



Overture



- [Resumo](#)
- [Cronograma](#)
- [Inspiração](#)
- [Apresentação do Aprendizado de Máquina](#)
- [Jargão do Aprendizado de Máquina](#)
- [Aprendizado de Máquina Supervisionado](#)
- [Conjuntos de Treinamento e de Teste](#)
- [Classificação Versus Regressão](#)
- [Aprendizado de Máquina Não Supervisionado](#)
- [Aprendizado de Máquina Semissupervisionado](#)
- [Projetos de Pesquisa](#)
- [Lista de Exercícios 01](#)
- [*Playlist*](#)
- [*Reading List*](#)
- [Autor](#)



Aprendizado de máquina representa uma das áreas mais ativas de pesquisa e desenvolvimento desta geração. Seu amplo leque de aplicações e seu rápido desenvolvimento atrai diversos interessados, desenvolvedores e usuários de suas técnicas. Os métodos de aprendizado de máquina aliados ao crescimento exponencial da geração de dados científicos possibilitam aos cientistas de todas as áreas criar modelos computacionais, bem como analisar quantidades de dados que eram impensáveis somente uma década atrás. A presente disciplina tem como objetivo aumentar o número de cientistas que usam aprendizado de máquina na sua pesquisa. Apresentarei durante o semestre as principais técnicas, destacando sempre as aplicações científicas. Não há requisitos para esta disciplina, além da curiosidade científica e a mente aberta para o novo. Boas vindas para o novo...

Que a luz da ciência acabe com as trevas do negacionismo.

Walter F. de Azevedo, Jr.



Março/2024

- 06 Apresentação da disciplina
- 13 Delineamento dos projetos
- 20 Modelos de aprendizado supervisionado I
- 27 Modelos de aprendizado supervisionado II

Abril/2024

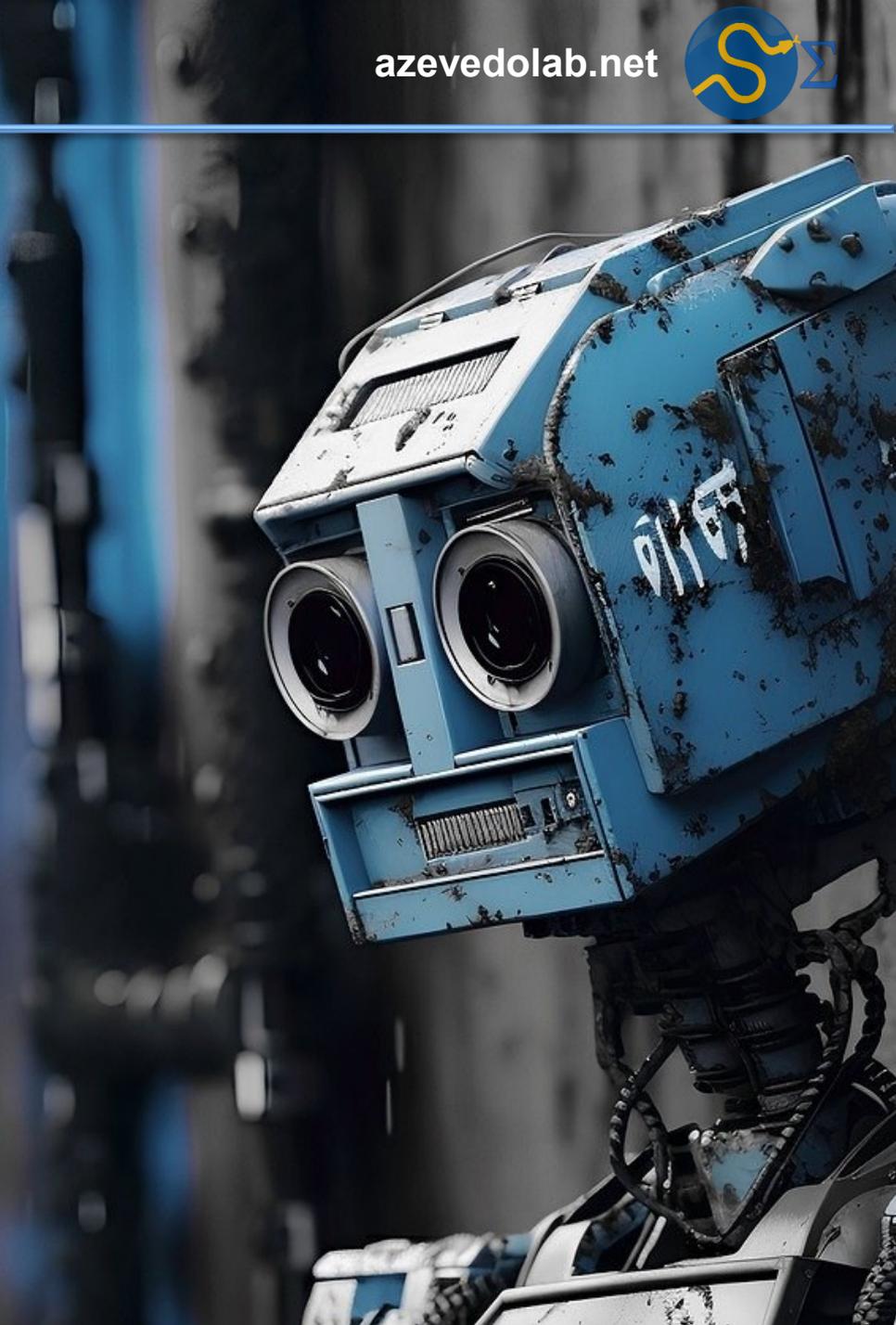
- 03 Modelos de aprendizado supervisionado III
- 10 Modelos de aprendizado supervisionado IV
- 17 Modelos de aprendizado supervisionado V
- 24 Modelos de aprendizado supervisionado VI

Maiο/2024

- 08 Aprendizado de máquina para classificação I
- 15 Aprendizado de máquina para classificação II
- 22 Estudo de redes neurais I
- 29 Estudo de redes neurais II

Junho/2024

- 05 Discussão de projetos de pesquisa
- 12 Discussão de projetos de pesquisa
- 19 Discussão de projetos de pesquisa
- 26 Entrega do Projetos**

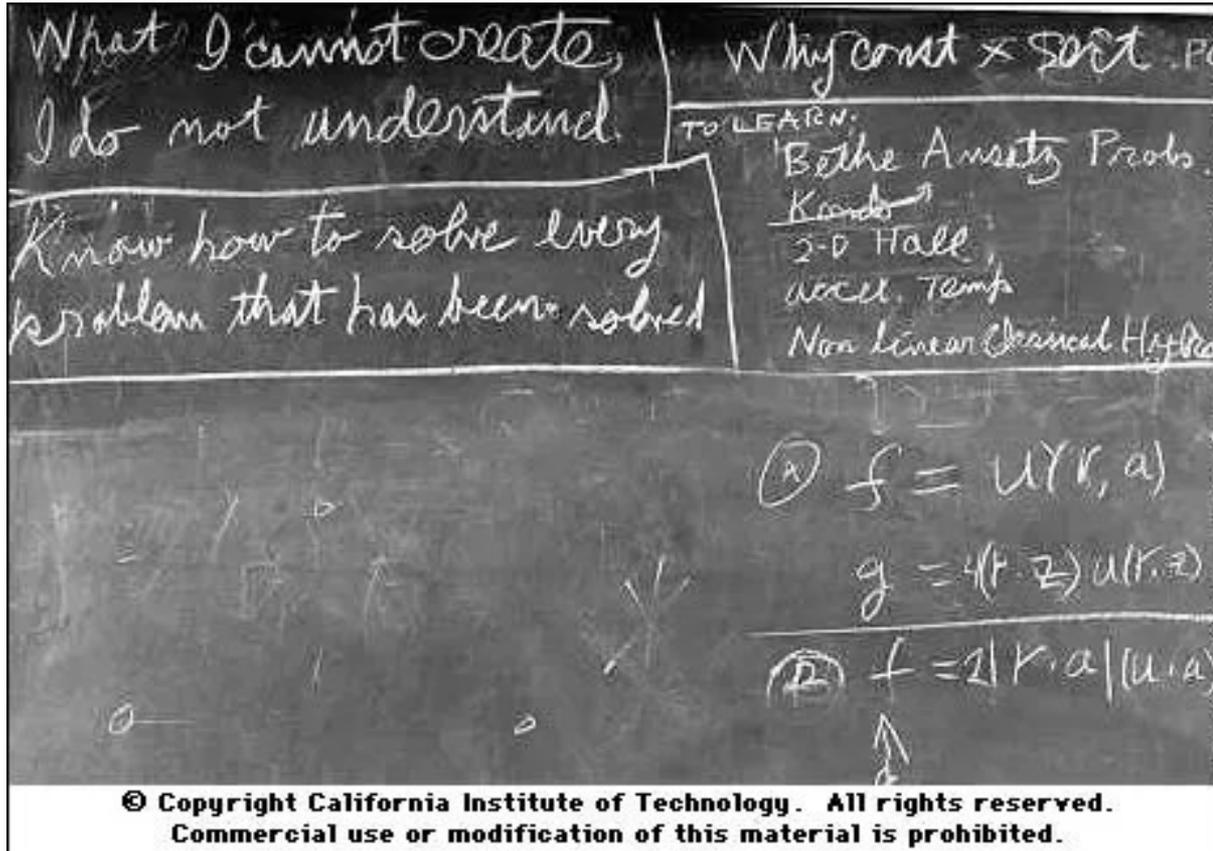




Richard Feynman

A poet once said, “The whole universe is in a glass of wine.” We will probably never know in what sense he meant that, for poets do not write to be understood. But it is true that if we look at a glass of wine closely enough we see the entire universe. There are the things of physics: the twisting liquid which evaporates depending on the wind and weather, the reflections in the glass, and our imagination adds the atoms. The glass is a distillation of the earth’s rocks, and in its composition we see the secrets of the universe’s age, and the evolution of stars. What strange array of chemicals are in the wine? How did they come to be? There are the ferments, the enzymes, the substrates, and the products. There in wine is found the great generalization: all life is fermentation. Nobody can discover the chemistry of wine without discovering, as did Louis Pasteur, the cause of much disease. How vivid is the claret, pressing its existence into the consciousness that watches it! If our small minds, for some convenience, divide this glass of wine, this universe, into parts—physics, biology, geology, astronomy, psychology, and so on—remember that nature does not know it! So let us put it all back together, not forgetting ultimately what it is for. Let it give us one more final pleasure: drink it and forget it all!

Fonte: [Feynman, Richard P.; Leighton, Robert B.; Sands, Matthew \(2015-09-28T22:58:59.000\). The Feynman Lectures on Physics, Vol. I: The New Millennium Edition: Mainly Mechanics, Radiation, and Heat . Basic Books. Edição do Kindle.](#)



"What I cannot create, I do not understand"
 Richard Feynman



Reconhecimento



Agrupamento



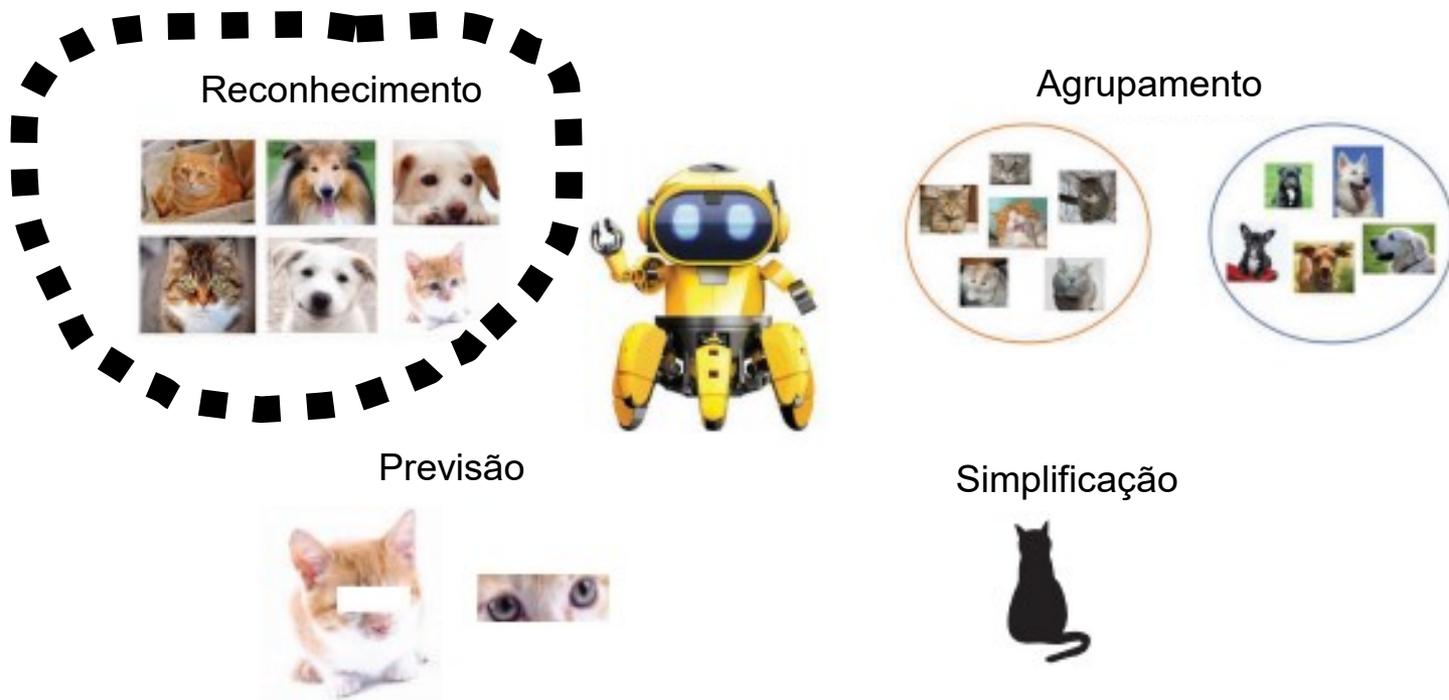
Previsão



Simplificação



A figura acima ilustra alguns processos vinculados ao estudo de aprendizado de máquina. De uma forma geral, podemos dizer que aprendizado de máquina é o processo de ensinamos o computador a tomar decisões baseadas em dados. Exemplos das decisões incluem reconhecer e caracterizar objetos baseados em similaridades ou diferenças, detecção de padrões em conjuntos de dados e distinguir informações relevantes de ruído em um conjunto de dados experimental. Com essa definição informal, iremos explorar alguns aspectos gerais da aplicação do aprendizado de máquina e depois retornaremos para definições mais formais.



Reconhecer uma instância de um certo tipo. Esse processo pode ser a identificação de diferentes tipos de animais a partir de imagens, ou reconhecer os moradores de um prédio para a partir da leitura de uma imagem da face.



Reconhecimento



Agrupamento



Previsão



Simplificação



Prever alguma propriedade ou informação baseada em outra dado. Por exemplo, podemos tentar prever o comportamento futuro baseado no comportamento anterior. Uma aplicação comum é a análise do comportamento da bolsa de valores em período anterior para prever o desempenho hoje. Outro uso é reconstruir um pedaço que falta de uma imagem usando dados de imagens similares.



Reconhecimento



Agrupamento



Previsão



Simplificação



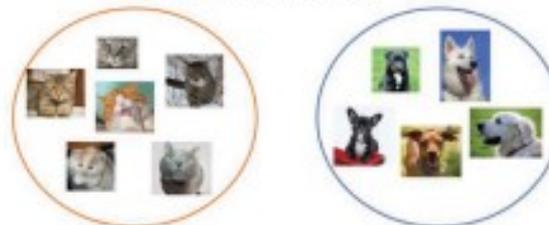
Agrupar objetos que são similares que podem também ser usados para identificar aqueles que são diferentes dos grupos (**outliers** na linguagem do aprendizado de máquina). Por exemplo, podemos tentar identificar quantos tipos diferentes de galáxias estão numa fotografia do céu diferenciando aqueles que podem ser artefatos da câmera. Esse tipo de aplicação chama-se agrupamento (*clustering*).



Reconhecimento



Agrupamento



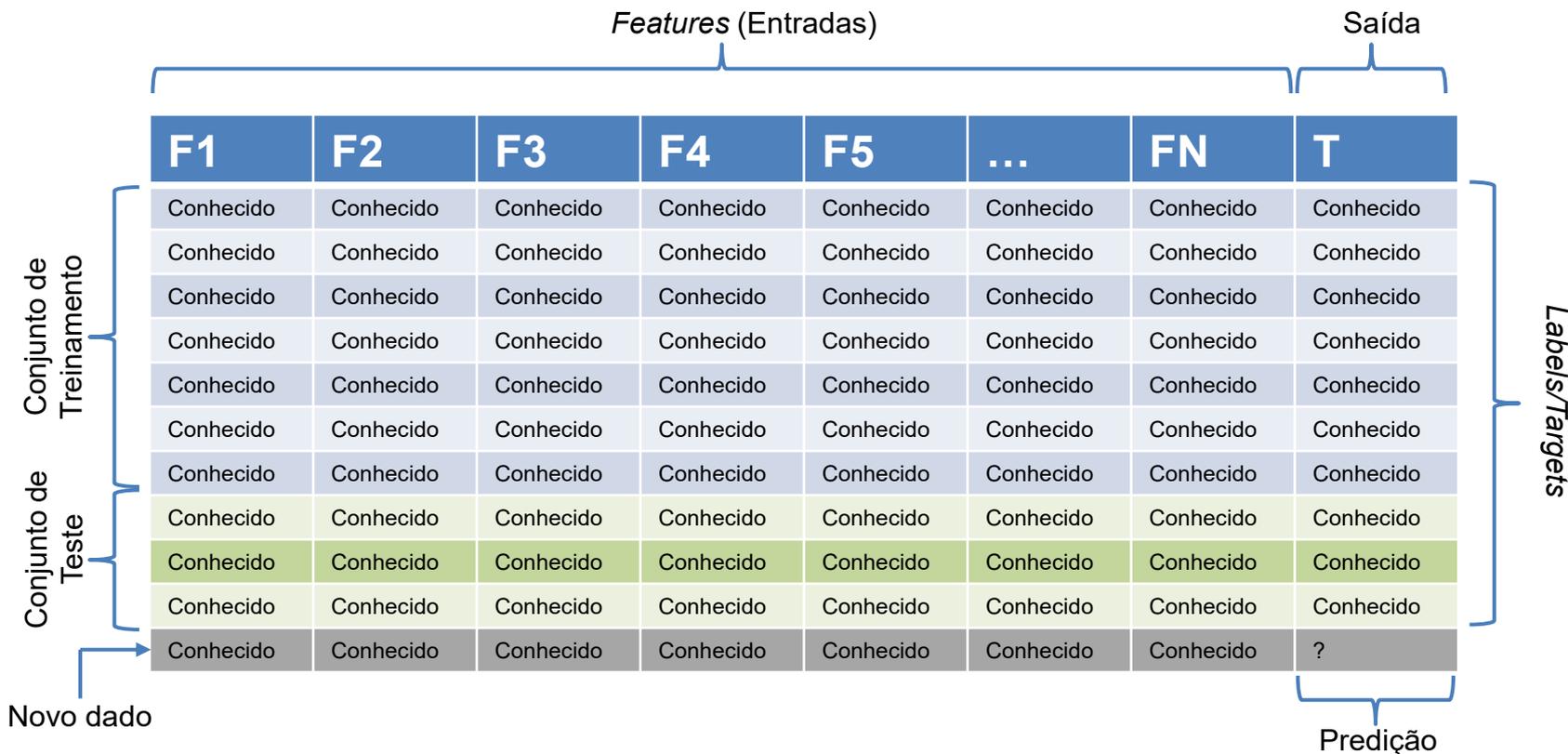
Previsão



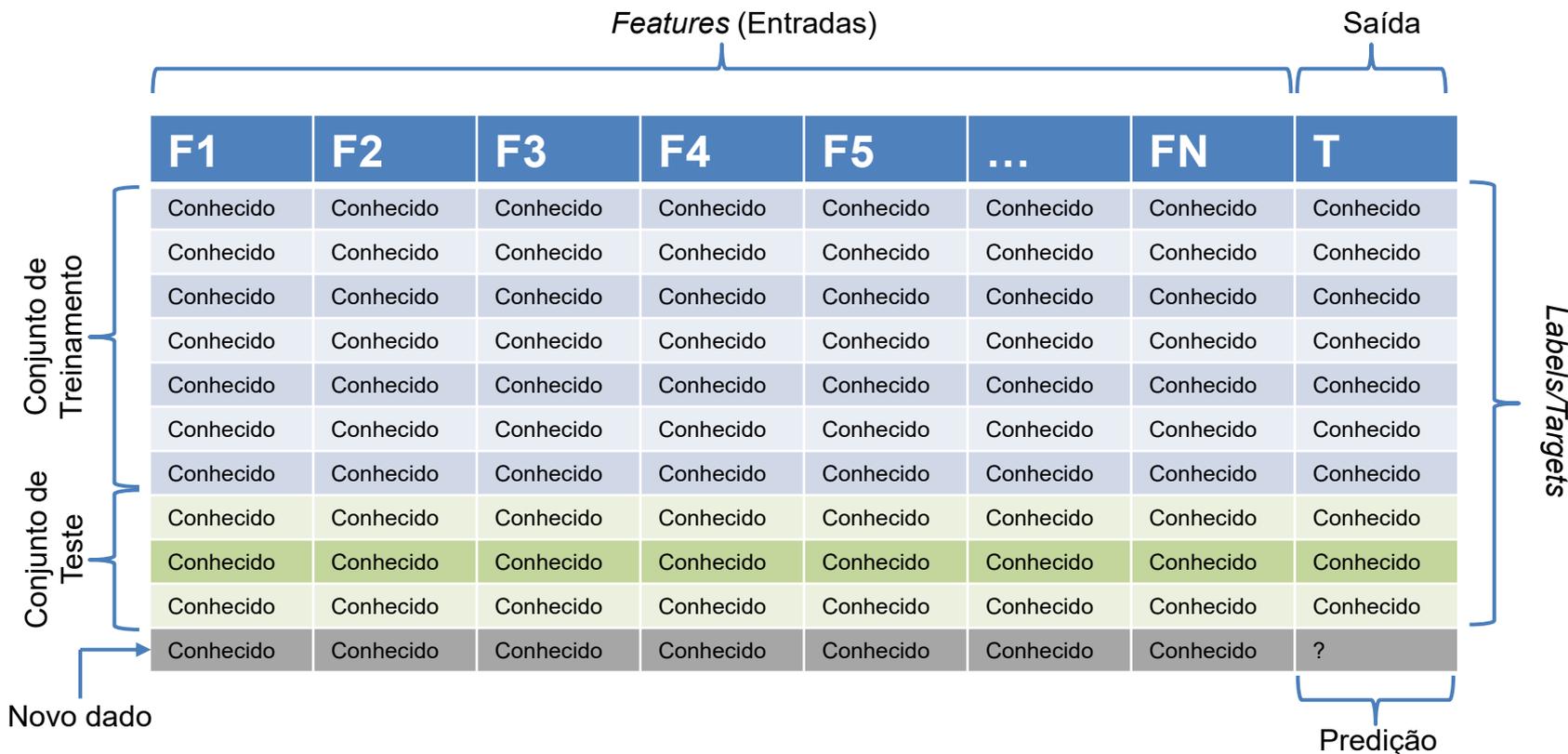
Simplificação



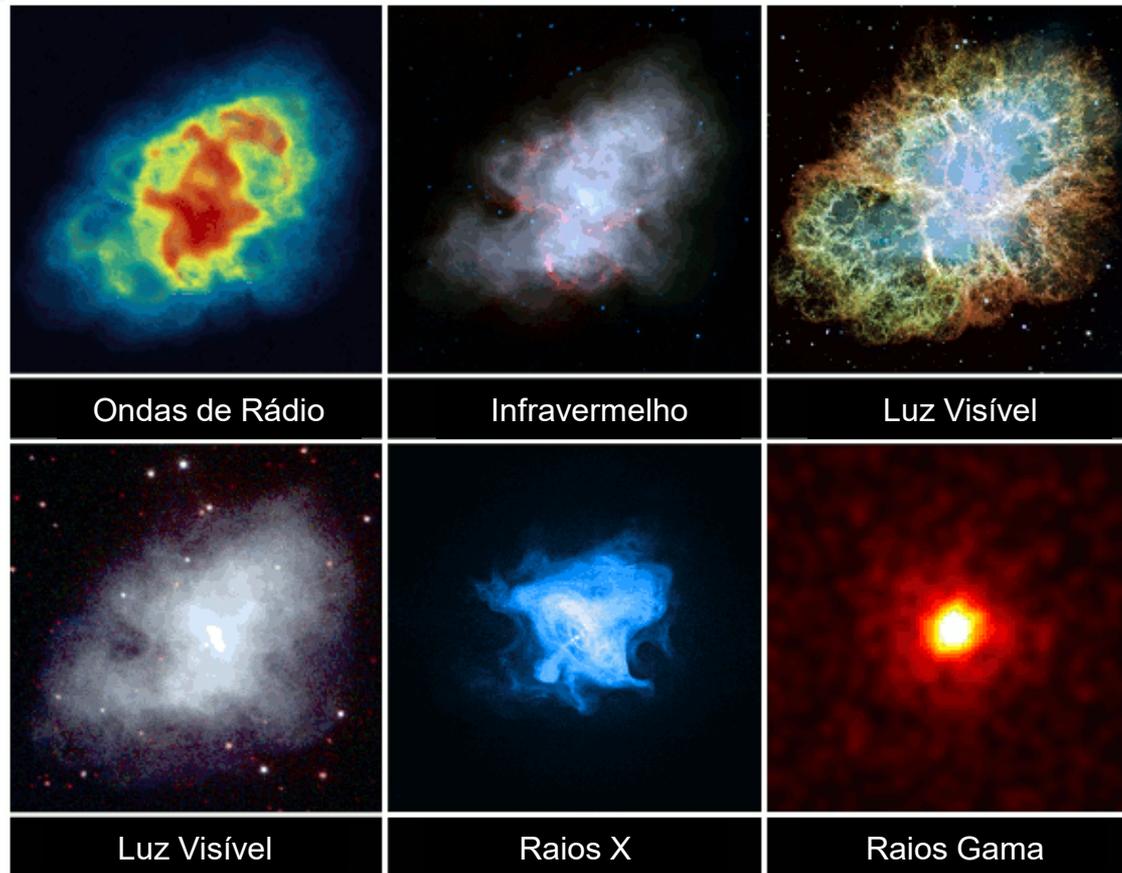
Simplificar a informação contida num conjunto de dados complicado para condensá-lo usando somente aspectos relevantes. Por exemplo, podemos perguntar qual a maneira mais simples de desenhar um gato de forma que ainda podemos reconhecê-lo como gato. Esse processo tem a vantagem adicional de reduzir o volume dos dados que precisamos manipular e ajuda a entendermos as propriedades fundamentais de uma dada categoria. Chamamos esses processos de técnicas de redução de dimensões.



Ao analisarmos um sistema, podemos realizar medições de alguma propriedade. Na tabela acima temos uma representação de um conjunto de dados. Elementos de um conjunto de dados são chamados de **instâncias** (*instances*) ou exemplos ou amostras (*samples*). Para cada instância temos algumas propriedades que são conhecidas (características que podem ser medidas, observadas num experimento, ou simuladas). Essa propriedades são chamadas de **features** (ou entradas). Em português alguns traduzem **features** como **características**.

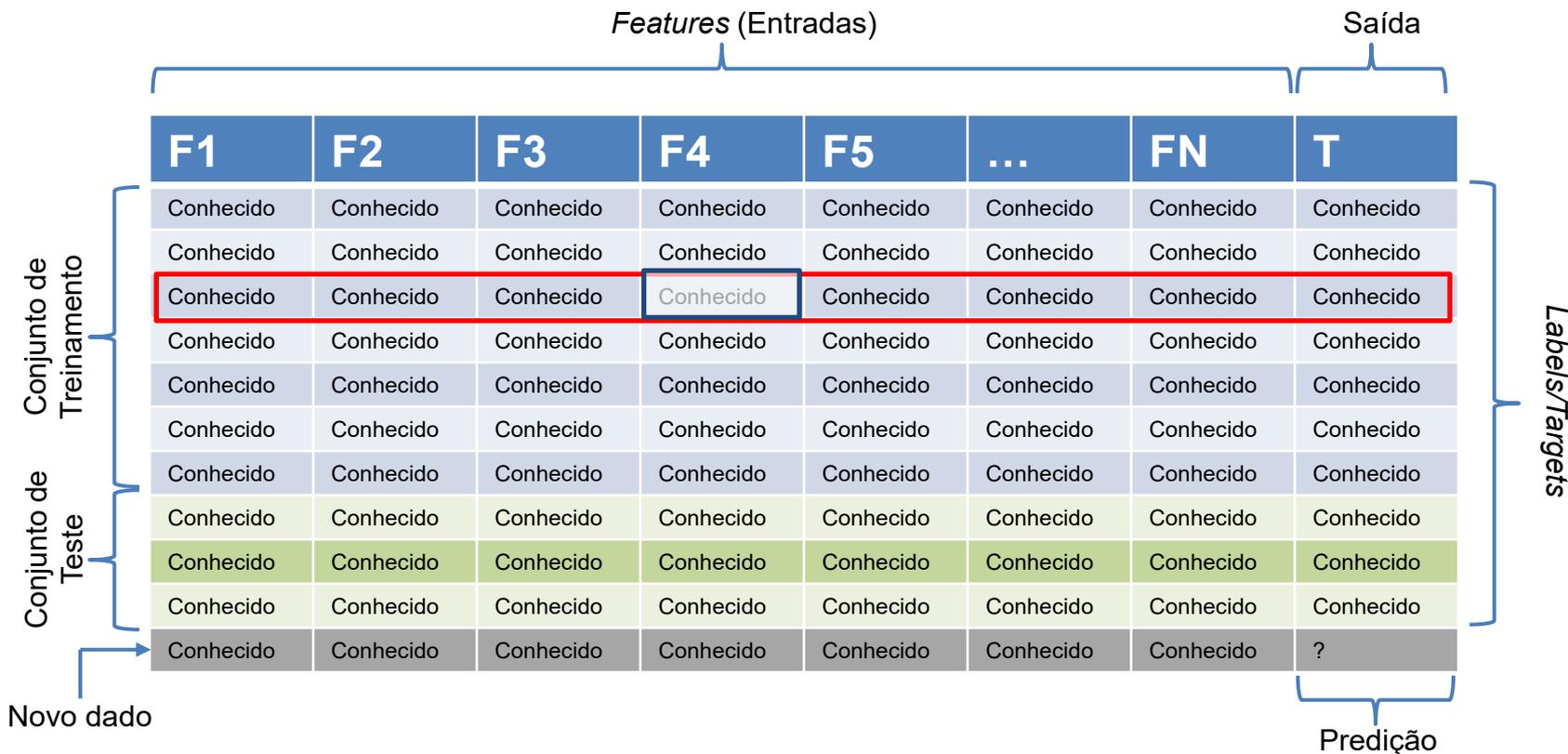


Como exemplo vamos considerar o estudo das leis de Newton, onde medimos a distância e o tempo de objetos em queda livre e queremos saber a velocidade final. Nesse sistema, os *features* são dois: distância e tempo. Consideremos agora o estudo de galáxias, onde realizamos observações em diferentes regiões do espectro de radiação eletromagnética (raios gama, raios X, ultravioleta, visível, infravermelho e ondas de rádio). Assim temos um ponto para cada medida e teremos seis *features*. Destacamos que os *features* não precisam ser numéricos.

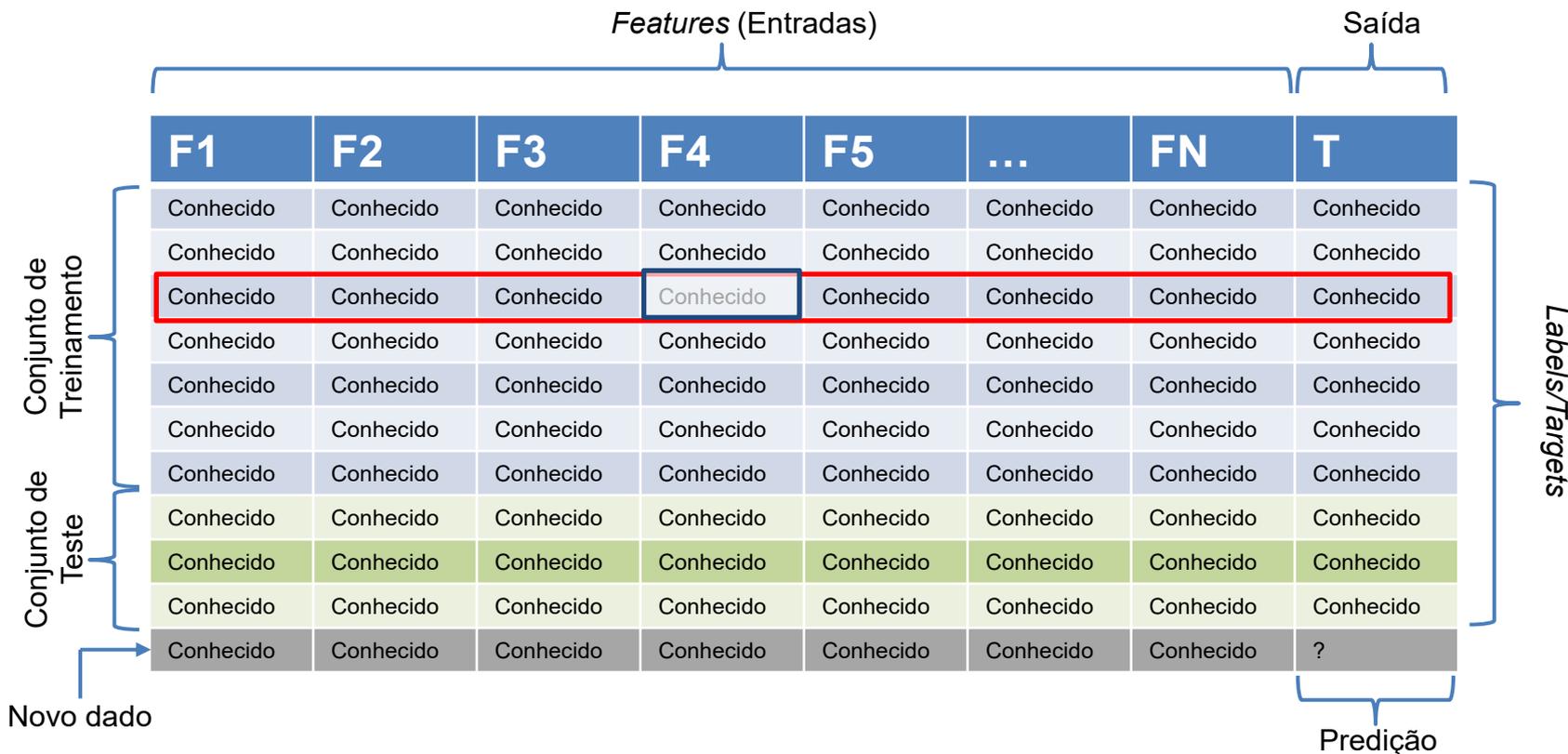


A imagem acima tem a nebulosa do caranguejo vista em múltiplas faixas do espectro de radiação eletromagnética, dos raios gama às ondas de rádio.

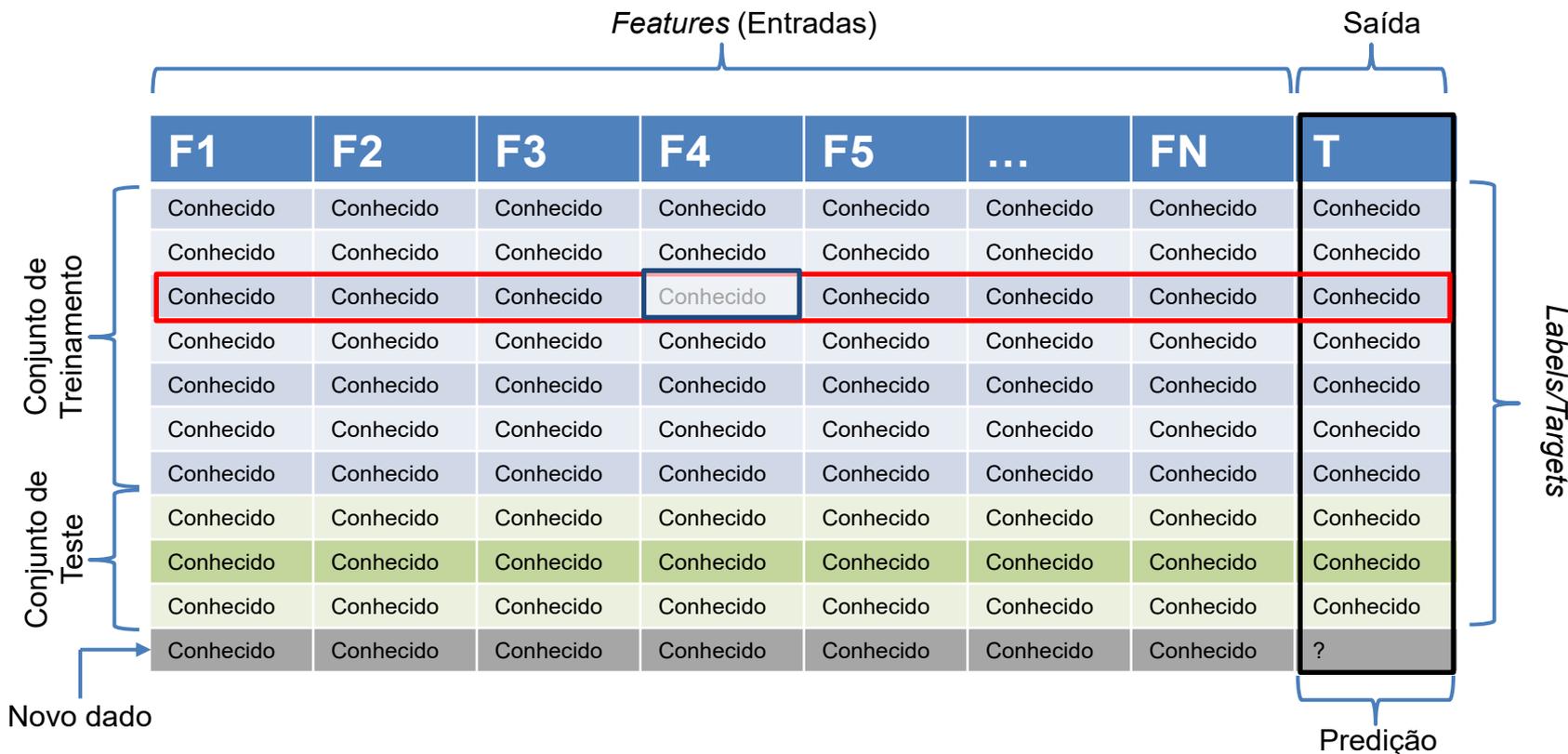
Credits: NRAO/AUI and M. Bietenholz; NRAO/AUI and J.M. Uson, T.J. Cornwell (radio); NASA/JPL-Caltech/R. Gehrz / University of Minnesota (infrared); NASA, ESA, J. Hester and A. Loll / Arizona State University (visible); NASA/Swift/E. Hoversten, PSU (ultraviolet); NASA/CXC/SAO/F.Seward et al.(X-rays); NASA/DOE/Fermi LAT/R. Buehler (gamma rays).



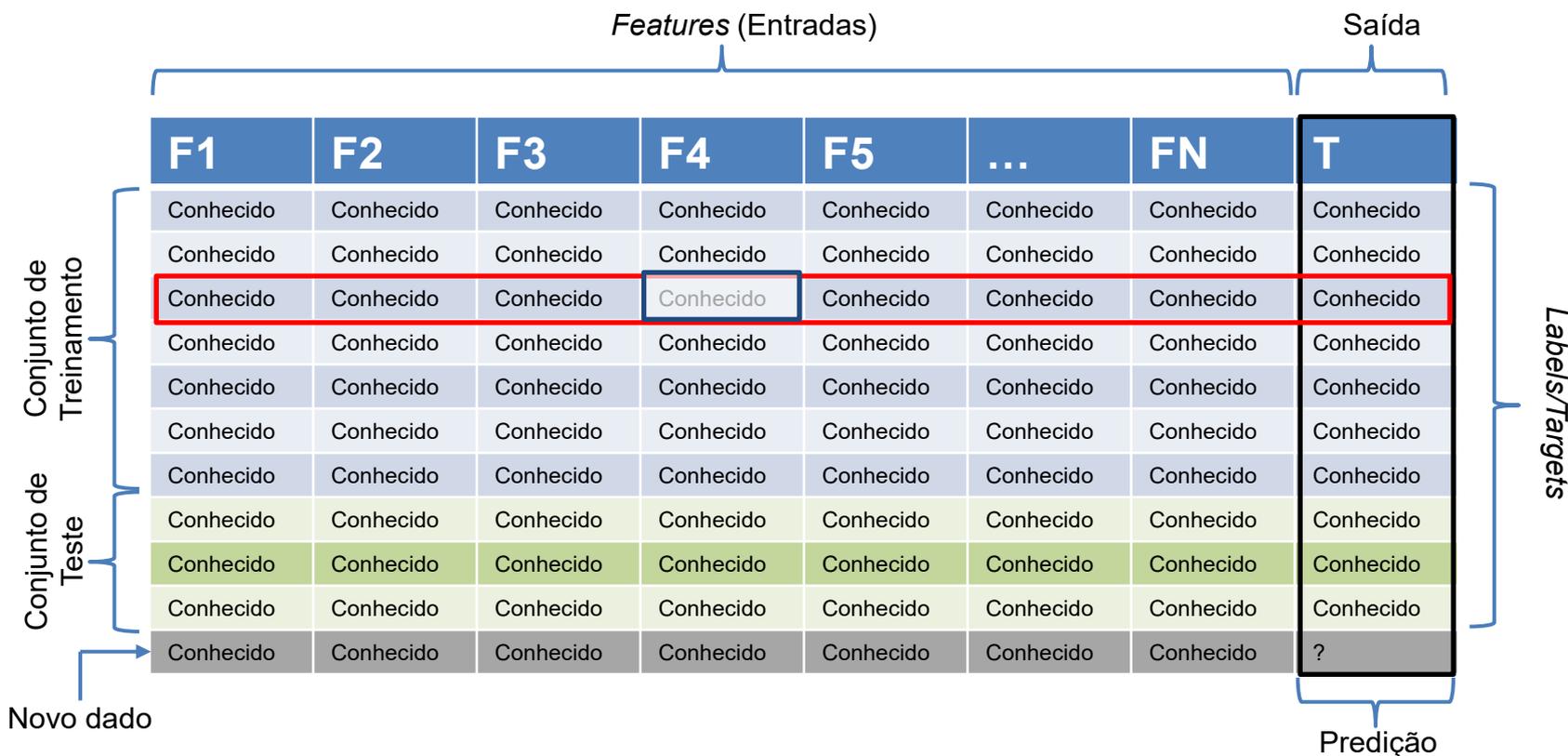
É comum organizarmos nossos conjuntos de dados em tabelas como a destacada acima. Na tabela acima, uma linha descreve uma instância (exemplo do retângulo **vermelho**) e cada coluna contém o valor medido de uma *feature* associado com aquela instância (*feature* F4 na terceira linha e quarta coluna em destaque). Por esta razão, usualmente o tamanho total do seu conjunto de dados será fornecido pelo número de instâncias (M) vezes o número de features (N).



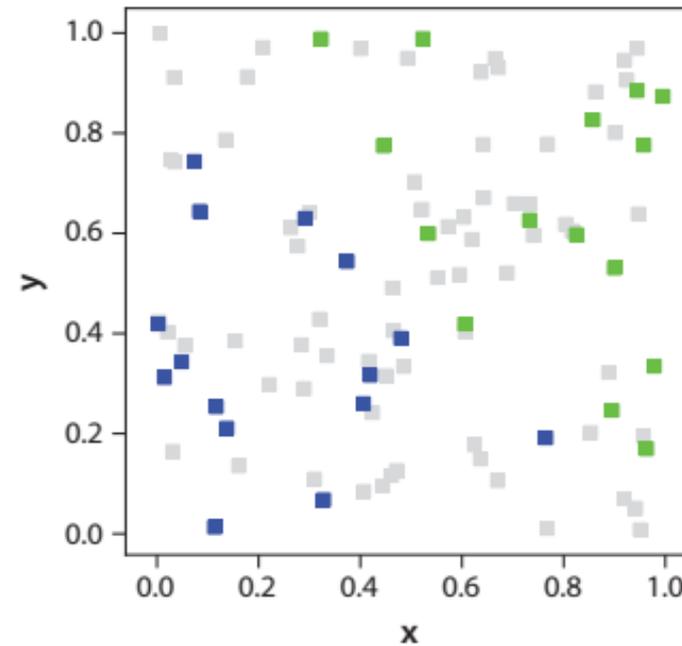
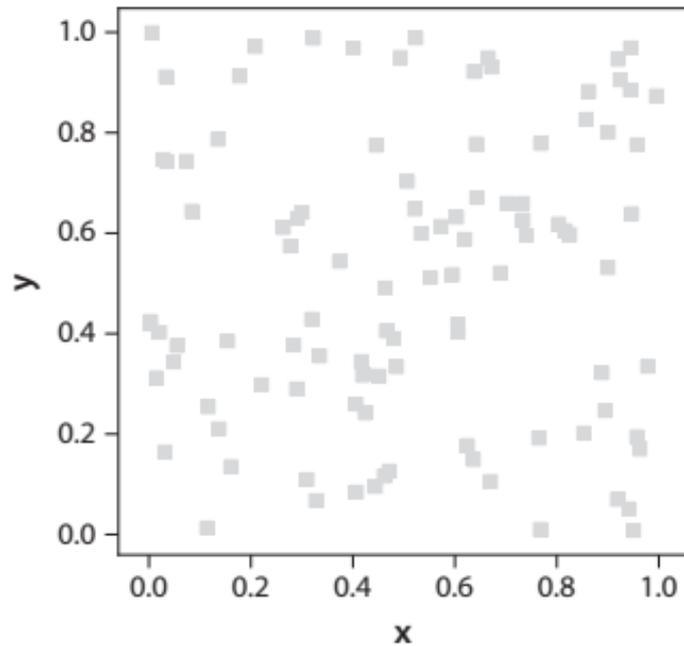
A representação da tabela $M \times N$ acima é facilmente transportada para uma notação matricial, onde os *features* e instâncias podem ser vetores (*arrays*). Podemos representar o conjunto de dados por uma matriz ($M \times N$). Operações envolvendo matrizes são centrais no estudo de aprendizado de máquina. Ao descrevermos os algoritmos de aprendizado de máquina faremos uso frequente de matrizes.



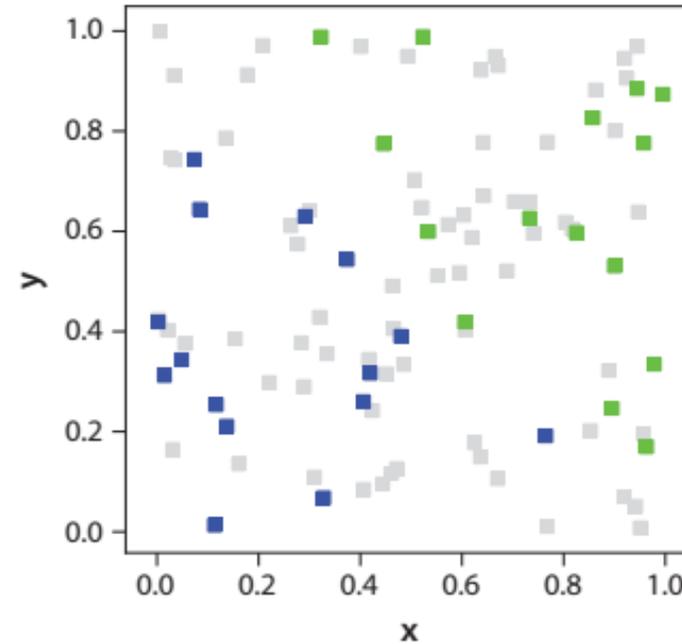
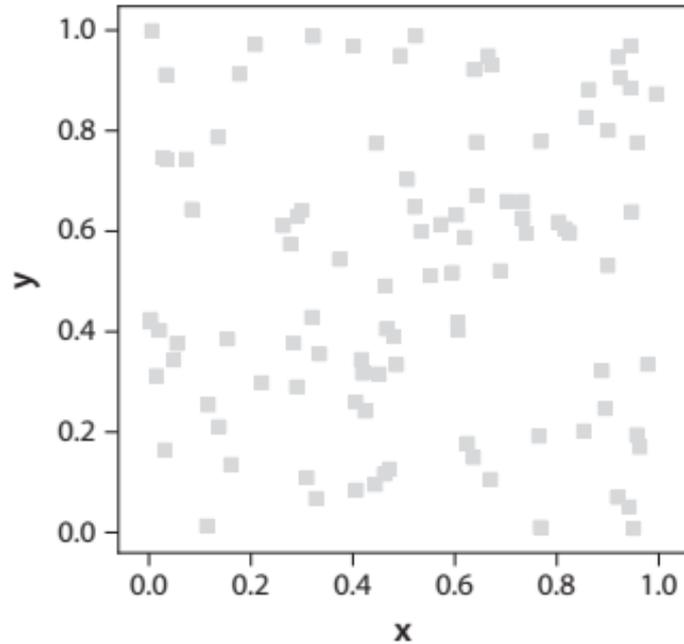
Na tabela acima a última coluna representa o alvo (*target*) ou rótulo (*label*) é o que queremos prever a partir dos *features*. No exemplo da queda livre a velocidade final é o rótulo (alvo) que podemos determinar a partir dos *features* (distância e tempo). Usualmente, a tarefa de aprendizado de máquina tem como objetivo construir uma relação entre as entradas (*features*) e a saída (alvo). Chamamos essa relação de **modelo de aprendizado de máquina** que nada mais que um objeto matemático que relaciona o espaço dos *features* com o espaço dos alvos. Para geramos o nosso modelo, usamos algoritmos de aprendizado de maquina como (ex. redes neurais).



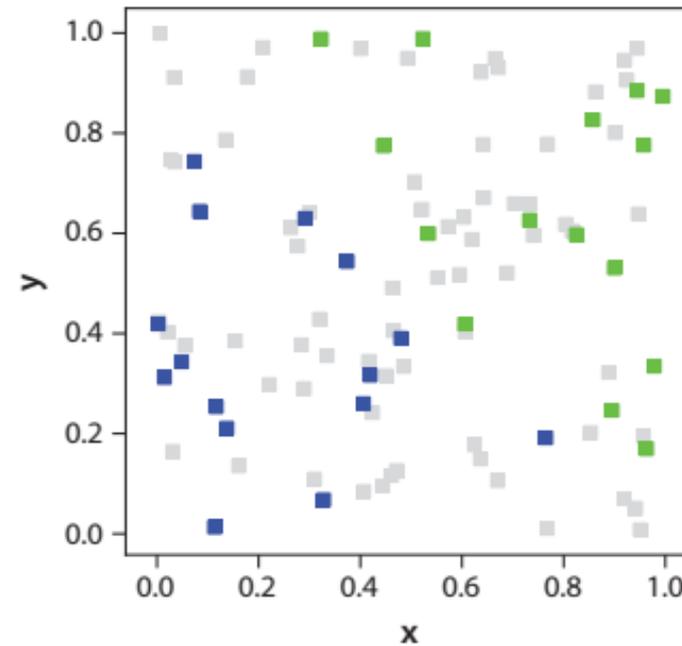
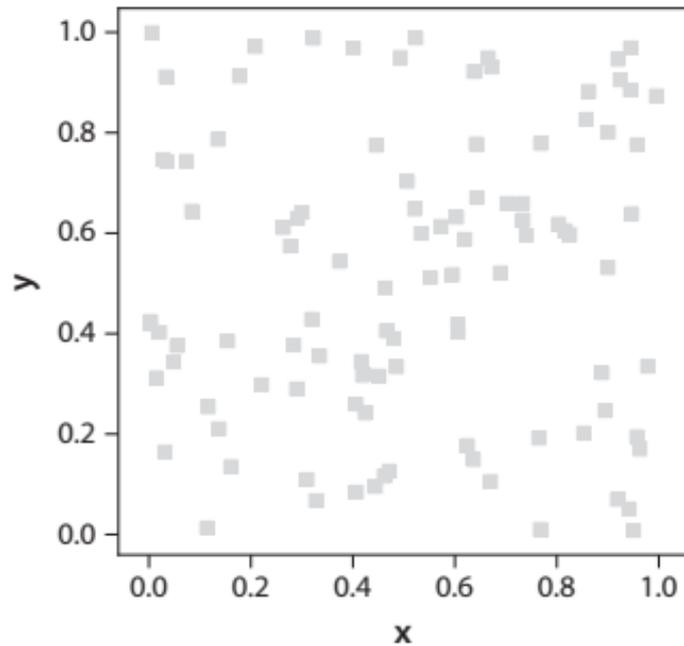
Os dados contidos na tabela acima foram divididos em dois subconjuntos, denominados de **conjunto de treinamento** e **de teste**. As instâncias do conjunto de treinamento são usadas para a construção do modelo de aprendizado de máquina. Os dados do conjunto de teste são empregados para aferir o poder de previsão do modelo de aprendizado de máquina. Iremos explorar mais esses tópicos à frente.



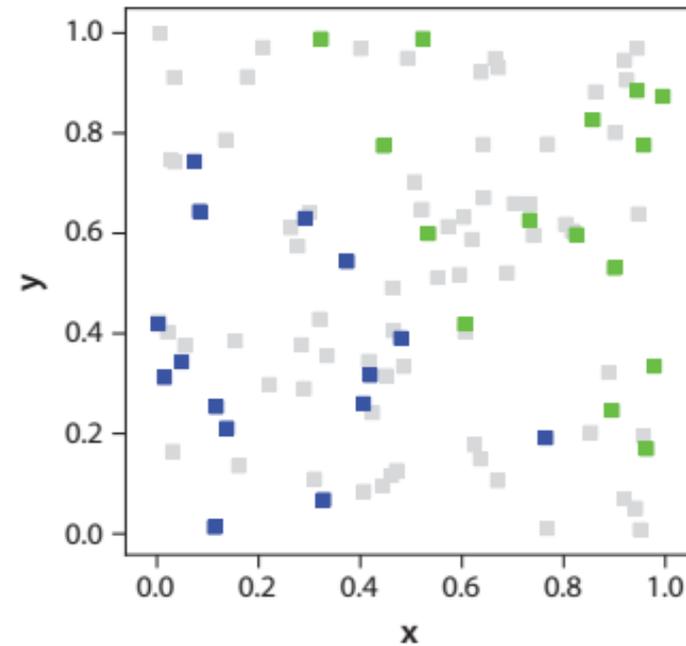
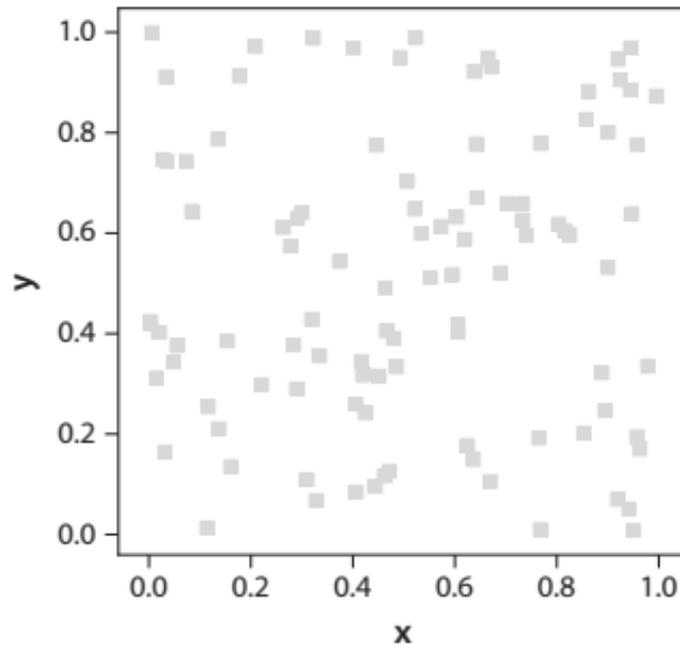
Os gráficos acima representam o mesmo conjunto de dados. À esquerda cada quadrado cinza indica uma instância do conjunto de dados. À direita destacamos parte dos dados com azul e verde que indicam o valor da variável alvo ou rótulo. Cada instância tem coordenadas x e y e a partir deste par de coordenadas (*features*) temos a cor associada à instância.



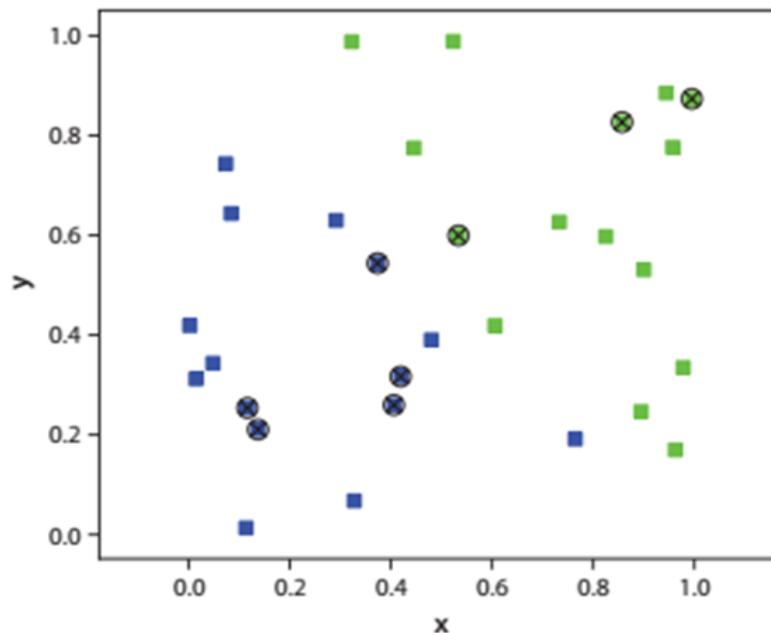
No **aprendizado de máquina supervisionado**, assumimos que há uma coleção de instâncias para as quais a propriedade alvo (rótulo) é conhecida. Assumimos que conhecemos todos os *features*. A coleção de *features* e valores conhecidos dos rótulos é chamada de **conjunto de aprendizado** (*learning set*). No conjunto representado acima, cada ponto no gráfico da esquerda tem uma cor associada a ele (azul ou verde). Nosso objetivo é aprender a cor a partir das coordenadas x e y .



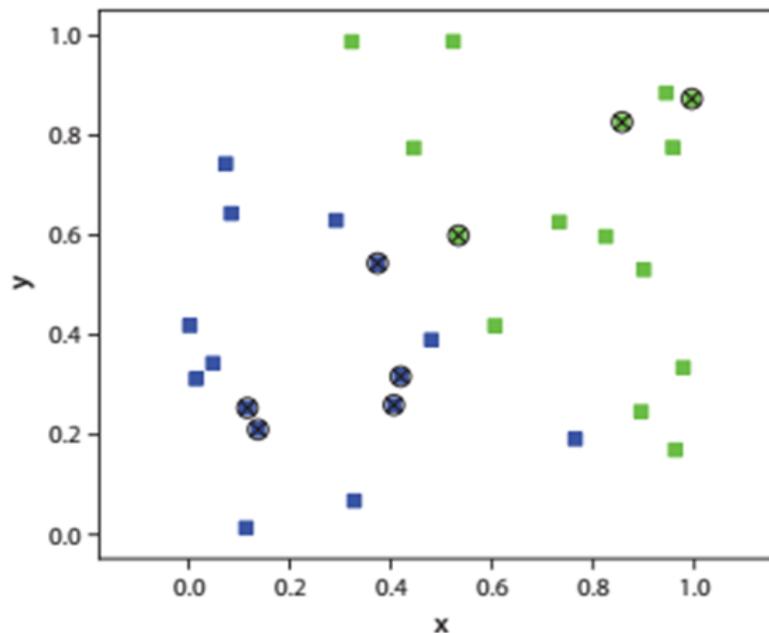
Aprendizado de máquina supervisionado consiste de aprender a partir de exemplos. É necessário que sejamos expostos a algumas instâncias para as quais a cor é conhecida, com o objetivo de desenvolvermos uma intuição, ou no linguajar do aprendizado de máquina, **construir um modelo** da relação entre as coordenadas x e y e a cor. O conjunto de aprendizado para esse processo é mostrado à direita da figura acima, nele temos 30 instâncias (o número de pontos coloridos).



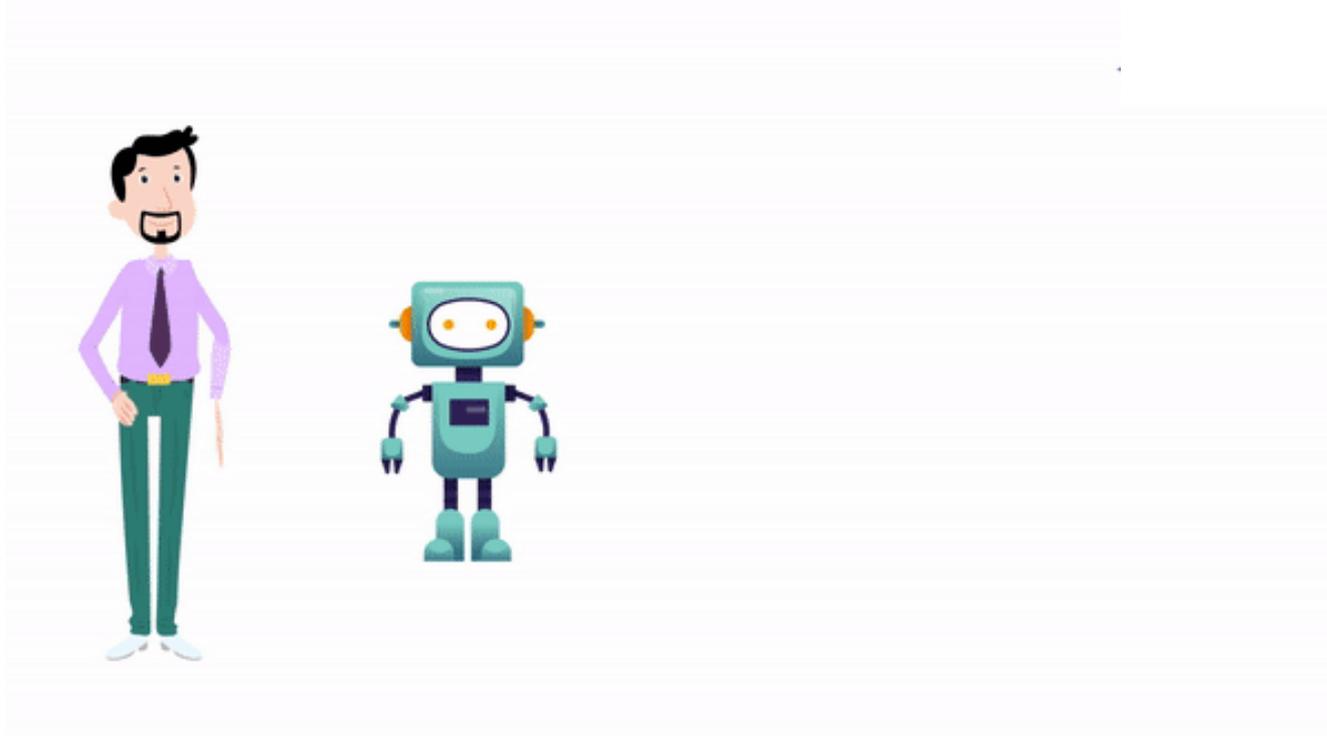
Para resolver o problema, um **algoritmo de aprendizado de máquina** tentará usar as instâncias no conjunto de aprendizado para inferir a regra que conecta as coordenadas à cor. Se tivermos sucesso, quando apresentarmos outro par de coordenadas x e y para os quais só conhecemos os *features*, o modelo construído será capaz de prever a propriedade alvo (a cor) de forma correta. Como regra geral, um modelo de aprendizado de máquina supervisionado é tão bom quanto os dados do seu conjunto de aprendizado. Um conjunto de treinamento com poucos exemplos não permitirá a construção de um modelo preciso.



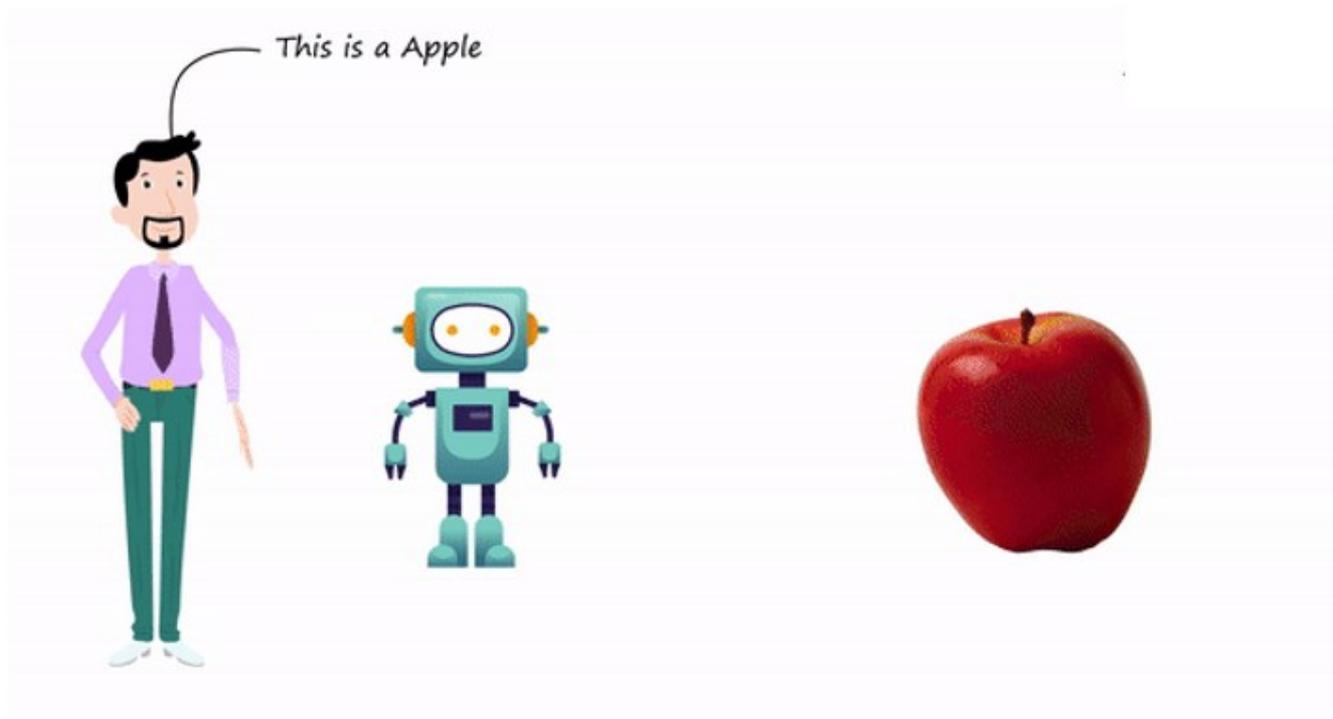
Considerando-se o conjunto de dados do slide anterior, temos agora uma divisão do conjunto de aprendizado (*learning set*) em dois subconjuntos. Na figura acima os pontos indicados com quadrados são as instâncias do **conjunto de treinamento** (*training set*) e os marcados com círculos com cruz representam instâncias do **conjunto de teste** (*test set*). As instâncias do conjunto de treinamento são usadas para construirmos nosso modelo e aprendizado de máquina, ou seja, são os *features* e os valores dos rótulos das instâncias do conjunto de treinamento que irão ditar os parâmetros do modelo de aprendizado de máquina.



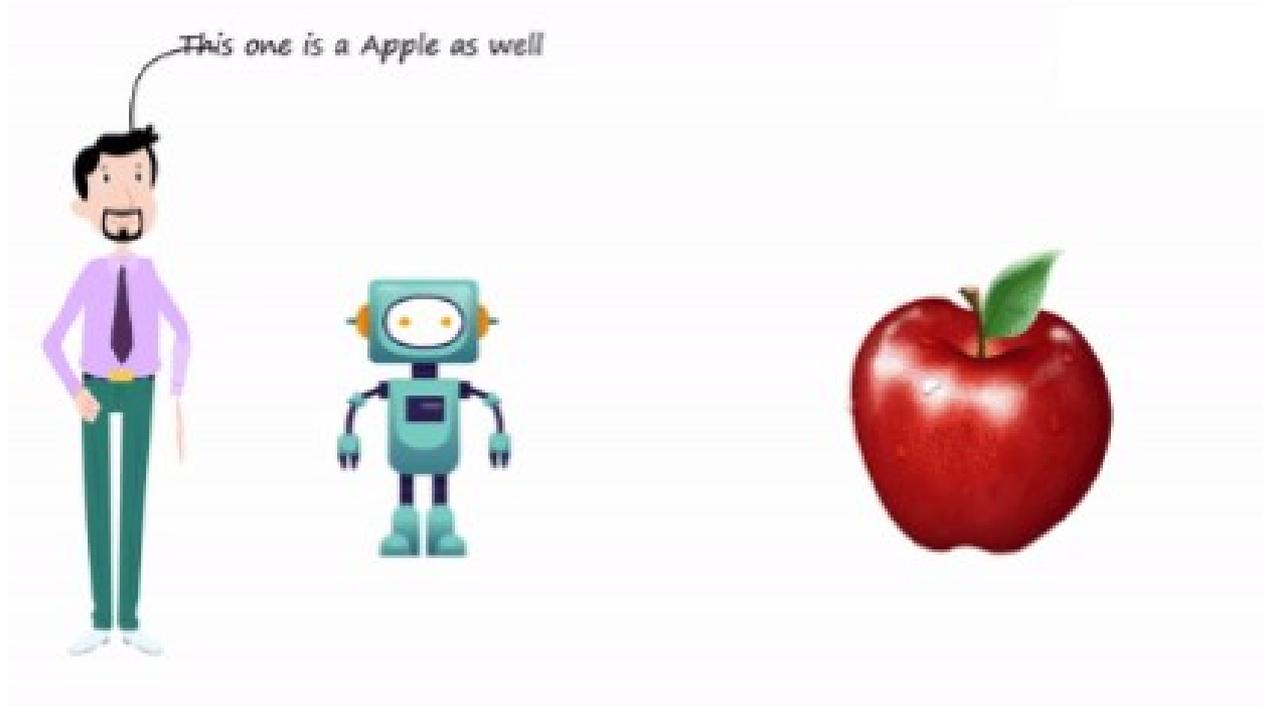
As instâncias do conjunto de teste são para verificar o poder de predição do modelo de aprendizado de máquina. Para um modelo coerente, espera-se uma boa concordância entre o previsto e o valor real. De uma forma geral, podemos avaliar o poder de predição de um modelo por meio da taxa de falhas da predição (erro) ou da taxa de sucessos (score). No jargão do aprendizado de máquina chamamos essas taxas de **métricas**. Há diversas métricas que são usadas para avaliação de modelos de aprendizado de máquina. Teremos aulas específicas, onde iremos discutir essas métricas com exemplos práticos.



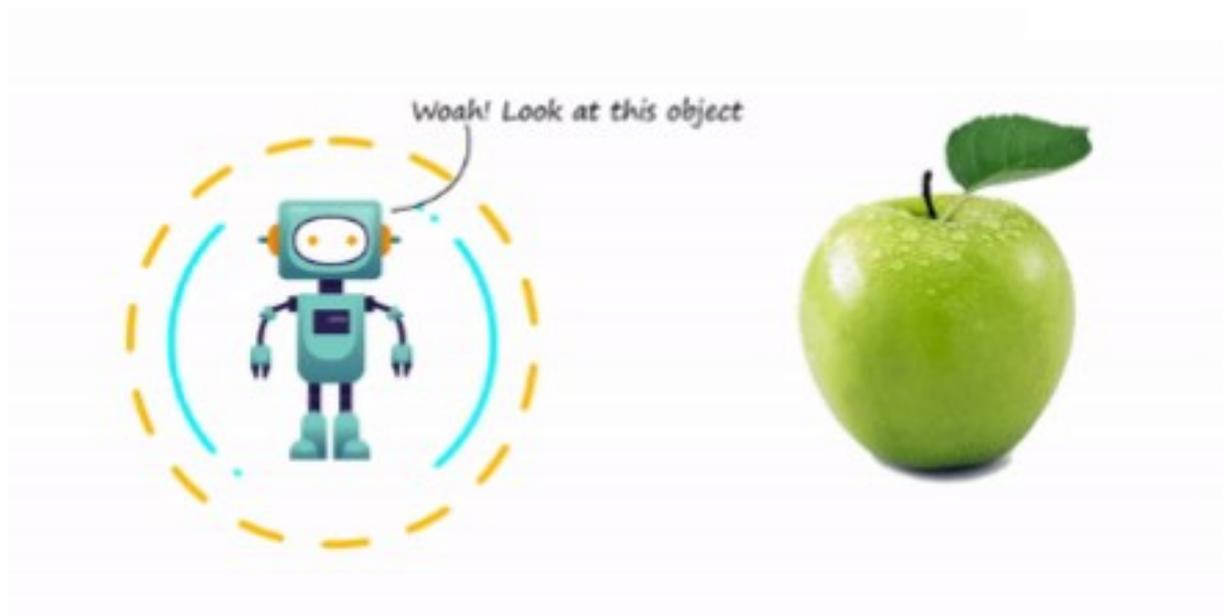
Quando nossa variável alvo (rótulos) não é contínua temos um problema de **classificação**. Vamos supor que queremos identificar se uma fruta é maçã ou não. Assim teremos como resultados possíveis é maçã ou não é maçã. Esse problema é um problema de **classificação**, onde podemos usar imagens que o modelo de aprendizado de máquina lerá e terá como resposta se o objeto é uma maçã ou não. Como temos duas opções somente no processo de classificação, chamamos esse processo de uma **classificação binária**.



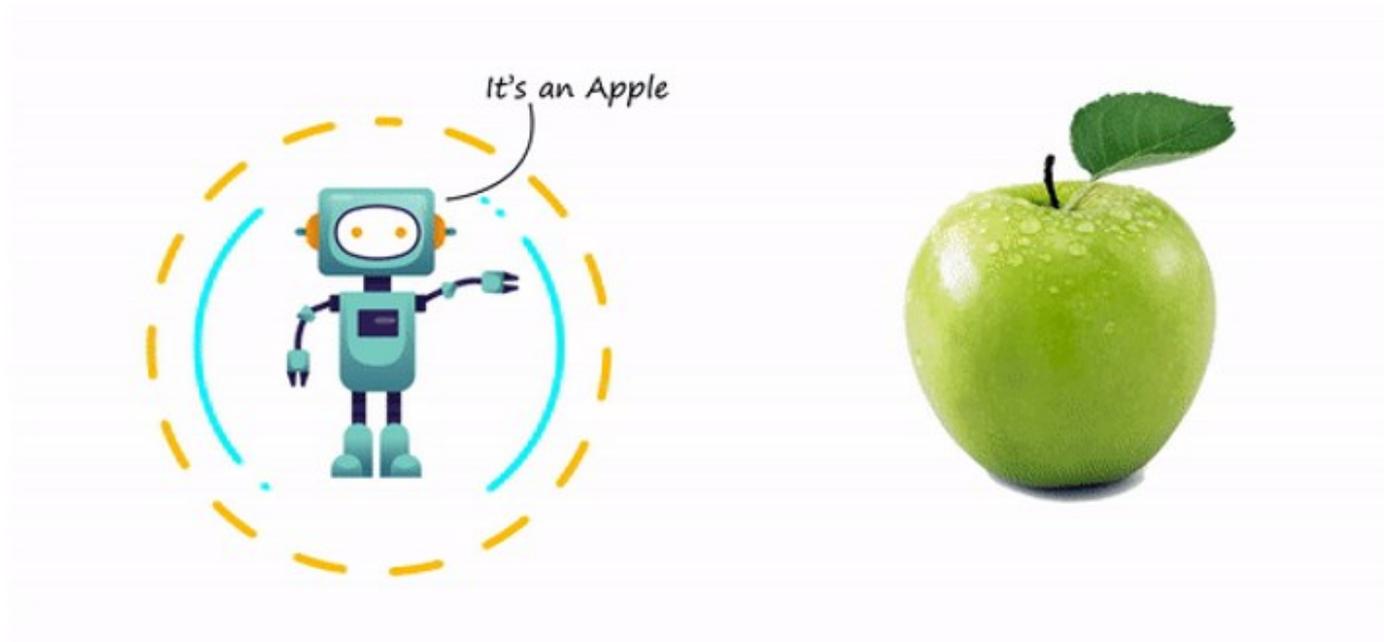
Para construirmos um modelo de aprendizado de máquina (nosso robô na figura) temos que inicialmente treiná-lo com as instâncias do conjunto de treinamento. Iremos mostrar várias maçãs diferentes, onde a imagem (ou as características desta) é o *feature* e a resposta é maçã ou não é a variável alvo. Na figura acima foi apresentada uma fruta e estabelecido que o rótulo da fruta é maçã.



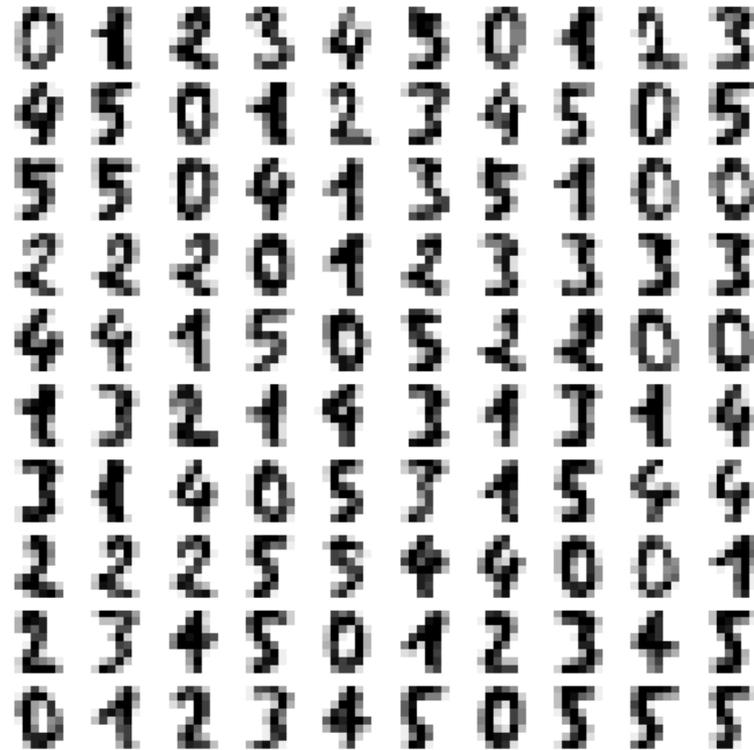
Necessitamos de várias instâncias no conjunto de treinamento para que processo seja efetivo. Agora foi apresentada outra fruta e destacado que está também é uma maçã. Poderíamos enriquecer o processo de treinamento do modelo mostrando exemplos de frutas que não são maçãs.



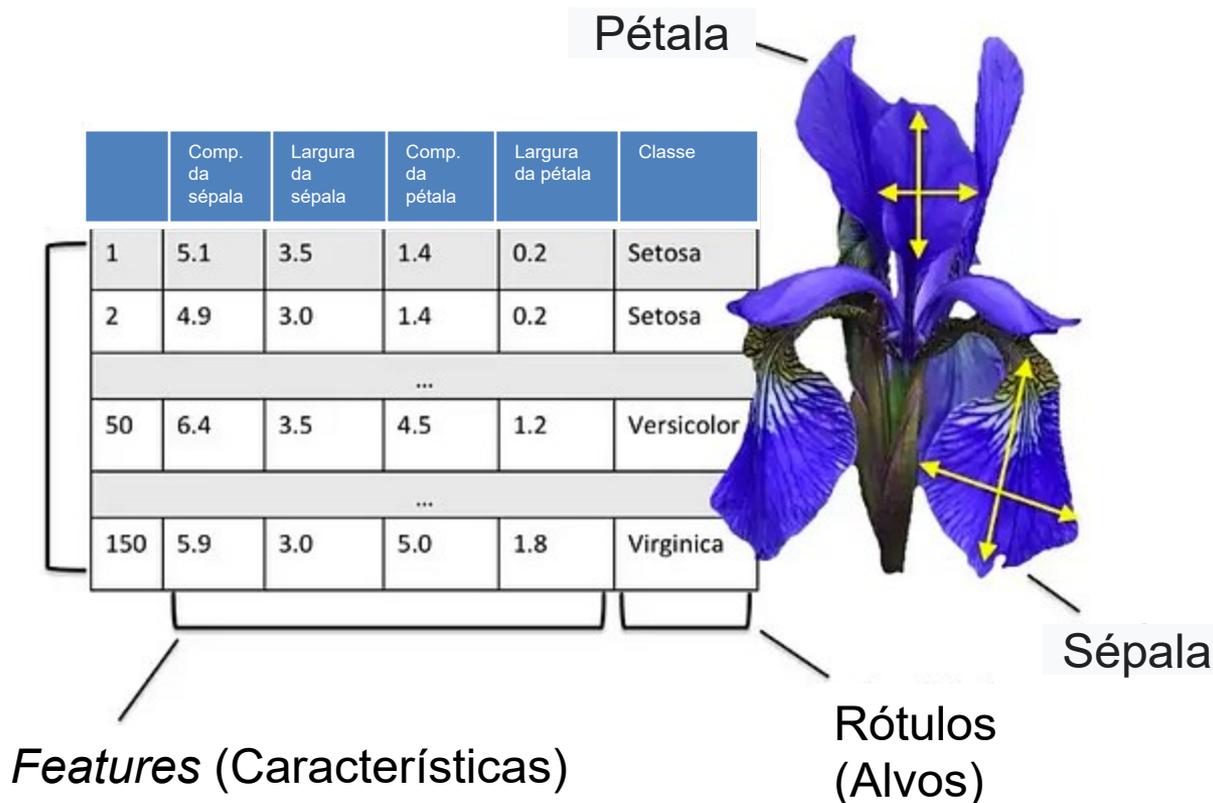
Uma vez treinado o modelo de aprendizado de máquina, usamos instâncias do conjunto de teste para verificar o seu poder de predição. A fruta mostrada não estava no conjunto de treinamento, ou seja, é um desafio para nosso modelo e aprendizado de máquina;



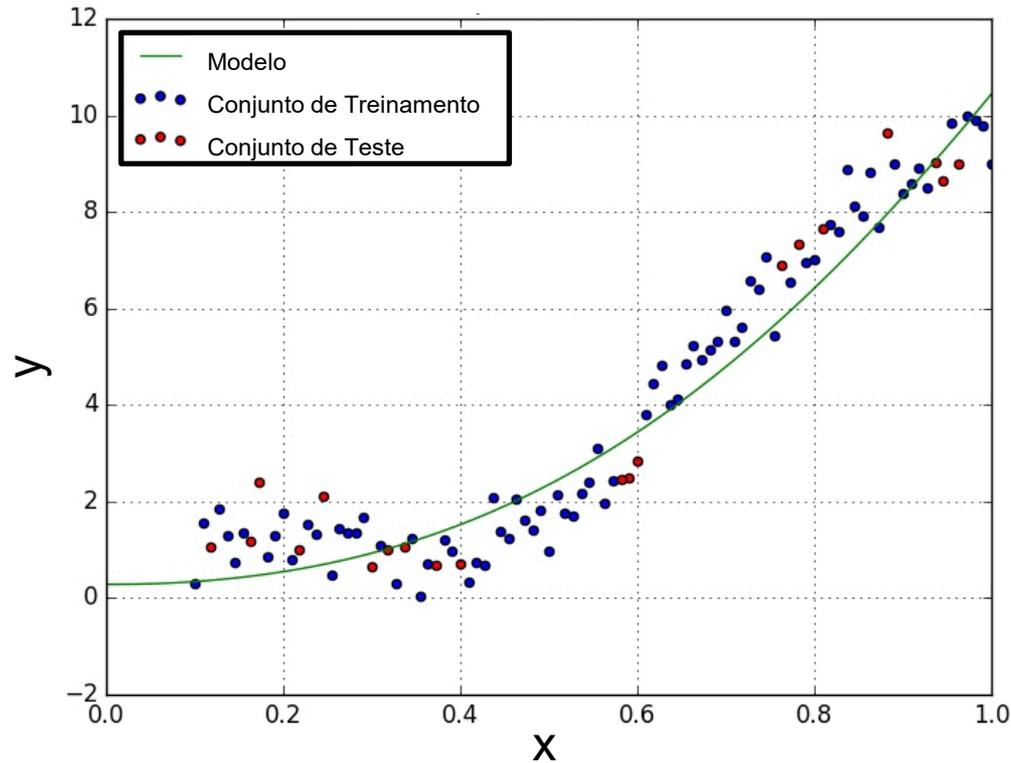
O nosso modelo de aprendizado de máquina teve sucesso na predição da instância acima.



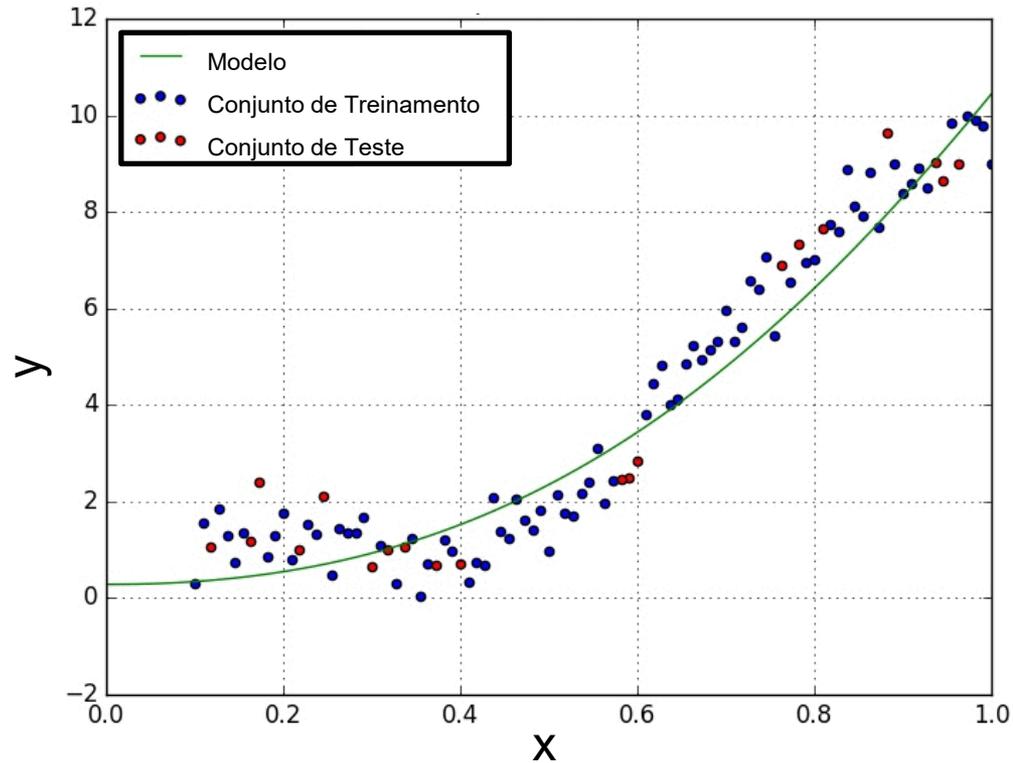
Um conjunto de dados normalmente usado no estudo de problemas de classificação é o **conjunto de dígitos**, onde temos imagens dos algarismos de zero a nove escritas à mão. O objetivo do modelo é ler a imagem do dígito escrito à mão e classificá-la em uma das dez classes possíveis. No desenvolvimento de um modelo para predição dos dígitos, usamos um conjunto de treinamento com dígitos escritos à mão, onde fornecemos os rótulos (qual dígito foi escrito). Depois testamos o poder de previsão com um conjunto de teste que não usamos para construir o modelo de aprendizado de máquina.



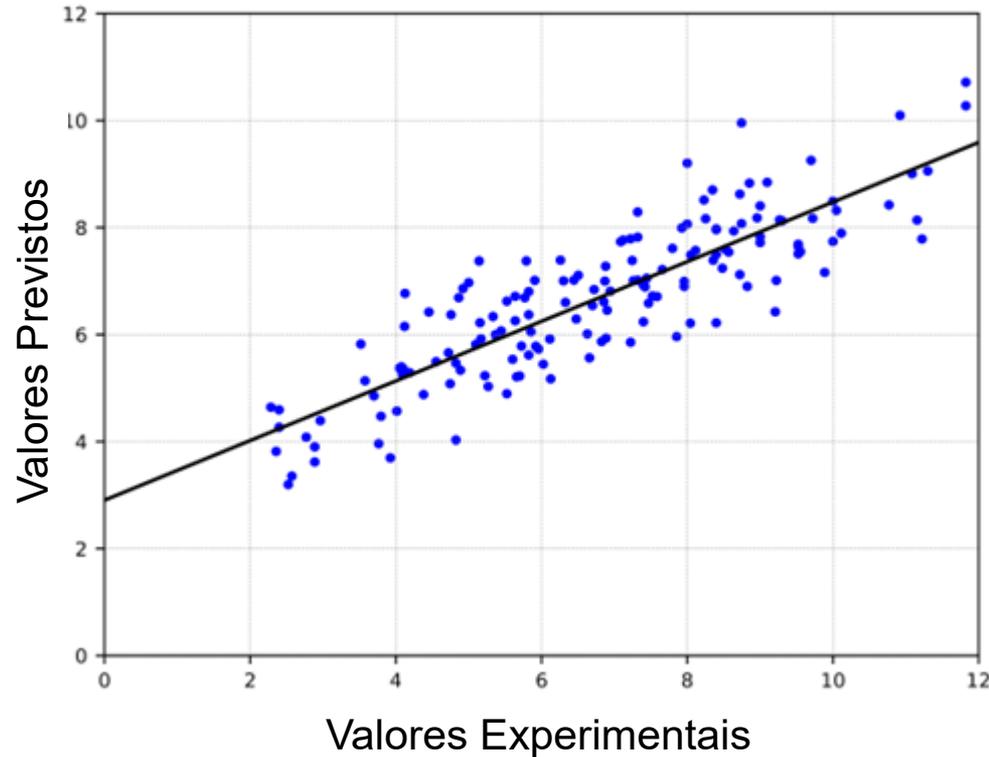
A biblioteca de ferramentas de aprendizado de máquina [Scikit-Learn](https://scikit-learn.org/) tem diversos conjuntos de dados disponíveis para download para treinamento de modelos de aprendizado de máquina para classificação e regressão. Entre os conjunto de dados usados para estudos de modelos de classificação, destaco o Iris. O conjunto de dados Iris tem características de 150 flores do gênero *Iris* (*features*) e a identificação das classes destas (rótulos). O objetivo é ler os *features* e prever qual é a classe da flor do gênero *Iris*.



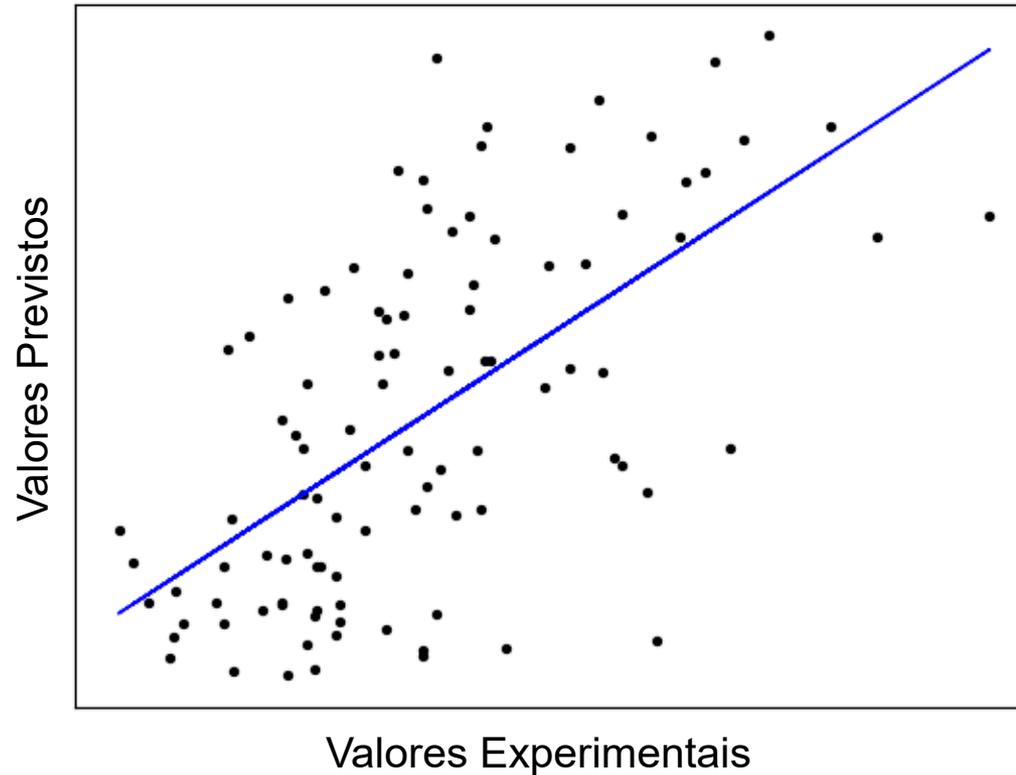
No conjunto de dados representado acima, temos medidas experimentais da variável alvo y (rótulo) para diferentes valores do *feature* x . Temos os pontos no conjunto de aprendizado, onde dividimos em conjuntos de treinamento (em azul) e de teste (em vermelho). O objetivo do modelo de aprendizado de máquina (linha verde) é prever o valor de y tendo como *feature* o x . Em problemas de regressão, nossa variável alvo (rótulo) é contínua, normalmente com números reais. Como na abordagem de treinamento supervisionado, dividimos o conjunto de aprendizado em conjuntos de treinamento e de teste. A linha verde ilustra diferentes modelos de aprendizado de máquina construídos a partir das instâncias do conjunto de treinamento.



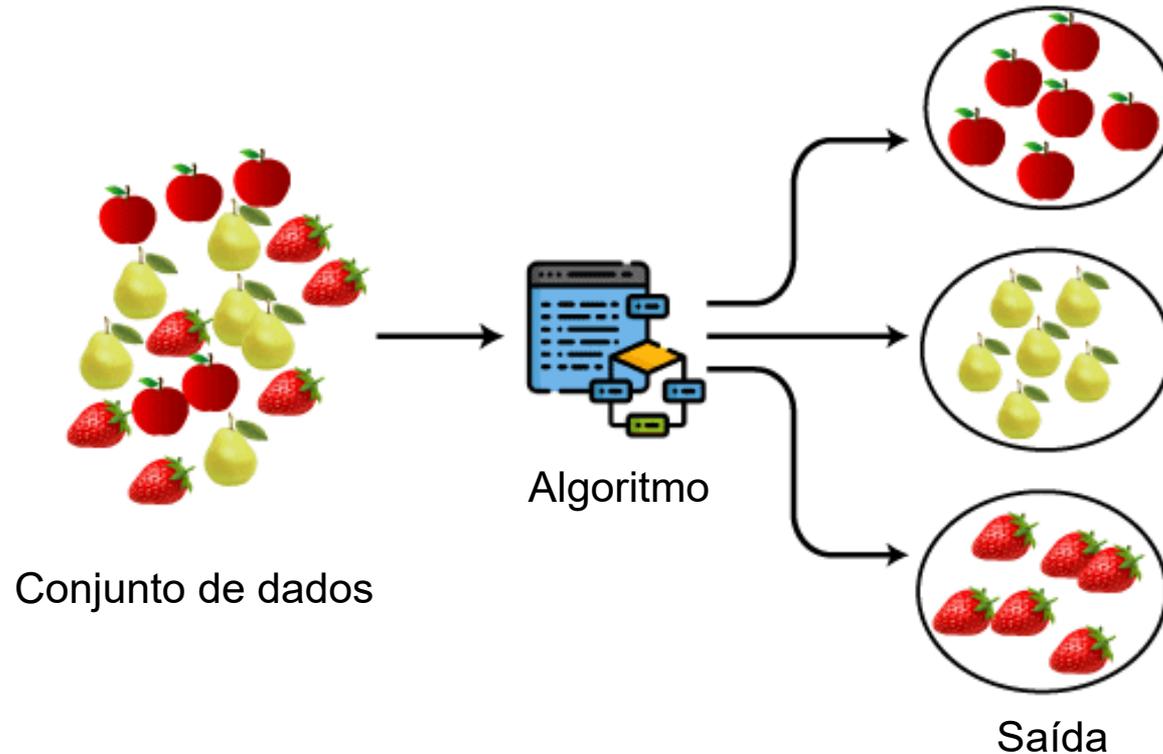
Quando analisamos o poder de previsão dos modelos de aprendizado de máquina, necessitamos de métricas que mostrem a concordância entre os valores previstos pelo modelo (linha verde) e os pontos. Como os pontos do conjunto de treinamento foram usados para construir os modelos, focamos nos pontos do conjunto de teste para aferir o poder de previsão dos modelos. Visualmente podemos inferir quais são os melhores modelos, mas cientificamente usamos um conjunto de equações chamadas métricas.



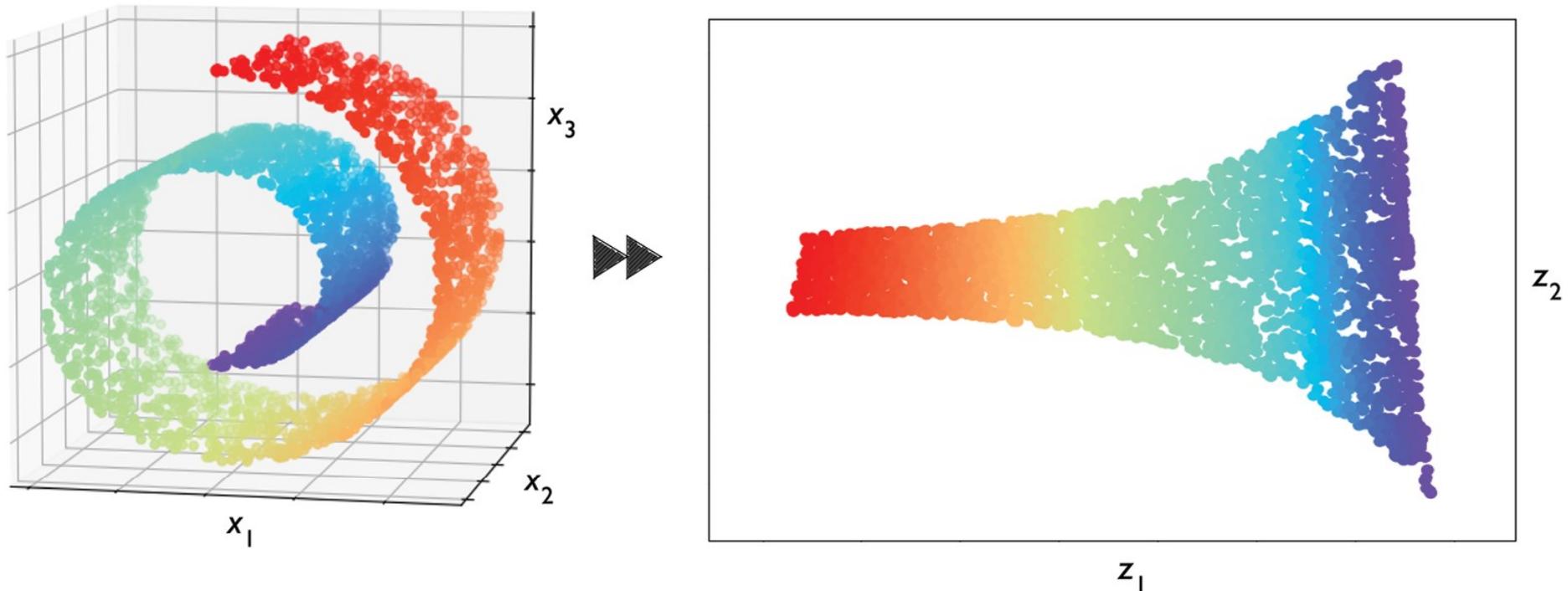
Na análise de modelos de regressão é comum usarmos **gráficos de dispersão** entre os pontos previstos e os experimentais, como o gráfico mostrado acima. No gráfico temos ao longo do eixo x os valores experimentais e no eixo y os valores previstos pelo modelo de aprendizado de máquina. Temos a indicação de uma reta de regressão linear entre os valores experimentais e previstos. Quanto mais próximos da reta estão os pontos, melhor é o modelo de aprendizado de máquina. Esses dados (valor previsto x valor experimental) são usados para o cálculo de métricas como os coeficientes de correlação.



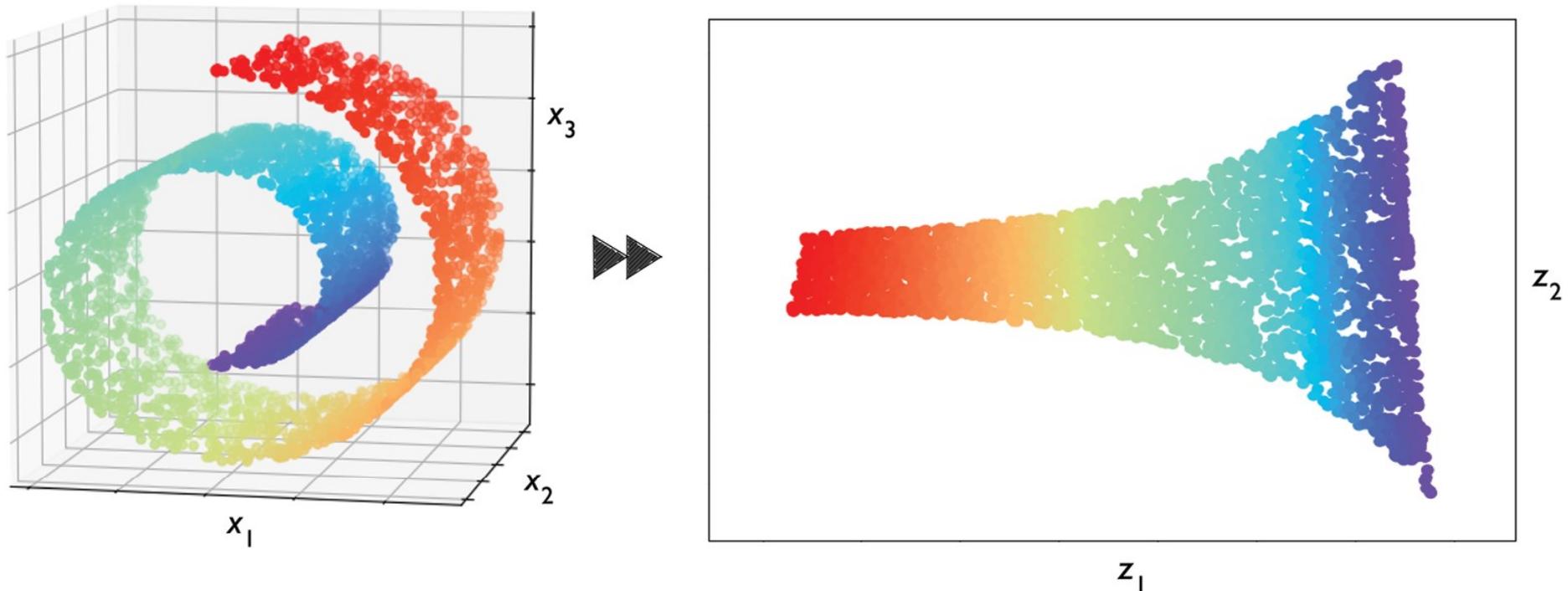
Já destaquei os conjuntos de dados disponíveis na biblioteca [Scikit-Learn](https://scikit-learn.org/) para estudos de classificação. O [Scikit-Learn](https://scikit-learn.org/) tem conjuntos de dados para análise de modelos de regressão. Entre eles um dos mais usados é o conjunto Diabetes. Esse conjunto de dados tem informações de 442 pacientes e dez variáveis (*features*) (idade, sexo, índice de massa corporal, pressão sanguínea média e seis medidas dos níveis séricos no sangue). O rótulo (variável alvo) é uma medida quantitativa da progressão da diabetes (número real) um ano após as medidas iniciais. O objetivo é construir um modelo de aprendizado de máquina para predição quantitativa da progressão da diabetes.



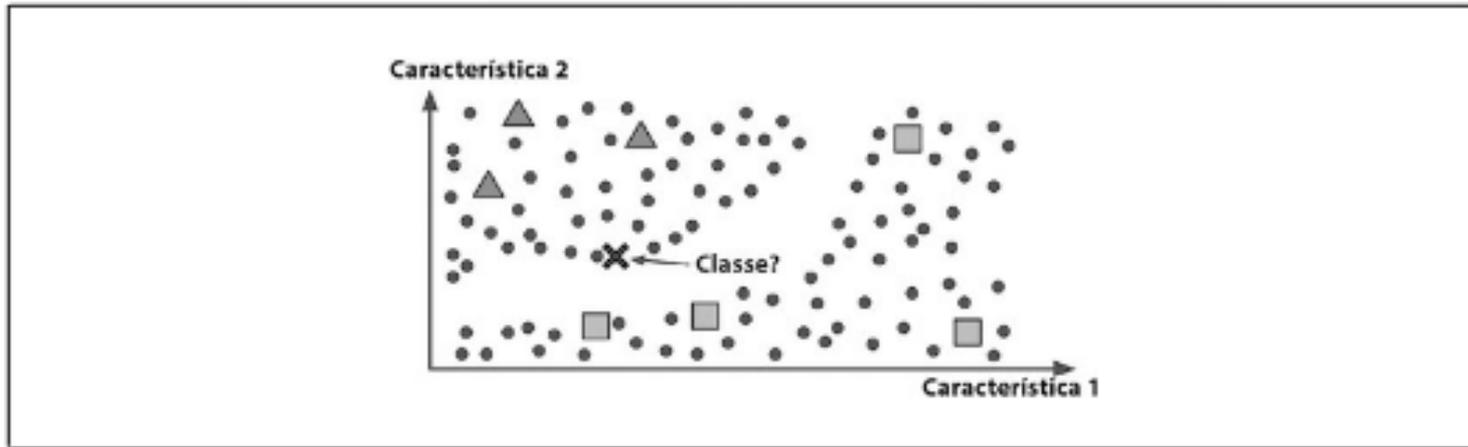
No **aprendizado não supervisionado** não temos uma variável alvo (rótulo). Uma das abordagens do aprendizado não supervisionado envolve clusterização, onde temos um conjunto de dados sem rótulos e o separamos em subconjuntos com características comuns. Na figura acima temos um conjunto de dados de frutas, o algoritmo de clusterização separa por cores. As frutas do conjunto de dados não apresentam rótulos e o processo de clusterização usa as informações contidas nos *features* para gerar a divisão do conjunto original.



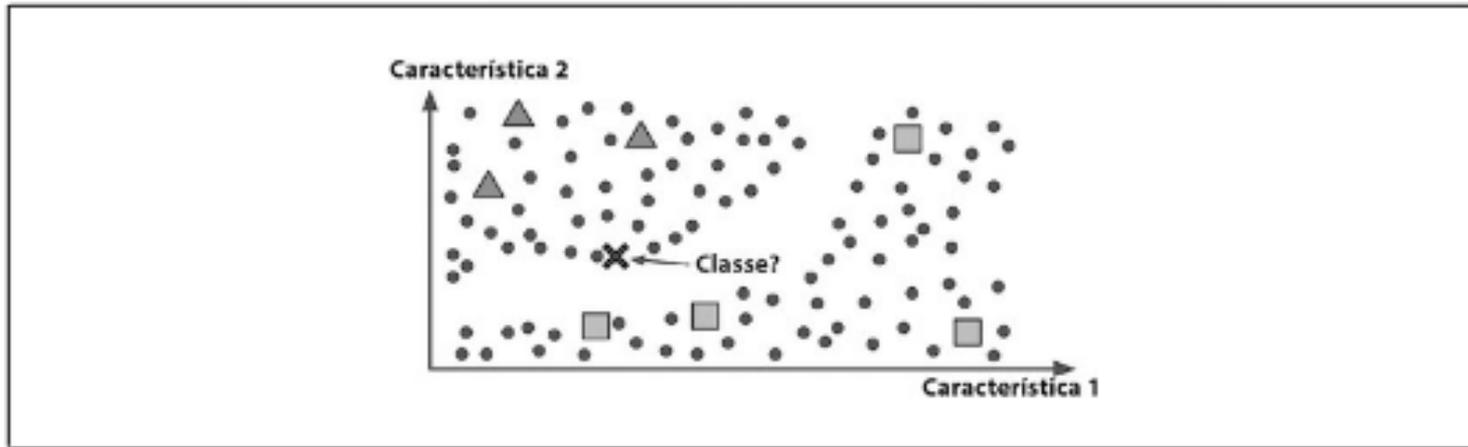
Outra tarefa comum do aprendizado não supervisionado envolve a **redução de dimensionalidade**. O que isso quer dizer? Vamos supor que num conjunto de dados com dez **features**, podemos destas dez considerar só as cinco mais representativas, normalmente com maior variância. Não iremos definir os termos estatísticos com equações neste primeiro contato. O conceito em si é simples, no contexto aqui discutido, a variância indica o quanto um *feature* muda. Espera-se que *features* que pouco variem (baixa variância relativa) contribuem pouco para a construção do modelo.



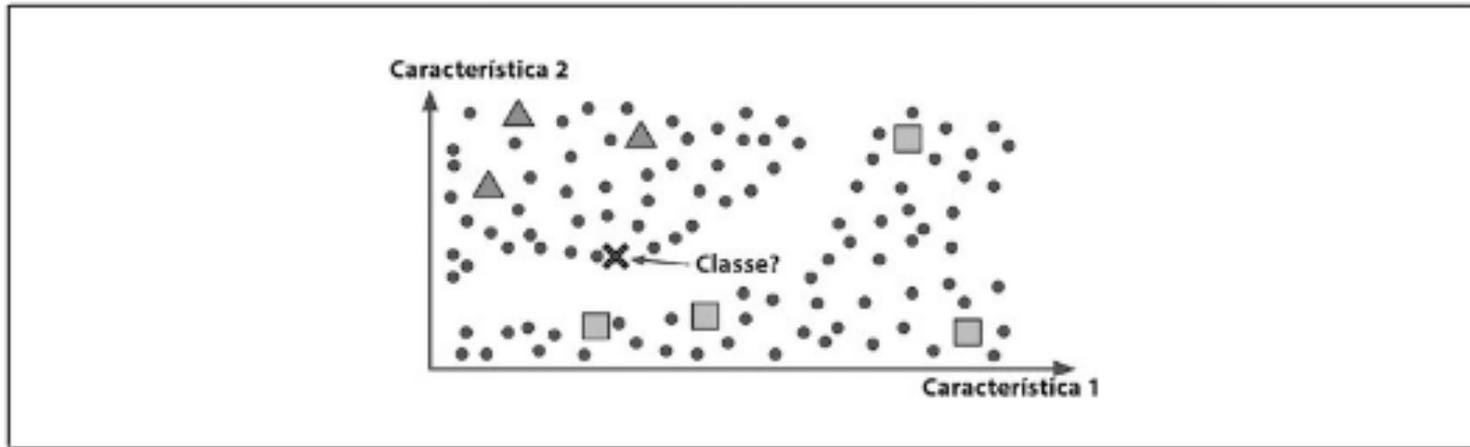
Ainda na redução da dimensionalidade, podemos fazer a fusão de dois ou mais *features* onde substituímos parte ou a totalidade dos *features* um número menor resultado da fusão destes. Na figura acima temos a redução de três para duas dimensões. A figura da esquerda tem três dimensões (três *features*) (x_1 , x_2 e x_3). Construímos dois novos *features* z_1 e z_2 baseados nos anteriores e geramos o gráfico da direita.



Conforme vemos aumentar o leque de opções entre os métodos de aprendizado de máquina, era de se esperar a possibilidade de fusão entre eles. É o caso do **aprendizado de máquina semissupervisionado**, onde temos um conjunto de dados parcialmente rotulados (triângulos e quadrados no gráfico acima) e o algoritmo de aprendizado de máquina semissupervisionado completa os rótulos.



Na figura acima temos um conjunto de dados parcialmente rotulado. Os triângulos e quadrados indicam instâncias rotuladas. Os pontos indicam dados sem rótulo. As instâncias não rotuladas (círculos) ajudam a classificar uma instância nova (a cruz) na classe triângulo em vez de na classe quadrado, ainda que esteja mais próxima dos quadrados rotulados.



Um exemplo de aprendizado de máquina semissupervisionado é o Google Fotos onde carregamos um conjunto de fotos não rotuladas e um algoritmo de clusterização (parte não supervisionada do algoritmo) separa em classes. Rotulamos uma das fotos em cada classe e o algoritmo de aprendizado de máquina semissupervisionado usa esses rótulos para identificar todas agrupadas sob a mesma classe.



AULA MAGNA

Programa de Pós-Graduação em Física
UNIFAL-MG / UFSJ

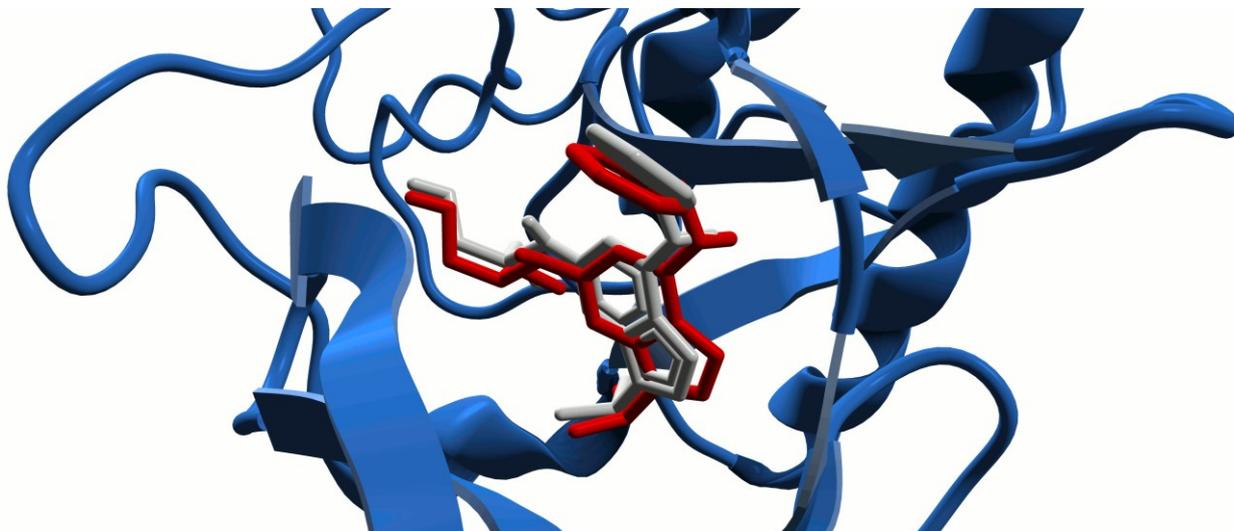
Aprendizado de Máquina para o Estudo de Interações Intermoleculares



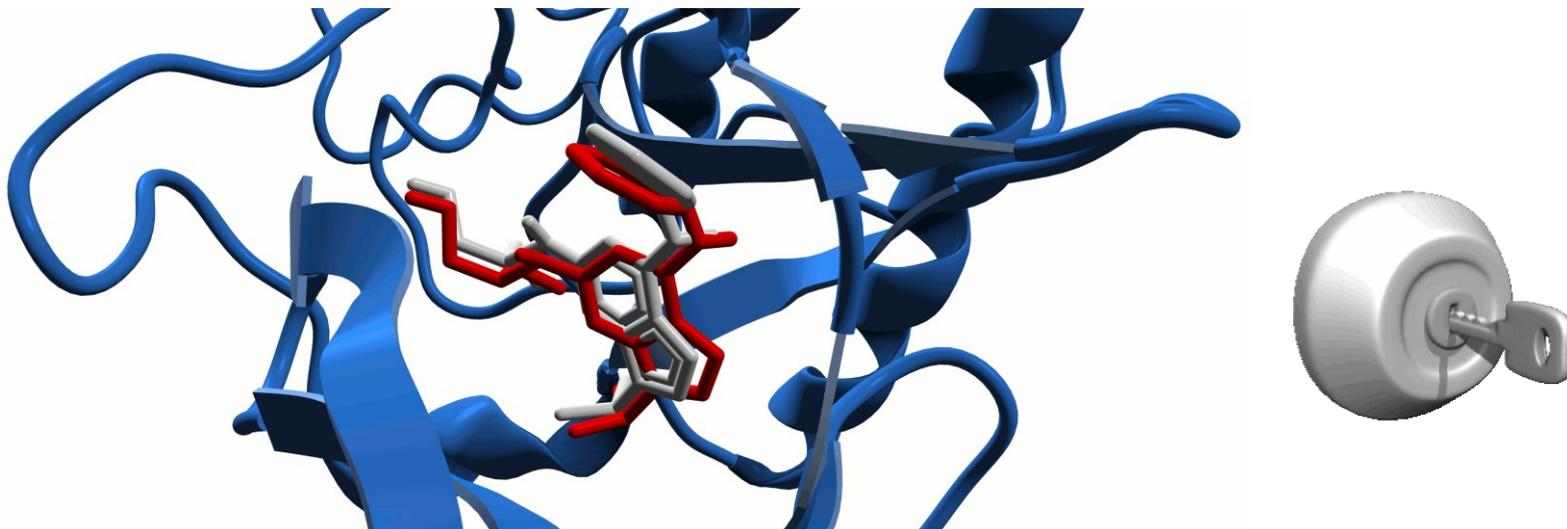
Dr. Walter Filgueira de Azevedo Jr.
Departamento de Física
Universidade Federal de Alfenas

Data: 14 de setembro de 2023
Horário: 14h00
<https://meet.google.com/oyg-yfsv-dac>

O foco desta disciplina está nas aplicações científicas e tecnológicas do aprendizado de máquina. Usaremos os métodos de aprendizado de máquina para gerarmos modelos para problemas científicos atuais. Minha ideia é aproveitar o que vocês já sabem e construir com a adição dos métodos de aprendizado de máquina. Assim aqueles que já estejam envolvidos com pesquisa científica (não importa o nível) podem desenvolver o projeto de pesquisa diretamente relacionado com a sua pesquisa. Obviamente se for possível. A segunda possibilidade é participar dos projetos que coordeno. Parte desses projetos estão descritos na Aula Magna que ministrei em setembro do ano passado (14 de setembro de 2023) disponível no link: <https://vimeo.com/868367654>.



Os projetos que tenho interesse envolvem a aplicação dos métodos de aprendizado de máquina para o estudo de sistemas biológicos. Especificamente a interação de proteínas com fármacos a partir de métodos de regressão onde são gerados modelos que preveem se uma molécula tem potencial farmacêutico ou não. Essa pesquisa é interdisciplinar e tem a participação de físicos, químicos, biólogos, matemáticos, engenheiros, cientistas de computação, farmacêuticos entre outros.



Na animação acima temos a interação de um fármaco em potencial (moléculas vermelhas e cinza) com uma proteína (em azul). Para começar a entender o processo, imagine que a pequena molécula é uma chave e procuramos encaixá-la em parte da proteína (a fechadura). Se você consegue imaginar uma chave encaixando numa fechadura, é tudo que eu preciso para que você entenda as bases do problemas que estou modelando.



1. Para cada um dos cenários indicados abaixo, responda as seguintes questões. A) É um problema de aprendizado de máquina supervisionado ou não supervisionado? B) É uma tarefa de classificação ou regressão? C) Quais seriam os *features* (características) úteis de serem coletados?

Cenário 1. Servidor de e-mails que direciona spam para uma pasta de spams.

Cenário 2. Listagem de preços de imóveis de uma cidade. Imagine que você é um corretor de imóveis e quer ter uma lista de preços dos imóveis baseada em dados.

Cenário 3. Previsão de preços de carros usados. Similar ao anterior só que agora você quer uma listagem dos preços dos carros usados.

Cenário 4. X (antigo Twitter) mostrando os tópicos de tendência (*trending topics*).

Cenário 5. Sistema de recomendação e tamanhos para venda de roupas online.

Cenário 6. Previsão de curso de graduação de um estudante baseado em suas informações pessoais.

Cenário 7. Contagem do número de pessoas que aparecem numa foto.

2. Serviços como o Netflix ou Spotify fazem sugestões de filmes e músicas para os clientes. Discuta similaridades e diferenças em como eles coletam informações, o que eles coletam e como elas usam as informações para fazerem sugestões.

3. Selecione um problema da sua área de pesquisa (ou de uma área de pesquisa em que você tenha interesse) que pode ser tratado com abordagens de aprendizado de máquina.

4. Faça uma pesquisa na internet e tente relacionar os seguintes termos em ordem hierárquica, do mais abrangente para o mais específico. Aprendizado de máquina, inteligência artificial, aprendizado profundo e ciência de dados.

5. A playlist a seguir tem uma canção que só foi possível graças ao uso intensivo de recursos de inteligência artificial. Identifique essa canção. Aproveite para curtir a seleção que será a trilha sonora do nosso curso...



Roger Waters - Comfortably Numb 2022

<https://www.youtube.com/watch?v=9JLN581QWxc>

The Alan Parsons Symphonic Project "Sirius" - "Eye In The Sky" (Live in Colombia)

<https://www.youtube.com/watch?v=Q6ZRGzx31iQ>

Bach - Brandenburg Concerto no. 2 in F major BWV 1047 - Sato | Netherlands Bach Society

<https://www.youtube.com/watch?v=t8rYzpcgk44>

Cosmos-Vangelis-Alpha

<https://www.youtube.com/watch?v=SynSKWw0vaA>

The Beatles - Now And Then (Official Music Video)

<https://www.youtube.com/watch?v=Opxhh9Oh3rg>

The Music of Cosmos: Heaven and Hell - Vangelis

https://www.youtube.com/watch?v=nl_IVInNp68

Jean Michel Jarre - Oxygene II (Live in your living room)

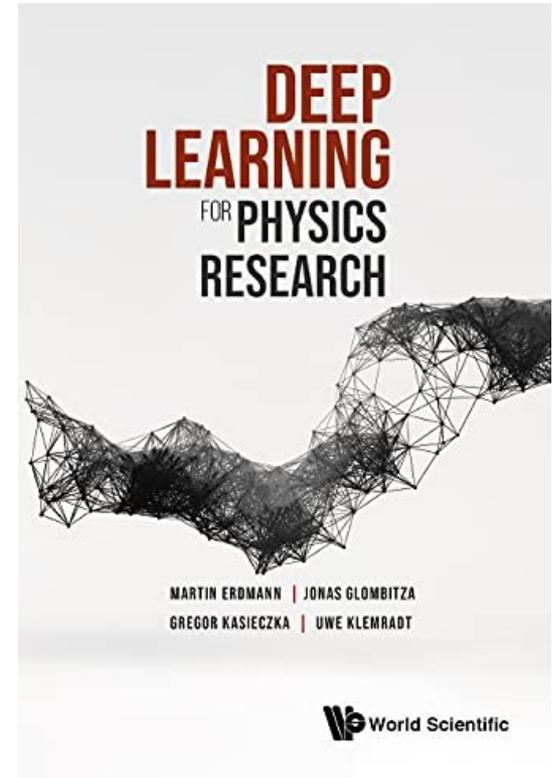
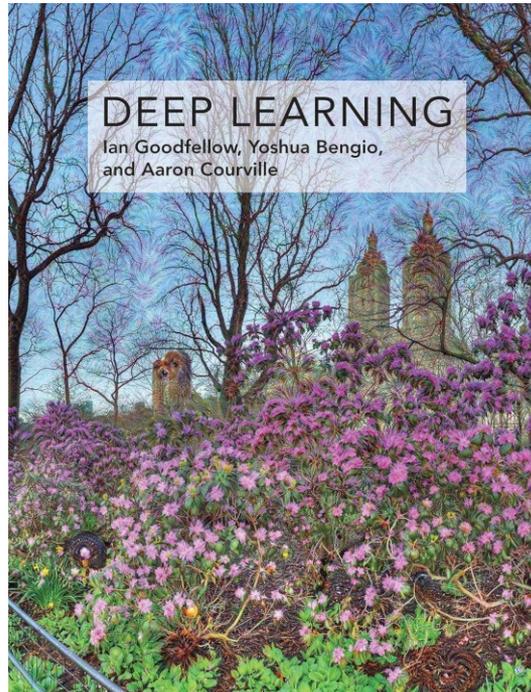
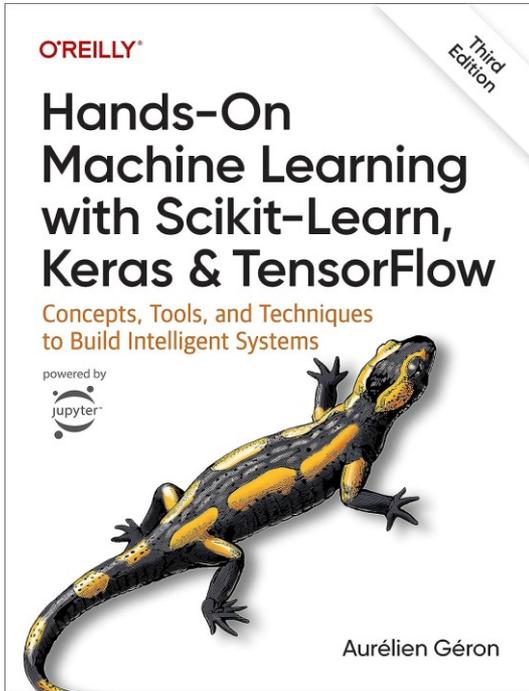
<https://www.youtube.com/watch?v=6HYtkbv5A6A>

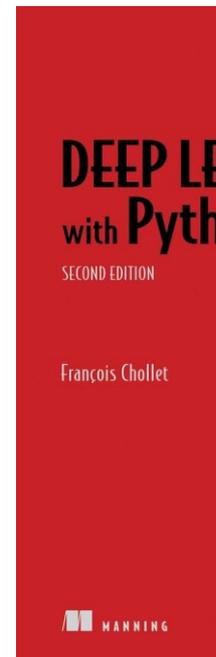
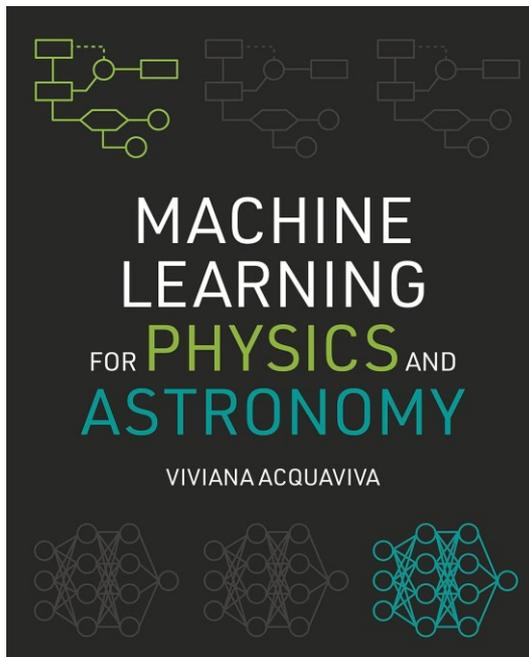
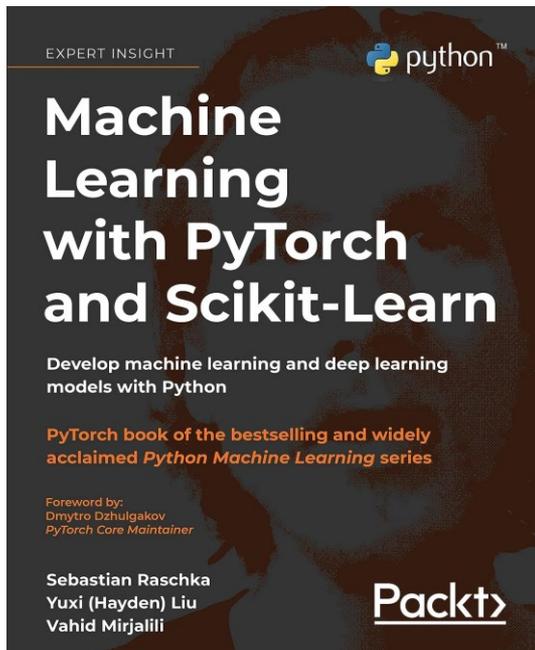
Kraftwerk The Hall of Mirrors

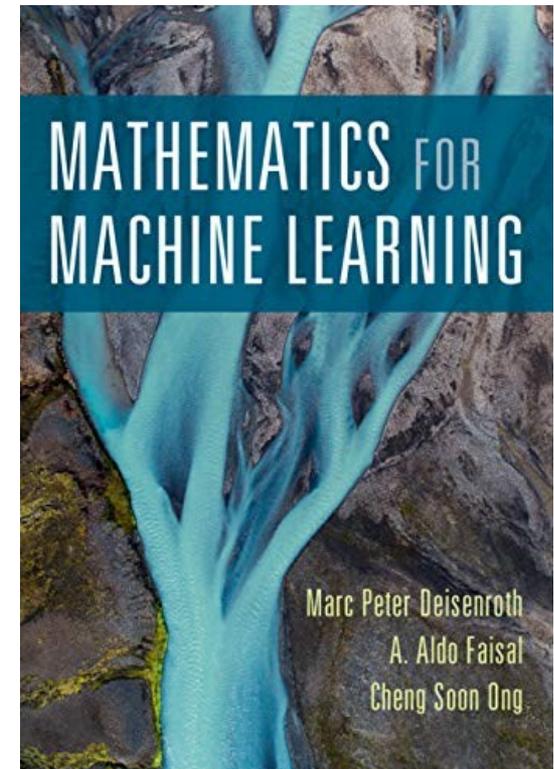
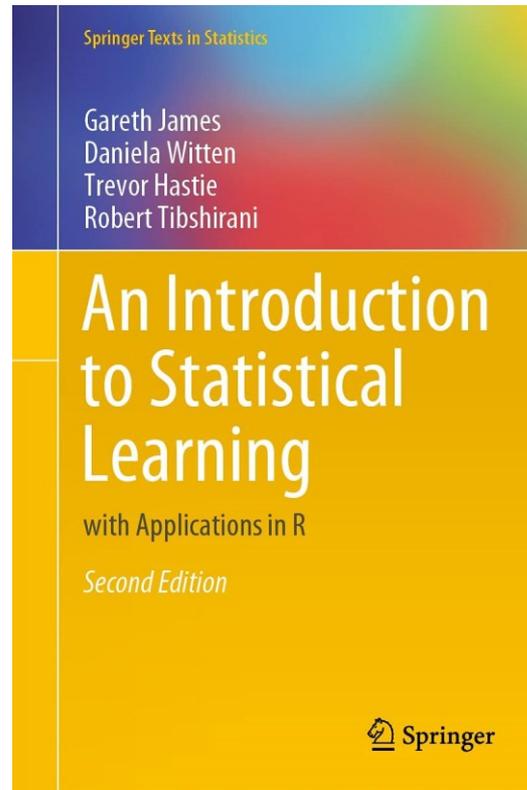
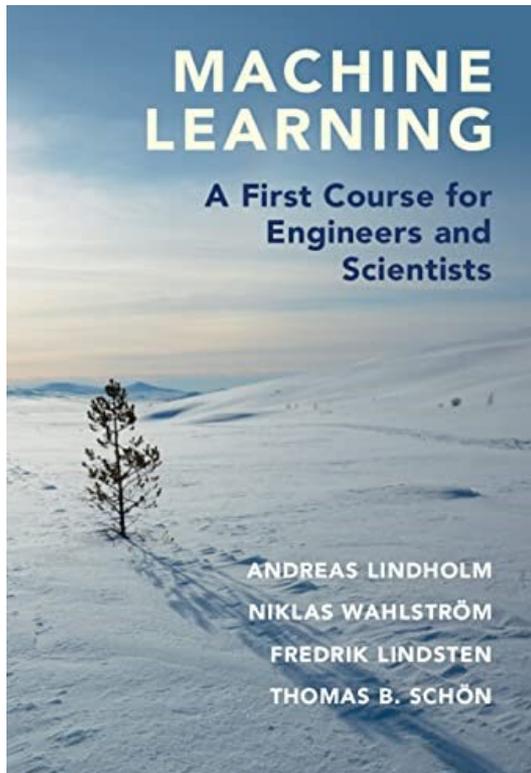
<https://www.youtube.com/watch?v=a15yqZ1uL9A>

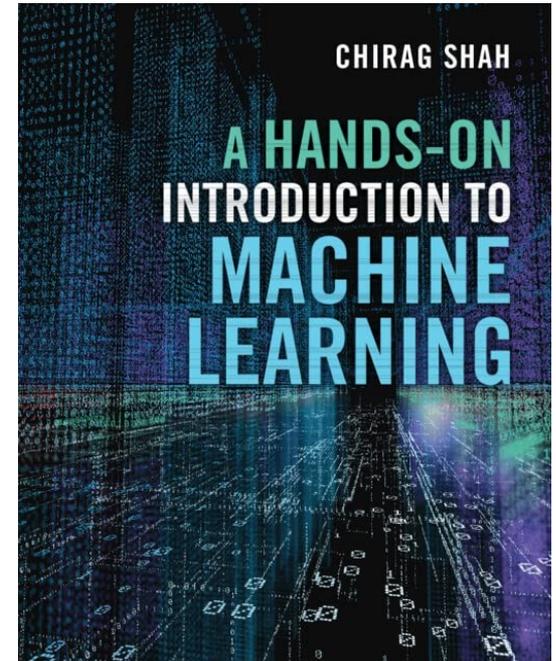
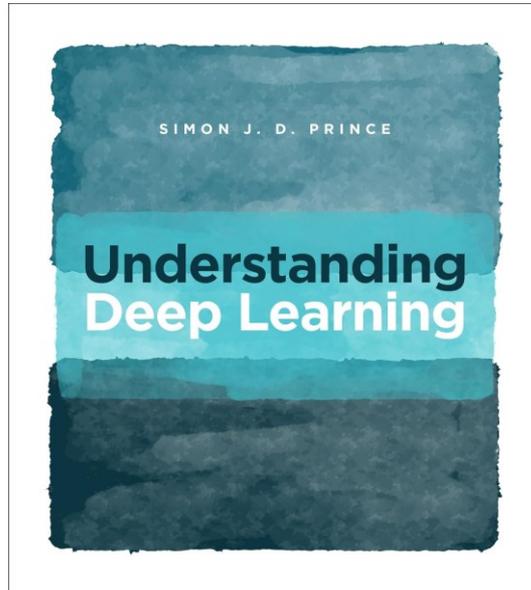
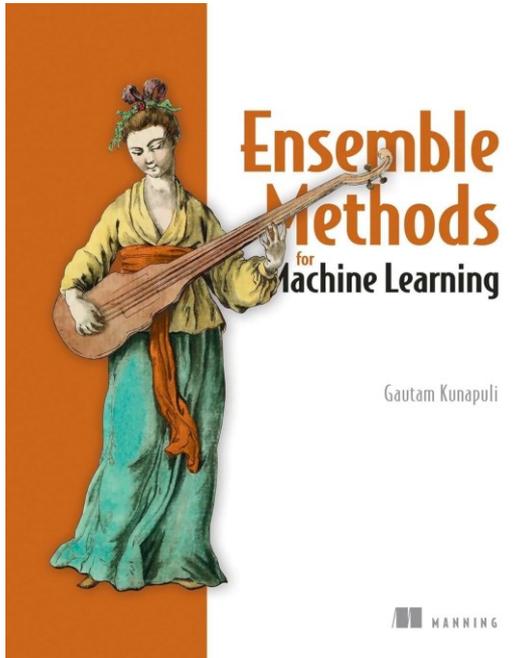
Dire Straits - Brothers In Arms (Official Music Video)

<https://www.youtube.com/watch?v=jhdFe3evXpk>











Biography

[Dr. Walter F. de Azevedo, Jr.](#) earned a BSc in Physics (1990), an MSc in Applied Physics (1992), and a DSc in Applied Physics (1997) from the University of São Paulo (Brazil). In his doctoral studies, Dr. Azevedo worked under the supervision of Prof. Yvonne Primerano Mascarenhas (University of São Paulo) and Prof. Sung-Hou Kim (University of California, Berkeley) on a split Doctoral program with a fellowship from the Brazilian Research Council (CNPq). During his first two years at Berkeley, he was under a CNPq fellowship (1993-95). Due to his performance, Prof. S.-H. Kim hired him as Visiting Researcher for the Department of Chemistry, University of California at Berkeley (1995-96). The work developed during these three years at Berkeley resulted in his thesis about the structure of Cyclin-Dependent Kinase 2 (CDK2) in complex with inhibitors (PDB access code: [2A4L](#)) ([de Azevedo et al., 1996](#); [de Azevedo et al., 1997](#)). Dr. Azevedo is the first author of both papers, and these publications gathered more than [1,000 citations on the Web of Science](#).

During 1997-98 he had a postdoc position at São Paulo State University (Unesp) with a [Fapesp](#) fellowship. He holds a habilitation degree in Physics (livre-docência) from the São Paulo State University (Unesp)(2004).

In 1998, Dr. Azevedo participated in a research project with NASA that sent proteins to crystallize in a microgravity environment onboard the Space Shuttle Discovery (STS-95). This research had coverage of Brazilian [TV networks](#). He published a book entitled "[Docking Screens for Drug Discovery](#)" with Springer Nature in 2019. This book sold 43,000 copies (January 2024) with over 2 million dollars in sales (<https://link.springer.com/book/10.1007/978-1-4939-9752-7>). In 2020, the [Journal Plos Biology](#) ranked Dr. Azevedo among the most influential researchers in the world (Fields: Biochemistry & Molecular Biology and Biophysics).

Dr. Azevedo has vast editorial experience. He is the frontiers section editor (Bioinformatics/Biophysics) for the [Current Drug Targets](#), section editor (Bioinformatics in Drug Design and Discovery) for the [Current Medicinal Chemistry](#), review editor for [Frontiers in Chemistry](#), associate editor for [Exploration of Drug Science](#), member of the editorial boards [Molecular Diversity](#) and the [Journal of Molecular Structures](#), and editor of Docking Screens for Drug Discovery (Methods of Molecular Biology)-Springer Nature. He is a reviewer for over 60 high-impact journals, including Nature Communications and Briefings in Bioinformatics.

His research interests are interdisciplinary, with three main emphases: machine learning, complex systems, and computational systems biology. Dr. Azevedo developed several [free software](#) to explore the concept of [Scoring Function Space](#). He has over 200 scientific publications about protein structures, computer models of complex systems, and simulations of protein systems. These workers have over 7000 citations on the Web of Science ([h-index: 47](#), [m-quotient: 1.7](#)), +7000 citations in Scopus ([h-index: 48](#)), and +9000 citations on Google Scholar ([h-index: 53](#)).