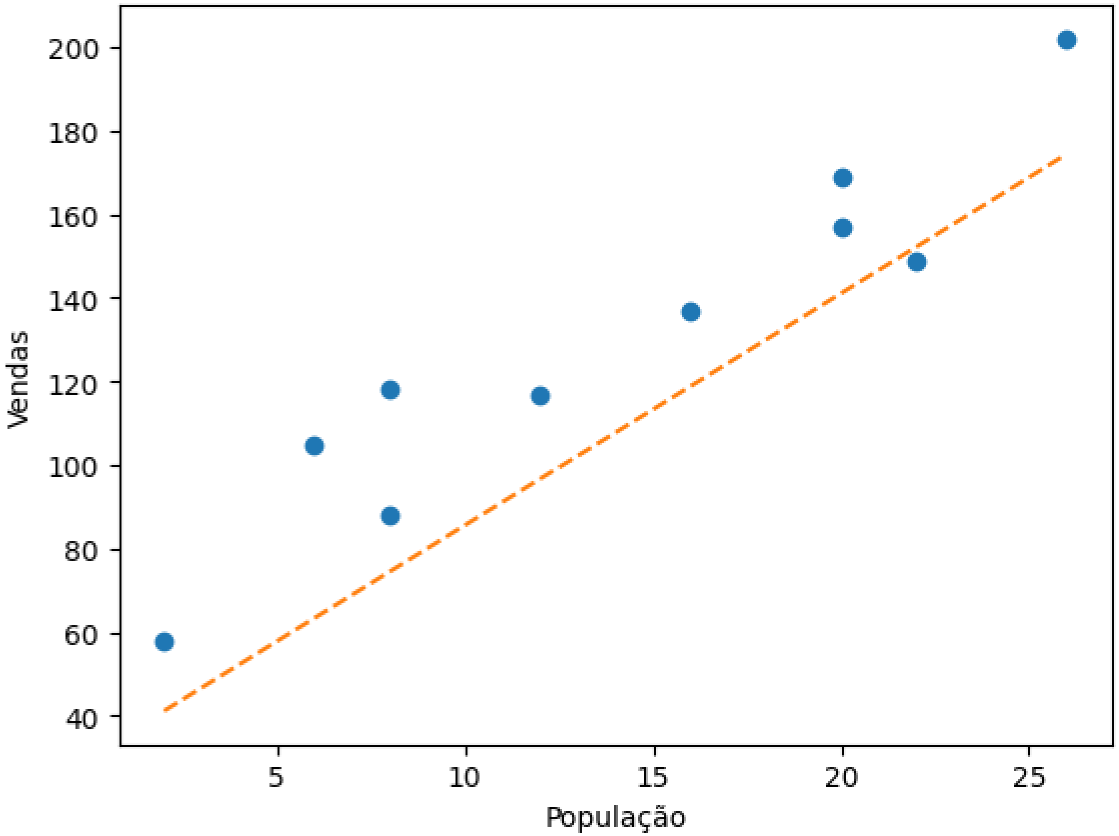


Exemplo

4ª Iteração

Restaurante	População
1	2
2	6
3	8
4	8
5	12
6	16
7	20
8	20
9	22
10	26

$$y = w_1 x + w_2$$
$$w_1 = 5.5495$$
$$w_2 = 30.1060$$

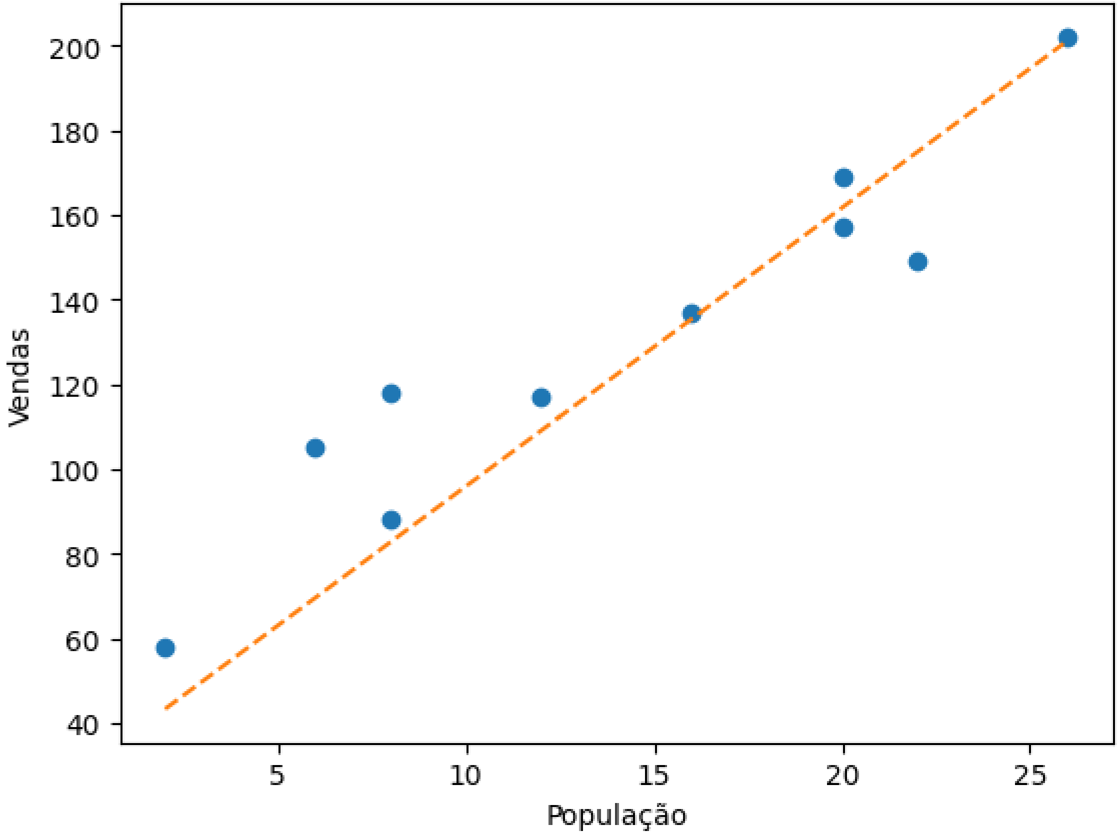


Exemplo

14ª Iteração

Restaurante	População
1	2
2	6
3	8
4	8
5	12
6	16
7	20
8	20
9	22
10	26

$$y = w_1 x + w_2$$
$$w_1 = 6.5708$$
$$w_2 = 30.2228$$

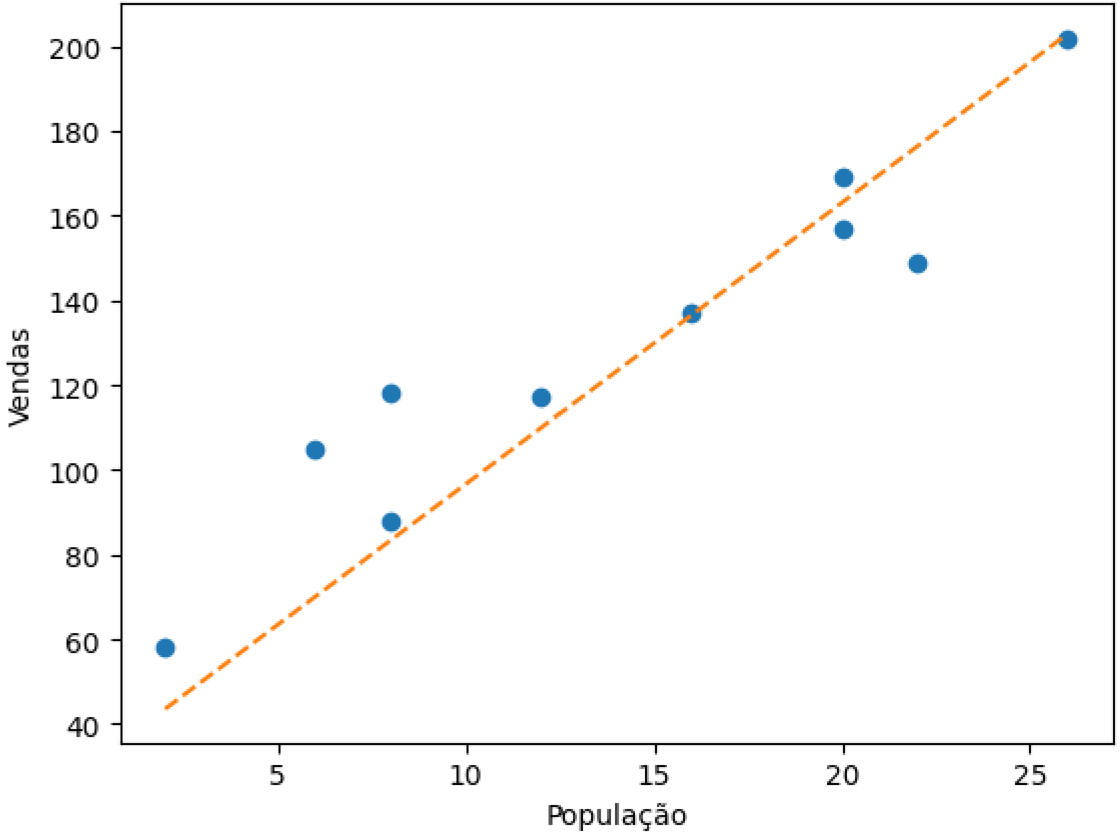


Exemplo

24ª Iteração

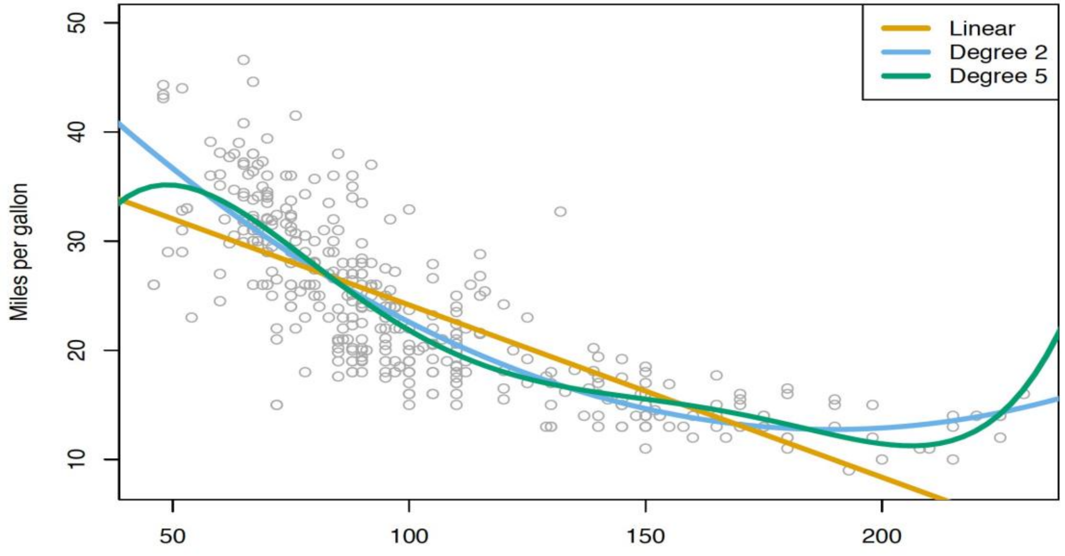
Restaurante	População
1	2
2	6
3	8
4	8
5	12
6	16
7	20
8	20
9	22
10	26

$$y = w_1 x + w_2$$
$$w_1 = 6.6412$$
$$w_2 = 30.2869$$



Extensão de Regressão Linear - Regressão Polinomial

- A regressão polinomial é uma extensão da regressão linear.
- Geralmente, a complexidade de um conjunto de dados excede a possibilidade de ajuste por uma linha reta. Ou seja, o subajuste (underfitting) óbvio ocorre se o modelo de regressão linear original for usado.
- A solução é usar regressão polinomial: $h_w(x) = w_1 x + w_2 x^2 + \cdots + w_n x^n + b$
- onde, a enésima potência **n** é uma dimensão de regressão polinomial (grau).
- A regressão polinomial pertence à regressão linear, pois a relação entre seus parâmetros de peso **w** ainda é linear, enquanto sua não linearidade é refletida na dimensão da feição.



Comparação entre regressão linear e regressão polinomial

Regressão Linear e Prevenção de Sobreajuste (overfitting)

- Os termos de regularização podem ser usados para reduzir o excesso de ajuste. O valor de w não pode ser muito grande ou muito pequeno no espaço da amostra. Você pode adicionar uma perda de soma quadrada na função de destino.

$$J(w) = \frac{1}{2m}\sum(h_w(x) - y)^2 + \lambda\sum\|w\|_2^2$$

- **Termos de regularização (norma)**: O termo de regularização aqui é chamado de **norma L2**. A regressão linear que usa essa função de perda também é chamada de **regressão de Ridge**.

$$J(w) = \frac{1}{2m}\sum(h_w(x) - y)^2 + \lambda\sum\|w\|_1$$

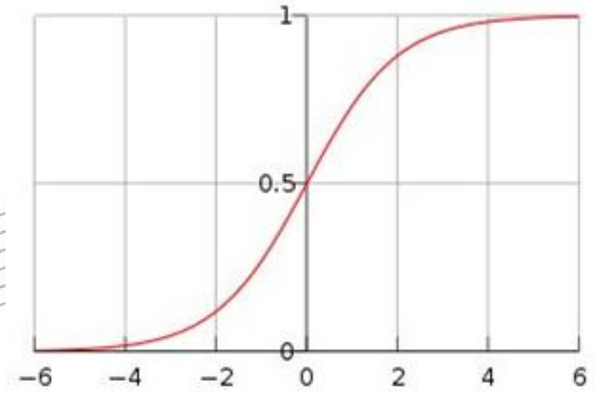
- A **regressão linear** com perda absoluta é chamada de **regressão de Lasso**.

Classificadores Supervisionados:

- Princípios sobre regressão
- Exemplo de regressão de uma função linear
- **Regressão como classificador**
- Árvore de decisão

IA Regressão Logística

- **Regressão logística**: O modelo de regressão logística é utilizado para resolver problemas de classificação. O modelo é definido da seguinte forma:
- $$P(Y=1|x)=\frac{e^{wx+b}}{1+e^{wx+b}}$$
  $$P(Y=0|x)=\frac{1}{1+e^{wx+b}}$$
- onde w indica o vetor de pesos, b indica o viés, e $wx+b$ é considerado como a função linear de x. Compare os dois valores de probabilidade anteriores. A classe com um valor de probabilidade mais alto é a classe de x.



Universidade Estadual do Ceará
Instituto Iracema Pesquisa e Inovação
MCTI Futuro
Softex
Ministério da Ciência, Tecnologia e Inovação
Governo Federal Brasil União e Reconstrução

IA Regressão Logística

- Tanto o modelo de regressão logística quanto o modelo de regressão linear são modelos lineares generalizados.
- A regressão logística introduz fatores não lineares (a função sigmoide) baseados na regressão linear e define limiares, para que possa lidar com problemas de classificação binária.
- De acordo com a função modelo de regressão logística, a função de perda da regressão logística pode ser estimada da seguinte forma, usando a estimativa de máxima verossimilhança:

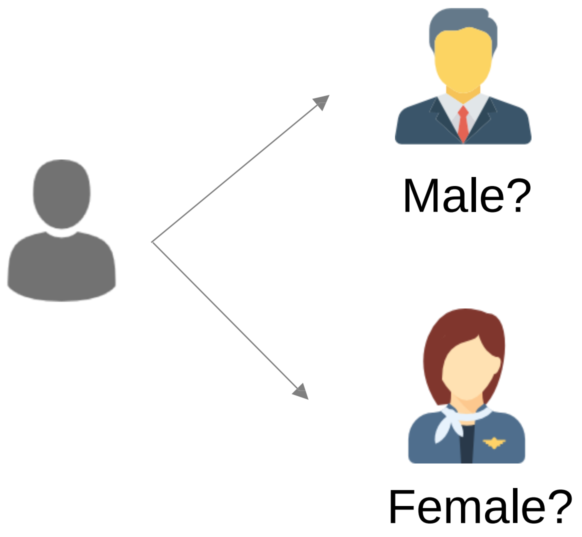
$$J(w) = -\frac{1}{m}\sum(y \ln h_w(x) + (1-y)\ln(1-h_w(x)))$$

- onde w indica o parâmetro de peso, m indica o número de amostras, x indica a amostra e y indica o valor real. Os valores de todos os parâmetros de peso w também podem ser obtidos através do algoritmo de gradiente descendente.

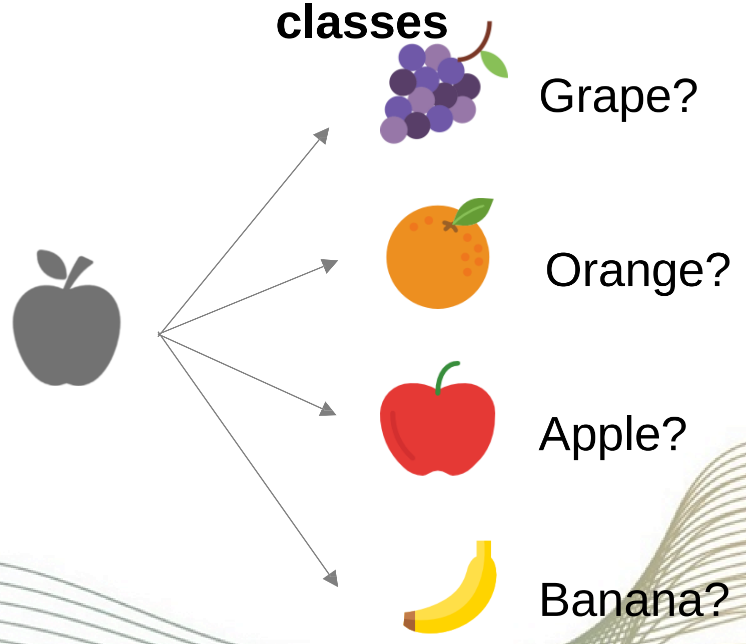
Função Softmax

- A regressão logística aplica-se apenas a problemas de classificação binária. Para problemas de classificação de várias classes, use a função Softmax.
-

Problema de classificação binária



Problema de classificação de várias classes



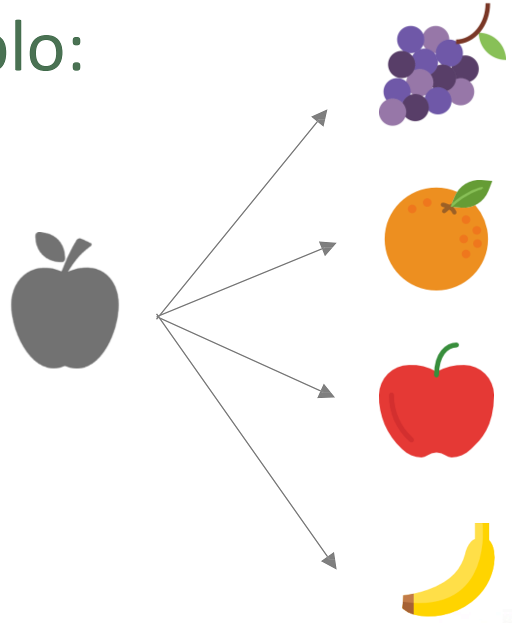
Função Softmax

- A regressão Softmax é uma generalização da regressão logística que podemos usar para a classificação da classe k.
- A função Softmax é usada para mapear um vetor k -dimensional de valores reais arbitrários para outro vetor k -dimensional de valores reais, onde cada elemento vetorial está no intervalo **(0, 1)**.
- A função de probabilidade de regressão de Softmax é a seguinte:

$$p(y=k|x;w)=\frac{e^{w_k^T x}}{\sum_{l=1}^{K} e^{w_l^T x}}, k=1,2\cdots,K$$

Função Softmax

- Softmax atribui uma probabilidade a cada classe em um problema de várias classes. Essas probabilidades devem somar 1.
- Softmax pode produzir uma forma pertencente a uma determinada classe.
- Exemplo:



Category	Probability
Uva?	0.09
Laranja?	0.22
Maçã?	0.68
Banana?	0.01

- **Soma de todas as probabilidades:**
  0,09 + 0,22 + 0,68 + 0,01 =1

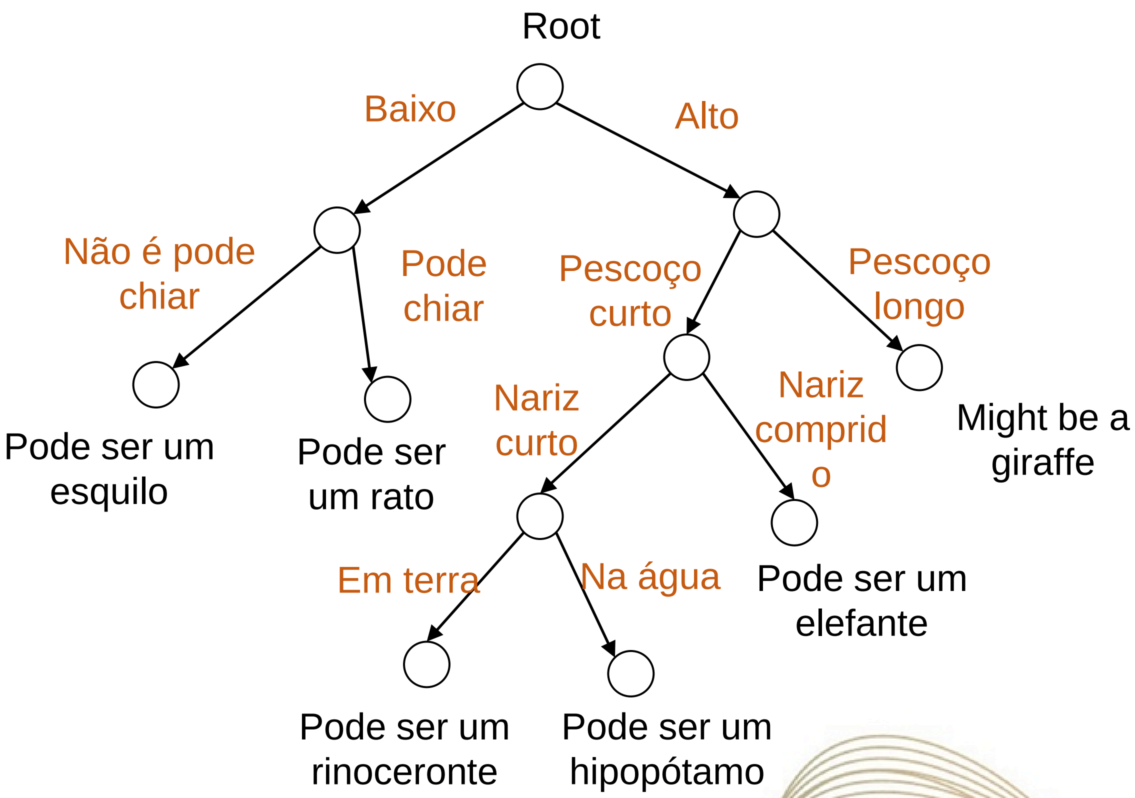
  Muito provavelmente, esta imagem é uma maçã.

Classificadores Supervisionados:

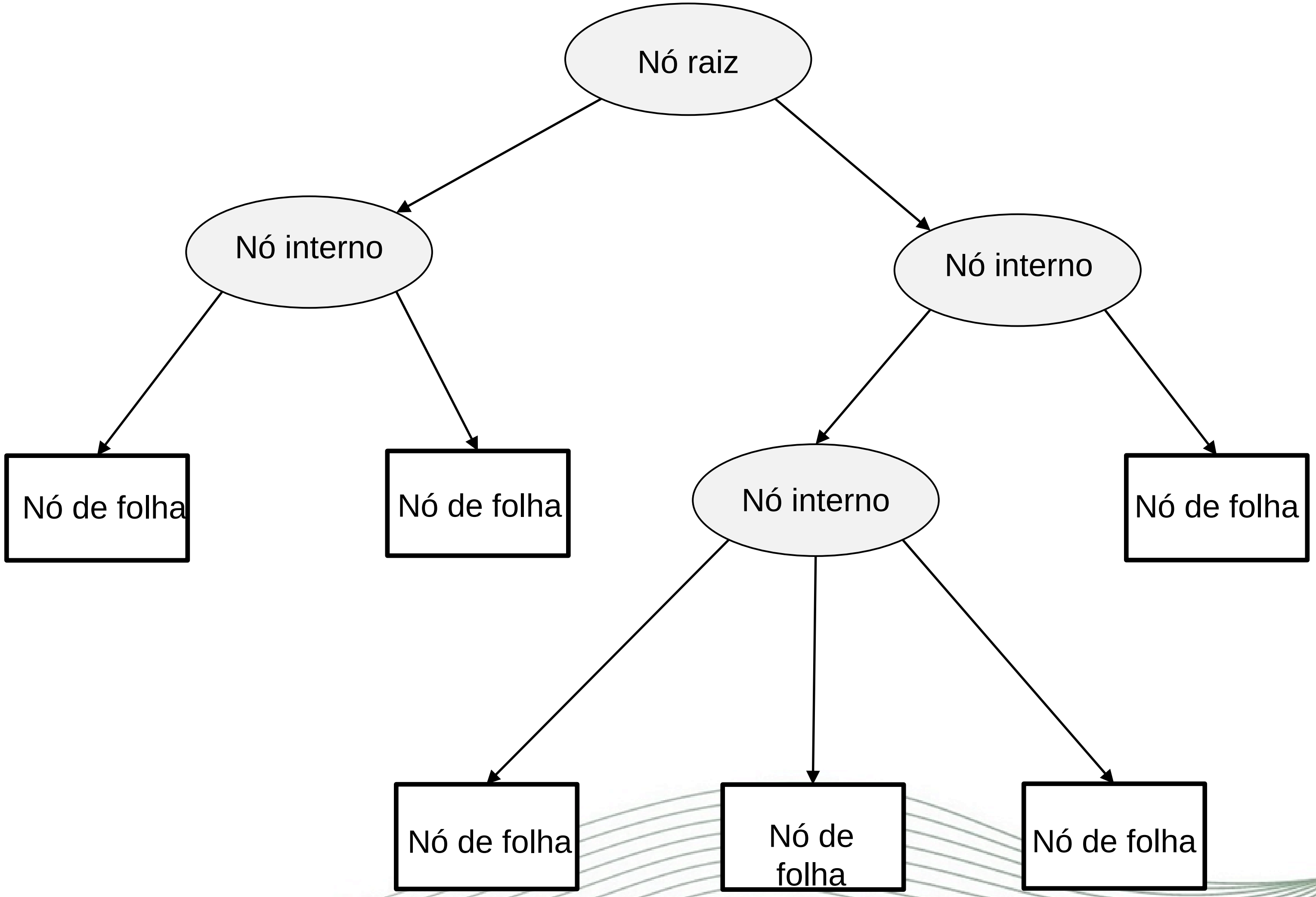
- Princípios sobre regressão
- Exemplo de regressão de uma função linear
- Regressão como classificador
- **Árvore de decisão**

IA Árvore de Decisão

- Uma árvore de decisão é uma estrutura de árvore (uma árvore binária ou uma árvore não binária).
  - Cada nó não folha representa um teste em um atributo de recurso.
  - Cada ramificação representa a saída de um atributo de recurso em um determinado intervalo de valores, e cada nó folha armazena uma categoria.
- Para usar a árvore de decisão, comece a partir do nó raiz, teste os atributos de recurso dos itens a serem classificados, selecione as ramificações de saída e use a categoria armazenada no nó folha como o resultado final.



IA Estrutura da Árvore de Decisão



Pontos-chave para construção da árvore de decisão

- Para criar uma árvore de decisão, precisamos selecionar atributos e determinar a estrutura da árvore entre os atributos de recurso. A principal etapa da construção de uma árvore de decisão é dividir os dados de todos os atributos de recursos, comparar os conjuntos de resultados em termos de 'pureza' e selecionar o atributo com a maior 'pureza' como o ponto de dados para a divisão do conjunto de dados.
- As métricas para quantificar a "pureza" incluem a entropia da informação e o Índice GINI. A fórmula é a seguinte:

$$H(X) = -\sum_{k=1}^{K} p_k \log_2(p_k) \qquad Gini = 1 - \sum_{k=1}^{K} p_k^2$$

- onde p_k indica a probabilidade de que a amostra pertença à classe k (existem classes K no total). Uma diferença maior entre a pureza antes da segmentação e a pós-segmentação indica uma melhor árvore de decisão.
- Algoritmos comuns de árvore de decisão incluem ID3, C4.5 e CART.

Processo de Construção da Árvore de Decisão

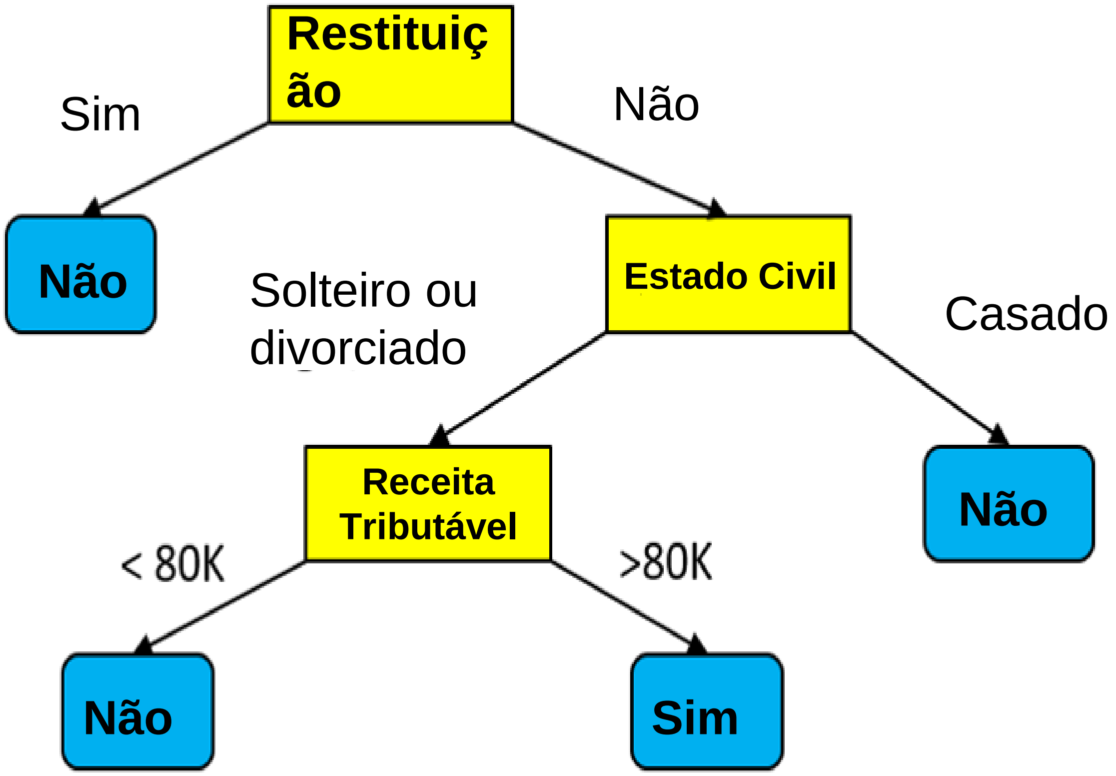
- **Seleção de características**: selecione um recurso dos recursos dos dados de treinamento como o padrão de divisão do nó atual. (Padrões diferentes geram algoritmos de árvore de decisão diferentes.)
- **Geração de árvore de decisão**: gere o nó interno de cabeça para baixo com base nos recursos selecionados e pare até que o conjunto de dados não possa mais ser dividido.
- **Poda (Pruning)**: A árvore de decisão pode facilmente se tornar superajustada (overfitting), a menos que a poda necessária (incluindo pré-poda e pós-poda) seja realizada para reduzir o tamanho da árvore e otimizar sua estrutura de nós.

IA

Árvore de Decisão

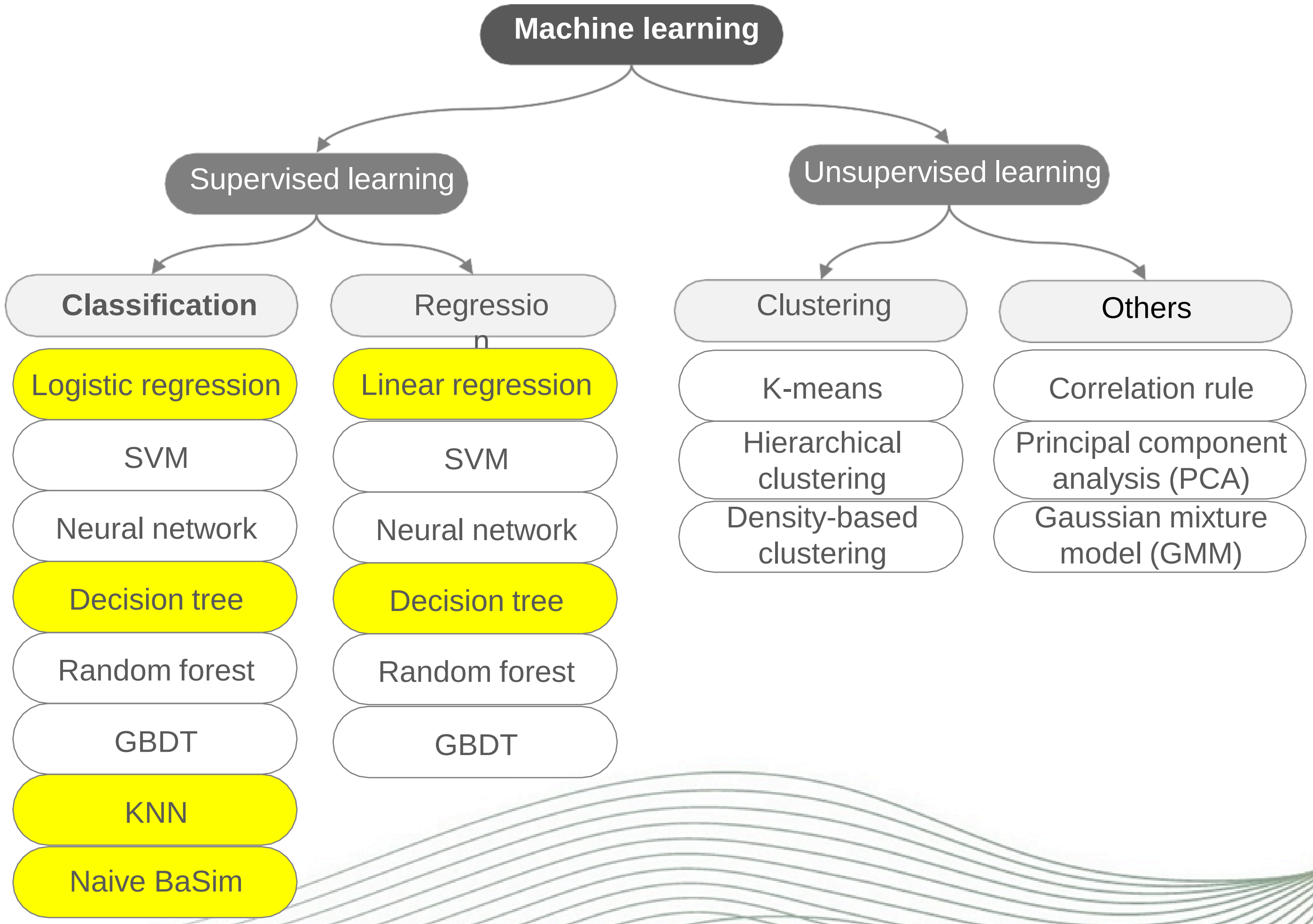
- A figura a seguir mostra uma classificação quando uma árvore de decisão é usada. O resultado da classificação é impactado por três atributos: **Restituição**, **Estado Civil** e **Renda Tributável**.

Tid	Restituição	Estado Civil	Receita Tributável	Fraude
1	Sim	Solteiro	125,000	Não
2	Não	Casado	100,000	Não
3	Não	Solteiro	70,000	Não
4	Sim	Casado	120,000	Não
5	Não	Divorciado	95,000	Sim
6	Não	Casado	60,000	Não
7	Sim	Divorciado	220,000	Não
8	Não	Solteiro	85,000	Sim
9	Não	Casado	75,000	Não
10	Não	Solteiro	90,000	Sim



IA

Revisão: Algoritmos de Machine Learning



Classificadores Supervisionados: próxima aula

- Prática de Regressão Logística e Árvore de decisão

Dúvidas?

Módulo de Inteligência Artificial