

FIAP

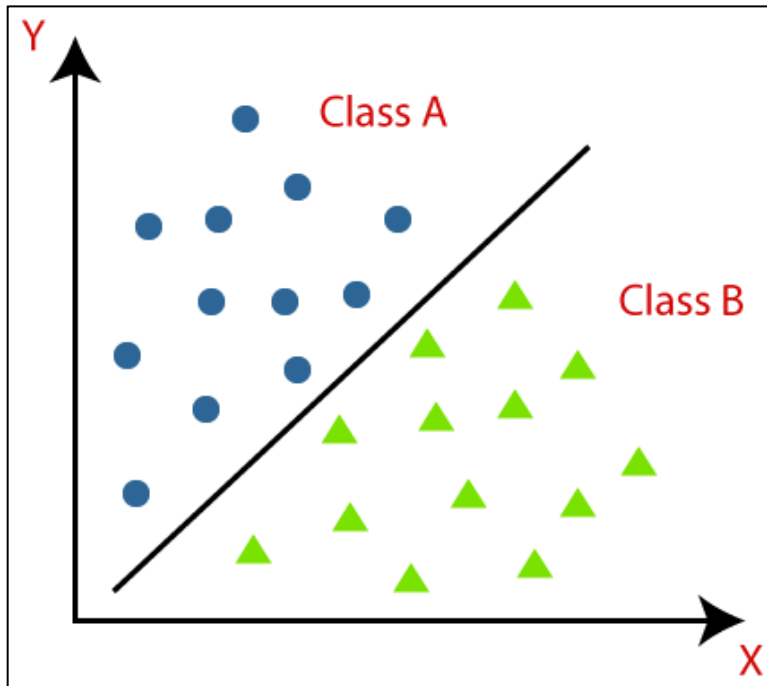
Aprendizado Supervisionado - Regressão

Objetivos da aula:

- Apresentar o conceito de Regressão
- Apresentar e utilizar algoritmo de Regressão linear
- Apresentar e discutir a matriz de correlação

Recapitulando....

O que são problemas de Classificação?



Fonte: <https://www.javatpoint.com/classification-algorithm-in-machine-learning>

1

Atribui rótulos ou categorias a dados de entrada baseados em suas características.

2

Neste tipo de problema, o Modelo recebe, como treinamento, um conjunto de dados chamados atributos, e uma classe, ou label, atrelada a esses dados. O Objetivo é fazer com que o Modelo aprenda a correlação entre dados e labels e generalize para dados que ele nunca viu.

3

Alguns exemplos de aplicação de Modelos de classificação são reconhecimento de imagens, detecção de SPAM, diagnóstico médico, detecção de fraudes, etc.

EXEMPLO

	Atributo A	Atributo B	Atributo C	Classe
Amostra 01	0,5	1,0	1,2	NÃO
Amostra 02	0,5	0,9	1,0	SIM
Amostra 03	1,0	2,5	5,0	NÃO
Amostra 04	1,0	0,8	2,5	SIM

Principais algoritmos de Classificação



Regressão Logística

Técnica da estatística que procura correlacionar dados e produzir uma saída binária do tipo 0 ou 1 (Sim ou Não)



Árvore de Decisão

Algoritmo que divide repetidamente os dados em subconjuntos com base nas suas características até que chegar a um subconjunto único.



K-Nearest Neighbors

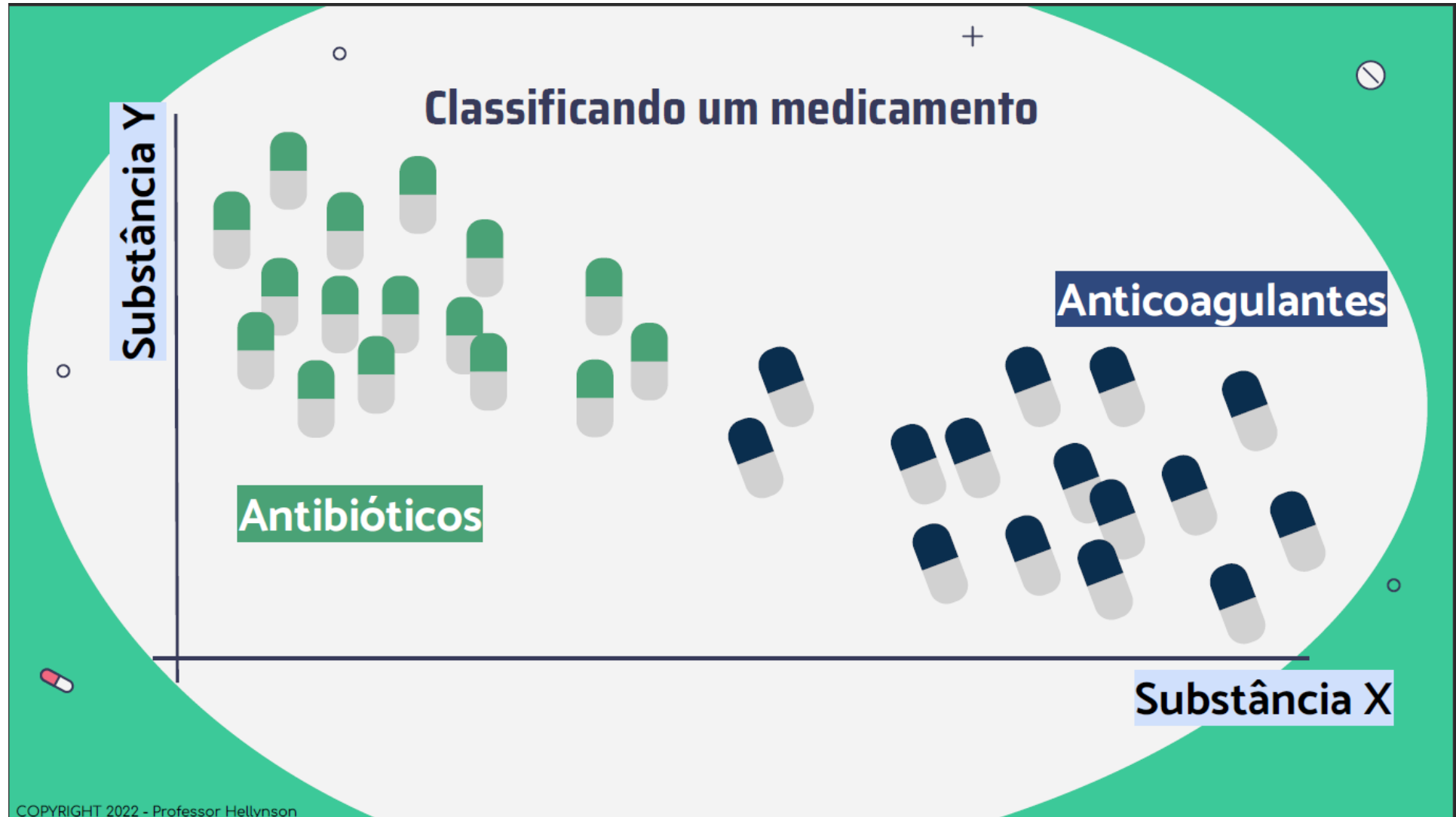
Algoritmo que classifica o dado em um conjunto conforme a semelhança entre os seus vizinhos mais próximos.



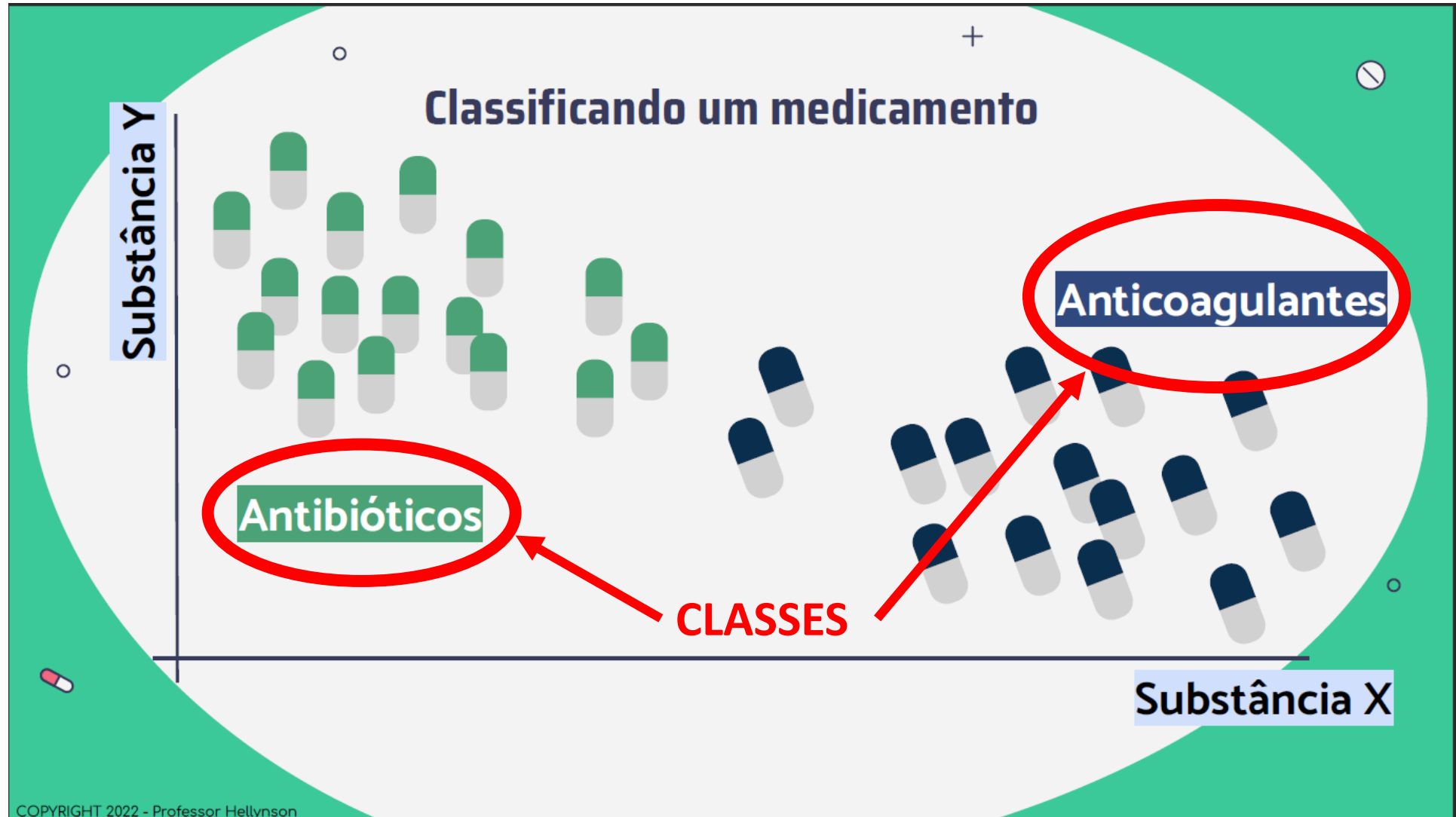
Naive Bayes

Usa o teorema de Bayes para estimar a probabilidade de um dado pertencer a um conjunto com base em suas características independentes.

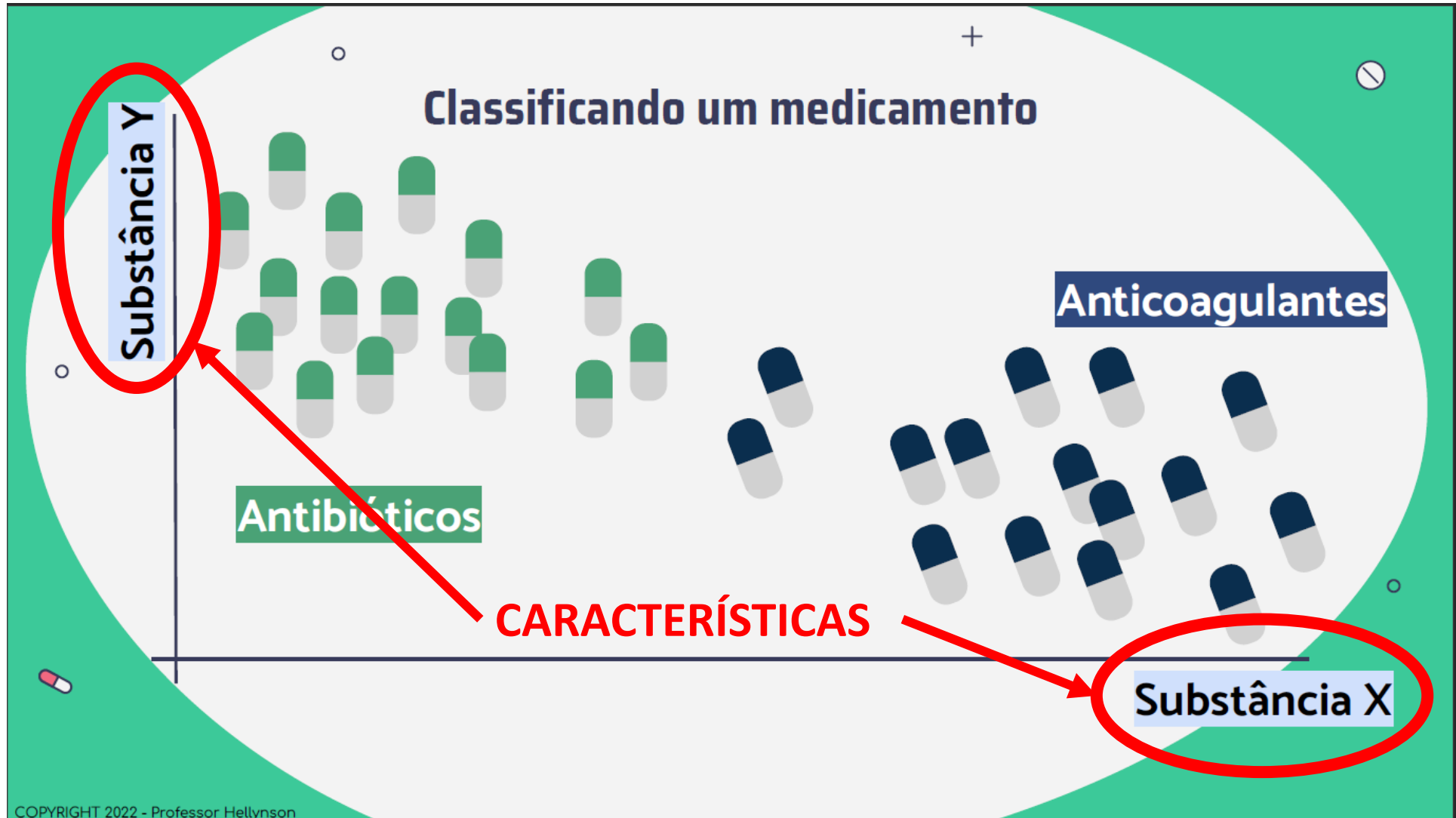
K-NN – K-Nearest Neighbors



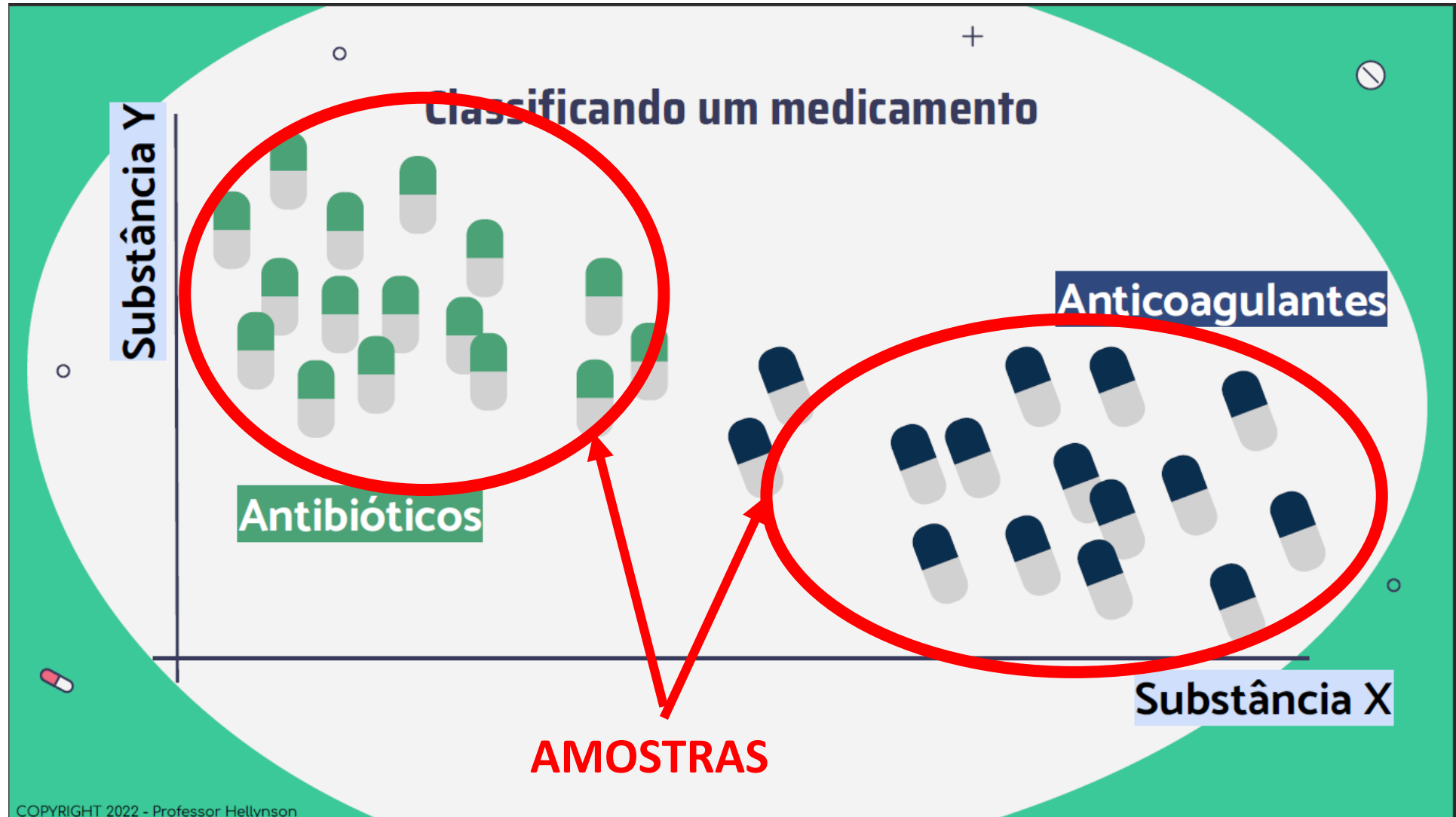
K-NN – K-Nearest Neighbors



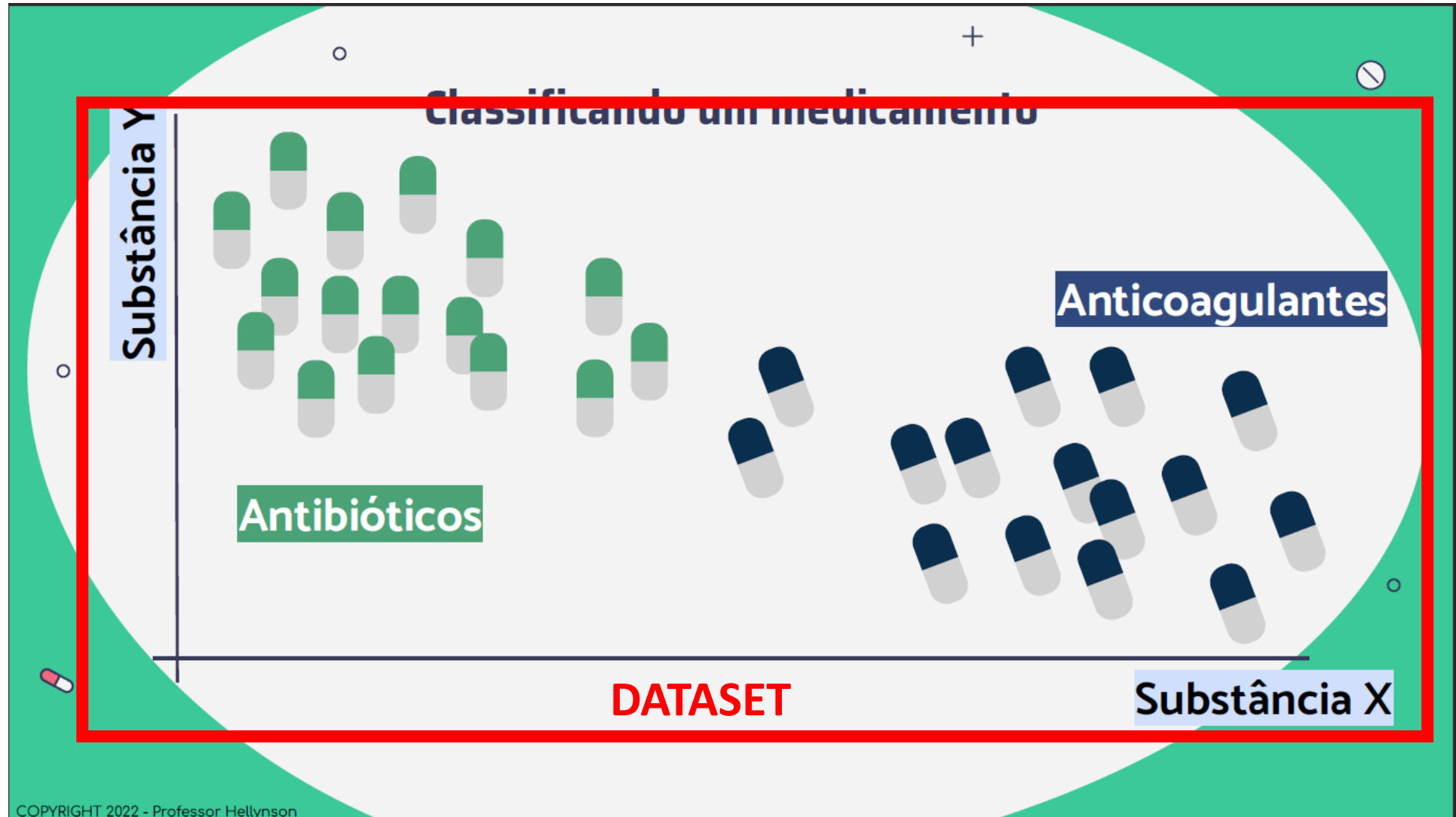
K-NN – K-Nearest Neighbors



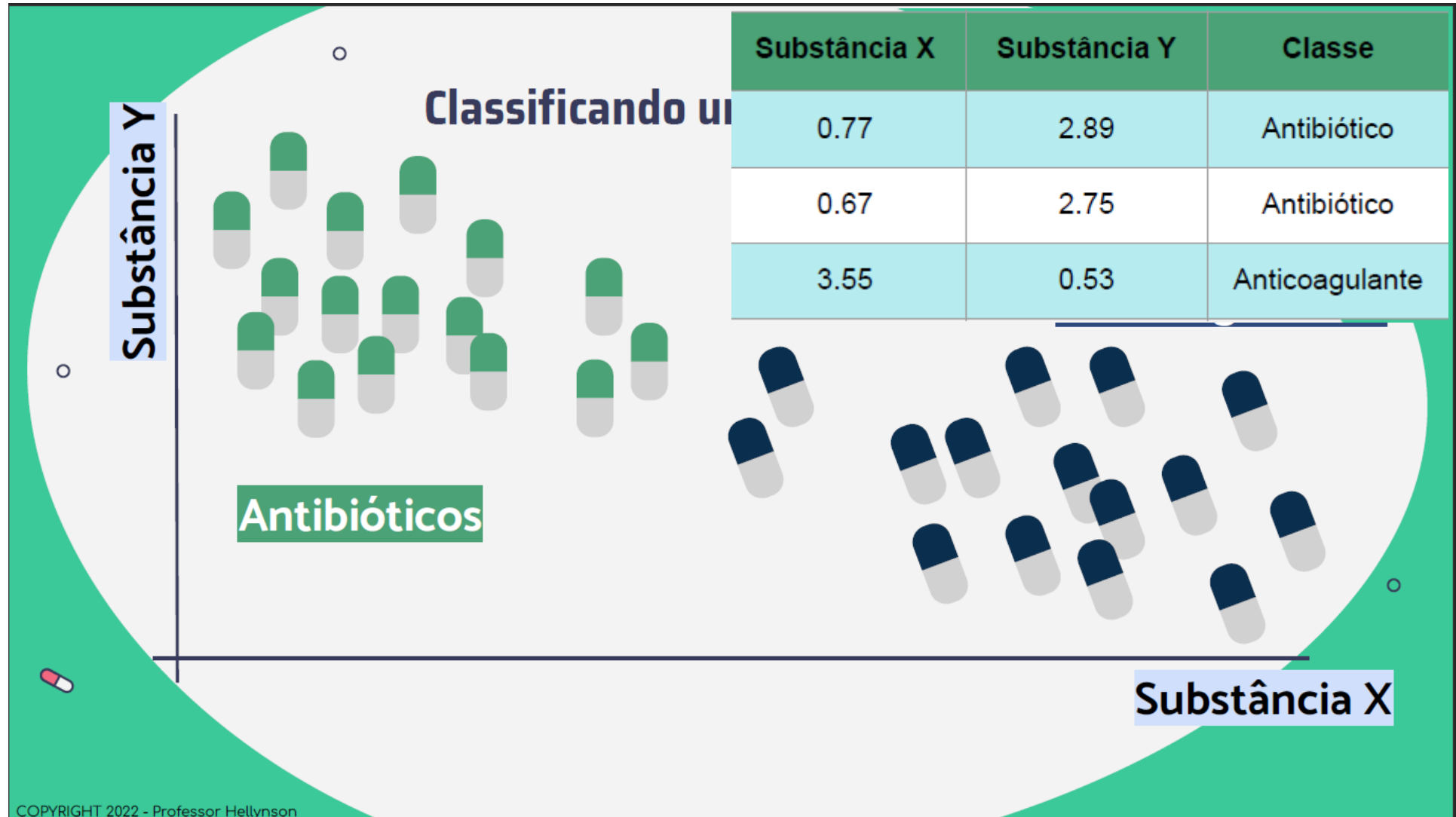
K-NN – K-Nearest Neighbors



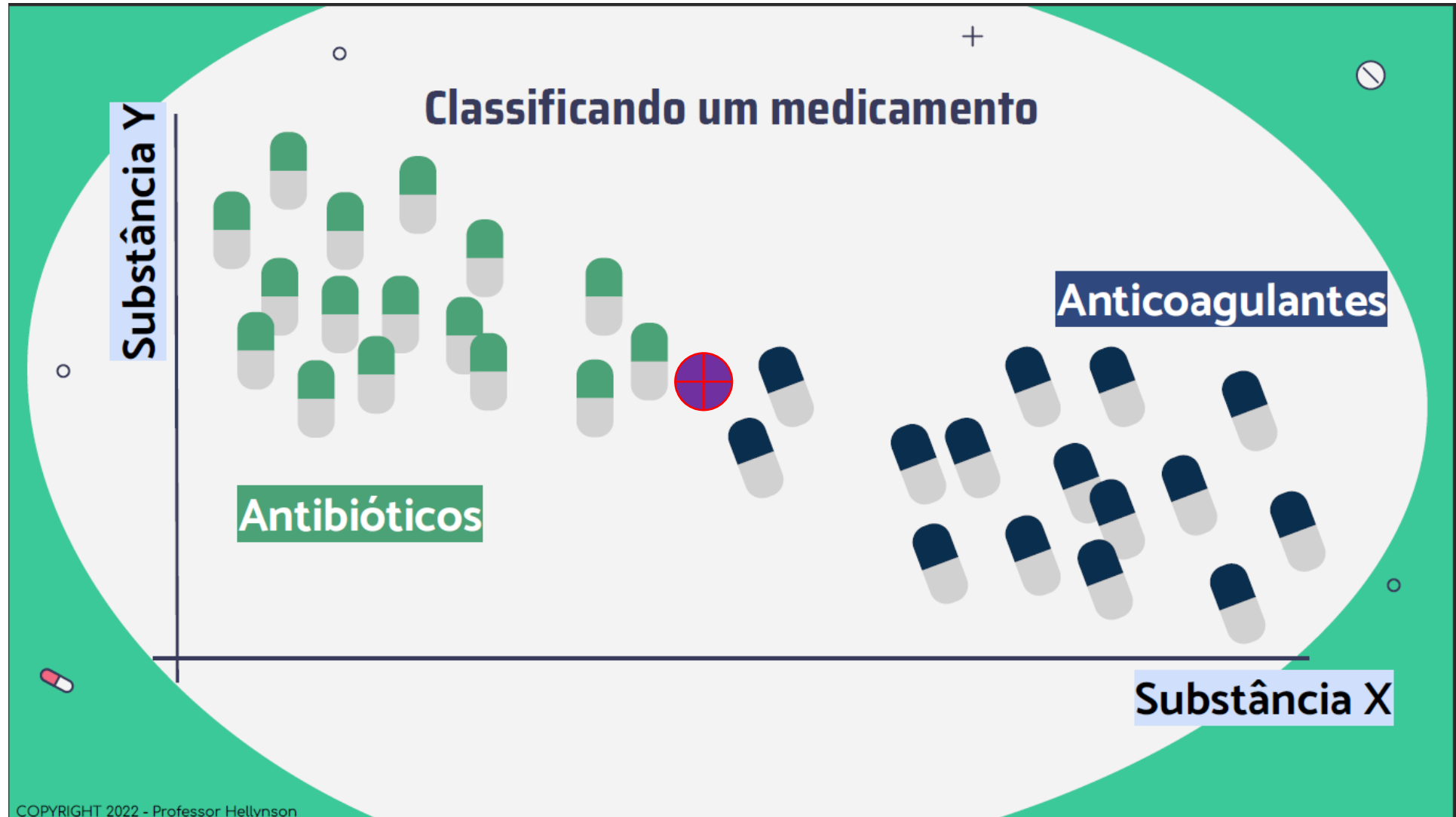
K-NN – K-Nearest Neighbors



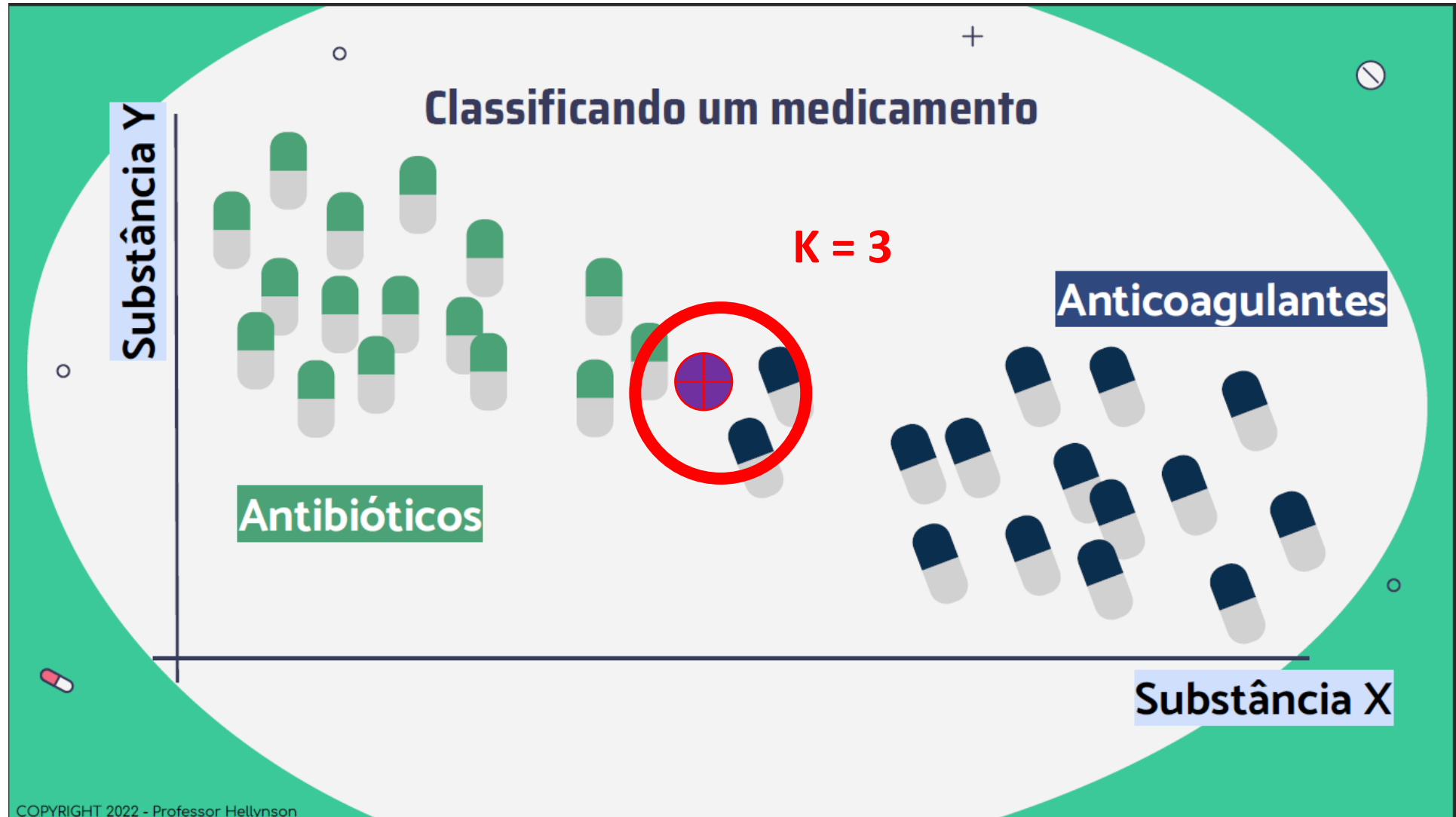
K-NN – K-Nearest Neighbors



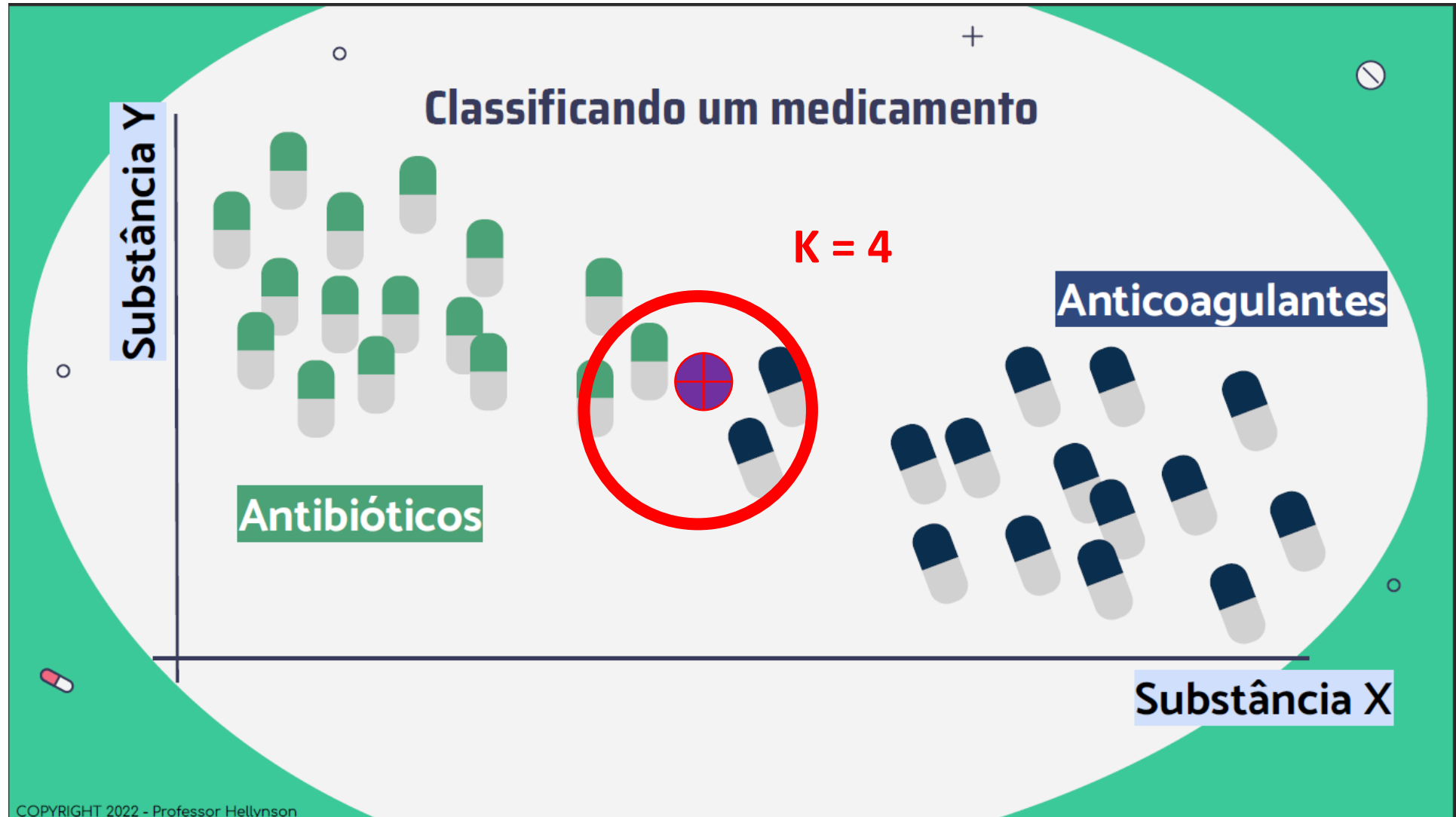
K-NN – K-Nearest Neighbors



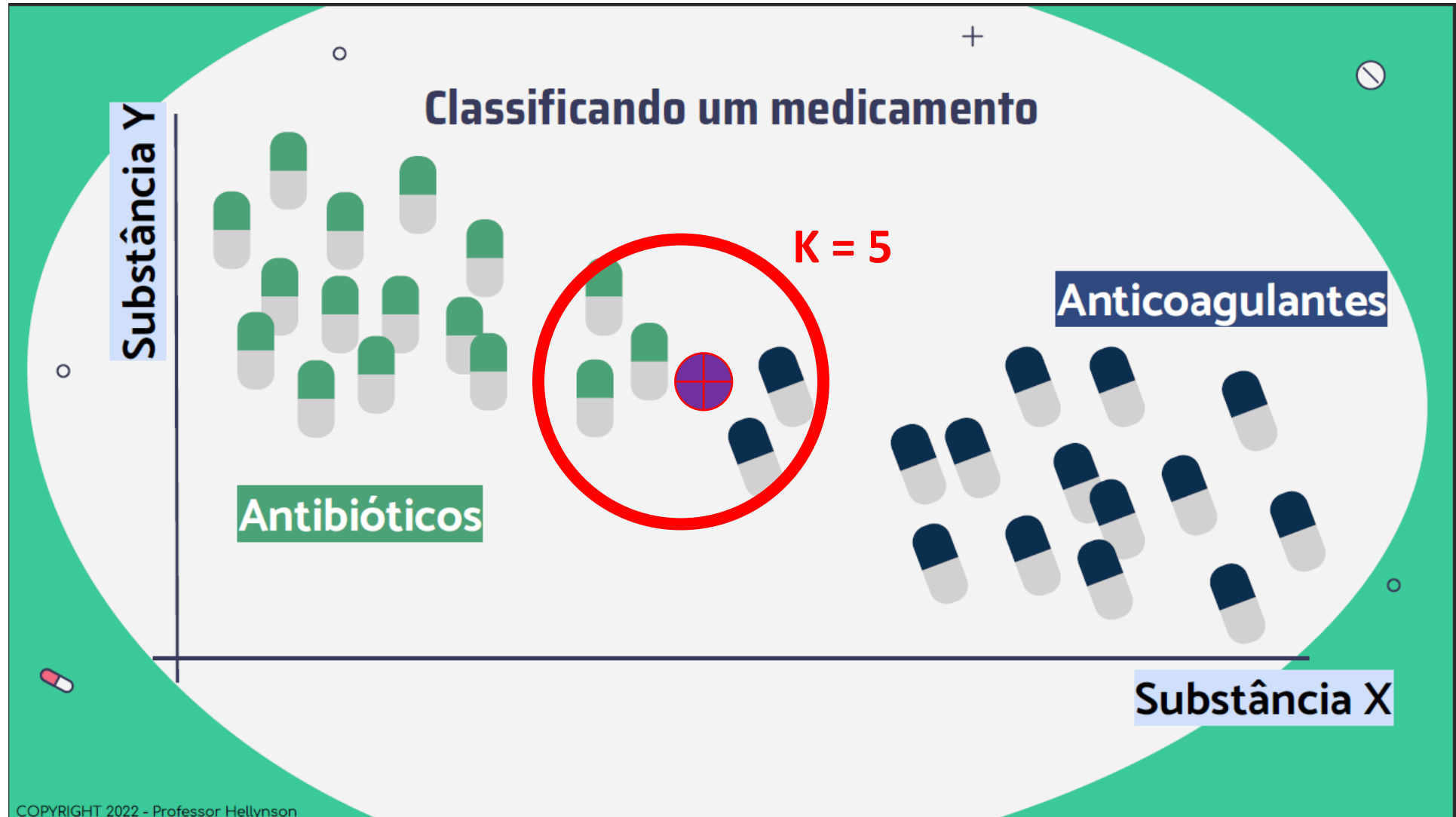
K-NN – K-Nearest Neighbors



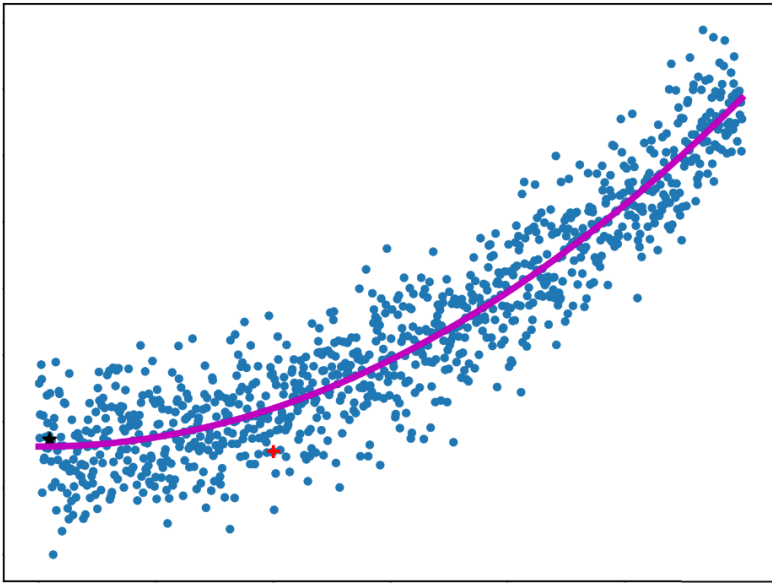
K-NN – K-Nearest Neighbors



K-NN – K-Nearest Neighbors



O que são problemas de Regressão?



Fonte: <https://www.engineersgarage.com/machine-learning-algorithms-classification/>

1

Referem-se a tarefas de aprendizado de máquina em que o objetivo é prever um valor contínuo com base em um conjunto de atributos ou variáveis independentes.

2

O objetivo principal em problemas de regressão é encontrar uma função que mapeie as variáveis independentes para uma resposta numérica contínua, a fim de realizar previsões precisas e úteis com base nos dados disponíveis

3

Alguns exemplos de aplicação de Modelos de Regressão são prever o preço de imóveis, realizar uma previsão de volume de vendas, prever demanda de produtos, prever valor de ações no mercado financeiro, etc.

EXEMPLO

	Atributo A	Atributo B	Atributo C	Valor
Amostra 01	0,5	1,0	1,2	1,5
Amostra 02	0,5	0,9	1,0	2,4
Amostra 03	1,0	2,5	5,0	0,56
Amostra 04	1,0	0,8	2,5	3,15

Principais algoritmos de Regressão



Regressão Linear

modelo de Regressão simples e popular que se ajusta a um conjunto de dados com uma linha reta. Ele é útil quando há uma relação linear entre as variáveis dependentes e independentes.



Regressão Polinomial

modelo de Regressão que se ajusta a um conjunto de dados com uma curva polinomial. Ele é útil quando há uma relação não linear entre as variáveis dependentes e independentes.



Support Vector Machine (SVM)

Modelo de Regressão que traça uma divisão entre os dados, considerando a inclinação que dá a maior distância entre os grupos, mas ao mesmo tempo, a menor distância entre os pontos da reta



Gradient Boosting

traça uma divisão entre eles, considerando a inclinação que dá a maior distância entre os grupos, mas ao mesmo tempo, a menor distância entre os pontos da reta

Como funciona a Regressão Linear?



Regressão Linear

É uma técnica que consiste em representar um conjunto de dados por meio de uma reta

Na matemática, aprendemos que a equação da reta é:

$$y = A + Bx$$

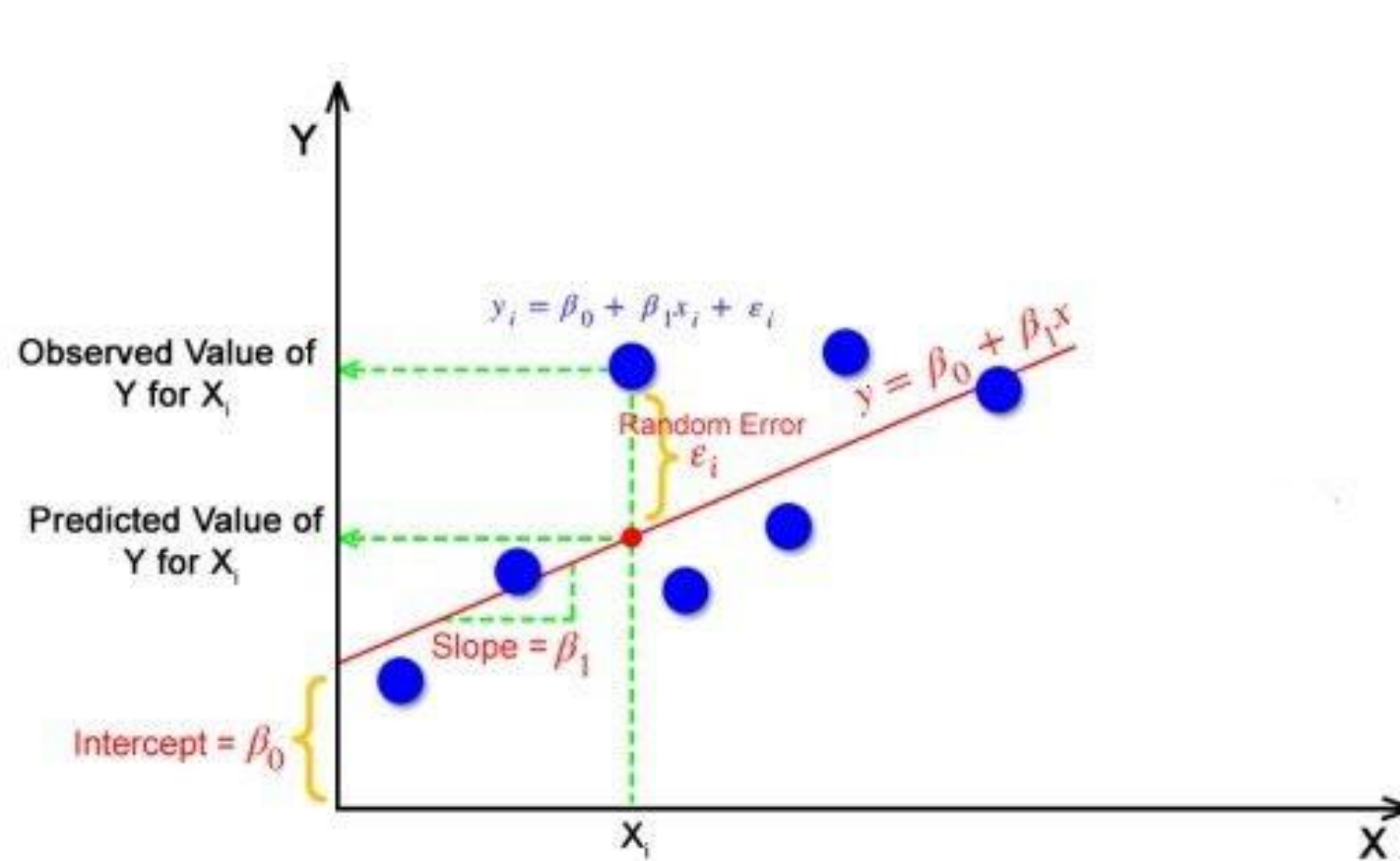
Em Machine Learning, aprendemos que uma Regressão Linear é:

$$y_{\text{predito}} = \beta_0 + \beta_1 x$$

β_0 e β_1 são parâmetros que determinam o peso e bias da rede. Para cada x temos um y_{predito} aproximado predito.

Modelos de regressão linear são intuitivos, fáceis de interpretar e se ajustam aos dados razoavelmente bem em muitos problemas

Regressão Linear

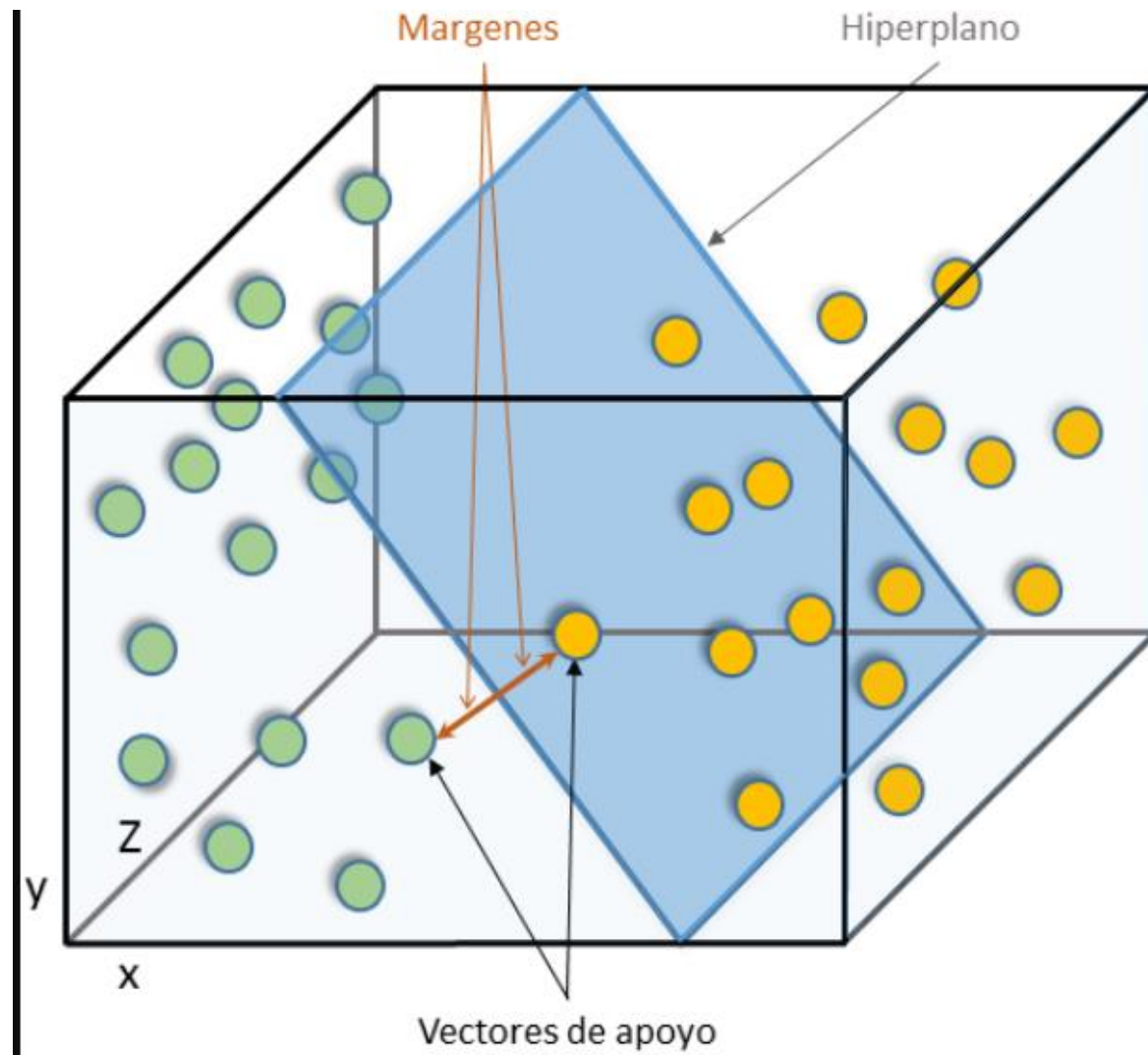


Fonte: <https://www.analyticsvidhya.com/blog/2021/10/everything-you-need-to-know-about-linear-regression/>

- Essa ideia se estende para mais de um parâmetro independente, mas nesse caso não estamos associando a uma reta e sim a um plano ou hiperplano:

$$Y_{\text{predito}} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Exemplo



Em outras palavras, modelos de regressão linear são intuitivos, fáceis de interpretar e se ajustam aos dados razoavelmente bem em muitos problemas.

Definição do Problema

Desenvolver um sistema de machine learning capaz de prever o valor de um imóvel na Califórnia. Para isso, vamos usar o dataset de informações do U.S. Census Service sobre habitação no estado da Califórnia, disponível no site Kaggle (<https://www.kaggle.com/datasets/camnugent/california-housing-prices>) .

Esse dataset contém informações derivadas do censo de 1990 sobre localização, número de quartos, média salarial, valor das casas, entre outras informações, estando concentradas em distritos da Califórnia.

California Housing Prices



Carregando o Dataset

- A biblioteca sklearn é uma biblioteca de aprendizado de máquina em Python amplamente utilizada para tarefas de análise de dados e modelagem estatística. Seu nome completo é scikit-learn e ela oferece uma ampla gama de algoritmos e ferramentas para tarefas de aprendizado de máquina, como classificação, regressão, agrupamento, redução de dimensionalidade, seleção de recursos e pré-processamento de dados.
- O sklearn possui alguns datasets já pré carregados, de modo que neste primeiro passo, vamos carregar o dataset California do sklearn.

Exemplo

```
# Inicialização das bibliotecas
%matplotlib inline

# Importando as bibliotecas necessárias
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Importando o dataset Iris
from sklearn.datasets import fetch_california_housing

# Carregar o dataset em um objeto
california = fetch_california_housing()
```

Explorando o Dataset

- A função `fetch_california_housing()` é chamada para carregar o conjunto de dados na variável `california`. Em seguida, você pode acessar os dados deste dataset chamando o método `keys()`. Note que o retorno desse método são `'data'`, `'target'`, `'frame'`, `'target_names'`, `'feature_names'` e `'DESCR'`. Este último vai descrever as características do dataset como um todo. Explore os outros e verifique o que cada um significa.

Keys()

```
#para conhecer o que foi importado do dataset  
california.keys()  
  
dict_keys(['data', 'target', 'frame', 'target_names', 'feature_names', 'DESCR'])
```

Informações dos atributos:

MedInc - Renda mediana no grupo de blocos

HouseAge - Idade média da casa no grupo de quarteirões

AveRooms - Número médio de quartos por domicílio

AveBedrms - Número médio de dormitórios por domicílio

Population - População do grupo de blocos

AveOccup - Número médio de membros da família

Latitude - Latitude do grupo de blocos

Longitude - Longitude do grupo de blocos



ML_Regressão.ipynb

Desafio 5

- Melhorando o Modelo de Aprendizado de Regressão Linear para Previsão de Preços de Imóveis.
 - Explorar diferentes implementações, caso necessário, pode utilizar outros algoritmos de regressão. Os alunos devem discutir suas alterações, justificando suas escolhas e avaliar o impacto das melhorias nas métricas de desempenho do modelo.

**Copyright © 2023 Prof. Arnaldo Jr/Yan
Coelho/Airton Toyofuku**

Todos direitos reservados. Reprodução ou divulgação total ou parcial deste documento é expressamente proibido sem o consentimento formal, por escrito, do Professor (autor).