

Técnicas de Machine Learning (Aprendizagem de Máquina)

PROFESSORES

Iara Carnevale de Almeida
Ronie Cesar Tokumoto
Janaína Aparecida de Freitas

ACESSE AQUI O SEU
LIVRO NA VERSÃO
DIGITAL!

EXPEDIENTE

DIREÇÃO UNICESUMAR

Reitor Wilson de Matos Silva **Vice-Reitor** Wilson de Matos Silva Filho **Pró-Reitor de Administração** Wilson de Matos Silva Filho **Pró-Reitor Executivo de EAD** William Victor Kendrick de Matos Silva **Pró-Reitor de Ensino de EAD** Janes Fidélis Tomelin **Presidente da Mantenedora** Cláudio Ferdinandi

NEAD - NÚCLEO DE EDUCAÇÃO A DISTÂNCIA

Diretoria Executiva Chrystiano Mincoff, James Prestes, Tiago Stachon **Diretoria de Graduação e Pós-graduação** Kátia Coelho **Diretoria de Cursos Híbridos** Fabricio Ricardo Lazilha **Diretoria de Permanência** Leonardo Spaine **Diretoria de Design Educacional** Paula Renata dos Santos Ferreira **Head de Graduação** Marcia de Souza **Head de Metodologias Ativas** Thuinie Medeiros Vilela Daros **Head de Tecnologia e Planejamento Educacional** Tania C. Yoshie Fukushima **Gerência de Planejamento e Design Educacional** Jislaine Cristina da Silva **Gerência de Tecnologia Educacional** Marcio Alexandre Wecker **Gerência de Produção Digital** Diogo Ribeiro Garcia **Gerência de Projetos Especiais** Edison Rodrigo Valim **Supervisora de Produção Digital** Daniele Correia

FICHA CATALOGRÁFICA

Coordenador(a) de Conteúdo

Flavia Lumi Matuzawa

Projeto Gráfico e Capa

André Morais, Arthur Cantareli e
Matheus Silva

Editoração

Adrian Marcareli e André Morais

Design Educacional

Bárbara Neves

Curadoria

Elziane Vieira Alencar

Revisão Textual

Nágela Neves

Ilustração

André Azevedo e Eduardo Aparecido
Alves

Fotos

Shutterstock

C397 CENTRO UNIVERSITÁRIO DE MARINGÁ.

Núcleo de Educação a Distância. **ALMEIDA**, Iara Carnevale; **TOKUMOTO**, Ronie Cesar; **FREITAS**; Janaína Aparecida.

Técnicas de Machine Learning (Aprendizagem de Máquina). Iara Carnevale de Almeida; Ronie Cesar Tokumoto; Janaína Aparecida de Freitas. Maringá - PR.: Unicesumar, 2021.

176 p.

ISBN 978-65-5615-606-4

"Graduação - EaD".

1. Técnicas 2. Machine 3. Learning. 4. EaD. I. Título.

CDD - 22 ed. 681.32

Impresso por:

Bibliotecário: João Vivaldo de Souza CRB-9-1679



NEAD - Núcleo de Educação a Distância

Av. Guedner, 1610, Bloco 4 Jd. Aclimação - Cep 87050-900 | Maringá - Paraná

www.unicesumar.edu.br | 0800 600 6360

BOAS-VINDAS

A UniCesumar celebra os seus 30 anos de história avançando a cada dia. Agora, enquanto Universidade, ampliamos a nossa autonomia e trabalhamos diariamente para que nossa educação à distância continue como uma das melhores do Brasil. Atuamos sobre quatro pilares que consolidam a visão abrangente do que é o conhecimento para nós: o intelectual, o profissional, o emocional e o espiritual.

A nossa missão é a de "Promover a educação de qualidade nas diferentes áreas do conhecimento, formando profissionais cidadãos que contribuam para o desenvolvimento de uma sociedade justa e solidária". Neste sentido, a UniCesumar tem um gênio importante para o cumprimento integral desta missão: o coletivo. São os nossos professores e equipe que produzem a cada dia uma inovação, uma transformação na forma de pensar e de aprender. É assim que fazemos juntos um novo conhecimento diariamente. São mais de 800 títulos de livros didáticos como este produzidos anualmente, com a distribuição de mais de 2 milhões de exemplares gratuitamente para nossos acadêmicos. Estamos presentes em mais de 700 polos EAD e cinco campi: Maringá, Curitiba, Londrina, Ponta Grossa e Corumbá, o que nos posiciona entre os 10 maiores grupos educacionais do país.

Aprendemos e escrevemos juntos esta belíssima história da jornada do conhecimento. Mário Quintana diz que "Livros não mudam o mundo, quem muda o mundo são as pessoas. Os livros só mudam as pessoas". Seja bem-vindo à oportunidade de fazer a sua mudança!

Reitor

Wilson de Matos Silva

*Tudo isso para honrarmos a
nossa missão, que é promover
a educação de qualidade nas
diferentes áreas do conhecimento,
formando profissionais
cidadãos que contribuam para
o desenvolvimento de uma
sociedade justa e solidária.*



MINHA HISTÓRIA

MEU CURRÍCULO



Lara Carnevale de Almeida

Olá, antes de falar da minha experiência profissional, gostaria de me apresentar, enquanto “pessoa lara”. Eu gosto muito de jardinagem, uma consequência disso é que minha área de serviço se tornou um pequeno jardim. Gosto muito de cozinhar, principalmente, a partir da culinária italiana. Faço questão de comer muita salada, verduras e frutas e, felizmente, passei isso para o meu filho. Às vezes, trabalho demais, mas o faço porque gosto muito de lecionar e pesquisar! Finalmente, queria salientar a minha satisfação de fazer parte da construção deste material.

Tenho doutorado em Informática pela Universidade de Évora (2011) com equivalência pela Universidade Federal do Rio Grande do Sul (2012). A minha tese de doutorado foi na área da Inteligência Artificial (IA), na qual propus um sistema multiagentes para argumentar e cooperar com as suas bases de conhecimento. A área da IA encanta-me muito e, por isso, tenho procurado manter-me atualizada. Assumo que, às vezes, não é fácil, devido à forte e rápida evolução dessa área!

Atualmente, estou como docente no Mestrado em Gestão do Conhecimento na Universidade Cesumar (Unicesumar) em Maringá-PR e leciono em cursos de graduação nas modalidades EaD e presencial nesta mesma instituição. Sou, também, bolsista de produtividade no Instituto Cesumar de Ciência, Tecnologia e Inovação (ICETI) na Unicesumar. Estou como líder do grupo de pesquisa CNPq “Educação e Conhecimento” pelo ICETI. Como se não bastasse, também leciono disciplina de Letramento Digital para o Fundamental II, no Colégio Objetivo de Maringá-PR. Acho fascinante e desafiador ensinar esta geração!

Desde maio de 2021, faço parte do Comitê da Área de Inovação da Fundação Araucária (Paraná) e tenho colaborado no grupo STEAM-IA para a disseminação do conhecimento sobre IA, em que procuro desmistificar essa área, como também prestar atenção aos cuidados que precisamos ter no que se refere à ética.

MINHA HISTÓRIA

MEU CURRÍCULO

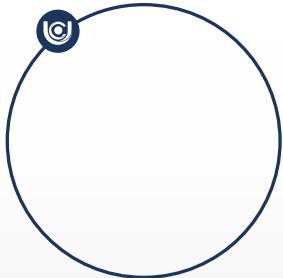


Janaína Aparecida de Freitas

O meu nome é Janaína, sou graduada em Informática, pela Universidade Estadual de Maringá (UEM/2010), e, depois, fiz uma Especialização MBA em Teste de Software, pela UniCeuma de Brasília. Na época, fiz parte da equipe de qualidade de uma empresa de Desenvolvimento de Sistemas; em seguida, trabalhei como Analista de Sistemas. Atuei no mercado gráfico como designer e diagramadora. Depois do mercado de trabalho, ingressei na carreira docente, cursando a pós em Docência no Ensino Superior: Tecnologias Educacionais e Inovação, pela Unicesumar, e outra pós, mas na área de Gerenciamento de Banco de Dados, pela mesma instituição.

Atualmente, sou graduanda em Letras e pós-graduada em Leitura e Produção Textual pela Unicesumar, como forma de desenvolver habilidades e competências pessoais, além de ser uma área que me atrai bem como me auxilia na carreira docente. Sou aluna não-regular do programa de Mestrado em Ciências da Computação, pela Universidade Estadual de Maringá (UEM).

Tenho experiência na área de Educação desde 1995, em escolas profissionalizantes de Informática, e, hoje, atuo no EaD-Unicesumar, desde junho de 2015, como professora mediadora. Atuo, também, como formadora e conteudista de disciplinas dos cursos de Engenharia de Software, Análise e Desenvolvimento de Sistemas, Sistemas para Internet e Gestão da Tecnologia.



Aqui você pode conhecer um pouco mais sobre mim, além das informações do meu currículo.

MINHA HISTÓRIA

MEU CURRÍCULO

Ronie Cesar Tokumoto

Graduei-me em Bacharelado em Informática, pela Universidade Federal do Paraná, em 2001. Depois, cursei pós-graduações em Docência no Ensino Superior pela Unicesumar, Tutoria em Educação a Distância pela Faculdade Eficaz e Gestão Escolar Integrada e Práticas Pedagógicas pela Faculdade Eficaz.

Recentemente, terminei minha segunda graduação como Tecnólogo em Gastronomia, como forma de desenvolver habilidades e competências pessoais e profissionais, além de estudar uma área que muito me atrai, desde muito tempo.

Tenho experiência na área de educação, desde 1994, em escolas profissionalizantes e colégios estaduais. Também possuo experiência em Coordenação de Curso Técnico, no ano de 2013, e participação em palestras e eventos organizados para o curso, durante esse mesmo ano.

Atualmente, encontro-me no EaD-Unicesumar, desde abril de 2014; primeiro, como Professor Mediador, até dezembro de 2017, e como tutor Pedagógico desde então.

Também, atuei como professor no curso de Engenharia de Software da Unifamma, de agosto de 2019 a dezembro de 2020.



PROVOCACÕES INICIAIS

TÉCNICAS DE MACHINE LEARNING

Imagine que você resolve encarar o desafio de trabalhar em uma organização comercial de grande porte. Essa organização necessita de um(a) profissional para lidar com o grande volume de dados gerados pelas suas vendas. Contudo esses dados devem ser analisados de forma rápida e eficiente, a fim de que os seus gestores possam tomar decisões e manter-se à frente da concorrência.

Ao aceitar esse desafio, você sabe que precisa conhecer as estratégias da gestão e, principalmente, como os dados são coletados. Um dos objetivos é sugerir ferramentas adequadas, baseadas em *Machine Learning* ou *Deep Learning*. Como você faria isto? Refletiu sobre qual deve ser o conhecimento necessário para determinar a melhor solução ao seu cliente?

Atualmente, muitas aplicações baseadas em IA estão disponíveis em sistemas computacionais e são utilizadas, pelas organizações, em diferentes setores: comercial, industrial e educacional. O *Machine Learning* (ML, em português, “Aprendizado de Máquina”) e o *Deep Learning* (DL, em português, “Aprendizado Profundo”) são subáreas da Inteligência Artificial (IA).

Reconhecemos o ML quando lidamos com um aplicativo que processa grande volume de dados (em inglês, Big Data), constrói o seu aprendizado com o apoio de especialistas humanos ou sem intervenção humana e consegue responder a questionamentos ou tirar conclusões sobre novos dados. Já o DL é uma subárea do ML, apresenta técnicas as quais ajudam a aprimorar a capacidade das máquinas em aprender tarefas, tais como: reconhecer, classificar, descrever e detectar parâmetros, ou seja, compreender.

Neste momento, imagino que você deve se questionar: quais são os modelos, métodos e algoritmos que devemos conhecer para construir soluções baseadas em *Machine Learning* e *Deep Learning*? Será que existem ferramentas prontas? Sim! Vamos conhecê-las?

Antes de conhecer algoritmos e métodos, precisamos aprender os tipos de modelo de aprendizagem: modelos preditivos, os quais possuem tarefas para encontrar uma função, um modelo ou uma hipótese para realizar uma previsão; e modelos descritivos, os quais possuem tarefas que são, basicamente, direcionadas a explorar ou a descrever um conjunto de dados. Além disso, existem diferentes tipos de análise de negócios: a análise descritiva trabalha com os dados

PROVOCACÕES INICIAIS

brutos, de forma a identificar o que aconteceu; a análise diagnóstica procura explicar porque aconteceu; a análise preditiva procura prever o que acontecerá no futuro; a análise prescritiva ajuda as empresas a alterar o futuro, por meio de recomendações inteligentes, a fim de gerar ou acelerar resultados desejados.

Existem diversos métodos e algoritmos com base em *Machine Learning*, por exemplo, os algoritmos: Máquinas de Vetores de Suporte (em inglês, *Support Vector Machines* – SVM); Árvores de Classificação e Regressão (em inglês, *Classification and Regression Trees* – CART); k-Vizinho Mais Próximo (em inglês, *k-Nearest Neighbors* – kNN); Redes Neurais Artificiais (em inglês, *Artificial Neural Networks* – ANN); Máquinas de Vetores de Suporte (em inglês, *Support Vector Machines* – SVM). Curioso(a) em conhecer as propostas desses algoritmos?

Imagino que você esteja curioso(a) acerca dos conceitos relacionados ao *Machine Learning*, estou certa? Que tal iniciar a reflexão a partir de uma área de atuação (educação, comércio ou indústria)? Em qualvocê acha que o *Machine Learning* pode ser empregado e, por consequência, melhorar a eficiência e a eficácia nessa área?

Esta provocação inicial tem como objetivo apontar que o(a) espera nesta disciplina, com vistas a mostrar que o universo das técnicas de *Machine Learning* é vasto. Ter conhecimento a respeito dessas técnicas permitirá que você, enquanto profissional, possa sugerir ferramentas adequadas para os seus clientes.

Procure identificar como o ML interage no seu dia a dia. Por exemplo, pergunte a um telefone celular e receba uma resposta. Caso você tenha um Iphone, fale: “olá, Siri” e pergunte: “qual a previsão do tempo para a minha cidade, amanhã?”.

O que mais você consegue detectar? Muitas outras opções existem, tais como: recomendação de vídeos do YouTube; sugestões de filmes disponíveis na Netflix; preenchimento automático quando está digitando; carros que se movem sem a necessidade de condutor; reconhecimento automático de rosto, lugar, comida, objeto e animal, apresentado em, por exemplo, Google Fotos ou Facebook.

Preparado(a) para estudar o universo das técnicas de *Machine Learning* e *Deep Learning*?

RECURSOS DE IMERSÃO



REALIDADE AUMENTADA

Sempre que encontrar esse ícone, esteja conectado à internet e inicie o aplicativo Unicesumar Experience. Aproxime seu dispositivo móvel da página indicada e veja os recursos em Realidade Aumentada. Explore as ferramentas do App para saber das possibilidades de interação de cada objeto.



RODA DE CONVERSA

Professores especialistas e convidados, ampliando as discussões sobre os temas.



PÍLULA DE APRENDIZAGEM

Uma dose extra de conhecimento é sempre bem-vinda. Posicionando seu leitor de QRCode sobre o código, você terá acesso aos vídeos que complementam o assunto discutido.



PENSANDO JUNTOS

Ao longo do livro, você será convidado(a) a refletir, questionar e transformar. Aproveite este momento.



EXPLORANDO IDEIAS

Com este elemento, você terá a oportunidade de explorar termos e palavras-chave do assunto discutido, de forma mais objetiva.



NOVAS DESCOBERTAS

Enquanto estuda, você pode acessar conteúdos online que ampliam a discussão sobre os assuntos de maneira interativa usando a tecnologia a seu favor.



OLHAR CONCEITUAL

Neste elemento, você encontrará diversas informações que serão apresentadas na forma de infográficos, esquemas e fluxogramas os quais te ajudarão no entendimento do conteúdo de forma rápida e clara

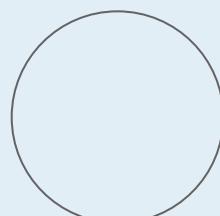
Quando identificar o ícone de QR-CODE, utilize o aplicativo Unicesumar Experience para ter acesso aos conteúdos on-line. O download do aplicativo está disponível nas plataformas:



Google Play



App Store



CAMINHOS DE APRENDIZAGEM

1
13
FUNDAMENTOS
DE MACHINE
LEARNING

2
47
MÉTODOS E
ALGORITMOS

3
83
MACHINE
LEARNING

4
113
DEEP
LEARNING

5
147
APROFUNDANDO
EM MODELOS,
MÉTODOS E
ALGORITMOS

Fundamentos de *Machine Learning*

Dra. Iara Carnevale de Almeida

OPORTUNIDADES DE APRENDIZAGEM

O objetivo desta unidade é apresentar os conceitos básicos de *Machine Learning*, (ML, em português, Aprendizado de Máquina) e *Deep Learning* (DL, em português, Aprendizado Profundo), levando em consideração aspectos do nosso cotidiano e procurando traçar um paralelo com *Artificial Intelligence* (em português, Inteligência Artificial - IA), por meio da teoria e da prática.



Olá! Seja bem-vindo(a)! Você já pensou o quanto o *Machine Learning* e o *Deep Learning* estão presentes em nossa rotina? Certamente, você já teve contato com alguma aplicação ou conhece alguma empresa que utiliza técnicas de aprendizagem por meio de aplicações práticas, executadas com sucesso. Os nomes que vieram à sua cabeça foram o Google ou, talvez, o Facebook? Em quais aplicações você pensou?

Você já pensou como é que o programa de vídeo game de, por exemplo, um jogo de futebol, usa tática defensiva para evitar que você marque os gols sempre da mesma maneira? E o reconhecimento facial, como será que a máquina entende e reconhece os rostos através das câmeras? E você já pensou como os carros autônomos conseguem desviar dos obstáculos, manter a velocidade e frear quando necessário?

Nesta unidade, discutiremos e coneceremos os conceitos e as principais aplicações que envolvem Inteligência Artificial baseada em aprendizado de máquina ou aprendizado profundo. Ambos estão presentes em várias áreas das nossas vidas! Certamente, você já se perguntou como o seu smartphone sabe as últimas pesquisas que você realizou em seu computador, pois, depois, sugere, justamente, os temas pesquisados, quando você acessa as suas redes sociais. As aplicações que trabalham com IA — cujo principal objetivo é simular a capacidade humana de “pensar” para analisar determinada resposta para um desafio — é formada por uma rede de técnicas e conceitos, tais como: algoritmos, sistemas de aprendizado e redes neurais.

Quais aplicações você conhece que trabalham com IA? Você consegue indicar quais dessas aplicações são baseadas em aprendizado? É muito importante compreender que a IA é muito mais abrangente que o ML e DL. Ambas são subáreas da IA. Reconhecemos o ML quando lidamos com um aplicativo que processa grande volume de dados (em inglês, *Big Data*), constrói o seu aprendizado com apoio de especialistas humanos ou, sem intervenção humana, consegue responder a questionamentos ou tirar conclusões sobre novos dados.

Será, porém, que somente essa diferença basta para entendermos essa relação? Não! Uma das características que expressam a diferença é que nem toda IA tem, necessariamente, aspectos ou características humanas, como vemos na maioria dos objetos com IA que fazem parte do nosso dia a dia. Entretanto não podemos esquecer que um humano nunca poderá calcular ou computar dados tanto na quantidade quanto na rapidez que um computador faz.

O que realmente representa a diferença entre ML ou DL com Inteligência Artificial é a capacidade de “aprender”, pois nem toda IA tem essa capacidade. Podemos conceituar a IA como algo mais amplo, que inclui outras tecnologias, além do aprendizado, tais como: processamento de linguagem natural, algoritmos de inferência e redes neurais, sempre com o objetivo de atuar e raciocinar como os humanos. Em outras palavras:

Nem toda inteligência artificial envolve aprendizado, mas toda tecnologia que envolve ML ou DL é um sistema de IA.

Por meio da simulação de redes neurais muito semelhantes às ligações do cérebro humano e trabalhando com um número enorme de informações, em um tempo muito pequeno, o ML traz para o computador a capacidade de evoluir

o seu aprendizado, a partir da realização de tarefas e aprender com seus erros, melhorando, assim, a execução da tarefa realizada.

Após essa breve contextualização, peço que você assista ao vídeo sobre veículos autônomos, suas características e benefícios, clicando no **QR Code**, a seguir.

Neste vídeo, são apresentados dois modelos de veículos autônomos que podem vir a fazer parte da nossa rotina, em um futuro não tão distante. Faça um comparativo entre o modelo da Google e o modelo da Tesla e, em seu Diário de Bordo, cite as características, diferenças e semelhanças desses dois modelos na utilização da IA e técnicas de aprendizado. Não se esqueça de citar outras aplicações que utilizam os princípios identificados no vídeo. Fique à vontade para visualizar outros vídeos sobre esse tema!

DIÁRIO DE BORDO

O dicionário Michaelis ([2021], on-line) define **Inteligência Artificial** como “projeto e desenvolvimento de programas de computador que simulam o pensamento humano, capaz de desenvolver um comportamento inteligente”. Por essa definição, podemos relacionar a IA às aplicações que executam atividades com uma certa “inteligência”. Quando as aplicações aprendem “sozinhas”, por meio da análise de grande volume de dados, falamos do ML.



A **Inteligência Artificial (IA - Artificial Intelligence)**, também, é um campo da ciência, cujo propósito é estudar, desenvolver e empregar máquinas para realizarem atividades humanas de maneira autônoma. Também, está ligada à robótica, ao aprendizado, ao reconhecimento de voz e de visão, entre outras tecnologias (O QUE..., 2019, on-line).

Não podemos esquecer que, por ter a habilidade de “aprender”, a IA necessita de sua base de dados sempre alimentada para que, assim, possa continuar e se desenvolver; como nós, para aumentarmos nosso conhecimento, precisamos sempre estudar. Antes de aprofundar o conhecimento sobre *Machine learning* e *Deep learning*, é importante que você compreenda melhor a área de Inteligência Artificial.

Na Figura 1, note que a IA contempla todas as teorias e tecnologias, abrangendo as duas categorias de ML e DL. Além disso, o DL é uma subárea do ML.



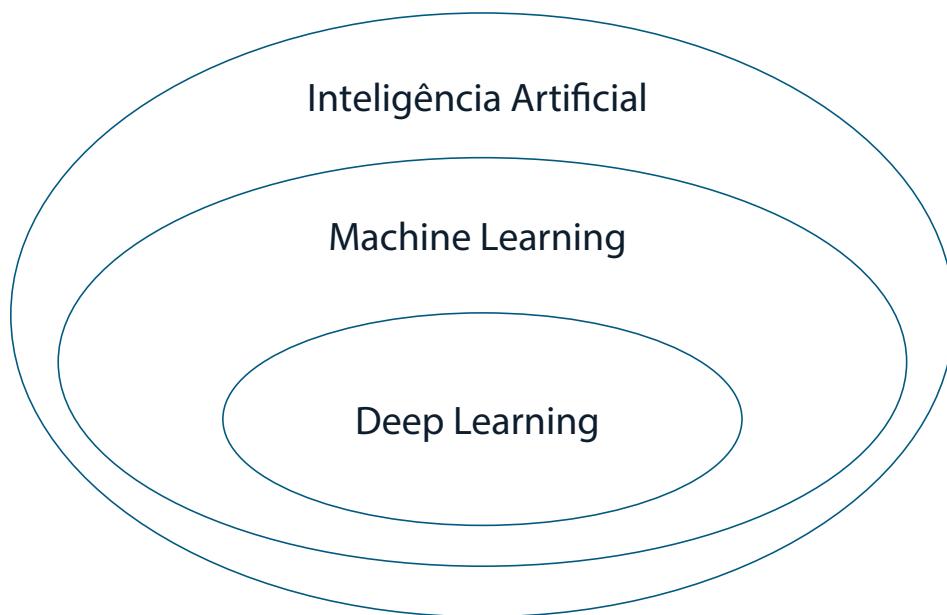


Figura 1 - Visão de alto nível dos principais componentes de IA / Fonte: Taulli (2020, p. 35).

Descrição da Imagem: a imagem apresenta três círculos, um dentro do outro. O menor é nomeado Deep Learning, este está dentro de um círculo maior, nomeado Machine Learning, que, por sua vez, está dentro de um círculo maior, nomeado Inteligência Artificial.



EXPLORANDO IDEIAS

O IBM Watson, desenvolvido por iniciativa da *International Business Machines*, é uma plataforma de inteligência artificial que combina as novas metodologias de aprendizado de máquina com o uso de supercomputadores, sendo utilizada em projetos de diversas áreas. O Watson se tornou muito conhecido, inicialmente, pela capacidade de análise linguística, ao ser o primeiro computador a vencer seus concorrentes humanos em um competitivo programa de perguntas e respostas norte-americano.

Fonte: Silva et al. (2019, p.158).

DENDRAL foi o primeiro programa especialista, baseado em conhecimento, escrito em 1967, que podia predizer as estruturas de compostos químicos desconhecidos, baseado em análises de rotinas. Com a evolução computacional, houve grande avanço na análise computacional e, por consequência, soluções baseadas em IA permitiram, por exemplo, que a máquina pudesse analisar e sintetizar a voz humana. Atualmente, a IA pode ser categorizada em:

- **Inteligência Artificial Estreita (IAE):** prevê o processamento restrito de habilidades. Os sistemas executam as tarefas atribuídas perfeitamente, mas não podem fazer qualquer tarefa que não seja atribuída com a mesma perfeição, tais como: jogar xadrez, reconhecer rostos em imagens ou gravações de vídeo e fazer a previsão de preços.
- **Inteligência Artificial Geral (IAG):** também é chamada de IA forte ou IA profunda. A IAG procura imitar a inteligência humana, com a capacidade de pensar, compreender, aprender e aplicar sua inteligência para resolver qualquer problema, assim como os humanos fazem em qualquer situação. Existem várias características que os sistemas IAG devem ter: senso comum, conhecimento prévio, aprendizagem por transferência, abstração e causalidade.

Neste material, partimos do pressuposto que a IA é um ramo da ciência da computação, envolvido na construção de máquinas inteligentes, capazes de realizar tarefas semelhantes às humanas. Salienta-se que a IA não atua somente quando solicitada, mas pode acompanhar, prever necessidades e lembrar de tarefas importantes. Além disso, o propósito da IA não é nos substituir, “a intenção é a de que o homem atue junto à máquina ganhando agilidade e escalabilidade. Assim, podemos oferecer mais qualidade de vida e eficiência para a vida das pessoas” (PIMENTA, 2021, on-line)

Russel e Norvig (2013) apresentam a IA em quatro dimensões distintas — pensar como humano, pensar racionalmente, agir como humano, agir racionalmente — e, quando relacionadas, tornam-se oito dimensões. Os mesmos autores afirmam que o Aprendizado de Máquina busca pela construção de algoritmos que, a partir de um certo conjunto de dados, podem aprender com as suas falhas e, possivelmente, prever novos dados. Cada vez mais, aplicações com a MA são utilizadas.

Em 1959, Arthur Samuel trabalhava na IBM e definiu o ML como um campo de estudo que dá a um computador a capacidade de aprender sem ser, explicitamente, programado para isso. Para demonstrar esta definição, Arthur Samuel (1959) teve

a seguinte ideia: como ele adorava jogar dama, decidiu escrever um programa para que o computador pudesse jogar contra ele. Notou, porém, que ele sempre ganhava da máquina; então, o que fez foi alterar o programa para que o computador pudesse aprender estratégias por meio das jogadas anteriores. Jogou, novamente, e perdeu, jogou mais uma e perdeu novamente, passando a perder em todas as outras jogadas. Surge, assim, o Aprendizado de Máquina!

Inicialmente, ML é definido por Mitchell (1997) como a capacidade de melhorar o desempenho na realização de alguma tarefa, por meio da experiência. Mais recentemente, ML é visto como a ciência (e a arte) de programar computadores de forma a aprender a partir de dados (GÉRON, 2017). Portanto, o ML pode ser definido como um sistema que altera o seu comportamento de forma autónoma, tendo como base a sua própria experiência. Esta alteração comportamental se faz a partir do estabelecimento de regras lógicas, cuja finalidade é melhorar o desempenho de uma determinada tarefa ou tomar a decisão mais apropriada para o momento, isso depende do tipo de aplicação.

Todo sistema baseado em ML é capaz de aprender, alterar ou executar atividades, tendo como base suas experiências. Note que isto simula o comportamento humano; após realizar várias vezes uma tarefa, nós aprendemos com essa repetição e realizamos alterações, se necessário for. Podemos, então, dizer que o uso de algoritmos para manipular dados é o ponto central do ML. Para ter sucesso, o ML deve utilizar um algoritmo apropriado para obter o resultado pretendido, assim como os dados devem servir para uma posterior análise.

Já dissemos que o ML trabalha com máquinas; estas são programadas para aprender com a experiência vivida e está relacionada ao aprendizado humano, mas você deve se perguntar “Como isso funciona?”. Funciona por meio de um princípio de inferência, chamado **indução**, que significa aprender um modelo ou padrão geral a partir de exemplos. Assim, obtém-se conclusões genéricas, a partir de um conjunto particular de exemplos, ou seja, algoritmos de ML aprendem a induzir uma hipótese ou função que possa resolver um problema, utilizando-se dos dados que representam partes do problema a ser solucionado, os quais são denominados conjunto de dados.

Cabe, aqui, citar que existe uma relação entre ML e mineração de dados (processo que realiza a extração automática de conhecimento, utilizando grandes bases de dados), que é a forma como o ML trabalha, sem nos esquecer que os algoritmos do ML, também, podem aprender a partir de interações com o ambiente ou com um simulador.



NOVAS DESCOBERTAS

Sugiro que você, aluno(a), assista ao vídeo “Dedução e Indução”, pois este traz uma visão didática da diferença entre esses termos.

Por que devemos estudar sobre aprendizado? Não é apenas para trabalhar com uma grande base de dados (principalmente, quando ocorre Armazenamento em Nuvem) e de recursos computacionais disponíveis. O ML pretende auxiliar o entendimento do aprendizado em humanos, por meio do desenvolvimento de sistemas complexos, que dificilmente podem ser construídos “manualmente” (isto é, apenas com intervenção humana), devido à necessidade de se ter conhecimento detalhado para a realização de tarefas específicas, tais como: extrair conhecimento de grandes bases de dados.



EXPLORANDO IDEIAS

Armazenamento em nuvem é a tecnologia que permite aos usuários e empresas armazenar, manter e acessar dados em servidores de alta disponibilidade, via internet. Geralmente, apresentado como um serviço, com ele qualquer computador ou dispositivo móvel pode enviar ou acessar informações, via internet.

É importante, também, sabermos **o que é e quais são os tipos de aprendizagem**. Para Rich e Knight (1991 *apud* CARVALHO, 2015, p. 64), a “aprendizagem é o processo pelo qual uma entidade adquire conhecimento”. Ginsberg (1993 *apud* CARVALHO, 2015, p. 64) define como “uma forma de inferência que objetiva começar com informações sobre o domínio e, então, estendê-las de alguma forma”. Já para Russel e Norvig (1994 *apud* CARVALHO, 2015, p. 64), “aprender significa se comportar melhor ao adquirir experiência. Todo aprendizado aprende a representação de uma função”. A aquisição de novos conhecimentos, o desenvolvimento de técnicas motoras e cognitivas, por meio de instrução ou prática, a generalização de conhecimentos adquiridos, a representação do conhecimento e a descoberta de novos fatos/teorias são alguns dos processos da aprendizagem.

Visto que vimos algumas definições e conceitos do *Machine Learning*, vamos, então, conhecer quais são os tipos e aplicações de aprendizagem no ML.

- **Aprendizagem Supervisionada:** trabalha com o objetivo de encontrar ótimos parâmetros que possam trabalhar com um modelo de conjunto de testes de rótulos desconhecidos. Dependendo do conjunto de rótulos, realiza-se classificação ou regressão. Se o rótulo é oriundo de um conjunto finito e não ordenado, chama-se **classificação**, realizando tarefas de, por exemplo, separar os e-mails em spam ou não spam ou, por meio da análise de imagens, indicar se um tumor é benigno ou maligno. Caso o rótulo seja um número real, realiza-se **regressão**, no qual se pode, por exemplo, estimar um valor para uma residência por meio de variáveis independentes, tais como número de quartos, banheiros, metragem, quantas vagas na garagem. Neste caso, a variável dependente é o valor do imóvel. A figura, a seguir, demonstra que existe uma relação entre o tamanho da casa e o seu valor.

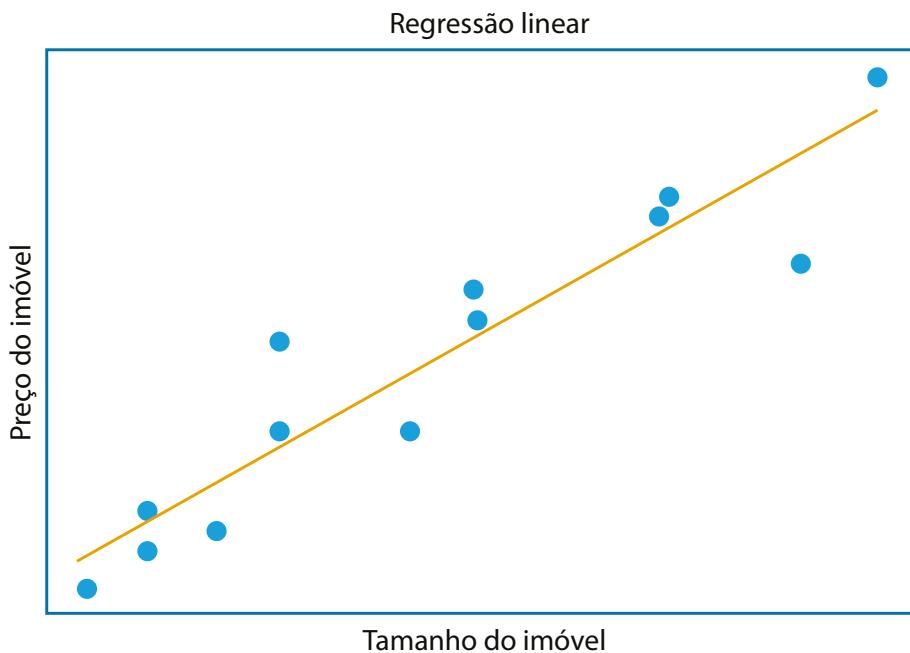


Figura 2 - Exemplo de regressão / Fonte: Budkewicz (2018, on-line).

Descrição da Imagem: a figura retrata o esquema de regressão, em que temos uma linha na diagonal que apresenta a relação entre o preço do imóvel e o seu tamanho, indicando que, quanto maior o tamanho do imóvel, maior será o seu valor.

A aprendizagem supervisionada ocorre em situações em que é possível perceber as entradas e as saídas, frequentemente, fornecidas por um supervisor (especialista) humano, envolvendo a aprendizagem de uma função, a partir dos exemplos obtidos pelas entradas e saídas. Para tal, dois conceitos são importantes: bias e variância. A complexidade do modelo aumenta, conforme ele se ajusta aos dados, essa é a definição de **bias**, o que faz com que ocorra a perda da sua capacidade de generalização, aumentando a sua **variância**, ou seja, a variabilidade das previsões do modelo. Para que possamos ter um equilíbrio entre o bias e a variância — ao que denominamos de **trade-off** —, devemos configurar os hiperparâmetros e, assim, equilibrar os níveis de assertividade e precisão da aprendizagem supervisionada.

Além disso, há diferenças entre os modelos de regressão linear e não linear. No modelo de regressão linear, a complexidade deve ser ajustada para que o bias e a variância tenham o menor valor possível; no modelo não linear, o equilíbrio é obtido conforme o bias for menor e a variância maior. Quando o ajuste ocorre de forma equilibrada, podem ocorrer: **underfitting** (subajuste), quando ocorre falta de capacidade do modelo, na representação dos dados; ou **overfitting** (superajuste), quando se ajusta muito os dados e se perde a capacidade de generalização, fazendo com que o erro seja muito alto, quando forem apresentadas as novas amostras. A seguir, um exemplo de um sistema para aumentar as vendas por meio de sugestões de marcas. A Figura 3 apresenta os três ajustes do modelo de dados e demonstra a diferença entre os modelos underfitting, overfitting e ideal.

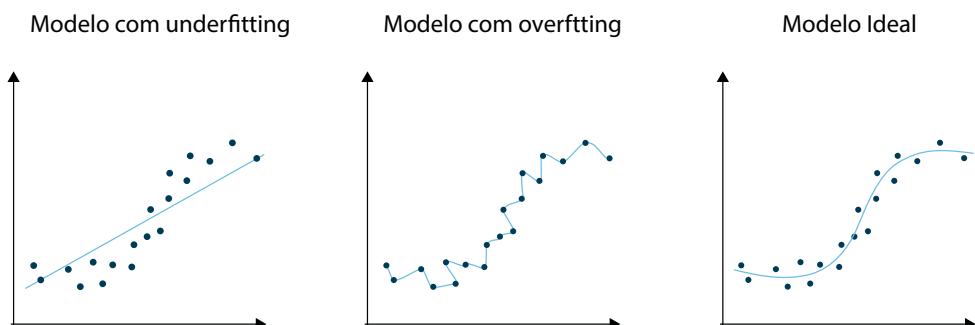


Figura 3 - Diferentes ajustes do modelo aos dados (sugestão de modelo)
Fonte: adaptada de Almeida, Carvalho e Menino ([2021], on-line).

Descrição da Imagem: na figura, há três gráficos, todos com duas linhas apontando o crescimento para cima e para a direita, representando diferentes ajustes do modelo de dados, sendo eles o Modelo Underfitting, Overfitting e o Modelo ideal. O primeiro é representado por uma linha reta, que cresce para cima e para a direita, o segundo traz uma linha irregular e o terceiro, uma linha curva que se direciona para cima e para a direita do gráfico.

Na literatura, fala-se também em **aprendizagem semissupervisionada**, um modelo que trabalha com o recebimento de uma grande quantidade de dados não identificados e uma pequena quantidade de dados identificados (considerados como a resposta certa). Diferencia-se da aprendizagem supervisionada, por ser mais econômica, devido ao menor custo dos dados recebidos.

- **Aprendizagem não supervisionada:** envolve a aprendizagem de padrões, conforme os valores de entrada, quando não são fornecidos os valores específicos de saída. Nesse modelo, o objetivo é analisar as semelhanças entre os objetos e agrupá-los. Ou seja, o **agrupamento** ocorre quando o modelo recebe um conjunto de dados não rotulados e tenta agrupá-los, mediante alguma característica semelhante. Salienta-se que se pode reduzir o número de variáveis, aplicando técnica para **redução de dimensionalidade**. Um exemplo desta técnica são as plataformas de streaming, que utilizam o agrupamento em seus sistemas de recomendações. Um exemplo para redução de dimensionalidade é a detecção de bordas no processamento de imagens digitais. Uma das aplicações do aprendizado não supervisionado é a detecção de transações fraudulentas, por meio da identificação de anomalias, em um grande conjunto de dados. Outro exemplo são os sistemas que apoiam a distribuição de produtos, de acordo com o comportamento de aquisições, para, assim, melhorar suas vendas.



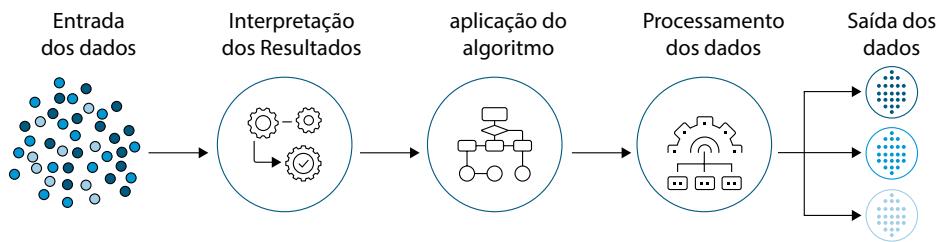


Figura 4 - Fluxo de execução do aprendizado não supervisionado (sugestão de modelo)

Fonte: adaptada de Almeida, Carvalho e Menino ([2021], on-line).

Descrição da Imagem: figura dividida em cinco partes, representando Fluxo de execução do aprendizado não supervisionado, sendo: a Entrada dos dados, representada por várias bolinhas de diferentes cores; a Interpretação dos resultados, a Aplicação do algoritmo, o processamento de dados e, por fim, a saída dos dados, em que temos a representação de três blocos de bolinhas coloridas, mas agora divididas por cores.

- **Aprendizagem por Reforço:** o agente deve aprender, a partir do reforço (ou seja, uma recompensa). É diferente dos tipos anteriores, pois não há um conjunto de treinamento, rotulado ou não. Este tipo de aprendizagem procura detectar como devem ser as ações de agentes de softwares em ambientes predeterminados. Esta aprendizagem realiza a interação entre dois elementos — **ambiente e agente de aprendizagem** —, em que o **ambiente** é o local onde o **agente de aprendizagem** interage ao tomar suas decisões. O agente de aprendizagem utiliza dois mecanismos de exploração: (1) **exploration**, quando o agente trabalha na tentativa e erro; e (2) **exploitation**, age conforme o conhecimento adquirido no meio ambiente ao qual o agente está inserido. Quando o agente de aprendizagem toma uma decisão: se as ações estão corretas, o ambiente **recompensa** este agente; caso contrário, aplica uma **penalidade**, sinalizando que as ações são negativas.

A recompensa, portanto, pode ser vista como uma ação positiva que reforça as ações do agente, permitindo que esse agente possa aprimorar o seu conhecimento para a seleção da próxima ação; uma penalidade do ambiente faz com o agente esqueça (ou seja, descarte) as ações realizadas. No início, o agente não possui nenhuma informação sobre o ambiente e obtém, aos poucos, o conhecimento por meio de suas experiências neste ambiente. Para tal, considera-se que o agente possui **estados**, que são as condições atuais do agente e do ambiente.

Desse modo, evoluções ocorrem nos seus **estados**, e todo estado é atualizado pelas recompensas e penalidades. Aprendizagem por reforço é utilizada em jogos. No jogo de xadrez, por exemplo, o **ambiente** é o tabuleiro; um **estado do agente** são as posições das peças em uma determinada jogada; a **ação** é o movimento executado em uma jogada; a **recompensa** é o eliminar uma peça do adversário; e a **penalidade** é a perda de uma peça para o adversário na jogada seguinte.

Outro exemplo de aplicação de aprendizagem por reforço é a condução automatizada em veículos autônomos, no qual o **ambiente** é o local físico onde o carro está; o **estado** é obtido pela percepção e localização de obstáculos capturados pelos sensores; as **ações** são os comandos de direção, freio e aceleração; a **recompensa** é a aproximação do destino; e a **penalidade** é a colisão com algum obstáculo. Outro exemplo, ainda, que pode ser citado, são os sistemas utilizados para treinar robôs de linhas de fabricação a fim de identificarem objetos para, por exemplo, separá-los. A figura seguinte apresenta a iteração entre os elementos anteriormente conceituados e exemplificados.

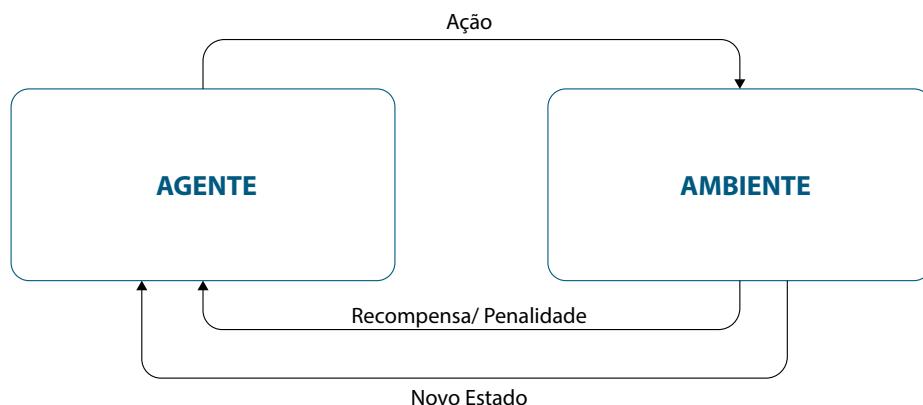


Figura 5 - Fluxo clássico do aprendizado por reforço (sugestão de modelo)
Fonte: adaptada de Almeida, Carvalho e Menino ([2021], on-line).

Descrição da Imagem: figura com dois quadrados, um nomeado “agente” e outro nomeado “ambiente” representam o fluxo clássico do aprendizado por reforço, em que o agente é ligado ao ambiente por uma seta, indicando ação, e o ambiente se liga ao agente por duas setas, indicando recompensa ou penalidade e novo estado.

Além dos tipos de aprendizagem apresentados, podemos, ainda, ter a seguinte classificação: (1) **aprendizagem por descoberta** é uma forma restrita de aprendizado, no qual o conhecimento é adquirido **sem** o auxílio de quem já possua

este conhecimento (por exemplo, um especialista, tutor, entre outros); (2) **aprendizagem dedutiva** ocorre quando as novas informações são consequências válidas de uma ação, processo ou atividade já conhecida; e (3) **aprendizagem indutiva** ocorre pela inferência indutiva de fatos fornecidos por um especialista ou pelo ambiente, em que, por exemplo, o programa aprende a classificar objetos baseados em rótulos que lhe são apresentados.

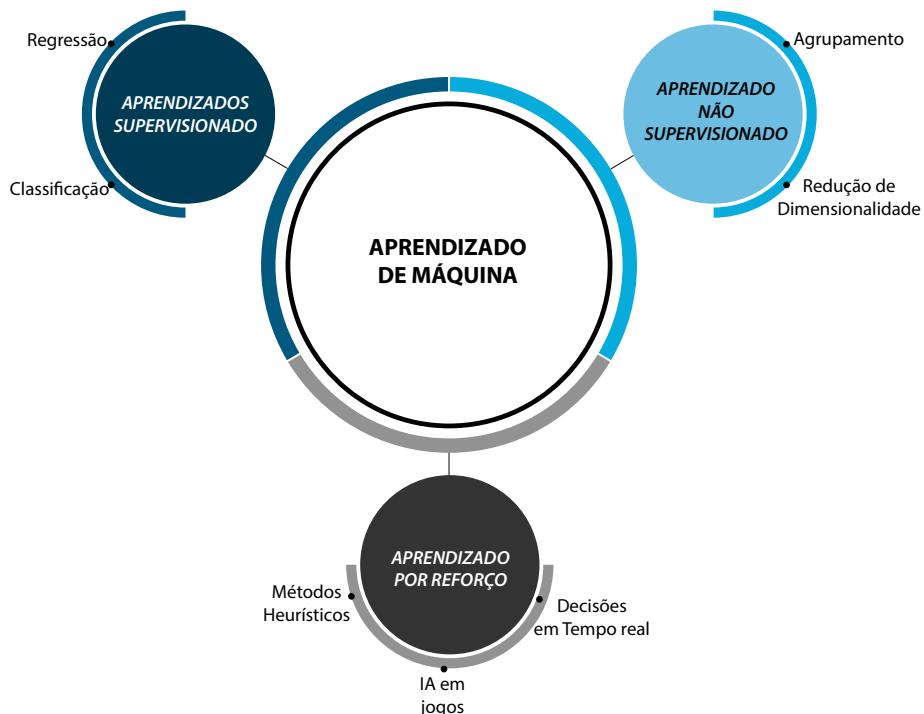


Figura 6 - Tipos de aprendizagem em ML (sugestão de modelo)
Fonte: adaptada de Almeida, Carvalho e Menino ([2021], on-line).

Descrição da Imagem: figura do diagrama dos tipos de aprendizagem em ML. Tem-se um círculo maior ao meio, representando o “aprendizado de máquina”, e, ao seu redor, outros três círculos: “aprendizado supervisionado” (regressão e classificação), o “aprendizado não supervisionado” (agrupamento e redução de dimensionalidade) e o “aprendizado por esforço” (métodos heurísticos, IA em jogos e decisões em tempo real).

Agora que você compreendeu as categorias de aprendizagem em ML, note que, no aprendizado supervisionado, existe a correspondência entre as variáveis de entrada e saída; já no aprendizado não supervisionado, fica evidente a técnica de ML, na qual não precisamos supervisionar o modelo, pois este descobrirá as informações. O Quadro 1 procura apresentar uma síntese dos tipos de aprendizagem.

Tipos de Aprendizagem - ML		
Supervisionada	Não supervisionada	Reforço
<ul style="list-style-type: none"> - Classificação / Regressão. Ex.: verificar se determinada imagem é de um ou de outro personagem (imagem do Homer e do Bart Simpson). 	<ul style="list-style-type: none"> - Associação / Agrupamento. - Analisar automaticamente os dados. - Necessita análise para determinar o significado dos padrões encontrados. Ex.: comparação de padrões de figuras geométricas. 	<ul style="list-style-type: none"> - Aprender com as interações com o ambiente (causa e efeito). - Aprender com sua própria experiência. Ex.: robô coletando lixo e aprendendo a andar em um ambiente ou controle automatizado de elevadores.

Quadro 1 - Tipos de aprendizagem de máquina / Fonte: a autora.


NOVAS DESCOBERTAS


Assista ao vídeo que trata da diferença entre aprendizado supervisionado e aprendizado não supervisionado. Neste vídeo, são apresentados exemplos práticos de aprendizagem supervisionada e aprendizagem não supervisionada. Procure refletir sobre semelhanças e diferenças entre estes dois tipos e, preferencialmente, pensar em novos exemplos.

Com a finalidade de evitar discussões filosóficas sobre o ML, Alan Turing propôs, em 1950, a criação de um teste que consistia na entrevista de um agente remoto para decidir se este era uma pessoa ou um programa de computador (TURING, 1950). Este mesmo autor assume que o ser humano é inteligente, mesmo sem uma definição precisa de inteligência, se um programa baseado em ML conseguisse responder como um ser humano, este programa tinha algum tipo de inteligência, mesmo sendo artificial. Para tal, foi proposto o Teste de Turing.

O **Teste de Turing** tem a seguinte premissa básica: se um ser humano conversa com uma máquina por cinco minutos, sem perceber que a máquina não é humana, ela passa no teste. O teste deve ser aplicado com duas entidades ocultas — um ser humano e um computador — que devem responder a questões feitas por um interrogador (humano) por meio de um teclado. Na sequência, baseado

nas respostas, este interrogador tem que decidir qual dos dois é o humano. Pretende-se que o respondente (ser humano ou computador) responda às questões, tais como: “têm algum hobby?” ou “como configuramos o tabuleiro de um jogo de damas?”.

O programa do computador, portanto, deve ser desenvolvido para responder como um ser humano. Ao final do teste, se o interrogador não distinguir quem é o humano, a conclusão é de que, segundo o Teste de Turing, o computador tem a capacidade de “pensar”. Claro que, para que este teste tenha sucesso, o programa não pode ser “perfeito”, por exemplo, na fala, e, por consequência, alguns erros na linguagem devem ser criados; outro exemplo são as questões matemáticas, pois o computador tem maior capacidade de processamento do que os seres humanos.

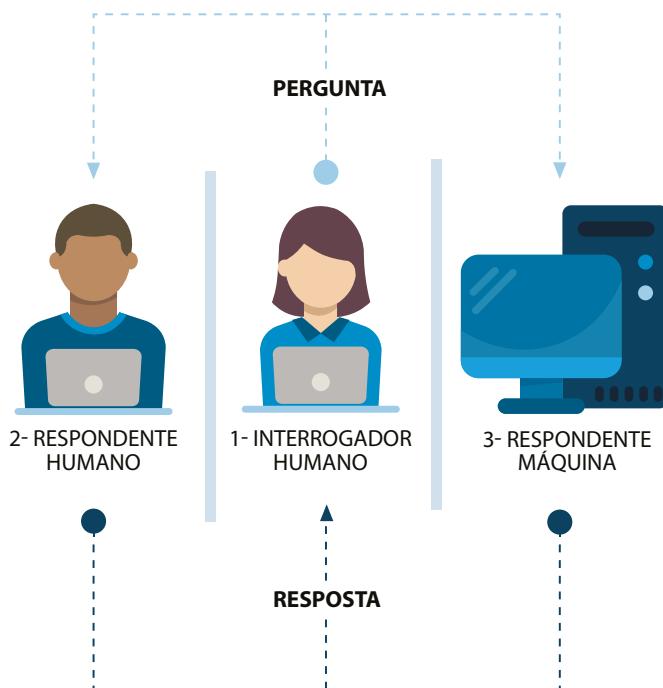
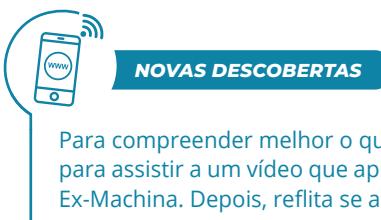


Figura 7 - Modelo para figura do esquema do teste de Turing clássico (gostaria que ela fosse o modelo)
Fonte: adaptada de Almeida, Carvalho e Menino ([2021], on-line).

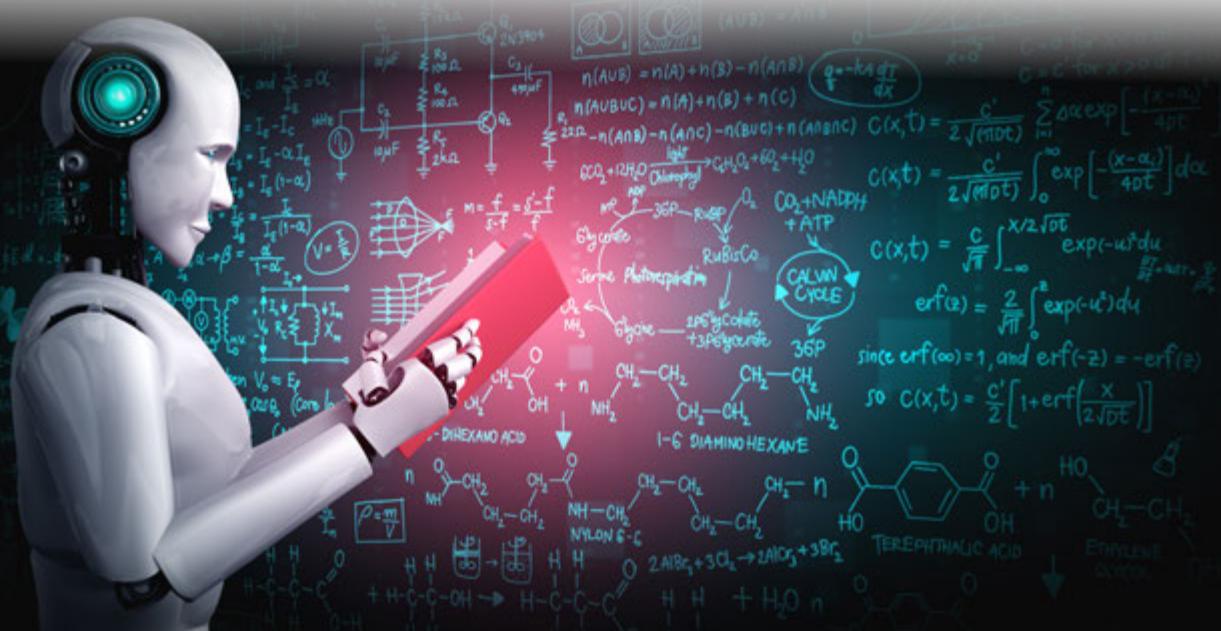
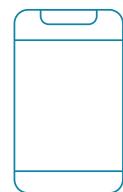
Descrição da Imagem: à esquerda da figura, temos a ilustração de um homem sentado à frente de computador (Respondente Humano); uma mulher, ao centro, sentada à frente de computador (Interrogador Humano), acima dela há o texto “Pergunta” que direciona uma seta para sua esquerda e outra para sua direita; abaixo deste mesmo personagem há uma seta de retorno com o texto “Resposta” vinda dos objetos em suas laterais; à direita temos um computador (respondente máquina).

Só porque o sistema consegue buscar uma resposta que parece compreensível, não quer dizer que ele a entenda de verdade. O interessante é que muitos dos sistemas que conseguiram convencer as pessoas de que eram humanos fizeram isso não por conversarem de maneira realista, mas por terem incluído erros de digitação ou de gramática. Note que as habilidades necessárias para que pareçam humanos e, por consequência, passem no teste de Turing, não estão necessariamente ligadas à “inteligência” ou ao “pensamento”.

Essa proposta de teste leva muitos cientistas e engenheiros a pensar o que nos faz ser “verdadeiramente humanos”. Também tem inspirado equipes de design de IA a criar sistemas que interagem de maneira mais natural e humana. Importante indicar que, em 2012, em uma competição em honra ao centenário de Alan Turing, uma máquina conseguiu o feito de ser aprovada no teste de Turing!



Para compreender melhor o que você acabou de ler, sugiro que clique no QR para assistir a um vídeo que apresenta uma explicação de um trecho do filme Ex-Machina. Depois, reflita se a máquina pode ter, ou não ter, consciência!



Na sequência, a Figura 8 apresenta uma linha de tempo que representa a evolução do Aprendizado de Máquina.

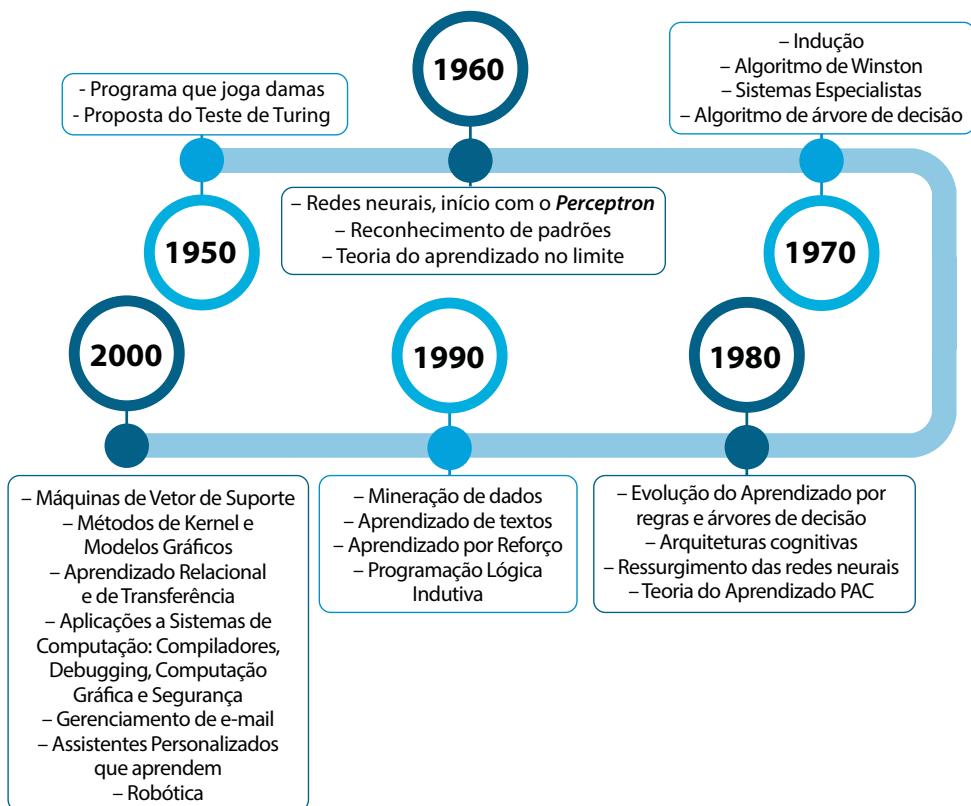


Figura 8 - Linha do tempo da Machine Learning / Fonte: a autora.

Descrição da Imagem: na figura, uma linha do tempo: anos de 1950: programa que jogava damas de Arthur Samuel, teste de Turing; anos de 1960: Redes neurais: início com o Perceptron, primeiros reconhecimentos de padrões, surge a Teoria do aprendizado no limite; anos de 1970: primeiros conceitos de Indução, algoritmo de Winston, sistemas especialistas, surge o Algoritmo de árvore de decisão; anos de 1980: Tem início a Mineração de dados, aprendizado de textos, surge o Aprendizado por reforço (RL), programação lógica indutiva (ILP); anos 2000: máquinas de Vetor de Suporte, métodos de Kernel e Modelos Gráficos, aprendizado Relacional e de Transferência, aplicações a Sistemas de Computação (compiladores, debugging, computação Gráfica, segurança), gerenciamento de e-mail, assistentes personalizados que aprendem, robótica.

Cada vez mais, as empresas têm prestado maior atenção ao ML, não só pelo aumento no poder de processamento, pelo crescente volume e variedade dos dados disponíveis, e pela facilidade de armazenamento destes dados mas também devido às vantagens que a ML pode adicionar aos negócios. Alguns exemplos de aplicação são:

- **Análise de dados** traz resultados relevantes para as empresas, no que diz respeito às previsões, apoio para tomada de decisões, detectar oportunidades e/ou mudanças nos diferentes setores destas empresas, possibilitando, assim, uma vantagem competitiva.
- **Recomendação de produtos** é um exemplo desta vantagem nos negócios, pois, por meio da análise dos dados de, por exemplo, hábitos de consumo de clientes de empresas de *E-commerce*, pode-se ter um serviço de recomendação de produtos que considere tanto o perfil quanto os desejos desses clientes, permitindo realizar marketing direcionado de produtos.
- **Análise de sentimentos**, o ML auxilia as empresas na detecção de pessoas que possam estar apresentando sentimentos bons (ou ruins) sobre suas marcas e/ou produtos, tendo por base os comentários nas redes sociais ou em e-mails.

Note que, devido ao fato do ML otimizar a análise de dados de forma mais assertiva e rápida, faz com que as empresas assumam o ML como uma solução tecnológica atual para os desafios destas empresas. Reforçamos que o ML ajuda empresas que buscam vantagem competitiva no mercado, pois permite que essas obtenham soluções rápidas e práticas, quando há grande volume de dados complexos que devem ser processados.



Dado que você aprendeu diversos conceitos relacionados ao Aprendizado de Máquina, que tal uma reflexão? Procure escrever sobre uma área de atuação (educação, comércio ou indústria) que você acha que o ML pode ser empregado e, por consequência, melhorar a eficiência e a eficácia nesta área.

Procure detectar, no seu dia a dia, o quanto o ML está presente em recursos virtuais, aplicativos, assistentes digitais, entre outros no mundo digital. Você consegue citar qual será a maior relevância do ML no mundo empresarial, nos próximos anos?

Uma das áreas que o ML tem mais auxiliado para que as empresas se destaquem no mercado é a **visualização de dados**. Esse auxílio vem por meio de ações que fazem com que a empresa entenda melhor o fluxo desses dados, de acordo com uma das suas principais ações, que é a de encontrar padrões em dados de uma forma

automática. Diminuir o tempo que seria necessário para a realização da análise dos mesmos, além de apresentar os dados mais relevantes e precisos, ajudando, assim, na tomada de decisões, gerando uma vantagem considerável.

Outra tarefa que pode ser simplificada pelo ML é o **gerenciamento e análise de conteúdo**, em que o ML pode ser utilizado para substituir os seres humanos em tarefas repetitivas e que exigem constante atenção para fazer uma identificação visual de, por exemplo, pessoas, logotipos, símbolos ou produtos. Imagine como seria mais fácil produzir um conteúdo na internet, visando ao marketing da sua marca. Se você souber a quantidade de consultas feitas sobre a sua marca ou seus produtos, saberá qual é o conteúdo mais visualizado e, por consequência, qual dos produtos possibilita maior número de vendas. Poderá, também, contabilizar quantas vezes a sua marca aparece em fotos ou postagens de eventos que você patrocina.

Atualmente, o ML tem contribuído para **detecção e prevenção de fraudes**, as empresas perdem uma parte significativa de suas receitas por causa de fraudes e falhas no sistema, no que se refere à proteção, segurança e integridade de dados. Você nunca recebeu a ligação da empresa do seu cartão de crédito, procurando confirmar uma compra? Se sim, provavelmente, esta compra se diferenciava das suas compras habituais. A criação de modelos de comportamento ideais, em tempo real e que analisa as transações anteriores (isto é, histórico, informações pessoais e dados que transitam nas redes do seu cliente), permite levantar e apontar anomalias que diferem do comportamento habitual deste cliente. Os grandes bancos utilizam o ML dessa maneira, com algoritmos que podem apontar e avaliar padrões de ações suspeitas em transações que, muitas vezes, os clientes não têm o hábito de realizar. Esta é, portanto, uma das grandes tendências que permite o aumento da confiabilidade e usabilidade em transações financeiras.

Referente à **customização de serviços** por meio da **análise de comportamento dos consumidores**, o ML auxilia a customizar serviços de acordo com as necessidades de cada cliente. Necessidades essas detectadas pela interação do cliente com os serviços. Isto permite a redução de custos e tempo neste processo de detecção. A capacidade do ML em analisar e aprender com os dados analisados faz com que o ML, utilizando-se dos seus algoritmos, trabalhe de forma mais eficiente e eficaz do que os humanos, em que o objetivo maior das empresas é aumentar suas vendas. Não podemos esquecer de citar, aqui, a utilização crescente de **chatbots** por empresas dos diversos setores — comercial, educacional e in-

dustrial —, que visam a se destacar no mercado, no que se refere ao atendimento ininterrupto aos seus clientes (isto é, 24h por dia e 7 dias por semana).

O ML pode auxiliar na diminuição de desafios da atual “vida moderna”, tais como: trânsito, correria e acidentes. Para tal, o ML realiza **previsões de padrões de tráfego** de forma a direcionar as pessoas para locais onde o tráfego seja menor em determinados horários, permitindo, assim, que haja uma economia de tempo e custo, além de evitar atrasos e longas esperas. Isto ocorre muito, devido à escolha e experiência das pessoas, pois permite que a máquina aprimore as suas soluções. Além disso, o ML também permite **otimizar agendamentos de serviços específicos** por meio de previsão da duração desses, estimada sobre trabalhos já realizados, e, assim, melhorar a produtividade dos serviços.

Será que levo o guarda-chuva? Será que visto uma roupa mais quente? Quantos de nós tiveram essas dúvidas? O ML tem atuado para melhorar o serviço de **previsão do tempo**. A maioria dos institutos de meteorologia já trabalham, ou desenvolvem modelos de previsão de tempo, que permitem a previsão do clima, tendo por base as informações históricas e os fatores sazonais. O mercado do agronegócio procura fazer uso desses modelos para previsões climáticas, o ML permite agilizar o processo de avaliação e identificação de quando certas atividades, tais como plantio ou colheita, podem ou não ser realizadas, evitando perdas, desperdício e, por consequência, custo acrescido.

O ML está presente na construção civil, mais especificamente na área de **mantenção preditiva**, monitorando o estado dos equipamentos de forma a prevenir problemas. Para tal, utiliza-se de dados coletados pela Internet das Coisas (IoT). Portanto, sem a intervenção humana, o ML pode antecipar a necessidade de reparos por meio de uma programação proativa. Com isso, o ML permite que as empresas evitem falhas que podem gerar altos custos e/ou interrupções de serviços que podem causar a perda de um cliente e/ou prejuízo financeiro.

Na área da saúde, as **operadoras de saúde** buscam o ML para auxiliar especialistas e profissionais da saúde a analisar dados, buscando o aperfeiçoamento em tratamentos e diagnósticos, para identificar possibilidades e tendências. Outra facilidade, também, facultada a *machine learning*, para essa área, são os dispositivos que liberam o acesso, em tempo real, dos médicos aos dados dos seus pacientes. Agilizar e otimizar a rede de distribuição de **petróleo e gás**, visando à eficiência e economia, além de análises minerais nos solos, auxiliar nas desco-

bertas de novas fontes de energia e atuar na prevenção de falhas nos sensores existentes nas refinarias, entre outras aplicações.

Uma das áreas que mais tem evoluído, fazendo uso da tecnologia do ML, é o **E-learning** (em português, Educação Online), a partir da utilização de algoritmos intuitivos e a automatização dos conteúdos *E-learning*. Ao prever resultados, os algoritmos do ML possibilitam, tendo como base os objetivos individuais de aprendizagem e o desempenho do aluno no passado, mais conteúdo personalizado de *E-learning*, ajustando o mapa do curso, de acordo com a análise desses dados; trabalha em cima das dificuldades apresentadas durante o curso; gera sugestões específicas para cada caso e, no que diz respeito às facilidades, indica tarefas para que essas habilidades sejam ainda mais desenvolvidas, determinando, assim, o ritmo de ensino, de acordo com o desempenho individual, e, portanto, mais rápido para uns do que para outros. Uma **melhor alocação de recursos**, também, é benefício da utilização do ML na *E-learning*, uma vez que os alunos passam a receber, de acordo com suas necessidades e objetivos, os recursos exatos para seu sucesso na Instituição.



No meio corporativo, os funcionários receberiam as informações que necessitam de forma personalizada, diminuindo, assim, o tempo de treinamento, sem impactar nas horas trabalhadas. Imagine que, após uma avaliação, realizada durante o processo de matrícula, o ML pode **automatizar o processo de agendamento e entrega de conteúdo**, personalizando todo o percurso do curso ou ajustá-lo, conforme novas demandas surgirem. Permitir ao aluno que cumpra o curso de acordo com o seu ritmo, de forma personalizada no seu cumprimento, mas comum no resultado final, conclusão: tende a **melhorar a motivação do aluno**, trocando o curso genérico pelo curso individualizado, fazendo com que ele utilize o tempo de aprendizado para cumprir seus objetivos e desenvolver suas habilidades necessárias, atingindo o seu potencial de forma mais definida, em um percurso único e individual.

E não é só isso! Imagine **criar programas de treinamento online mais efetivos**, eficazes, com o ML auxiliando e sugerindo aos alunos mentores que possam trabalhar suas habilidades específicas ou oferecer alternativas de melhoria, caso o aluno não esteja rendendo o esperado, a partir da previsão de resultados que a tecnologia da *machine learning* permite com o uso dos seus algoritmos.

Qual empresa nos dias atuais pode se dar ao luxo de, nos tempos em que as comunicações são em sua maioria em tempo real — aplicativos, redes sociais — de não aproveitar a tecnologia do *machine learning*? Com certeza, a área do **contact center** não é! O ML é uma grande aliada dessa área, trazendo ferramentas com respostas mais rápidas e integradas aos desejos de cada consumidor. Assim, utilizando-se da análise da grande quantidade de dados gerados nessa área, obtidas por meio das várias formas de comunicação que ela realiza diariamente, automatiza-se esse processo e apresenta respostas direcionadas para cada questão da clientela, promovendo, assim, uma melhoria da personalização, a redução no volume de chamadas de retorno para o *contact center* e o aumento no rendimento, por meio de vendas personalizadas.

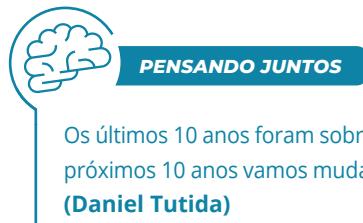
Nem tudo, porém, são flores! Ao se apresentar como uma tecnologia que aprende cada vez mais, ao realizar tarefas que seriam de humanos, o ML também atrai pessoas que querem utilizar essa tecnologia com más intenções. Por trazer a possibilidade de automatização total de diversas atividades, faz-se necessário a utilização de recursos para barrar os maliciosos. Enganar algoritmos de reconhecimento facial, por meio da alimentação com dados

maliciosos, cuja finalidade é a análise errônea, é, portanto, a principal função do **aprendizado de máquina adversário**, ação cada vez mais presente.

O objetivo é simples: fraudar autenticação biométrica por meio da substituição do rosto correto pelo rosto do fraudador, acessando, assim, sua conta e seus dados. Outra força desse tipo de fraude ocorre com a alteração das URLs de sites bancários, facilitado por geralmente o usuário não verificar que o endereço eletrônico fraudulento é bem maior que o normal. Nesse sentido, o ML não é a solução para todos os problemas, e a sua principal desvantagem é como as pessoas o utilizam, não podendo esquecer também do talento e da empatia humana, que, ainda, não podem ser copiados ou aprendidos pelos robôs.

Geralmente, imagens ou gravações de rostos não utilizam criptografia ao serem gravados na nuvem, o que os deixa bastante vulneráveis para algoritmos maliciosos e é, nessa área, que a *machine learning* para **preservação da identidade** pode atuar, uma vez que o reconhecimento facial é uma das formas seguras utilizadas para login, e o ML auxiliaria como uma forma de segurança extra ao criptografar esses dados.

Trabalhar juntos e conectados, agilizando a implementação de novos algoritmos, com eficiência em todas as etapas de desenvolvimento, com compartilhamento real de informações, seleção de algoritmos e preparação de dados é o papel do **aprendizado de máquina fim a fim**, tendo por objetivo reduzir tempo e procedimentos, trazendo uma resposta mais ágil para futuros ataques que porventura possam ocorrer. O aprendizado ativo acontece quando os algoritmos do ML rotulam, automaticamente, a maior parte dos dados, fazendo com que uma pequena parte apenas seja responsabilidade humana, esse tipo é bastante utilizado na área acadêmica e de pesquisa, economizando tempo e recursos.



Os últimos 10 anos foram sobre a construção de um mundo que é móvel em primeiro lugar. Nos próximos 10 anos vamos mudar para um mundo que vai ser movido por Inteligência Artificial.
(Daniel Tutida)

E o futuro? Futuro nada, ML já faz parte do seu dia a dia! Imagine que você pede sempre o mesmo prato ao almoçar em um determinado restaurante e, conforme o tempo passa e você continua visitando esse mesmo restaurante, o garçom passará a lhe oferecer esse prato, antes mesmo da sua escolha. Aposto que você já se perguntou como é que o computador sabe e traz a propaganda de determinado produto que você buscou na rede, em um site que não tem nada a ver com o produto pesquisado! Campanhas de mídia paga mais segmentadas, conteúdo personalizado, *chatbots* e utilização de *voice search* nas buscas são algumas das ações presentes em nosso dia a dia que contam com a tecnologia do ML. Essas são realizadas graças às atividades preditivas que a *machine learning* realiza, apresentando previsões com resultados mais eficientes.



Figura 9 - Interação entre diferentes profissionais para ML / Fonte: Norman (2018, on-line).

Descrição da Imagem: a figura representa a interação entre profissionais da informação (simbolizado com livros), da inteligência (simbolizado por um chip) e da imaginação (simbolizado por uma lâmpada em um balão de imaginação).



**EXPLORANDO IDEIAS**

Veículos autônomos incorporam aprendizado de máquina para melhorar sua segurança e confiabilidade. Um carro autônomo que usa inteligência artificial tradicional pode responder a qualquer condição de estrada que tenha sido programada para lidar. No entanto se o sistema não reconhecer uma entrada, o carro pode ativar uma medida de segurança, tais como: desacelerar, parar ou alertar para que o condutor assuma a condução.

O aprendizado de máquina pode permitir que o veículo reconheça eventos e objetos que não foram explicitamente programados no código fonte do seu sistema. Por exemplo, um carro pode ser programado para reconhecer luzes da rua, mas não luzes intermitentes nas barricadas de uma construção. O Sistema pode aprender, possivelmente, registrando o comportamento do condutor (ser humano), e o carro começará a reconhecer as barreiras de construção e a reagir de forma similar ao condutor.

Ressalta-se que a tecnologia que apoia o sistema de aprendizagem de máquina é o que permite aos veículos autônomos diferenciar os objetos em uma estrada, tais como: carros, motos, seres humanos e animais. Esta tecnologia são sensores que fornecem constante atualização do ambiente, de forma que estes automóveis possam andar com segurança, em condições climáticas difíceis (nevoeiro, chuva intensa, entre outros) ou estradas onde a demarcação do chão não esteja muito visível. O objetivo é permitir que esses veículos dirijam similar aos seres humanos, evitando reproduzir erros humanos (condução imprudente, entre outros).

Sobre o mercado de trabalho, para quem quiser trabalhar com ML, ele é muito recente e com certeza sofrerá, ainda, muitas mudanças e adaptações, mas alguns cargos já podem ser encontrados hoje em dia, como o **Cientista de Dados** (em inglês, *Data Scientist*), que é o mais conhecido e mais antigo e, com a evolução do ML, vem alterando o seu foco para um papel mais atuante na criação de modelos estatísticos com ML e não só na análise descritiva e no desenho de experimentos como era anteriormente. Tem alguma semelhança com o **Engenheiro de Dados** (em inglês, *Data Engineer*). A Figura 10 apresenta estes profissionais, salientando suas habilidades (em inglês, *skills*) e o conhecimento técnico necessário para exercer suas atividades.

DATA ENGINEER	DATA SCIENTIST	DATA ANALYST
 <p>Construir e otimizar a infraestrutura de dados que irá permitir que o Cientista de Dados e o Analista de Dados rodem suas análises.</p> <p>SKILLS Programação, bancos de dados, infraestrutura cloud, big data.</p> <p>TECHS SQL, Python, Cloud (Azure, AWS, Google).</p>	 <p>Utiliza estatística, aprendizado de máquina e algoritmos matemáticos para realizar previsões, identificar padrões e responder perguntas de negócios.</p> <p>SKILLS Programação, bancos de dados, estatística, matemática.</p> <p>TECHS SQL, Python, R., Cloud.</p>	 <p>Compreende questões de negócios, consolida e elabora dashboards e relatórios para comunicar resultados.</p> <p>SKILLS Comunicação, UX, conhecimentos de negócio, banco de dados.</p> <p>TECHS SQL, Power BI, Tableau, Qlik, Vlew, Excel.</p>

Figura 10 - Habilidades e conhecimento técnicos / Fonte: a autora.

Descrição da Imagem: figura constando as habilidades e conhecimentos técnicos do Data Engineer, Data Scientist e Data Analyst. Têm-se três colunas, sendo a primeira relativa ao Data Engineer, com a seguinte descrição: Construir e otimizar a infraestrutura de dados que irá permitir que o Cientista de Dados e o Analista de Dados rodem suas análises. Suas habilidades são programação, bancos de dados, infraestrutura cloud, big data e seus conhecimentos técnicos são SQL, Python, Cloud (Azure, AWS, Google). A segunda coluna apresenta as habilidades do Data Scientist, que utiliza estatística, aprendizado de máquina e algoritmos matemáticos para realizar previsões, identificar padrões e responder perguntas de negócios. Suas habilidades são programação, bancos de dados, estatística, matemática, e seus conhecimentos técnicos são SQL, Python, R., Cloud. A terceira, e última coluna, traz as características do Data Analyst, que comprehende questões de negócios, consolida e elabora dashboards e relatórios para comunicar resultados. Suas habilidades são comunicação, UX, conhecimentos de negócio, banco de dados e seus conhecimentos técnicos são SQL, Power BI, Tableau, Qlik, Vlew, Excel.

É necessário que haja integração entre o ML e os sistemas existentes, para que sua análise seja completa e apresente os resultados esperados, é aí que entra o **engenheiro de ML/MLOps** (MLOps = conjunto de práticas que combina Machine Learning, DevOps e Engenharia de Dados), papel que ganha mais destaque, conforme os algoritmos de ML desenvolvem sua maturidade e ganham automatização.

Não basta você, simplesmente, acordar um dia e falar que, a partir de hoje, utilizará o ML na sua empresa. Esse é um processo que precisa ser estudado, quantificado, e esse é um dos papéis do **analista**, que apesar de não ter uma formação específica, geralmente é oriundo da área das exatas, tais como engenharia, TI ou economia, e que entenda a regra de negócio da empresa em questão.

Não podemos esquecer, também, que ainda existe muito o que se aprender e se discutir sobre o ML e, para isso, temos o **pesquisador de ML**, que fará esse trabalho de pesquisa sobre aplicações e, também, sobre algoritmos que possam surgir, auxiliando na conexão entre desafio e solução.

Agora que já chegamos até aqui, vamos reforçar o conhecimento obtido por meio de uma ação que ocorre, constantemente, no nosso dia a dia: você recebeu uma ligação da sua operadora de cartão de crédito, perguntando se você fez uma determinada compra. A operadora não deseja saber sobre seus gastos ou gostos, mas sim alertar você sobre possíveis gastos que outra pessoa poderia ter feito, utilizando o seu cartão. A IA e o ML em conjunto com o sistema da operadora do cartão detectou um padrão de gasto diferente do usual e a alertou para que você fosse avisado. Siga a linha de raciocínio desta unidade e trace um paralelo entre os conceitos apresentados e os conceitos que você identificou no caso acima. O que você entendeu dessa ação? Quais conceitos podemos encontrar nessa ação que foram desenvolvidos durante essa unidade? Cite outros exemplos em que é possível observarmos aplicações que utilizam a IA e o ML de forma semelhante à apresentada. Lembre-se, ainda, de que temos muito a aprender sobre o ML. Aguardo você na próxima unidade!



Venha absorver mais conhecimento com o nosso podcast!
Dê o play e descubra mais sobre aplicações e técnicas de ML
na agricultura, construção civil e na área da saúde!

AGORA É COM VOCÊ



1. Por diversas vezes, os conceitos de Inteligência Artificial (IA) e Machine Learning (ML) se confundem. De acordo com o que foi visto nessa unidade, conceitue Machine Learning e explique a principal diferença entre a IA e o ML.

2. Ocorre nas situações em que é possível perceber tanto as entradas como as saídas, que são frequentemente fornecidas por um supervisor (especialista) humano e envolve aprendizagem de uma função a partir de exemplos de suas entradas e saídas. Essa é a definição de:
 - a) Aprendizagem por reforço.
 - b) Aprendizagem não supervisionada.
 - c) Regressão.
 - d) Aprendizagem supervisionada.
 - e) Variância.

3. Aprendizagem por reforço é uma área de aprendizagem de máquina que investiga como agentes de software devem agir em determinados ambientes, de modo a maximizar alguma noção de recompensa cumulativa. Esse tipo modela a interação entre dois elementos: o ambiente e o agente de aprendizagem. O agente de aprendizagem alavanca dois mecanismos. Assinale quais são:
 - a) Regressão e Classificação.
 - b) Ambiente e Agente de Aprendizagem.
 - c) Exploration e Exploitation.
 - d) Dedução e Indução.
 - e) Deep Learning e Over Learning.



4. “A Inteligência Artificial é composta por códigos e dados, sendo que os primeiros são responsáveis pela leitura e interpretação dos segundos. No entanto a IA é mais do que apenas análise de dados e, para cumprir uma infinidade complexa de comandos que resultam na capacidade de imitar os humanos, ela inclui diversas tecnologias”.

CONTENT, R. Inteligência Artificial: entenda o que é e como ela funciona. **Rockcontent**, [S. l.], 2019. Disponível em: <https://rockcontent.com/br/blog/inteligencia-artificial/>. Acesso em: 25 nov. 2021.

Considerando o trecho apresentado, as três principais são:

- f) Classificação, Regressão e Indução.
- g) Machine Learning, Deep Learning e Processamento de Linguagem Natural.
- h) Classificação, Regressão e Dedução.
- i) Indução, Redução e Classificação.
- j) Classificação, Machine Learning e Dedução.

5. Considerado o pai da computação. Inglês, matemático e criptógrafo, suas ideias foram a base para o desenvolvimento dos primeiros computadores. Teve, também, um papel importantíssimo durante a segunda guerra mundial, utilizando seus conhecimentos de criptografia para decifrar os códigos da máquina Enigma. Além disso, criou um teste cujo objetivo era verificar se o computador poderia imitar e pensar como o nosso cérebro, como se fosse uma IA e esse teste leva o seu nome. Estamos falando de:

- a) Alan Turing.
- b) Charles Babbage.
- c) Blaise Pascal.
- d) John Eniac.
- e) Jack Boss.

AGORA É COM VOCÊ



6. Existem várias áreas de pesquisa em que o ML tem contribuído diretamente para seu avanço, como na neurociência, probabilidades, teoria da informação ou estatística, entre outras, apesar de estar sempre associada à inteligência artificial. Probabilidade e estatística, também, são muito importantes em ML. E por mais que pareça algo surreal, o ML está presente em várias aplicações que buscam solucionar problemas do nosso dia a dia. Descreva três aplicações ou processos que você conhece e que utilizam o ML.
7. A aquisição de novos conhecimentos, o desenvolvimento de técnicas motoras e cognitivas, por meio de instrução ou prática; a generalização de conhecimentos adquiridos; a representação do conhecimento e a descoberta de novos fatos/teorias são alguns dos processos da aprendizagem. Faça a relação entre a aplicação e o tipo de aprendizagem:
- I - Aprendizagem supervisionada
 - II - Aprendizagem não supervisionada
 - III - Aprendizagem por reforço
- A - Amplamente utilizada em games.
B - Detecção de transações fraudulentas.
C - Descobrir o ano de fabricação de um veículo.
- a) I - A, II - C e III - B.
 - b) I - A, II - B e III - C.
 - c) I - B, II - C e III - A.
 - d) I - C, II - B e III - A.
 - e) I - B, II - A e III - C.

MEU ESPAÇO



Métodos e Algoritmos

Dra. Lara Carnevale de Almeida

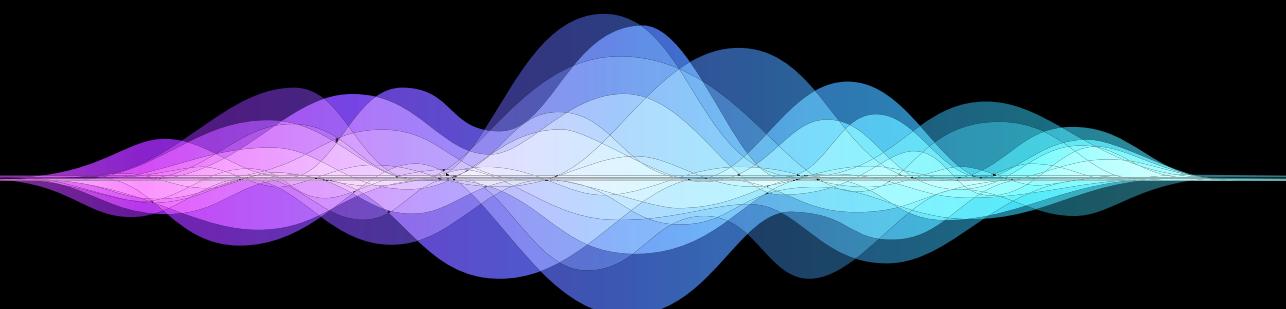
Esp. Ronie Cesar Tokumoto

OPORTUNIDADES DE APRENDIZAGEM

O objetivo desta unidade é apresentar conceitos iniciais sobre métodos e algoritmos, baseados em Inteligência Artificial, que permitem aplicar processos de análises estatísticas e predições, tanto para *Machine learning* (em português, Aprendizado de Máquina - AM) quanto para *Deep learning* - DL (em português, Aprendizado profundo).

Olá, Aluno(a)! Agora que você já conhece os princípios de Inteligência Artificial (IA), percebe como a IA tem estado presente em diversas soluções atuais do mercado. Atualmente, podemos observar soluções que pareciam ser futuristas cada vez mais presentes em nossas vidas, muitas delas graças ao avanço desta área.

Acreditamos que você já esteja curioso(a) para, por exemplo, perguntar para um telefone celular e receber uma resposta. Caso você tenha um Iphone, fale “Olá, Siri” e pergunte “Qual a previsão do tempo para a minha cidade amanhã?”. Se você possui um dispositivo com o sistema operacional Android e o Google Assistente instalado, basta substituir o “Olá, Siri” por “Olá, Google”. Tanto a Siri quanto o Google irão reconhecer a sua voz, descobrir a sua cidade (a partir dos dados de sua geolocalização ou de suas preferências de localização), procurar pela previsão do tempo e vão lhe responder, com uma voz nada robótica. Todo este processo é muito rápido!



As aplicações que usam IA não param por aí, pois podemos encontrá-las em sistemas computacionais para setores comercial, industrial e educacional. No setor educacional, por exemplo, os algoritmos podem prever as chances de um acadêmico desistir de uma faculdade ou identificar quais conhecimentos, competências, habilidades e atitudes este aluno possui dificuldades ou que precisa desenvolver. Outros exemplos são sites de busca, carros autônomos, análises preditivas, sistemas sofisticados de diagnósticos médicos, gêmeos digitais que simulam as mais diversas atividades industriais, da agricultura e da pecuária, além de sistemas de apoio às pesquisas e naves espaciais, todos esses são algumas aplicações do extenso universo da IA, muitas delas se baseiam em *Machine learning* ou *Deep learning*.

Acreditamos também que você tenha algumas dúvidas, tais como: “Já comprehendi que o uso de IA permite que sejam apresentadas soluções para uma tomada de decisão mais eficaz e eficiente, mas como isto é feito?”, “Como analisar grandes volumes de dados, por exemplo, em *Big Data*?”, “Como podemos separar áreas de interesse em um sistema computação baseado em IA?”, “Afinal, quais são as diferenças entre *Machine learning* e *Deep learning*?“

Pois bem, para responder às perguntas anteriores, você precisa compreender os processos que estão associados ao reconhecimento de padrões e classificações de dados. A classificação de dados consiste em prever e separar os dados em categorias com base na observação de dados. Note a importância de agrupar os dados pelas suas características comuns para, por exemplo, classificar os pacientes, no Pronto Atendimento de um hospital, com base nos sintomas que esses pacientes apresentam. Pode-se, assim, enquadrar os novos pacientes nos grupos já existentes desse hospital com base nas informações inicialmente adquiridas.

Tudo isto é possível por meio da aplicação de métodos e algoritmos. Para tal, conto que leia com atenção a esta unidade e, principalmente, procure realizar as atividades propostas durante a leitura deste material.

É um fato a utilização da IA para modificar ou reinventar processos e atividades. A aplicação de IA em uma organização, em um processo ou uma atividade pode seguir duas abordagens ou dois caminhos: o primeiro é a utilização de programas e/ou plataformas de terceiros, uma das abordagens mais utilizadas; a segunda é o desenvolvimento interno de modelos e programas. Independentemente da abordagem escolhida pela organização, a implantação da IA deve começar sempre pela educação e treinamento dos envolvidos no processo.

Gordon Moore, cofundador da Intel, previu, em 1965, que a cada dois anos a capacidade de processamento dos computadores dobraria. Tal afirmação ficou conhecida como a lei de Moore, que mais tarde teve o tempo diminuído para 1,5 anos. A miniaturização dos materiais proporcionou não só o aumento do poder computacional, mas também o aumento da capacidade e a diminuição dos custos com o armazenamento de dados.

Assuma que um **algoritmo** é um conjunto de instruções que são dadas a um computador para a execução de uma determinada tarefa. Quanto melhor ficam os computadores, melhores, mais eficientes e inteligentes ficam os algoritmos, realizando tarefas que, antes, eram realizadas exclusivamente pelos seres humanos. Além disso, fatores que impulsionaram a IA moderna: a Internet, o crescimento explosivo dos conjuntos de dados (em inglês, *datasets*), a infraestrutura e ferramentas oferecidas por empresas, tais como: Google, Amazon, IBM e Microsoft, dentre outras.

Assuma também que **infraestrutura** é um conjunto de equipamentos físicos e meios de comunicação que são disponibilizados para os usuários. Além disso, **ferramentas** são os algoritmos e programas disponibilizados para os usuários. Ao utilizar a infraestrutura e as ferramentas de terceiros, o tempo e o custo de implantação de soluções de IA podem ser reduzidos. Portanto, os **métodos e algoritmos** são utilizados em sistemas que visam a previsões e classificações, com base em informações previamente inseridas. Estes dados podem ser o perfil dos alunos em uma universidade, produtos ofertados por empresas que têm maior busca por clientes, tipos de veículos em via pública em um determinado tempo, dentre outras possibilidades.

Lembre-se do exemplo do Pronto Atendimento de um hospital: quando um paciente chega para uma consulta, deve informar os dados referentes ao seu estado de saúde, no momento da triagem. Com base em informações previamente armazenadas no banco de dados, o sistema pode cruzar informações e determinar os possíveis diagnósticos, que serão analisados pelo médico que atenderá esse paciente. Logo, o aprendizado sobre métodos e algoritmos de classificação permitirá que você possa estruturar sistemas para analisar dados, após separação e classificação desses dados!

Vamos compreender a aplicação de um sistema de classificação? Pense em um software baseado em *Machine learning* que, com base nas informações iniciais obtidas no cadastro do usuário (tais como nacionalidade, idade, sexo), o sistema consegue identificar o perfil deste usuário e classificá-lo com base nas suas ações, neste sistema.

Consegue pensar em um sistema que você já utilize? A Netflix é um serviço de Streaming que aplica o conceito de classificação de dados do usuário a todo momento, para assim recomendar filmes e séries que possam ser do seu interesse. Sugiro que você leia “[Como funciona o sistema de recomendações da netflix](#)”.



Armazenar, adquirir, representar e manipular conhecimento são ações que um sistema IA pode realizar, representando a manipulação, por meio de dedução ou inferência de novos conhecimentos e relações entre fatos e conceitos. Para tal, utiliza-se de métodos de manipulação e representação, partindo de conhecimento existente e construindo uma solução para o questionamento e/ou desafio apresentado.

Relembrando, o *Machine Learning* está presente em várias aplicações que buscam solucionar problemas do nosso dia a dia. Além do reconhecimento de voz e da condução de automóveis autônomos, temos outras aplicações baseadas em ML, tais como:

- **Banco de dados autônomo:** o ML permite automatizar várias ações que eram realizadas por um Administrador de Banco de Dados (em inglês, *Database administrator* - DBA), minimizando assim os problemas causados pela indisponibilidade do banco de dados devido à falha humana, liberando o DBA para outras atividades.
- **Combate a fraudes em sistemas de pagamento:** o ML permite barrar grande parte das ações relacionadas a fraudes com cartões de crédito ou outros métodos de pagamento.
- **Tradução de textos:** o ML permite melhorar a precisão dos tradutores automáticos, levando em consideração o contexto, expressões regionais e outros parâmetros, evitando assim que a tradução não seja puramente literal.
- **Recomendação de conteúdo:** o ML possibilita a análise das preferências dos usuários para que plataformas de *streaming* de vídeo e áudio possam recomendar conteúdo condizente com estas preferências.
- **Engenharia Mecânica:** embora um pouco restrito, é observado nas áreas de manufatura, com a indústria 4.0, e também da Internet das Coisas, nas quais se utilizam e enviam informações em tempo real, substituindo regras estáticas e prevendo comportamentos complexos das máquinas, por meio de manutenção preditiva, um tipo de manutenção que preza pela prevenção de danos e falhas.

Muitos problemas em áreas de pesquisa — que envolvem neurociência, probabilidade, teoria da informação, estatística, entre outros — são resolvidos com soluções baseadas em ML. Procure prestar mais atenção nas soluções divulgadas, baseadas em Inteligência Artificial; algumas vezes, essa solução está baseada em Aprendizado de Máquina!

Devemos, também, salientar duas áreas que têm contribuído com o ML, probabilidade e estatística. Estas duas áreas são, frequentemente, utilizadas para compreensão de detalhes gráficos e análise exploratória de dados. Contudo essas áreas podem também auxiliar na construção de modelos de ML para realizar: análise de conjunto e tipo de dados, inferências, validação de suposições, resumo de dados e compreensão de histogramas.

Tenha atenção, pois sistemas baseados em ML estão presentes em nosso cotidiano. De forma positiva, estas funcionalidades nos ajudam e são, em sua maioria, assertivas. Contudo temos que evitar que esses sistemas façam classificações errôneas dos dados, pois isto provocará uma tomada de decisão inadequada. Portanto, aprimorar os seus conhecimentos para lidar com esses sistemas será, com certeza, um diferencial, tornando você um especialista capaz de aplicar estes conhecimentos de forma adequada.

DIÁRIO DE BORDO

Um sistema baseado em IA apresenta componentes inter-relacionados, conforme apresentado na Figura 1. As ações de aquisição, representação e manipulação (por meio da dedução e pesquisa) de conhecimento permitem estruturar: (1) a máquina de inferência (isto é, a estratégia de controle) para determinar os itens de conhecimento que devem ser pesquisados, (2) a definição das ordens dos passos e (3) as deduções que devem ser feitas.

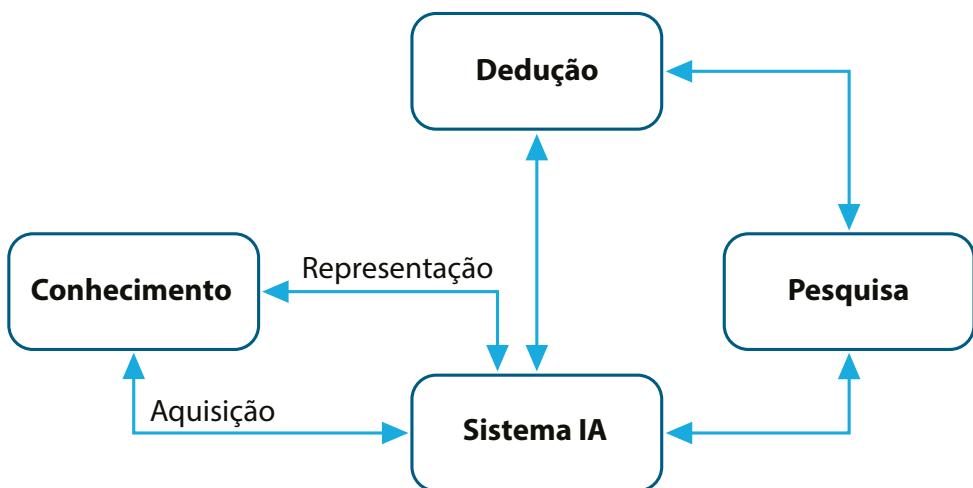


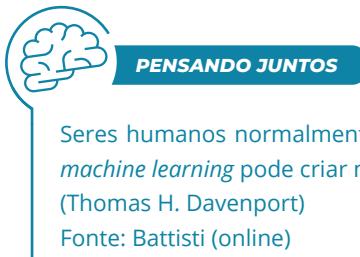
Figura 1 - Visão Geral sobre Inteligência Artificial / Fonte: a autora.

Descrição da Imagem: a figura retrata, por meio de um esquema, os principais componentes de um sistema que utiliza Inteligência Artificial, representado por caixas nomeadas como Conhecimento, Dedução, Pesquisa e Sistema IA. As caixas Conhecimento, Representação e Aquisição ligam-se, por meio de setas, ao Sistema IA. As caixas Pesquisa e Dedução estão ligadas entre si.

O *Machine Learning* tem o papel de tornar as operações menos dependentes de ações humanas e mais autônomas a partir, por exemplo, da utilização de sensores e algoritmos, procurando antecipar e alertar sobre possíveis falhas.

Algoritmos baseado em ML tem a função de encontrar padrões em grande volume de dados, permitindo uma tomada de decisão mais assertiva e com a possibilidade de fazer previsões sobre uma nova carga de dados. Quanto melhor for o algoritmo, melhor serão as tomadas de decisão e as previsões.

Para tal, técnicas de mineração de dados devem ser aplicadas, pois visam à descoberta de propriedades desconhecidas de dados. É um recurso importante para garantir o bom andamento do ML, visto que este trabalha com predição e análise de dados, utilizando técnicas de agrupamento de dados (em inglês, *clustering*), verificação de anomalias e regras de associação.



Seres humanos normalmente conseguem criar um ou dois bons modelos por semana; *machine learning* pode criar milhares.

(Thomas H. Davenport)

Fonte: Battisti (online)

Você deve compreender **algoritmos** e, por consequência, saber criar sequências lógicas. Algoritmos possibilitam a interpretação de informações matemáticas, são implementados na forma de scripts, e são sequências de operações que visam à execução de uma tarefa. Podemos fazer uma analogia com a sequência lógica (o algoritmo) que você usa para ir ao seu trabalho ou para executar uma tarefa doméstica.

Algoritmos de aprendizado, baseado em ML, permitem que computadores aprendam determinada função ou tarefa sem a necessidade de serem explicitamente programados. Conforme Sejnowski (2019), o **Machine learning** permite que os computadores tenham a habilidade de aprender, sem terem sido programados especificamente, e esta área explora a construção de algoritmos que podem aprender com os seus erros e, na sequência, fazer previsões sobre dados de entrada, permitindo apresentar decisões/diagnósticos considerados resultados confiáveis e repetíveis. Este mesmo autor indica que o **Deep learning** é uma modalidade de aprendizado de máquina, em que os computadores podem aprender conceitos mais complexos, a partir de sua própria experiência, e envolver conceitos mais simples.

Podemos dizer que aplicando conceitos de estatística, um programa baseado em aprendizado pode ser treinado para executar tarefas e, por consequência, aprimorar as suas previsões. A Figura 3 procura apresentar as diferenças entre as soluções baseadas em ML e DL.

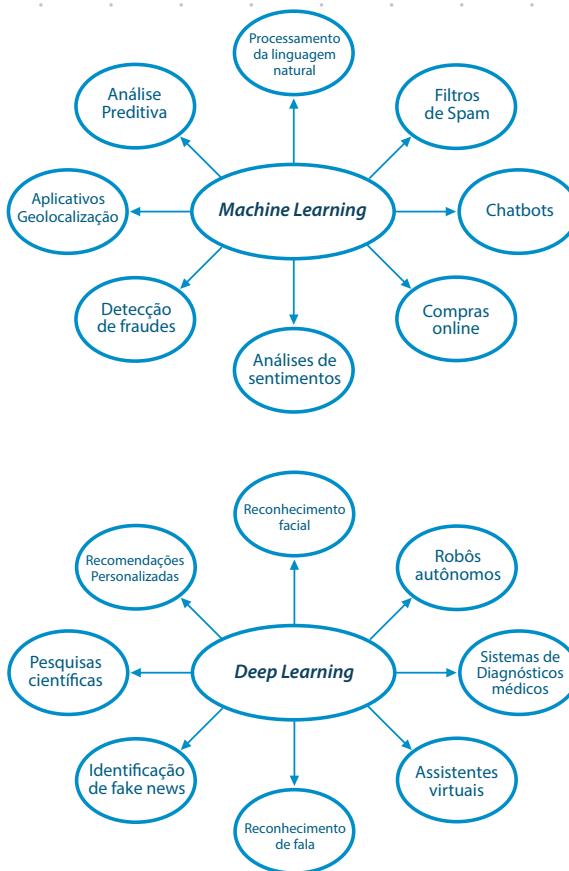
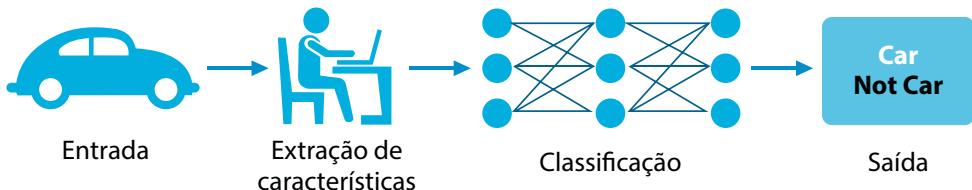


Figura 2 - Inteligência Artificial aplicada utilizando ML e DL / Fonte: adaptado de Taulli (2020).

Descrição da Imagem: a figura apresenta duas imagens. A imagem da esquerda apresenta um círculo central escrito "Machine Learning", dele saem oito outros círculos ligados por setas. São eles, em sentido horário: "Processamento da linguagem natural", "Filtros de Spam", "Chatbots", "Compras online", "Análises de sentimentos", "Detectação de fraudes", "Aplicativos Geolocalização" e "Análise Preditiva". A imagem da direita apresenta um círculo central escrito "Deep Learning", dele saem oito também círculos ligados por setas. São eles, em sentido horário: "Reconhecimento facial", "Robôs autônomos", "Sistemas de Diagnósticos médicos", "Assistentes virtuais", "Reconhecimento de fala", "Identificação de fake news", "Pesquisas científicas" e "Recomendações Personalizadas".

A Figura 3 salienta os processos envolvidos para aplicar Machine Learning e Deep Learning. Note que os algoritmos baseados em ML devem, primeiro, extrair características dos dados de entrada para, depois, realizar a classificação. No Deep Learning, a extração de características e classificação são realizadas na mesma etapa. Isso é possível por meio dos princípios de redes neurais, em que os dados passam por várias camadas de processamento não lineares, sendo possível, por exemplo, reconhecer imagens e voz.

Machine Learning



Deep Learning



Figura 3 - Processos envolvidos no Machine Learning X Deep Learning

Fonte: adaptada e traduzido de *Semi Engineering* (2018, on-line).

Descrição da Imagem: a figura demonstra a diferença entre Machine Learning e Deep Learning. O Machine Learning tem seu processo representado por Entrada (simbolizada por um ícone de carro); Extração de características (simbolizada pelo ícone de uma pessoa sentada no computador); Classificação (simbolizada por círculos interligados entre si) e Saída (simbolizada por um quadrado com os dizeres "car - not car". Já o Deep Learning é representado por Entrada (simbolizada por um ícone de carro); Extração de características + Classificação (simbolizada por círculos interligados entre si) e Saída (simbolizada por um quadrado com os dizeres "car - not car").



EXPLORANDO IDEIAS

O **Machine Learning**, em si, busca mecanismos para que a máquina possa, de certa forma, evoluir a partir de seu software progressivamente, sem que isto seja totalmente implementado pelo desenvolvedor, mas o software foi projetado para que possa realizar de forma autônoma, inclusive, podendo haver falhas no aprendizado como ocorre com o ser humano. Já o **Deep Learning** se utiliza de conceitos de redes neurais para que atividades mais sofisticadas possam ser realizadas como o aprendizado de caráter mais abstrato, a partir de um determinado conjunto de dados, trabalhando sobre camadas que alimentam o aprendizado gradual para camadas posteriores.

Venha absorver mais conhecimento com o nosso podcast! Dê o play para compreender melhor as diferenças nas aplicações baseadas em aprendizado!

Conforme Peremulter (2020), o *Deep Learning* está na vanguarda da IA. Muitas empresas adotaram o DL para o reconhecimento de padrões de fala e voz, por exemplo, a ferramenta do *Google Translate* (em português, Google Tradutor). Em reconhecimento de imagem, DL é usado para fazer a descrição automática da cena. Veja a seguir outros exemplos de aplicação:

- **Comportamento do cliente:** quando o cliente acessa sites de e-commerce, por exemplo, muitos dados são coletados. E quanto mais tempo e experiência no site, mais chances de compras e, com isso, compreensão do que os clientes preferem.
- **Sistemas de recomendação:** empresas, como Amazon e Netflix, usam o comportamento anterior do cliente para ter boas chances de acertar no que o cliente pode estar interessado ou em suas preferências em múltiplas plataformas.
- **Suporte técnico ao cliente:** algumas empresas optam por um suporte técnico personalizado automático para prestar serviços de assistência técnica remota a um cliente.
- **Reconhecimento facial:** muitas aplicações fazem uso dessa técnica, como o Facebook, para identificar os usuários nas publicações. A Polícia Militar utiliza para reconhecimento de foragidos e empresas de segurança em seus sistemas de segurança. O reconhecimento de imagem tem sido usado para identificar áreas e pessoas presentes na cena do crime ocorrido, a partir das fotos tiradas das cenas ou por meio do uso de câmeras com tecnologia de 360°.
- **Diagnósticos de doenças:** redução de erros nos diagnósticos de doenças, como o câncer, ajudando na identificação das células cancerígenas por meio das imagens.
- **Classificação de doenças:** com a alimentação de uma base de dados e da análise de imagens, é possível auxiliar, por exemplo, na identificação

de exames médicos para muitas doenças que podem auxiliar médicos, como retinas afetadas por diabetes.

- **Carros autônomos:** tecnologia usada para que o carro seja conduzido por uma rota mais curta e sem trânsito, com o uso de mapas e paradas.

OLHAR CONCEITUAL

A Internet das Coisas (em inglês, Internet of Things - IoT) está presente em várias áreas, da medicina à área da agronomia. O IoT é um conceito de integração de dispositivos embarcados, em que cada equipamento se conecta a outros dispositivos com o intuito de coletar e transmitir informações.

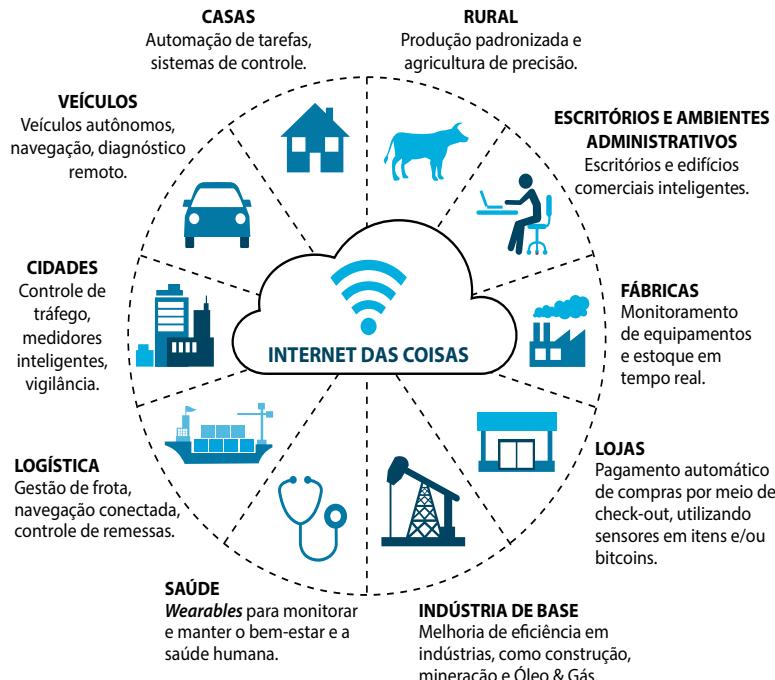


Figura 4 - Internet das Coisas / Fonte: a autora.

Descrição da Imagem: a figura em que temos, no centro, um círculo nomeado Internet das Coisas e, ao redor, simbolizada por ícones referentes aos seus títulos, temos Casas - automação de tarefas, sistemas de controle; Rural - produção padronizada e agricultura de precisão; Escritórios e ambientes administrativos - escritórios e edifícios comerciais inteligentes; Fábricas - monitoramento de equipamentos e estoque em tempo real; Lojas - pagamento automático de compras por meio de check-out, utilizando sensores em itens ou/e beacons; Indústria de base - melhoria da eficiência em indústrias, como construção, mineração, óleo e gás; Saúde - wearables para monitorar e manter o bem-estar e a saúde humana; Logística - gestão de frota, navegação conectada, controle de remessas; Cidades - controle de tráfego, medidores inteligentes, vigilância; e Veículos inserir espaço- veículos autônomos, navegação, diagnóstico remoto.

Um dos maiores desafios para que as ações do ML obtenham sucesso é a qualidade dos dados que serão analisados, pois a maioria desses dados, ainda, não são sólidos (ou não estão prontos) para serem analisados, agrupados, entre outras atividades. Devemos apresentar ao programa a maior quantidade de dados possível ao algoritmo de aprendizado para que, baseado nestes dados, o algoritmo tire conclusões e/ou faça previsões automaticamente.

Programas baseados em Machine Learning têm o objetivo de descobrir o relacionamento entre as variáveis do sistema (entrada/saída) a partir de uma amostra de dados. Note que a amostra não é necessária, quando os relacionamentos entre todas as variáveis do problema (entrada/saída) são compreendidas. Para tal, elementos principais para o aprendizado supervisionado de ML são:

1. Dados de treinamento: são os dados de entrada/saída, em que deve ocorrer um mapeamento funcional da entrada para a saída.
2. Espaço de hipóteses: é um conjunto particular, ou uma classe de funções candidatas.
3. Algoritmo de aprendizado: recebe dados de treinamento como argumento de entrada para, posteriormente, selecionar uma hipótese pertencente ao espaço de hipóteses predefinidas.

Peremulter (2019) indica que **programas baseados em Deep learning** podem ser representados pela equação **dados + resposta = regras**. Este autor salienta que programas de computadores clássicos foram, e ainda são em sua maioria, baseados na **equação regras + dados = resposta ou algoritmos + dados = informação**, em que: os dados de entrada são processados de acordo com as regras ou algoritmos; e a saída é vista como a resposta ou informação, ou dito de outra forma, o resultado do processamento. Note que o DL é uma tecnologia que permite, por meio do processamento e da análise de uma enorme quantidade de dados, relacionar os dados e encontrar padrões que os seres humanos têm dificuldade, ou não conseguem, pelo tempo necessário para processar e analisar esses dados (PEREMULTER, 2019).

Conforme Mitchell (1997), para compreender melhor como o aprendizado pode ser aplicado, você deve assumir que um programa de computador **aprende** a respeito de uma tarefa **T**, por meio de uma experiência **E**, em relação a uma métrica de desempenho **P**, se o seu **desempenho** na tarefa **T**, quando medido pela métrica **P**, **aumenta** com a experiência **E**.

Procure imaginar que estamos desenvolvendo um sistema, um programa de meteorologia, com a finalidade de prever, na próxima hora, o acumulado de precipitação, tendo por base dados anteriores, a tarefa T seria estimar a precipitação das chuvas (acumulado) para a hora seguinte, o P — medida de desempenho — seria a diferença entre o previsto e o realizado (medida de erro) e o E — experiência — seria as tentativas de realização da previsão. Conforme as previsões se aproximam do valor observado durante as experiências, acontece o aprendizado de máquina.

Muito abstrato? Veja, então, mais alguns exemplos:

- Diagnóstico de câncer (tumor maligno ou benigno):

Tarefa T: determinar, com base em imagens de exames, se um tumor é maligno ou benigno.

Medida de Desempenho P: percentual de diagnósticos corretos.

Experiência de Treinamento E: imagens de tumores obtidas de exames médicos (raio X, ultrassom, tomografia, ressonância magnética etc.).

- Navegação de um veículo autônomo:

Tarefa T: navegar em uma rodovia de cinco pistas usando sensores de visão.

Medida de Desempenho P: distância média viajada antes de um erro ocorrer.

Experiência de Treinamento E: uma sequência de imagens e comandos de direção registrados por meio da observação de um motorista humano.

- Robô de investimentos em bolsa de valores:

Tarefa T: vender e comprar ações de modo a maximizar o lucro obtido em um determinado prazo.

Medida de Desempenho P: lucro obtido pelo sistema ao longo do prazo determinado.

Experiência de Treinamento E: uma sequência histórica de notícias políticas e econômicas, dados financeiros das empresas negociadas e as variações de preço de mercado ao longo do tempo.

Reforçamos o discurso que, para a criação de uma solução IA, utilizando ML ou DL, é importante seguir um processo ou utilizar um framework, de forma a aplicar um conjunto de técnicas e ferramentas que tem uma sequência lógica e uma abordagem sistemática dos passos necessários para se criar a solução.

Grandes organizações — tais como Microsoft, Google, IBM e Amazon — já possuem os seus próprios modelos, algoritmos e infraestrutura. Contudo existem propostas de processo genéricas que é importante conhecer. Taulli (2020) generaliza e diz que, para implementar uma solução baseada em IA, deve-se:

- Identificar e definir, com o máximo de precisão, o problema a resolver.
- Selecionar ferramentas e plataformas adequadas.
- Criar o modelo de IA em conformidade com as ferramentas e a plataforma utilizadas.
- Implantar o modelo.
- Monitorar os resultados e, caso não sejam satisfatórios, fazer ajustes no modelo.
- Ficar atento à qualidade e os tipos de dados que serão utilizados.



Os avanços das técnicas de inteligência artificial permitem que sistemas computacionais aprendam a inferir e a extrapolar com base em situações que são apresentadas e ensinadas ao longo do tempo. E o desenvolvimento de técnicas para a análise de grande quantidade de dados (Big Data) alavanca de forma exponencial a capacidade das máquinas para processar informações e gerar recomendações (PERELMUTER, 2019, p. 48).



EXPLORANDO IDEIAS

É preciso compreender que muitos dos dados obtidos em uma fase de pré-processamento possuem alguns problemas, tais como:

- Incompletude: ausência de alguns valores ou atributos.
- Inconsistência: quando os dados extraídos apresentam diferentes e conflitantes versões de um mesmo dado em locais diferentes.
- Ruído: variações inexplicáveis e indesejadas. Por exemplo, uma imagem borrada ou com baixa nitidez.

Esses problemas precisam ser tratados antes que as atividades de análise e classificação sejam realizadas. Existem diversas formas para tratamento de dados, sendo as principais:

- Limpeza: exclusão de ruídos e correção de inconsistências;
- Integração: união de dados de diferentes fontes em um único local.

Fonte: Castro e Ferrari (2016)

Relembre que a aprendizagem supervisionada tem o objetivo de encontrar ótimos parâmetros que possam ser trabalhados em um modelo de conjunto de testes de rótulos desconhecidos. Já a aprendizagem não supervisionada envolve a aprendizagem de padrões na entrada, quando não são fornecidos valores de saída específicos, em que o objetivo é analisar as semelhanças entre os objetos e agrupá-los. Existem muitos desafios que podem ser solucionados com apoio de modelos ML. Atualmente, os mais conhecidos são:

- **Problemas de classificação** precisam que a solução possa escolher um rótulo dentre os rótulos existentes, como identificar um objeto específico (por exemplo, gato, carro, placa de trânsito, entre tantos outros) em uma imagem. Ambos os modelos, assim como outros, baseiam-se na seguinte premissa: aprender padrões relevantes sobre o problema que deve ser resolvido, por meio da observação de dados, de forma a permitir a solução deste, sem que haja necessidade de programação explícita.
- **Problemas de regressão** necessitam que a solução possa predizer um valor numérico pertencente a um intervalo contínuo, por exemplo, na determinação do valor de uma casa, considerando a data de construção desta, seu tamanho, entre outros parâmetros imobiliários.

A principal diferença é que a classificação trabalha com rótulos e a regressão lida com números, a Figura 5 retrata esta diferença.

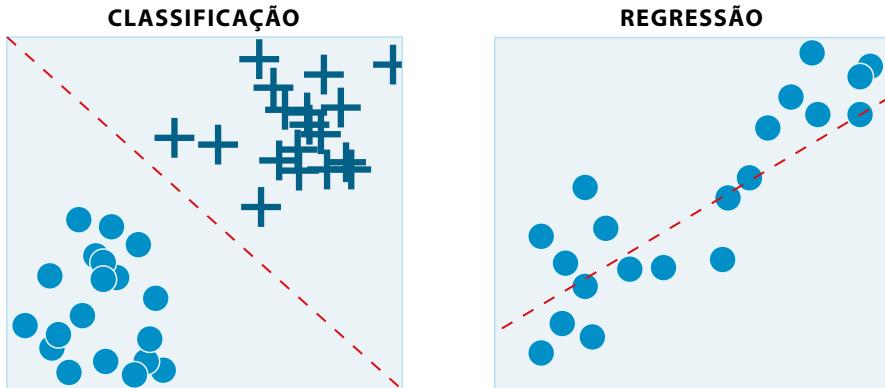


Figura 5 - Diferença entre classificação e regressão

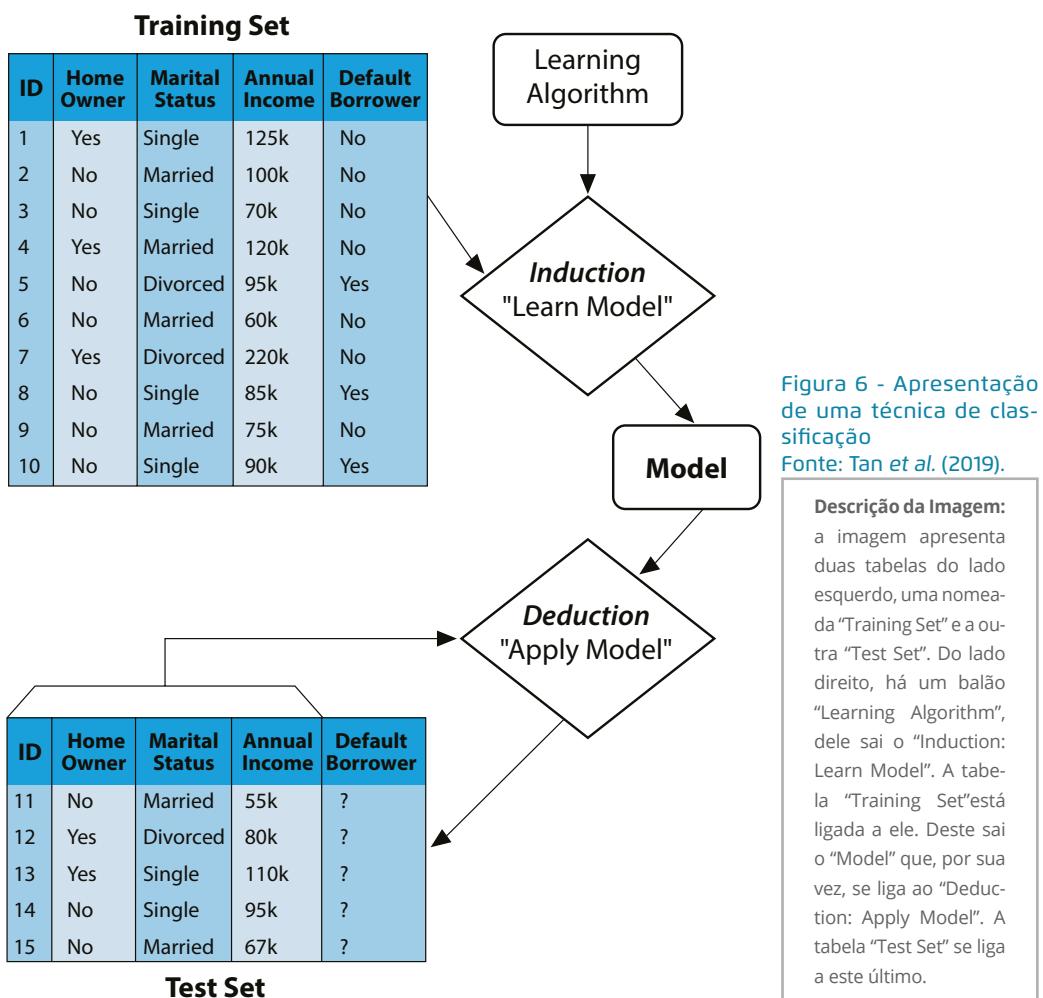
Descrição da Imagem: duas figuras, lado a lado, exemplificam e mostram a diferença entre classificação e regressão. Na primeira, simbolizando a Classificação, temos um quadrado dividido ao meio por uma linha tracejada, na diagonal. Na parte de cima, temos sinais de adição e, na parte de baixo, temos círculos, ambos distribuídos aleatoriamente em cada metade do quadrado. Na Regressão, temos um quadrado também dividido por uma linha tracejada, com círculos distribuídos aleatoriamente em cima e ao redor da linha tracejada.

Se a tarefa tem o rótulo oriundo de um conjunto finito e não ordenado, temos uma classificação; caso o rótulo seja um número real, temos regressão. Em outras palavras, a **classificação** é a atribuição de casos ou instâncias de dados a uma ou mais possíveis classes, e a **regressão** é a estimativa do valor de uma variável baseada em exemplos.

Finalmente, existem duas abordagens diferentes para a classificação de dados: a primeira considera apenas uma distinção dicotômica entre as duas classes e atribui rótulos de classe 0 ou 1 para um item de dados desconhecido. A segunda procura modelar $P(y|x)$, produzindo não apenas um rótulo de classe para um item de dados, mas também uma probabilidade de pertencer à classe. Para a primeira abordagem, tem-se o algoritmo de Máquinas de vetores de suporte (em inglês, *Support Vector Machines* - SVM). Já os algoritmos Regressão logística (em inglês, *Logistic Regression* - LR), Redes neurais artificiais (em inglês, *Artificial Neural Networks* - ANN), k-vizinho mais próximo (k-Nearest Neighbors - kNN) e Árvores de Decisão (em inglês, *Decision Trees* - DT) pertencem a segunda abordagem, embora estes variem consideravelmente na construção de uma aproximação para $P(y|x)$, a partir dos dados. Esses algoritmos serão apresentados nas próximas unidades.

Algoritmo de classificação

Para compreender o funcionamento de um **algoritmo de classificação**, observe na Figura 6 um framework genérico para criação de um *Model* (em português, Modelo). Conforme Tan, Steinbach e Kumar (2009), a **técnica de classificação** é “uma abordagem sistemática para construção de modelos de classificação a partir de um conjunto de dados de entrada”. O objetivo desta classificação é criar um modelo que consiga “aprender” se uma pessoa é considerada *Default Borrower* (em português, Mutuário padrão), a partir de informações desta pessoa.



No exemplo apresentado na Figura 6, tem-se um modelo de aprendizagem de máquina supervisionado. Na primeira tabela desta figura, os registros contêm os atributos *ID*, *Home Owner*, *Marital Status*, *Annual Income* e *Defaulted Borrower* (em português, ID, Proprietário, Estado civil, Rendimento anual e Mutuário inadimplente). Essa tabela é denominada como *Training Set* (em português, conjunto de treino), pois os seus dados servirão de base para a construção do *Model*. Dessa tabela, considere que os atributos *ID*, *Home Owner*, *Marital Status* e *Annual Income* são atributos independentes; contudo o atributo *Defaulted Borrower* é dependente. Ou seja, o modelo de classificação aprenderá qual será o valor do atributo dependente, a partir de atributos independentes.



Na sequência, o *Training set* será aplicado como entrada para um algoritmo de aprendizagem de máquina que irá aprender com o *Teste Set* (em português, conjunto de treino) e gerar como saída um *Model*. O *Training set*, como pode ser observado na Figura 6, não possui valores para o atributo dependente *Defaulted Borrower*. Este modelo desenvolvido recebe como entrada de dados o conjunto de testes e gera como saída o valor classificado para o atributo dependente. Finalmente, a Figura 7 apresenta uma árvore de decisão modelada para classificar os dados, a partir de um conjunto de treino.

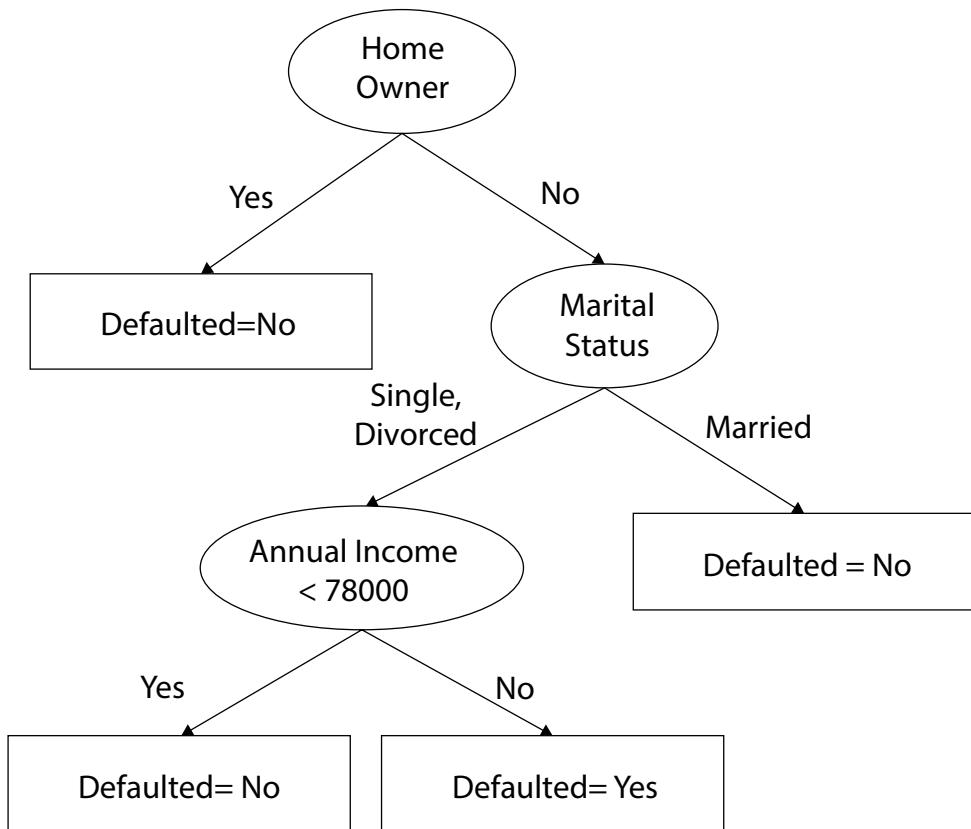


Figura 7 - Apresentação de uma técnica de classificação / Fonte: Tan et al. (2019).

Descrição da Imagem: a imagem é um fluxograma. Em primeiro, temos “Home Owner”, dele saem duas opções: se a resposta for sim “Defaulted = No”, e se for não “Marital Status”. Deste último saem duas opções: se a resposta for Single, Divorced “Annual Income < 78000”, e se for Married “Defaulted = No”. Do “Annual Income < 78000” saem outras duas opções: se sim, “Defaulted = No”, se não “Defaulted = Yes”.



PENSANDO JUNTOS

Há muitas aplicações para algoritmos de classificação. Tomando como base o exemplo apresentado nas Figuras 6 e 7, podemos imaginar sistemas bancários que utilizam Inteligência artificial e dados de seus clientes para decidir quando conceder ou não empréstimos. As mesmas técnicas também podem ser aplicadas em assistentes virtuais que reconhecem o que uma pessoa está dizendo, processam a solicitação para, ao final, agir ou dar uma resposta. Aplicações, como o Google Now, a Siri da Apple, a Alexia da Amazon e o Cortana da Microsoft, utilizam tipos especializados de classificação com processamento de linguagem natural para responder perguntas simples e/ou realizar ações, tais como agendar reuniões ou fazer compras.

A tarefa de classificar dados é decidir a associação de classe y' de um item de dados desconhecido x_0 com base em um conjunto de dados $D = (x_1, y_1), \dots, (x_n, y_n)$ de itens de dados x' com associações de classe conhecidas y_i . Para facilitar a compreensão, vamos considerar problemas de classificação dicotômica, em que os rótulos de classe y são 0 ou 1. Os x_i são geralmente vetores m-dimensionais, os componentes são chamados de covariáveis e variáveis independentes (uma linguagem estatística) ou variáveis de entrada (conforme comunidade de aprendizado de máquina).

Na maioria dos domínios do problema, não há relação funcional $y = f(x)$ entre y e x . Neste caso, a relação entre x e y deve ser descrita de uma forma mais geral: distribuição de probabilidade $P(x, y)$, em que se assume que o conjunto de dados D contém amostras independentes de P . Conforme a teoria de decisão estatística, uma decisão ótima de pertencimento à classe deve escolher o rótulo de classe y que maximiza a distribuição posterior $P(y|x)$.

Algoritmo de Regressão

A **regressão** é um método preditivo, ou seja, baseado em um conjunto de atributos, pode-se fazer algum tipo de previsão de algo que pode acontecer no futuro. A regressão faz parte de muitas áreas das nossas vidas, as quais a gente nem imagina! Por exemplo, para fazermos a previsão da velocidade do vento, temos como variáveis o valor da temperatura da umidade e da pressão do ar. Para realizar a previsão do limite de crédito do cartão de uma pessoa, temos como variáveis os gastos desse cartão e o histórico de compras desse cliente. Podemos utilizar as variáveis idade de uma pessoa e a possibilidade de doenças pré-existentes para fazer a previsão do custo de um plano de saúde.

O que podemos notar em comum nos exemplos citados? Todos trabalham com, no mínimo, duas variáveis para que a previsão possa ser executada. Portanto, ao falarmos de regressão, estamos falando de previsões numéricas. Modelos de regressão são muito utilizados em várias áreas, tais como: computação, estatística, administração, engenharias, biologia, agronomia e saúde, entre outras.

Na **regressão linear**, você deve modelar a relação entre as variáveis numéricas. Temos uma **variável dependente Y** e **variáveis exploratórias X** as quais fazem parte deste processo, pois a variável Y é o motivo pelo qual queremos fazer a previsão e as variáveis exploratórias X. Para facilitar o entendimento, veja os exemplos para fazer a previsão:

- da velocidade do vento por meio da variável Y, têm-se as variáveis exploratórias X com valores da temperatura, da umidade e da pressão do ar;
- do limite do cartão por meio da variável Y, têm-se as variáveis exploratórias X com gastos no cartão de crédito e o histórico do cliente;
- do custo de um plano de saúde a partir da variável Y, têm-se as variáveis exploratórias X com idade da pessoa.

Chamamos de **regressão linear simples**, quando temos apenas uma variável exploratória — como no exemplo da especificação do plano de saúde — que utiliza apenas a idade para prever esse custo.



EXPLORANDO IDEIAS

Para que uma regressão linear seja confiável, os dados precisam atender aos seguintes critérios:

- A relação entre as variáveis deve ser linear.
- Independente dos valores das variáveis preditoras, os termos de erro devem ter uma variância constante.
- Os erros nas variáveis preditoras não devem ser independentes, ou seja, não devem estar correlacionados.
- As variáveis preditoras não podem ser próximas de uma correlação perfeita.
- Os valores das variáveis preditoras não devem estar contaminados com erros de medida, pois erros de medida podem levar a estimativas inconsistentes e uma superestimação dos coeficientes de regressão.

Um exemplo clássico para a regressão linear simples é baseado na idade das pessoas, porque esse tipo de regressão trabalha com apenas uma variável exploratória. Assuma que sabemos a idade de uma pessoa e o custo do plano de saúde. Assuma as idades 18, 32, 42, 54 e 66 anos com, respectivamente, o custo mensal para o plano de saúde de R\$800, R\$1.600, R\$2.400, R\$3.200 e R\$4.000. Ou seja, uma pessoa com 17 anos tem o custo do seu plano de saúde por volta de 800 reais, enquanto a pessoa que tem mais de 66 anos tem por volta de 4000 reais como valor pelo seu plano de saúde. Com base na idade da pessoa, pretende-se realizar a previsão do custo do plano de saúde dessa pessoa. Quando analisamos as intersecções de X com Y, verificamos que existe uma relação linear entre essas duas variáveis, ou seja, quanto maior a idade, maior o custo do plano de saúde. Um outro exemplo seria verificar se existe relação entre o índice de desemprego de uma determinada cidade e os seus números de criminalidade, ou seja, verificar se houver mais desemprego, haverá mais crimes.

Atenção que uma regressão linear nem sempre pode ser utilizada para encontrar a melhor função dentro de um determinado conjunto de dados. Para indicarmos a direção e a força do relacionamento linear entre dois atributos, temos a correlação. Note que isto não implica em causalidade, ou seja, duas variáveis podem estar correlacionadas, mas não existe relação de causa-efeito entre elas.

Resumindo, o objetivo da **análise de regressão** é verificar a possibilidade de ajuste de um modelo que demonstra essa relação, após explorar a natureza dessa relação. Chamamos de **modelo de regressão simples** o modelo no qual existe apenas uma variável preditora, e **modelo de regressão múltiplo** quando têm-se mais de uma variável preditora. A **análise de regressão linear** tem por finalidade determinar o modelo que demonstre esta relação (equação de regressão), enquanto que a **análise de correlação linear** busca apresentar o grau de relacionamento entre duas variáveis.



Estatística: variância amostral

Muitas vezes, o estudo de técnicas, métodos e algoritmos se baseia em conteúdos de outras áreas do conhecimento, e, no caso da Ciência de Dados, é muito frequente o uso de conceitos estatísticos. Este vídeo traz uma perspectiva da aplicação da chamada variância, utilizada em técnicas estudadas, nesta unidade, de uma forma bastante interessante.

A análise de regressão constrói um modelo matemático para representar dois atributos (variáveis) X e Y , em que:

- $Y = f(X)$ tal que $f(\cdot)$ é a função que relaciona X e Y .
- X é a variável independente da equação.
- $Y = f(X)$ é a variável dependente das variações de X .

Para apresentar os dados, é utilizado o **gráfico de dispersão** (em inglês, *scatterplot*) do tipo cartesiano com os pares de informação X e Y , referente a cada observação. Forma-se uma “nuvem” de pontos que define um eixo (ou direção) e caracteriza um padrão de relacionamento entre as variáveis X e Y , conforme apresentado na Figura 8. Observe que existe uma tendência (ou eixo linear) entre a nuvem dos pontos.

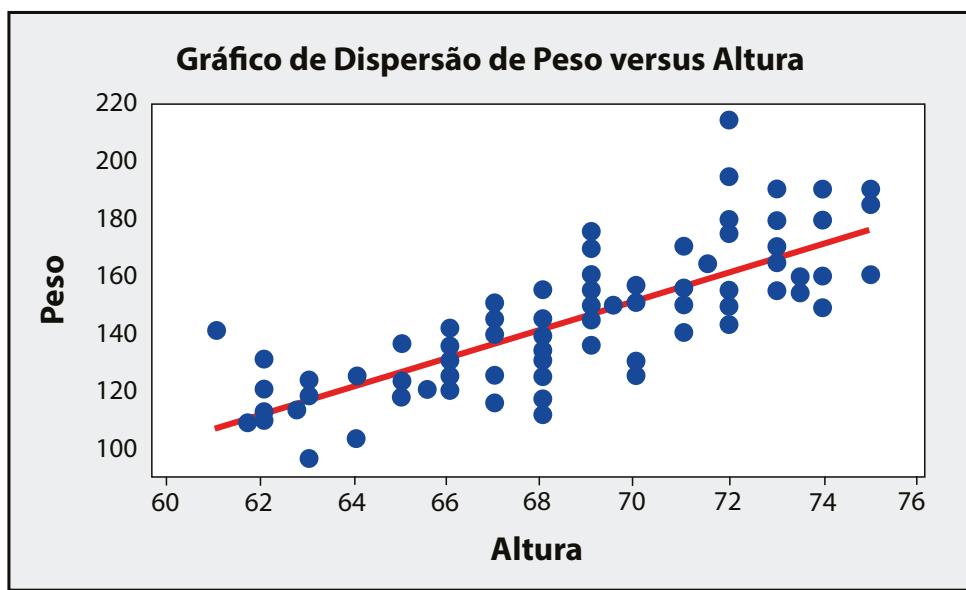


Figura 8 - Exemplo de gráfico de dispersão / Fonte: Minitab (online)

Descrição: modelo de exemplo de gráfico de disperso comparando e relacionando peso, na vertical (100,120,140,160,180,200,220) com altura, na horizontal (60,62,64,66,68,70,72,74,76); uma linha crescente na diagonal, com círculos distribuídos aleatoriamente em cima e ao redor da linha.

Na **regressão simples**, uma linha reta pode representar uma relação de causa e efeito, por meio da relação entre duas variáveis, possibilitando a previsão de valores da variável dependente a partir dos resultados da variável independente. A equação é:

$$Y = \beta_0 + \beta_1 X + E$$

onde:

- Y é a variável dependente (ou resposta).
- X é a variável independente (ou explicativa).
- β_0 é o intercepto em Y ; é o valor esperado de Y quando $X = 0$.
- β_1 é a inclinação (ou taxa de mudança); é o aumento esperado em Y quando X aumenta uma unidade.
- E é um erro aleatório, utilizado para incluir todas as influências no comportamento da variável Y que não podem ser explicadas linearmente pelo comportamento da variável X ; o valor de E pertence a $N(0, \sigma^2)$.

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + E$$

A **regressão múltipla** ocorre quando não podemos explicar a variável de resposta com apenas uma variável preditora, sendo o caso de incorporarmos várias variáveis independentes para chegarmos à solução. Essa é a sua equação:

onde Y é a variável dependente (ou resposta).

- x_1, \dots, x_k são as variáveis independentes (ou explicativas).
- β_0, \dots, β_k são os parâmetros desconhecidos do modelo a ser estimado.
- β_0 é o valor esperado de y , quando todas variáveis são iguais a zero.
- β_j , para todo j diferente de 0, representa a variação em y , quando x_j aumenta em uma unidade (mantendo constantes todos x_i com exceção do x_j). Se $\beta_j > 0$, y aumenta; caso contrário, y diminui.
- E é um erro aleatório, no qual se procuram incluir todas as influências no comportamento da variável y que não podem ser explicadas linearmente pelo comportamento das variáveis x_1, \dots, x_k . E pertence à $N(0, \sigma^2)$.

Por exemplo, assuma a relação entre os números da criminalidade, o índice de desemprego e a escolaridade dos seus moradores. Prevê-se que se houver mais desemprego, haverá mais crimes e menor escolaridade.

Finalmente, é importante que você perceba que: (1) sempre que os resultados vão decrescendo conforme temos valores (em quantidade, tempo, entre outros), estamos diante de uma relação marginalmente decrescente; (2) fenômenos marginalmente decrescentes são extremamente comuns e não são lineares! Assim, para utilizar técnicas de AM, devemos entender melhor estas relações. Para tal, vamos aprender relações não lineares na próxima unidade.

Existe um esforço muito grande para a popularização dos recursos de IA. Uma das iniciativas que se destacam pertence ao ecossistema *Google*, nomeado como [Google Cloud AutoML](#). O *AutoML* permite que até as pessoas que não são programadoras e que tenham o conhecimento não aprofundado em ML possam ser capazes de construir modelos de alta qualidade em alguns minutos, utilizando toda a infraestrutura disponibilizada pela *Google*. A linha AutoML é formada por seis vertentes:

- *Vertex AI*: uma plataforma centralizada e integrada para apoiar a criação, implantação e escalonamento de modelos de IA.
- *AutoML Vision*: criação de modelos para detecção e classificação de objetos.
- *AutoML Video Intelligence*: permite o processamento e a análise com IA aplicada em vídeos para extração de informações.
- *AutoML Natural Language*: criação de IA para análise da estrutura e significado de textos.
- *AutoML Translation*: detecção e tradução dinâmica de conteúdo de um idioma para outro.
- *AutoML Tables*: ambiente para criação e implantação automática de modelos de *machine learning* com dados estruturados.

Essas soluções possuem interfaces simples para a construção de modelos, podem ser utilizadas para aplicação em ambientes complexos, por meio da API oferecida com o modelo treinado ou utilizando a execução em nuvem e interação WEB. No exemplo a seguir, utilizamos a plataforma *AutoML Vision*.

Vamos construir uma IA usando a plataforma AutoML Vision. Embora seja uma interface simplificada e intuitiva, essa plataforma faz uso da TensorFlow,

uma das mais utilizadas bibliotecas de IA. Os algoritmos são disponibilizados na plataforma AutoML Vision em uma interface simples, mas que permite a criação de aplicações reais para serem utilizadas em ambiente real. Neste exemplo, pretendemos construir uma IA capaz de identificar pessoas e classificar o seu nome. Tal aplicação pode ser empregada em diversos ambientes, permitindo identificar os rostos ou controlar o fluxo de pessoas.

Vale ressaltar que o objetivo desta prática pretende demonstrar uma possível aplicação real do uso de IA. Temas muito importantes, tais como: éticas, morais e legais, devem ser considerados na construção de uma aplicação real.

Siga os passos para construir a sua própria IA capaz de identificar e classificar o nome de pessoas. Para tal, você precisa ter cadastro no ambiente Google. Se tiver, já pode efetuar o seu login no Google.

AVISO IMPORTANTE: A plataforma Google AutoML Vision é um ambiente avançado, comercial e pago, utilizado por muitos produtos. Neste exemplo, serão utilizados os créditos gratuitos oferecidos pelo Google no modelo Free Trial. No momento da escrita, estes créditos se limitam à 40 horas de nó (em inglês, *Node Hours*) como pode ser visto em <https://cloud.google.com/vision/automl/pricing> e pode ser utilizado apenas uma vez. Para evitar surpresas, lembre-se de remover os modelos após a execução e teste.

1. **Aquisição dos dados:** esta etapa é uma das mais importantes para qualquer IA. Vamos utilizar um conjunto *opensource* de imagens de celebridades que pode ser adquirido na [plataforma Kaggle](#). Ao acessar este link, escolha a opção “Download (3MB)” para iniciar o download. Pode ser necessário efetuar o login para realizá-lo, mas você pode utilizar seu próprio e-mail da Google para isso, basta escolher a opção “Sign in with Google”. Após efetuar o download, salve em uma pasta de seu computador e descompacte o arquivo ZIP. Você verá uma estrutura composta por 3 pastas. Você irá utilizar apenas a pasta *TRAIN*, que apresenta imagens de celebridades. Para facilitar o processo, procure utilizar imagens de apenas duas celebridades. Em cada pasta, é possível notar entre 14 e 22 imagens, que representam um pequeno conjunto de dados para aplicação de reconhecimento, porém para esta atividade será suficiente. Salientamos que este processo se aplica para fotos de outras pessoas, com qualquer quantidade.

2. **Registro na plataforma Google AutoML:** acesse o site da [Google Cloud](#) para criar seu projeto. Faça o login com seu e-mail Google; caso não o tenha, faça o seu cadastro gratuitamente. Salientamos, aqui, que o Google oferece \$300 créditos gratuitos, por um período de 3 meses.

Selecione o botão ATIVAR e, nas próximas telas, insira as informações: país e número de telefone celular (para realizar a verificação). Neste passo, será necessário inserir dados do seu cartão de crédito para confirmar sua identidade. Apenas após ativar o período de avaliação, será possível prosseguir. Destacamos que os créditos são válidos por 90 dias!

Defina um nome para o projeto. Pode ser necessário incluir uma conta de faturamento para prosseguir; porém não será realizada a cobrança ao utilizar neste exemplo, pois serão utilizados os recursos gratuitos. Após preencher, selecione a opção CRIAR.

Certifique-se de que o projeto que acabou de criar está selecionado; caso não esteja, basta selecioná-lo. Você pode buscar pelo recurso AutoML Vision ou, então, acessar pelo [link](#). Selecione a opção ATIVAR A API AUTOML. O processo de ativação iniciará e pode levar alguns segundos para a sua conclusão.

3. **Inserção do conjunto de amostras:** após a ativação da API AutoML, selecione o item CONJUNTO DE DADOS no painel, à esquerda, conforme Figura 9.

Vision		Conjuntos de dados		NOVO CONJUNTO DE DADOS
Painel		<input checked="" type="radio"/> Nome Tipo Total de Imagens Imagens rotuladas Última atualização		
Conjuntos de dados		Nenhuma linha a ser exibida		
Modelos				

Figura 9 - Item Conjunto de Dados / Fonte: a autora

Descrição da imagem: trata-se da mesma página anterior, só que agora a opção “conjuntos de dados”, localizada do lado esquerdo, está selecionada. Do lado direito, aparece “Conjunto de dados” e a opção de “Novo conjunto de dados”. Abaixo, está localizado o seguinte: “Nome”, “Tipo”, “Total de Imagens”, “Imagens rotuladas” e “Última atualização”.

Selecione a opção NOVO CONJUNTO DE DADOS no cabeçalho da página. Será aberta uma nova janela, você deve definir um nome para o novo conjunto de dados. Após, selecionar a opção CLASSIFICAÇÃO DE RÓTULO ÚNICO para permitir identificar e classificar um único objeto nas imagens. Neste caso, o rosto e nome das pessoas. Na sequência, selecione CRIAR CONJUNTO DE DADOS.

Na sequência, você deve fazer o upload das imagens de uma classe. Faça o *upload* da classe, uma por vez, para que seja possível definir os rótulos de cada uma delas. Clique em SELECCIONAR ARQUIVOS e escolha todas as imagens de uma única pessoa. Também é necessário definir um dispositivo de armazenamento em nuvem do Google, então clique em *BROWSE*.

Defina um nome para o **Bucket** (em português, balde), e clique na opção AVANÇAR nas telas a seguir. Na segunda etapa, para escolher o local para armazenar seus dados, selecione o tipo de Local (em inglês, “Region”) e defina como “us-central1 (Iowa)” e, finalmente, avance selecionando CONTINUAR. Uma configuração incorreta do local pode ocasionar erro na etapa seguinte. Embora existam outras configurações, não precisa alterá-las para esse exemplo.

Após definir o nome, selecione CRIAR. A janela anterior será exibida, basta escolher a opção SELECCIONAR. Ao voltar à tela inicial, a opção CONTINUAR estará habilitada para ser escolhida. Ao clicar em SELECCIONAR, o *upload* será realizado, pode demorar alguns segundos. Após um período de tempo, será possível escolher a opção IMAGENS, para verificar as imagens que você inseriu. Caso não tenha finalizado, aguarde a conclusão. Em alguns casos, pode ser necessário recarregar a página (pressione F5, *refresh* ou recarregue a página).



Figura 10 - Opção "Todas as Imagens" / Fonte: a autora.

Descrição: a imagem apresenta ao topo as opções “Importar”, “Imagens”, “Treinar”, “Avaliar” e “Testar e Usar”. O fundo da página é branco. Do lado esquerdo, há uma lista, sendo “Todas as imagens” com o número 17 na frente; “Com rótulo” e o número 0; e “Sem rótulo” e o número 17. A seguir, consta a opção de “Filtro” e um botão clicável “Adicionar novo rótulo”. Do lado direito, aparece a opção de “Selecionar tudo” e 3 fotos de Elton John.

Pode ser que seja disparado um erro ao final deste processo (similar ao apresentado na Figura 11), informando que não é possível mover arquivos entre regiões e solicitando que seja utilizado um *Bucket* regional na mesma localização que está sendo executado o Google AutoML. Caso este erro ocorra, retorne ao passo anterior. Na aba IMPORTAR, selecione a opção BROWSE para criar um novo Bucket. Desta vez, na etapa para escolha do local para armazenar os dados, selecionar o tipo de local “Region” (em português, região) e definir o local exato indicado no erro, e avance com CONTINUAR.

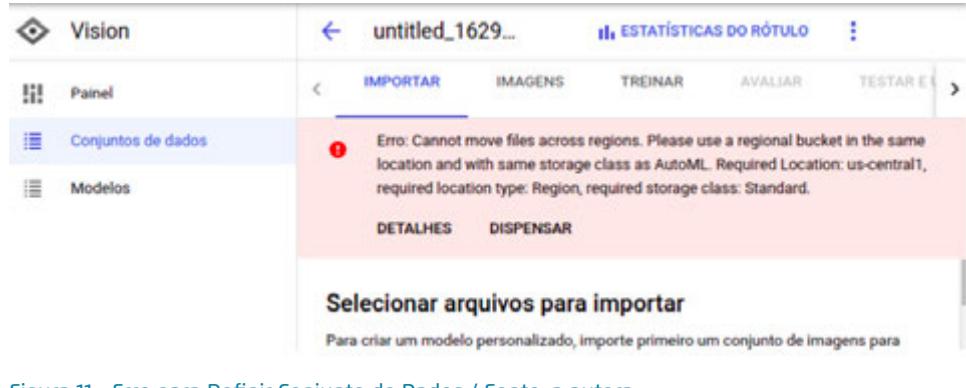


Figura 11 - Erro para Definir Conjunto de Dados / Fonte: a autora.

Descrição: o fundo da página é branco. Do lado esquerdo há uma lista de opções, sendo “Vision”, “Painel”, “Conjunto de dados” e “Modelos”. A opção “conjuntos de dados” está selecionada. Do lado direito aparece “untitled_1629” e a opção de “Estatística do Rótulo”. Abaixo está localizado o seguinte: “Importar”, “Imagens”, “Treinar”, “Avaliar” e “Testar e Usar”. Abaixo, encontra-se um sinal vermelho indicando erro.

Atenção que cada imagem deve pertencer a uma classe específica, basta selecionar a opção ATRIBUIR RÓTULOS. Procure criar dois rótulos para esta tarefa. Cuidado que os rótulos não devem ter acentos, espaço ou caracteres especiais.

Após finalizar a criação dos rótulos e fazer o *upload*, você deve selecionar a opção SEM RÓTULO e, finalmente, escolher a opção SELEÇÃO TUDO. Com isso, ficará visível a opção ATRIBUIR RÓTULOS. Será exibida uma lista com todos os rótulos para que seja selecionado aquele correspondente às imagens. Marque o rótulo correto e selecione SALVAR. Faça o mesmo para todas as demais imagens, repetindo o processo de *upload*, selecionando imagens sem rótulos, selecionando todas e atribuindo o rótulo. Clique na opção IMPORTAR no menu horizontal. Clique em SELEÇÃO ARQUIVOS e selecione outras imagens.

Clique em BROWSE para definir o *Bucket* do Google Cloud a ser utilizado e selecione a mesma pasta utilizada no passo anterior. Para concluir, pressione SELECT. Agora já é possível clicar em CONTINUAR para iniciar o processo de upload das imagens. Na sequência, aperte no item do menu IMAGENS e aguarde o término da importação. O processo pode levar alguns instantes. Em alguns casos, pode ser necessário recarregar a página (pressione F5, refresh ou recarregue a página). Observe se todas as imagens estão rotuladas. Caso não estejam, repita o processo para todas as imagens sem rótulos.

4. **Treinar a IA:** depois que todas as imagens foram inseridas e devidamente rotuladas, pode-se iniciar o processo de treinamento da IA. Selecione a opção TREINAR no menu horizontal.

Atenção que, neste exemplo, para facilitar o acompanhamento, será utilizado um conjunto de dados pequeno, com poucas amostras. O recomendável é que exista, pelo menos, 100 imagens de cada classificação para conseguir resultados mais precisos e robustos.



Figura 12 - Marcar Imagens para Treinamento / Fonte: a autora.

Descrição da imagem: no topo, há três opções, sendo: "rostos_celebridades", "Estatísticas do rótulo" e "Exportar dados". Abaixo, encontram-se "Importar", "Imagens", "Treinar", "Avaliar", em azul, e "Testar Usar". O fundo da página é branco. Abaixo há o título "Marque mais imagens antes do treinamento", a seguir, temos "Imagens não rotuladas não são usadas. Seu conjunto de dados será dividido automaticamente em conjuntos de treino, validação e teste" e "O ideal é que cada rótulo tenha pelo menos 10 imagens. Menos imagens normalmente resultam em uma precisão e um recall não exatos. Você também precisa ter pelo menos 8, 1, 1 imagens cada atribuídas aos conjuntos de treino, validação e teste". Abaixo encontram-se indicadores de "marcadores", "imagens", "treinar", "validação" e "testes", sendo Elton John com 17, 13, 1 e 3, respectivamente, e Madonna 19, 15, 1 e 3. Ao final, encontra-se um botão clicável "Iniciar treinamento".

Serão exibidas recomendações quanto à distribuição dos conjuntos de treinamento, validação e teste. Selecione INICIAR TREINAMENTO. Uma janela para configuração do ambiente de processamento será aberta. Utilize os valores padrões no modo "Cloud Hosted". Na sequência, será solicitada uma quantidade máxima de NODE HOURS (em português, horas por nó) para realizar o processamento.

Para a versão trial, terá 40 horas por nó de processamento gratuito. Mantenha 16 horas por nó, e marque a opção IMPLANTAR MODELO EM UM NÓ APÓS O TREINAMENTO para que fique disponível para uso assim que você finalizar o processo. Após essas configurações, selecione INICIAR TREINAMENTO.

O processo de treinamento e processamento será iniciado na sequência. Como é um processo que pode demorar, não é necessário aguardar, será enviada uma notificação de finalização no e-mail cadastrado. O e-mail conterá um link para voltar ao ambiente Google Auto ML. Na página, será possível observar parâmetros sobre a qualidade do modelo, tais como: índices de precisão e recall, os quais são melhores quando se aproximam de 100%.

Para testar, você precisa clicar em TESTAR e USAR. A infraestrutura Google, que foi utilizada para implantar a solução, será responsável por receber novas imagens e executar a classificação, utilizando o modelo de IA construído. Vale ressaltar que é possível submeter quaisquer imagens para testar, o resultado será sempre um dos rótulos utilizados para o treinamento da IA. Clique em UPLOAD IMAGES para testar com novas imagens.



AGORA É COM VOCÊ



Caro aluno(a), espero que você tenha compreendido o conteúdo apresentado nesta unidade. Sabemos que o assunto é complexo e que apresenta muitos conceitos. Sugerimos, então, que você construa um Mapa Conceitual para relembrar os diversos conceitos apresentados e, principalmente, determinar a relação entre esses conceitos. O Mapa conceitual é muito utilizado para sintetizar os conceitos e torná-los representativos visualmente, permitindo que você exerçite o seu pensamento crítico! Uma sugestão de ferramenta gratuita é o www.goconqr.com.

MEU ESPAÇO





Machine Learning

Dra. Iara Carnevale de Almeida

Esp. Ronie Cesar Tokumoto

OPORTUNIDADES DE APRENDIZAGEM

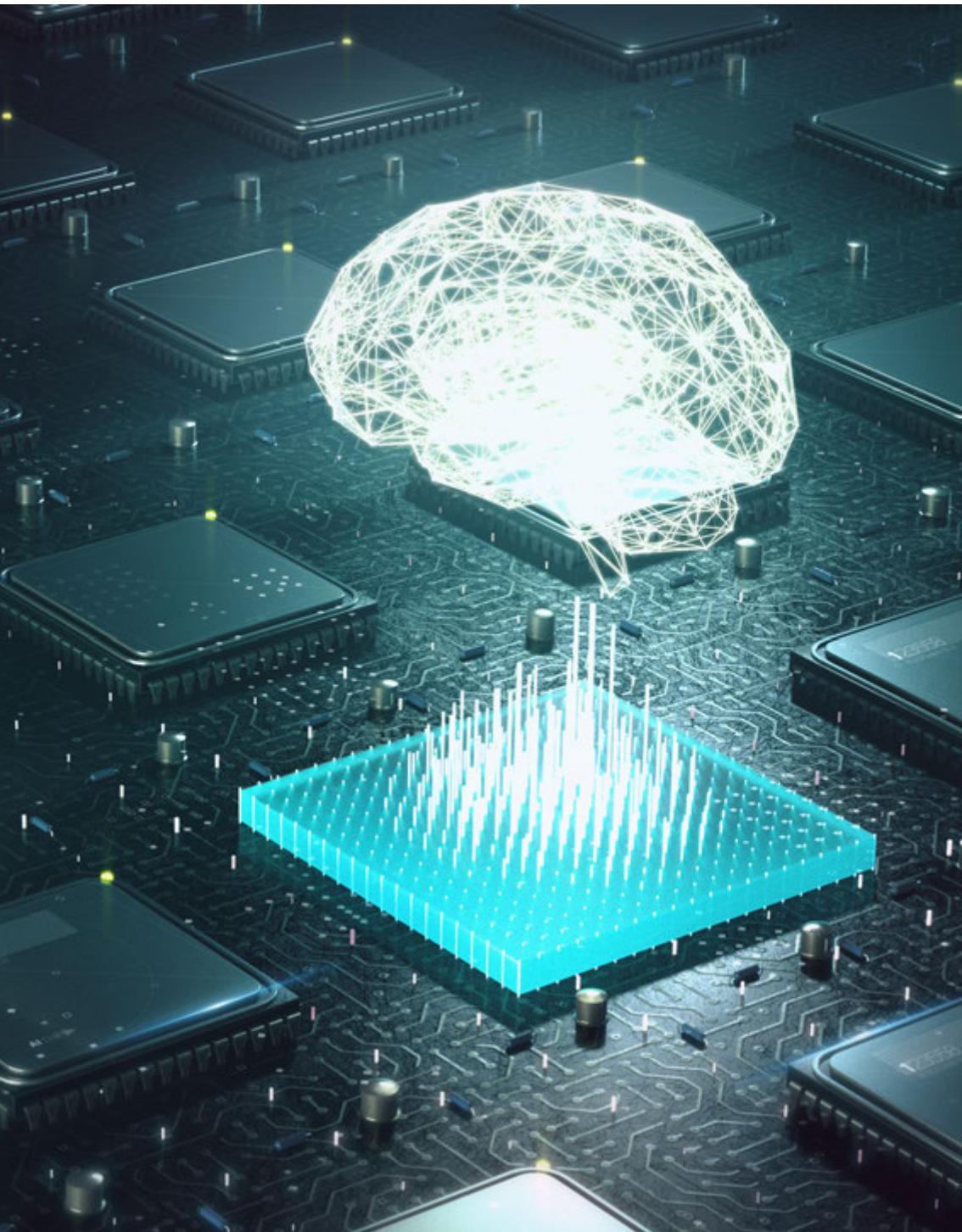
Quando se estudam os tópicos relacionados à área de inteligência artificial, imagina-se logo que serão criados mecanismos para que as máquinas possam ser automatizadas. Este processo, porém, pode não ser tão simples, pois, assim, como a mente humana, existem diferentes níveis de automatização que dependem de processos que envolvem aprendizado para, posteriormente, poder realizar uma tomada de decisão sobre os dados coletados. Esta unidade visa a apresentar conteúdo referente às técnicas utilizadas no aprendizado não supervisionado, técnicas essas aplicadas na área do *Machine Learning*.

Imagine que você está em casa, e a televisão liga (sozinha) no canal que você gosta de assistir um determinado programa. Ou, então, você está dirigindo, seu telefone está fora de alcance e você precisa fazer uma ligação urgente. O que pode fazer nesta situação? Acionar o seu telefone via comando de voz. Estas ações já são, provavelmente, parte do seu dia a dia e aprimoradas pelas técnicas e algoritmos de *Machine Learning* — ML (em português, Aprendizado de Máquina).

O aprendizado de máquina baseia-se no reconhecimento de padrões, sem a necessidade de a máquina ser programada, previamente, para que possa realizar tarefas específicas. O ML surgiu pelo interesse de pesquisadores da área de Inteligência Artificial — IA (em inglês, *Artificial Intelligence*), os quais pretendiam que os computadores aprendessem, por meio dos dados de entrada. A interação é um aspecto importante do ML, pois a máquina é exposta a novos dados, aprende com esses dados e adapta-se de forma independente. Com isso, a máquina pode tomar decisões confiáveis!

Importante ressaltar que a base do ML é a compreensão sobre a estrutura de dados, por meio de seus algoritmos. De forma geral, o ML segue a premissa de ser um software e de respeitar um conjunto de regras. A partir dessas regras, o ML pode construir estruturas que representam o aprendizado adquirido. Será, porém, que existem técnicas que possibilitam o aprendizado em condições desfavoráveis, tais como: os dados serem poucos ou que não atendam minimamente aos padrões estabelecidos, de forma que o processo de aprendizado possa ocorrer de forma satisfatória?





O *Machine Learning* tem contribuído na automatização de processos que são realizados, manualmente, pelo ser humano. Relembre algumas funcionalidades presentes no nosso dia a dia que utilizam essa tecnologia:

- Recomendação de vídeos do YouTube, conforme o seu gosto.
- Sugestões de filmes disponíveis no Netflix, conforme suas escolhas prévias.
- Preenchimento automático, quando você está digitando em, por exemplo, aplicativos de mensagem ou em buscadores online.
- Carros que se movem sem a necessidade de condutor, conhecidos como carros autónomos.
- Reconhecimento automático de rosto, lugar, comida, objeto e animal apresentado em, por exemplo, Google Fotos ou Facebook.

Cada vez mais, detecta-se forte evolução em diversas áreas que compõem a área da informática para que a automatização de processos ocorra, tais como: robótica, inteligência artificial, engenharia de sistemas, entre outras.

As soluções apresentadas exigem o processamento de grande volume de dados! As pessoas são capazes de realizá-las. Contudo, quanto maior o número de dados, maior o tempo necessário para executá-las!

Lembre-se de que, na Unidade 1, deste livro (página 21), foi salientado que “a aquisição de novos conhecimentos, o desenvolvimento de técnicas motoras e cognitivas, por meio de instrução ou prática, a generalização de conhecimentos adquiridos, a representação do conhecimento e a descoberta de novos fatos/teorias são alguns dos processos da aprendizagem”. Além disso, foram apresentadas técnicas de aprendizado nesta mesma unidade, as quais salientamos:

- Aprendizado supervisionado: trabalha com o objetivo de encontrar ótimos parâmetros que possam trabalhar com um modelo de conjunto de testes de rótulos desconhecidos.
- Aprendizado não supervisionado: trabalha quando não são fornecidos os valores específicos de saída, objetiva analisar semelhanças entre objetos e agrupá-los.

Comparativamente, o aprendizado supervisionado procura prever variáveis dependentes com base nas variações das variáveis independentes, por exemplo, prever a venda de produtos (dependente), de acordo com a sazonalidade ou o clima (independentes). No caso do aprendizado não supervisionado, a análise é mais abstrata, por exemplo, em organizações do setor comercial, tais como: supermercados ou lojas de departamentos, onde os clientes podem ser agrupados conforme os seus perfis de compra e, assim, fazer marketing direcionado.

Note que o aprendizado não supervisionado permite que ocorra o aprendizado com dados não tratados, incompletos ou pouco adequados. Vamos, então, compreender como isso ocorre?

Desafio! Procure compreender como o aprendizado não supervisionado consegue buscar por dados (de forma aleatória, sem filtros ou critérios de busca) e, a partir desses dados, gerar novos dados que sejam relevantes para o contexto em que esses dados foram obtidos.

Sugiro que o faça na organização onde você trabalha, procure detectar a fonte dos dados. São de diferentes fontes, tais como: relatórios, blogs, e-mails, entre outros. A partir desses dados e de forma intuitiva, procure detectar os dados que não estão explícitos (ou seja, que não se obtém diretamente), tais como: tendências e probabilidades, apenas com o objetivo de se compreender como os algoritmos podem ser criados, a partir de percepções humanas sobre dados.

Procure refletir sobre os aspectos éticos que estão relacionados ao Aprendizado de Máquina, visando a compreender como deve ser aplicado o AM, assumindo que esta atividade ocorre em paralelo com a atividade humana. Pense nos benefícios que um sistema baseado em AM pode trazer para a sociedade. Não deixe, porém, de refletir sobre os possíveis transtornos nessa mesma sociedade.

Um dos pontos positivos a se considerar é o fato de que o AM possibilita a automatização de processos nos mais diferentes setores — industrial, comercial e educacional —, poupando as pessoas de realizar atividades repetitivas. No entanto considero importante pensar em soluções para que essas pessoas possam assumir novas atividades.

DIÁRIO DE BORDO

Assume-se, antão, que um **processo**, conforme Davenport (1994), é visto como um conjunto de atividades estruturadas e medidas, que devem ter como resultado um produto específico, o qual atenda um cliente, uma organização ou mercado. Além disso, para este mesmo autor, o processo deve ter suas atividades de trabalho organizadas no tempo e no espaço, com especificação clara tanto em relação ao início e fim quanto da identificação clara de suas entradas e saídas.

A **automação de processos** procura, de forma gradativa, que as atividades rotineiras e manuais dos processos sejam apoiadas ou, até mesmo, substituídas, por recursos tecnológicos que facilitem a operacionalização destes processos. Relembre que a automatização de processos ocorre desde quando os seres humanos buscam adaptar e aprimorar as suas atividades, com apoio das evoluções tecnológicas. Essas evoluções permitem a diminuição no tempo da realização das atividades e, muitas vezes, dos custos envolvidos.

Nesta unidade, focaremos em métodos e técnicas baseados no aprendizado não supervisionado que, de forma geral, entende que o aprendizado

ocorre, por meio da análise de padrões, conceitos ou outros aspectos, sem a necessidade de que o processo seja supervisionado. Esta ramificação do *Machine Learning* procura aproveitar dados de testes, que não tenham sido tratados, de forma a classificá-los, categorizá-los ou rotulá-los. Portanto, o aprendizado não supervisionado atua, de forma mais independente, na avaliação de novos dados, pois procura por similaridades nos dados já analisados, de forma que os novos dados possam gerar aprendizado.

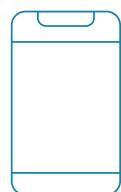
Note que o aprendizado não supervisionado é aplicado em situações em que não se tem uma base completa para realizar uma avaliação precisa sobre os resultados obtidos. Isso, normalmente, ocorre na prestação de serviços no setor do comércio, em que é importante obter referências sobre a concorrência (para poder avaliar os impactos no mercado e as operações necessárias), realizar previsão de vendas, identificar a sazonalidade e aceitação dos clientes, entre outros. Essa análise sobre os dados permite apoiar uma melhor tomada de decisões, afetando positivamente as campanhas de marketing e os setores de logística, produção e demais setores envolvidos.

Para trabalhar com dados na área de Ciência de Dados, devemos considerar que, quanto mais os dados estiverem organizados e “limpos”, melhor será o posterior processamento e, por consequência, os resultados obtidos. Para que isto ocorra, existem métodos que possibilitam preparar os dados das amostras.



NOVAS DESCOBERTAS

Este livro (em português, *Princípios básicos do aprendizado de máquina, continuação: construindo seu modelo de aprendizado da primeira máquina*) foi projetado a ser um guia para servir de suporte e referência para praticantes iniciantes que desejam começar a construir seu primeiro modelo preditivo de Machine Learning. Ao final do guia prático, espera-se que os leitores sejam capazes de compreender: (a) as principais considerações, ao preparar um conjunto de dados para o aprendizado de máquina; (b) valor da modelagem rápida, usando Dataiku AutoML, e como ela pode ser a base de seu primeiro modelo de aprendizado de máquina; (c) conceitos para avaliar e ajustar um modelo na Plataforma Dataiku. Clique aqui para saber mais detalhes sobre esta plataforma.



Na sequência, conforme Aliguer (2019, on-line), apresentamos algumas técnicas utilizadas no aprendizado não supervisionado:

- **Clusterização:** procura dividir um conjunto de dados em grupos, de acordo com alguma similaridade. Contudo, muitas vezes, esta técnica superestima a similaridade entre os grupos, com base em um algoritmo programado, não tratando os pontos de dados como individuais. Portanto, a clusterização não supervisionada é mais indicada para tarefas de segmentação desses dados.
- **Detecção de anomalias:** permite descobrir quais são os pontos de dados que não são comuns em um conjunto de dados. Esta técnica é útil, pois possibilita identificar transações fraudulentas, descobrir peças de hardware que estão defeituosas ou identificar um erro humano, durante a entrada de informações.
- **Mineração de dados de regras de associação:** permite identificar conjuntos de itens que ocorrem juntos e com frequência no seu conjunto de dados. Uma aplicação desta técnica ocorre junto aos varejistas para que possam analisar seus produtos inseridos nos carrinhos de compra. Esta análise permite descobrir os itens que são comprados juntos e com frequência, possibilitando, assim, desenvolver estratégias eficazes para promover a venda desses produtos.
- **Modelos variáveis latentes:** possibilita o pré-processamento dos dados, permitindo a redução do número de recursos em um determinado conjunto de dados ou, até mesmo, a decomposição deste em vários componentes.

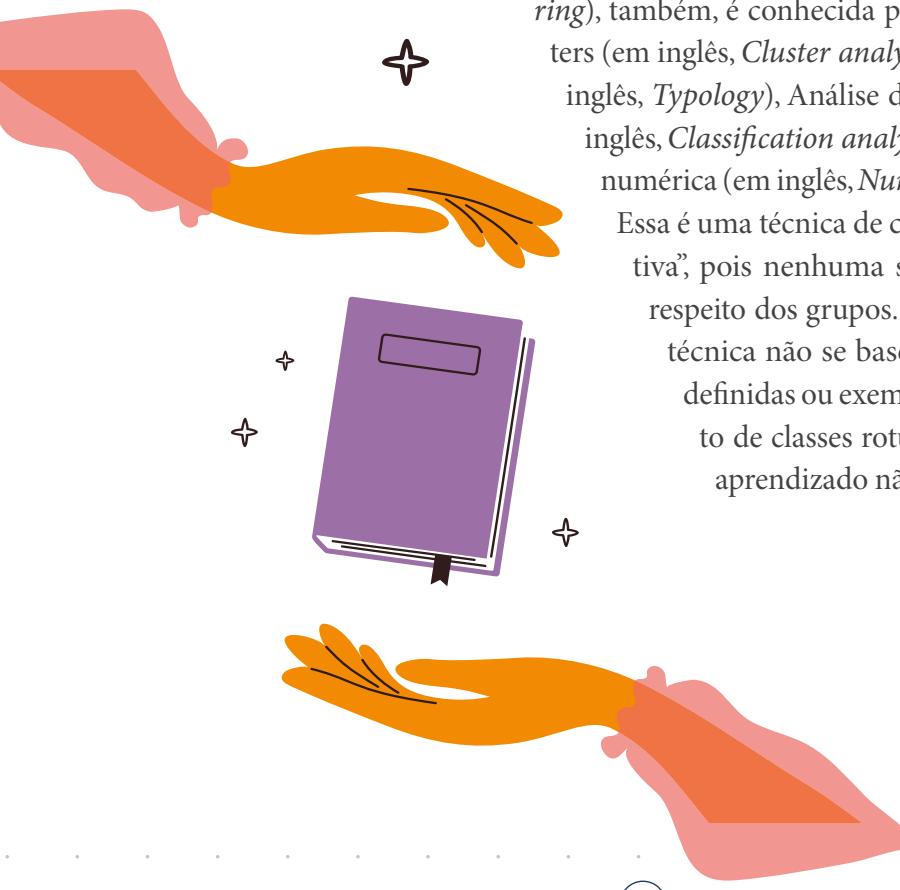
As técnicas, aqui, apresentadas permitem descobrir novos padrões, podendo melhorar o processamento sobre um grande volume de dados para, posteriormente, haver uma melhor tomada de decisão. São técnicas que envolvem categorização e classificação desses dados. Na sequência, detalhamos os algoritmos que permitem o implementar dessas técnicas. Vamos, então, conhecer um pouco mais sobre essas técnicas?

**EXPLORANDO IDEIAS**

Mitchell (1997) faz uma reflexão sobre o real funcionamento do aprendizado de máquina. O autor salienta que, no período de 1987-1997, o AM evoluiu de um campo de demonstrações de laboratório para um campo de valor comercial significativo. Algoritmos de AM aprenderam a detectar fraudes de cartão de crédito por mineração de dados em transações anteriores; a dirigir veículos, autonomamente, em rodovias públicas a 70 milhas por hora; e os interesses de leitura de muitos indivíduos para montar eletrônicos personalizados.

Esse mesmo autor afirma que surge uma nova teoria computacional de aprendizagem que começava a lançar luz sobre questões fundamentais, tais como: compensação entre o número de exemplos de treinamento disponíveis, o número de hipóteses consideradas e a precisão provável da hipótese aprendida. Atualmente, cada vez mais, pesquisas exploram muitas outras questões relacionadas à aprendizagem de máquina!

Fonte: adaptado de Mitchell (1997).



Na literatura, **Clusterização** (em inglês, *Clustering*), também, é conhecida por Análise de clusters (em inglês, *Cluster analysis*), Tipologia (em inglês, *Typology*), Análise de classificação (em inglês, *Classification analysis*) ou Taxonomia numérica (em inglês, *Numerical taxonomy*).

Essa é uma técnica de classificação “primitiva”, pois nenhuma suposição é feita a respeito dos grupos. Relembre que essa técnica não se baseia em classes pre-definidas ou exemplos de treinamento de classes rotuladas, por realizar aprendizado não supervisionado.

Qualquer método de clusterização é definido por um algoritmo específico que determina como será feita a divisão de casos em clusters distintos. Os métodos propostos na literatura são baseados na ideia de “distância” ou “similaridade entre as observações”. Procura-se definir a pertinência dos objetos em cada cluster, de forma a detectar o que cada elemento tem de similar em relação aos outros do mesmo grupo.

De que forma isso é feito? Os elementos que compõem um cluster devem apresentar forte similaridade, ou seja, são elementos muito parecidos ou que seguem um mesmo padrão) e não ter similaridade com elementos de outros clusters. Resumindo, toda clusterização pretende maximizar tanto a homogeneidade dentro de cada cluster quanto a heterogeneidade entre os clusters.

Uma das vantagens de aplicar clusterização é conseguir descrever, de forma mais eficiente e eficaz, as características inerentes de cada um dos clusters. Isso permite que haja maior entendimento sobre o conjunto de dados original, além de permitir o desenvolvimento de esquemas de classificação para dados novos e correlacionar os atributos desses dados que, muitas vezes, não são facilmente visualizados por um ser humano. Importante ressaltar que esta técnica pode realizar o pré-processamento para algoritmos de caracterização e classificação.

Hruschka e Ebecken (2001) apresentam uma definição formal do problema de clusterização. Considere um conjunto X de n objetos tal que $X = \{X_1, X_2, \dots, X_n\}$ é um vetor de p medidas reais que dimensionam as características do objeto. Estes objetos devem ser clusterizados em k clusters disjuntos $C = \{C_1, C_2, \dots, C_k\}$, de forma que as seguintes condições sejam respeitadas:

1. $C_1 \cup C_2 \cup \dots \cup C_k = X$;
2. $C_i \neq \emptyset, \forall i, 1 \leq i \leq k$;
3. $C_1 \cap C_2 = \emptyset, \forall i \neq j, 1 \leq i \leq k, 1 \leq j \leq k$.

Enfatiza-se que, pelas condições apresentadas, um objeto não pode pertencer a mais de um cluster (ou seja, os grupos são disjuntos) e cada cluster tem que ter, ao menos, um objeto. Além disso, as aplicações de clusterização pretendem cumprir, pelo menos, um dos seguintes objetivos principais:

- Identificação da estrutura subjacente: intuir sobre os dados, determinar hipóteses, detectar anomalias e determinar características marcantes destes dados.
- Classificação natural: identificar a semelhança entre os objetos.
- Compressão: organizar os dados e especificar protótipos do cluster.

Normalmente, o valor de k (o número de clusters) é desconhecido; caso contrário, o problema passa a ser referido como k-Clusterização. Ressalta-se que o problema de encontrar o melhor agrupamento é NP-completo e não é computacionalmente possível encontrá-lo, a não ser que n (o número de objetos) e k sejam extremamente pequenos, visto que o número de partições distintas, em que podemos dividir n objetos em k clusters, aumenta aproximadamente como $kn / n!$

Importante que você compreenda que existem diferentes tipos de agrupamento de cluster ou grupos, exemplificados nas Figuras 1 e 2.

1. ***Exclusive Cluster*** (em português, Grupo Exclusivo): agrupamento em que os registros são exclusivos, ou seja, cada registro pertence a um único grupo.
2. ***Overlapping Cluster*** (em português, Cluster Sobreposto): agrupamento em que os registros podem pertencer a mais de um grupo ou cluster.
3. ***Complete or Partial Cluster*** (em português, Cluster completo ou parcial): agrupamento que consegue separar (ou não) todos os registros.
4. ***Homogeneous or Heterogeneous Cluster*** (em português, Cluster homogêneo ou heterogêneo): agrupamento que apresenta (ou não) tamanhos, formas e densidades similares.
5. ***Hierarchical Cluster*** (em português, Cluster hierárquico): agrupamento que apresenta uma hierarquia entre os grupos que podem conter subgrupos e outros registros.

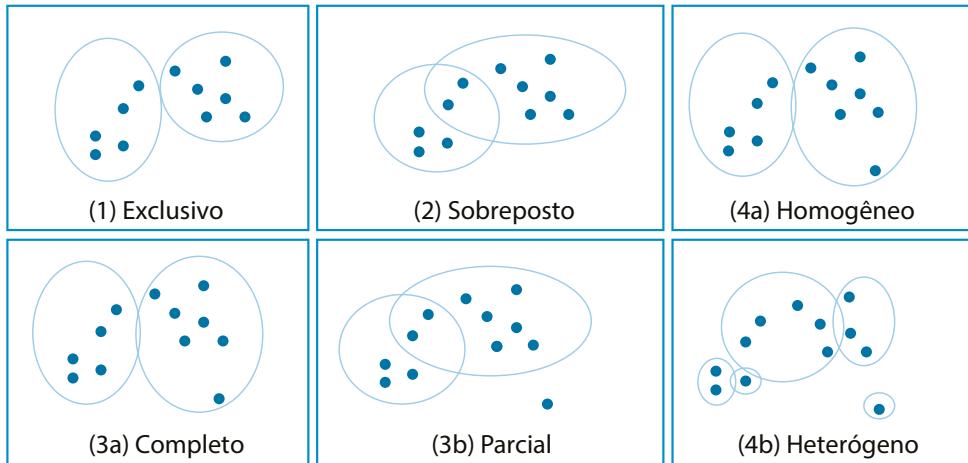


Figura 1 - Tipos de Agrupamento / Fonte: os autores.

Descrição da Imagem: a figura pretende apresentar os tipos de agrupamento, para tal são apresentadas seis imagens que possuem sempre dois elementos: círculos, que representam os clusters, e pontos, que representam os elementos dos clusters. A imagem (1) apresenta um exemplo do tipo exclusivo, por isso, mostra dois círculos, cada um deles com pontos dentro. Os círculos estão encostados. A imagem (2) apresenta um exemplo do tipo sobreposto, esta imagem possui dois círculos que possuem alguns pontos. Estes círculos se interseccionam (tal como a intersecção de conjunto da matemática) e alguns pontos estão, também, na interseção. A imagem (3a) apresenta o exemplo do tipo completo, a figura é similar ao exemplo do tipo exclusivo. Salienta-se que todos os pontos estão dentro dos clusters. A imagem (3b) apresenta o exemplo do tipo parcial. O que esta imagem salienta é que um ponto está fora dos círculos, indicando que um elemento não pertence a nenhum dos clusters. A imagem (4a) apresenta o exemplo do tipo homogêneo. Nesta figura, temos os pontos distribuídos de forma homogênea nos dois clusters. Já a imagem (4b) apresenta o exemplo do tipo heterogêneo. Nesta, temos cinco clusters, cada cluster tem um número diferente de números.

Note que a Figura 1 exemplifica os tipos de cluster que são, relativamente, intuitivos de se compreender. Já a Figura 2 apresenta o cluster hierárquico. A imagem, à esquerda dessa figura, apresenta o cluster hierárquico, note que B e C compõem um cluster. Cluster esse que faz parte de um cluster maior, ao qual A também percebe. E, assim, por diante. A imagem, à direita, indica como esses clusters podem ser melhor visualizados, no formato de uma árvore.

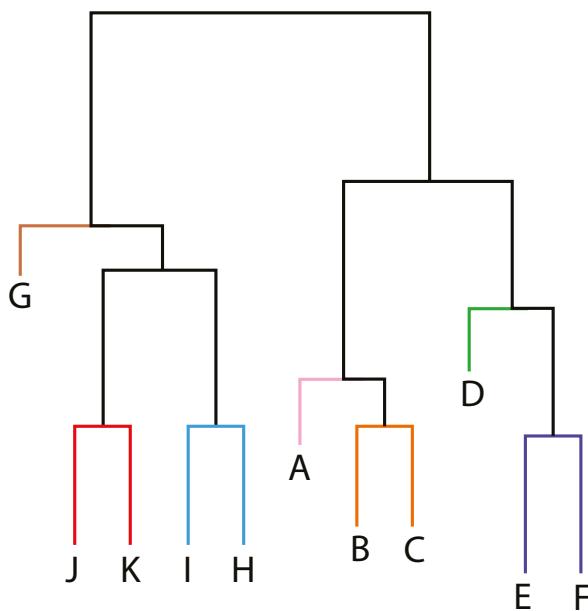
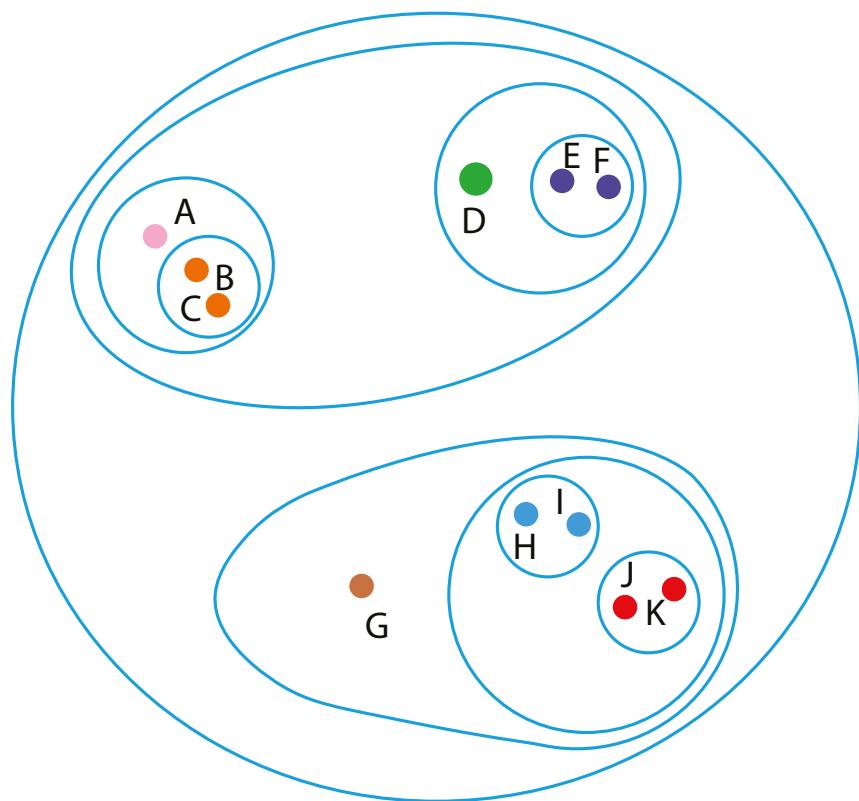


Figura 2 - Tipos de Agrupamento / Fonte: Estatística... ([2021], on-line).

Descrição da Imagem: na figura temos uma ilustração representando os tipos de agrupamentos, em que, à esquerda, há um círculo grande com outros círculos menores em seu interior, na parte superior deste círculo está o primeiro agrupamento com outros dois novos círculos médios (o que está localizado à esquerda e o que está localizado à direita), no círculo à esquerda há um ponto (rosa), identificado como "A", e em um círculo menor há outros dois pontos (alaranjados), identificados como "B" e "C". No círculo à direita há um ponto (verde), identificado como "D", e em um círculo menor há outros dois pontos (azul marinho) identificados como "E" e "F". Na parte inferior, está o segundo agrupamento com um ponto marrom identificado como "G", e em um círculo médio há outros dois novos círculos pequenos, o primeiro com dois pontos (azul claro) identificados como "H" e "I" e o segundo com dois pontos (vermelhos) identificados como "J" e "K". À direita da figura, está o organograma desses clusters, em que há uma linha no topo de onde se originam duas ramificações, à esquerda e outra à direita. A primeira cria outros dois novos caminhos, o primeiro refere-se a "G" e é apresentado na cor correspondente a este (marrom), e o outro origina outros dois caminhos, o primeiro leva a "J e K" apresentados na cor vermelha e o segundo leva a "I e H", apresentados na cor azul claro. A segunda ramificação com origem na linha "topo" cria outros dois novos caminhos, o primeiro leva a "A" e é apresentado na cor correspondente a este (rosa) e o outro leva a "B e C" apresentados na cor alaranjada e o segundo leva a "D" que é apresentado na cor verde e o outro leva a "E e F" apresentados na cor azul marinho.

Para exemplificar o método de Clusterização, temos o algoritmo K-means. A seguir, uma versão em linguagem natural desse algoritmo em que, para compreender o processo do algoritmo, são ignorados detalhes de semântica e de sintaxe das linguagens de programação. Na sequência, a Figura 3 apresenta a evolução desse algoritmo.

1. Leitura da base de dados.
2. Organização dos dados em colunas de uma estrutura de dados.
3. Divisão dos dados em dependentes e independentes.
4. Aplicação do método K-means nos dados das variáveis dependentes.
5. Os resultados de cada iteração podem ser armazenados em uma matriz, esta atribui uma característica adequada aos dados, isto é, uma indicação a qual cluster será agregado. Por exemplo, cada marca de carro em um diferente cluster, em uma base de análise de vendas de veículos, usado em determinada região, por exemplo.
6. Ao final das iterações, quando a condição de parada de distância entre novos centróides for atendida, a matriz conterá o estado final dos clusters organizados e melhores centróides encontrados, podendo então ser utilizada a matriz em novo algoritmo para análise de dados, geração de modelo gráfico representativo etc.

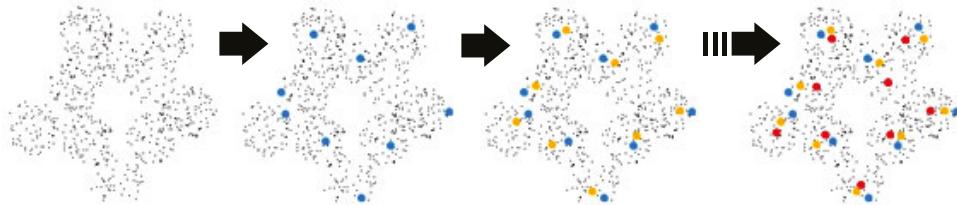


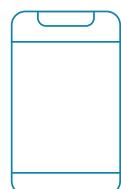
Figura 3 - Exemplo ilustrativo de execução do algoritmo K-means / Fonte: os autores.

Descrição da Imagem: a imagem traz uma sequência de passos do algoritmo iniciando com um conjunto de pontos. No passo seguinte, alguns pontos estão destacados em azul e, depois, outros próximos aos azuis, destacados em amarelo. Por fim, no último passo, outros pontos estão destacados em vermelho, além dos azuis e amarelos mantidos dos passos anteriores.

Existem muitos outros algoritmos que implementam o método Clusterização. Para compreender melhor esses algoritmos, sugiro que leia com atenção o material disponibilizado em *The 5 Clustering Algorithms Data Scientists Need to Know* (em português, *Os cinco algoritmos de clusterização que os cientistas de dados precisam saber*, clique [aqui](#) para acessar este material). Nele, você verá uma simulação para cada um dos algoritmos e, ao final, uma comparação da execução desses algoritmos para diferentes situações!



Além disso, procure praticar para melhor compreender os métodos e algoritmos, em alguma linguagem de programação como Python ou em outra linguagem de sua preferência. Para tal, sugerimos que busque por conteúdo complementar. Uma sugestão é o site Scikit-learn, disponível em: <https://scikit-learn.org/stable/modules/classes.html#module-sklearn.cluster>. Este site contém exemplos na linguagem Python, aplicando diversos algoritmos de clusterização baseados em diferentes algoritmos que possuem diferenças entre si, tornando-os mais adequados a diferentes situações no tratamento de dados para o aprendizado de máquina.



Vamos, agora, dissertar sobre a técnica de **detecção de anomalias** baseada em ML? Essa técnica pretende identificar ocorrências raras que acontecem em um conjunto de dados. Você sabe, porém, conceituar o que é anomalia? É uma observação muito diferente de outras observações e, por isso, suscita suspeita de que ela tenha sido criada por um recurso diferente das outras observações.

Os estudos, nesta área, são baseados na exploração dos dados e/ou treinamento de algoritmos, para que se possa distinguir dados normais de anomalias. A técnica utilizada depende não apenas do tipo de problema, mas também da natureza dos dados usados. Além de ser desafiadora para a área da ciência da computação, a detecção de anomalias pode ser aplicada para resolução de problemas nos mais diversos setores de negócio, tais como: utilities, meios de pagamentos, saúde e indústria. Portanto, há muita pesquisa e interesse em desenvolver algoritmos capazes de realizar essa tarefa. Ressalta-se que, similar ao aprendizado, as técnicas para detecção de anomalias são classificadas em:

- **Treinamento supervisionado:** assume-se a disponibilidade de um conjunto de dados de treinamento, classificados como normais ou anomalias. Os resultados são, então, mapeados, e o algoritmo replica a categorização dos dados entre anomalias e dados normais.
- **Treinamento semi-supervisionado:** tem-se um pequeno conjunto de dados, classificados previamente como anomalias e dados normais, e um conjunto maior de dados não classificados. A ideia é observar as similaridades entre os dados já classificados com os não classificados, colocando os dados similares nos mesmos grupos.
- **Treinamento não supervisionado:** essa técnica não exige dados classificados previamente. São utilizadas características intrínsecas dos dados para a identificação de dados normais e anomalias.

Na sequência, temos a **Mineração de dados de Regras de associação** (em inglês, *Association rules Data mining*). A descoberta de regras de associação em bancos de dados relacionais ou em data warehouses é uma das tarefas da linha de pesquisa denominada Mineração de dados (em inglês, Data mining - DM). A DM pertence ao campo da Ciência da Computação que tem por objetivo oferecer estratégias automatizadas para a análise de grande volume de dados gerados por organizações, procurando extrair informações implícitas e/ou desconhecidas, mas potencialmente úteis.

O conhecimento obtido pelas técnicas de DM é, geralmente, expresso na forma de regras e padrões. As **regras de associação** são um dos mais importantes tipos de conhecimento que podem ser minerados nas bases de dados; estas regras representam **padrões** de relacionamento entre os diferentes itens de uma base de dados. As atuais ferramentas para mineração de regras de associação funcionam da seguinte forma:

1. O usuário especifica a base de dados que deseja minerar.
2. O usuário especifica valores mínimos para as medidas de interesse, tais como: suporte, confiança e lift.
3. A ferramenta executa um algoritmo que analisa a base de dados e gera como saída um conjunto de regras de associação com valores de suporte e confiança superiores aos valores mínimos especificados pelo usuário.

Para compreender um pouco melhor sobre essa área, sugiro que leia o material “Data Mining de Regras de Associação” ([clique aqui para acessar este material](#)).

Finalmente, apresentaremos a técnica referente a **modelos variáveis latentes**. Antes de mais nada, precisamos compreender o que significa “variável latente”. A definição mais comum é: uma variável que não pode ser medida diretamente. Algumas aplicações desses modelos são:

- Avaliação ou medição estatística: pretende avaliar o desempenho, a capacidade e/ou a qualidade.
- Avaliação em ciências humanas e sociais: pretende a descrição de fenômenos humanos e sociais e a discriminação de pessoas ou de instituições.
- Comparação e/ou qualificação de organizações nos seus diferentes setores (industrial, comercial e educacional).

Portanto, salientamos que essas aplicações podem satisfazer alguns destes objetivos:

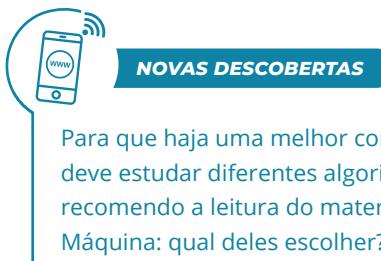
- Certificação ou qualificação de indivíduos.
- Acompanhamento e direcionamento de atividades, aprendizagem e de projetos.

Para uma melhor compreensão de como o aprendizado por máquina pode contribuir a essa temática, por exemplo, imagine o que se pretende quando indicamos que uma pessoa é prudente? De forma geral, significa que esta pessoa faz, por exemplo, seguro contra riscos, que não participa de jogos de azar, que

presta atenção por onde anda, entre outros cuidados. Se pudermos definir que estes hábitos caracterizam essa pessoa, por conveniência, podemos classificá-la como uma pessoa “prudente”, procurando adotar uma abstração para seus atos de forma a obter:

Imagen (variável latente, abstrata, teórica) → indicadores ditados pela experiência cotidiana.

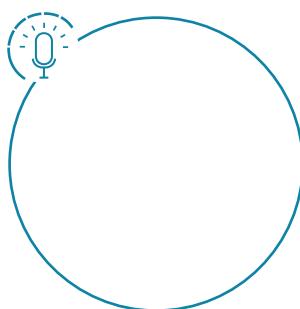
Note que, na verdade, não se espera que uma pessoa prudente sempre faça seguro contra riscos ou que evite os jogos de azar, mas que a **probabilidade** dessa pessoa em executar esses atos seja maior do que ocorreria com uma pessoa imprudente.



Para que haja uma melhor compreensão do *Machine Learning*, você deve estudar diferentes algoritmos e suas aplicações. Para iniciar, recomendo a leitura do material "Algoritmos de Aprendizagem de Máquina: qual deles escolher?".



Caro(a) aluno(a), esperamos que, com a apresentação de diferentes técnicas de aprendizado não supervisionado, você perceba as imensas oportunidades de aplicação. Salientamos que, nesta unidade, muitas vezes, as técnicas de *Deep Learning* estão implícitas, pois recorde que o *Deep Learning* aplica um conjunto dessas técnicas, mas com a pretensão de uma ter uma maior independência no processamento. Na Unidade 4, você compreenderá que aplicações são essas.



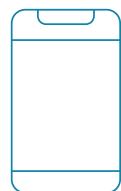
Caro(a) aluno(a), dentre os diversos assuntos desafiadores e relevantes desta unidade, nesta roda de conversa, pretendemos refletir sobre o uso de dados públicos e privados, visando à sua importância nos dias atuais. De forma simples e objetiva, pretendemos apresentar alguns usos práticos que salientam a importância dos dados, o quanto esses podem ser relevantes para uma melhor tomada de decisão. Reforçamos, ainda, a importância do desenvolvimento de aplicações baseadas em algoritmos estatísticos que estão disponíveis para as organizações de diferentes setores.

Um sistema pode aprender e melhorar o processamento dos dados, sem haver a necessidade de explicitar a programação para que isto ocorra. Para tal, existem algoritmos de *Machine Learning* que, por exemplo, utilizam estatística avançada para fazer previsões precisas.

Existem bibliotecas e linguagens de programação que facilitam o desenvolvimento de algoritmos de inteligência artificial e, mais especificamente, aprendizagem de máquina. A maioria dos algoritmos tem exigido competências e habilidades que, normalmente, os especialistas em desenvolvimento na área de Tecnologia da Informação possuem. Mais recentemente, porém, são disponibilizados ambientes de desenvolvimento que permitem rápido e descomplicado uso de IA, via interações simples. Um desses ambientes que possibilita o aprendizado e a construção de prática, nesta trilha de aprendizagem, é a plataforma Teachable Machine, desenvolvida pela Google.

O Teachable Machine visa a que qualquer pessoa possa criar um sistema de forma que possa reconhecer sua imagem, sons ou, até mesmo, gestos. Toda experiência é realizada por meio de uma interface amigável e intuitiva, permitindo a criação de modelos complexos que podem, posteriormente, ser utilizados e integrados com sites e aplicativos, por exemplo. Vamos, então, experienciar essa plataforma? Procure realizar os passos apresentados no tutorial a seguir.

PASSO 1 - Acesse o site, não precisa fazer nenhum cadastro, nem mesmo instalar software ou plugin. O processo é simples, rápido e descomplicado. A imagem, a seguir, demonstra os recursos e as possibilidades que você verá na plataforma.



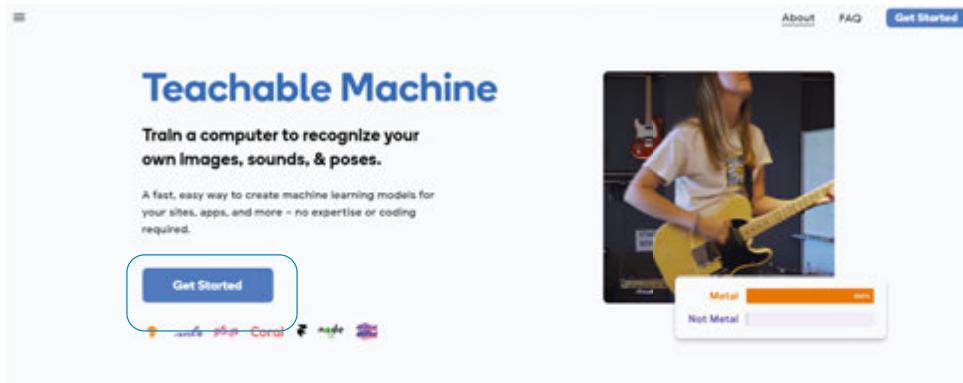


Figura 4 - Exemplo: Passo 1 / Fonte: os autores.

Descrição da Imagem: trata-se do print de uma página em que, ao lado direito, há uma moça de camiseta branca segurando uma guitarra amarela com detalhe preto. Ao lado esquerdo, há escrito em azul "Teachable Machine"; abaix, de cor preta em negrito, encontra-se "Train a computer to recognize your own images, sounds, & poses". Abaixo, encontra-se, escrito em preto "a fast, easy way to create machine learning models for your sites, apps, and more - no expertise or coding required". Mais abaixo, encontra-se um botão clicável azul escrito "Get Started".

PASSO 2 - Procure explorar a plataforma, clique no botão Azul denominado ***Get Started*** (em português, **Iniciar**), conforme destacado na imagem a seguir.

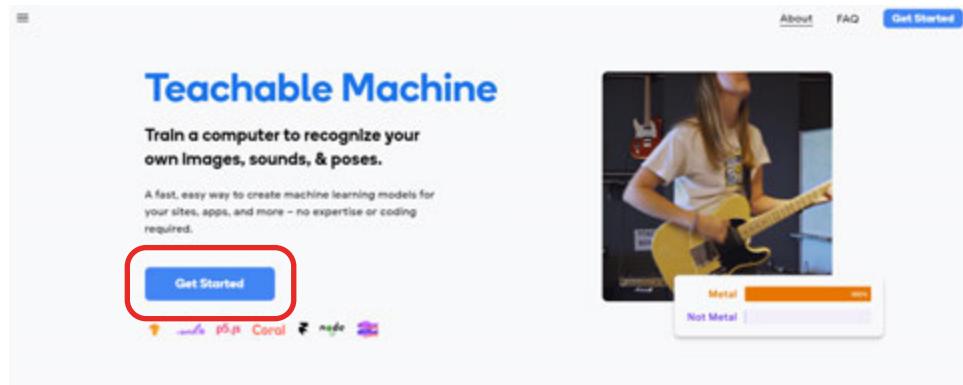


Figura 5 - Exemplo: Passo 2 / Fonte: os autores.

Descrição da Imagem: trata-se da mesma imagem anterior, mas, desta vez, há um retângulo vermelho em volta do botão clicável azul, destacando-o.

PASSO 3 - Na sequência, apresenta-se uma tela com três opções. Note que cada uma delas se refere à categoria de informação que será utilizada no processo de treinamento e, depois, no processo de classificação. Lembre-se de que no processo de treinamento, como o próprio nome indica, a IA entenderá as características de um conjunto de amostras apresentadas e com rótulo ou classe conhecida.

No processo de classificação, a IA utilizará seus conhecimentos adquiridos para definir o rótulo ou classe de novos objetos, ou seja, que ela ainda não conhece, mas utilizará seu conhecimento treinado para classificar. Clique na opção **Imagen Project** (em português, **Projeto Imagem**), conforme destacado na imagem a seguir. Salienta-se que você pode usar a opção para alterar o idioma, disponível na barra inferior, à direita. Caso deseje, escolha a opção de Português Europeu; caso contrário, se desejar ficar com opção de inglês, os passos apresentados são simples e intuitivos.

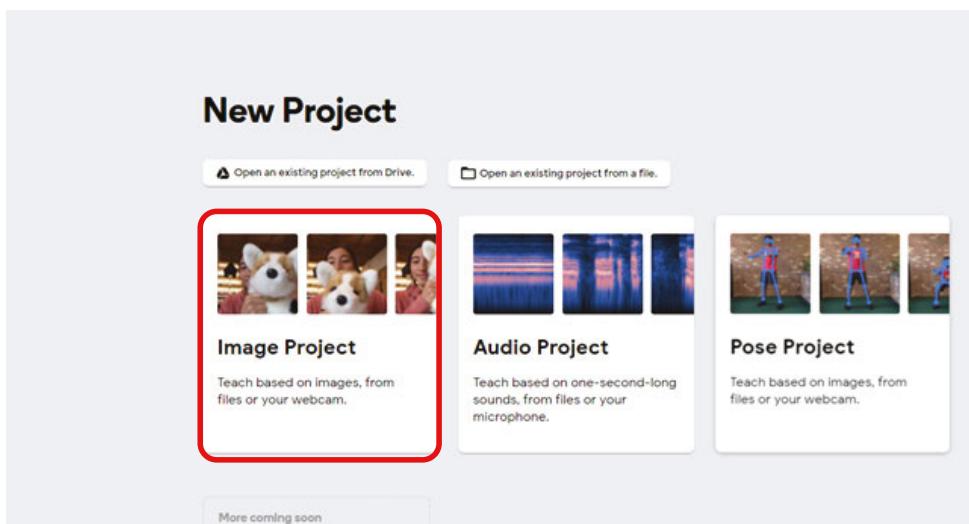


Figura 6 - Exemplo: Passo 3 / Fonte: os autores.

Descrição da Imagem: a imagem possui como título “New Project”, escrito na cor preta com negrito. Abaixo, é possível ver duas opções de escolha, uma ao lado da outra, sendo “Open an existing project from Drive” e “Open an existing project from a file”. A seguir, há mais três opções em quadrados grandes, dispostos um ao lado do outro. No primeiro, há três imagens de uma menina segurando um cachorrinho de pelúcia em diferentes ângulos, abaixo está escrito “Image Project” e, a seguir, “Teach based on images, from files or your webcam”. No segundo, há três imagens de uma paisagem em tons de preto, rosa e azul, também em diferentes ângulos, abaixo está escrito “Audio Project” e, a seguir, “Teach based on one-second-long sounds, from files or your microphone”. No último, há três imagens de um homem vestindo shorts preto e camiseta vermelha, abaixo está escrito “Pose Project” e, a seguir, “Teach based on images, from files or your webcam”.

PASSO 4 - É possível desenvolver modelos para serem utilizados com microcontroladores que, geralmente, possuem hardware limitado. Trabalharemos, porém, com um projeto padrão convencional. Para tal, clique na opção **Standard Image Model** (em português, **Modelo de Imagem Padrão**), conforme destacado na imagem a seguir.

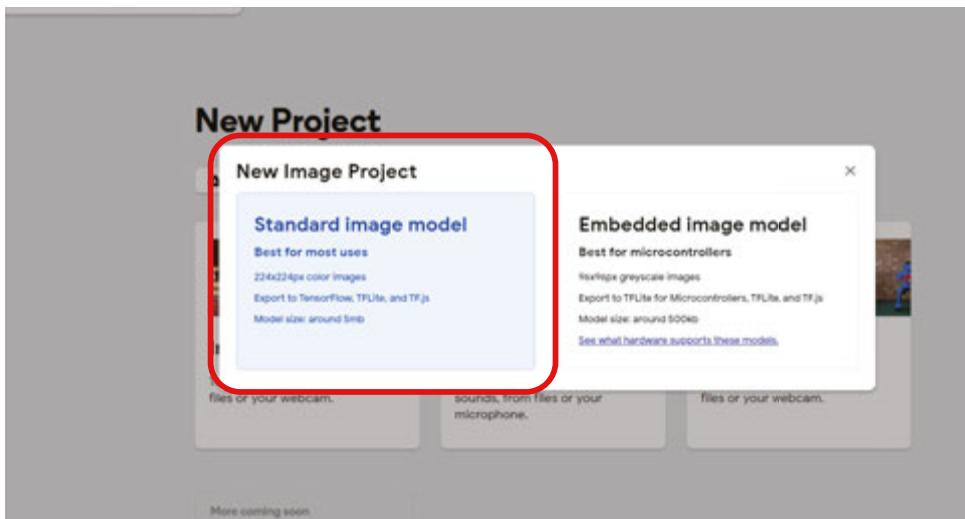


Figura 7 - Exemplo: Passo 4 / Fonte: os autores.

Descrição da Imagem: na imagem, há um retângulo vermelho dando destaque a uma caixa de texto, em que se encontra “New Image Project” e abaixo “Standard image model”, do lado direito encontra-se escrito “Embedded image model”. Há outras frases escritas, mas elas são ilegíveis.

PASSO 5 - Observe as opções disponíveis na página principal da plataforma, conforme o destaque das cores:

- **VERDE:** esta cor indica as classes de cada objeto. Essas classes representam um grupo específico de objetos que se queira identificar. Para cada classe, devem ser definidas amostras que representam o objeto, as imagens serão capturadas pelo webcam ou será feito upload de arquivos já existentes. Essas amostras farão com que o sistema possa aprender, em um processo de treinamento, por meio das características e padrões extraídos destas amostras. Toda classe deve ter um nome significativo, pois este será atribuído ao objeto que representa.
- **AZUL:** nesta cor, é apresentada a opção para adicionar mais classes. Devem ser adicionadas tantas classes quanto forem necessárias, conforme necessidade de classificação.

- **AMARELO:** nesta cor, tem-se a opção de iniciar o treinamento da IA. Ao utilizar essa opção, todas as amostras serão lidas e utilizadas para aprender cada uma das classes previamente definidas. Será possível configurar outras variáveis, para customizar o processo de aprendizagem e treinamento da IA. Salientamos que os valores apresentados como padrão já produzem bons resultados. Neste bloco, será possível acompanhar o andamento do processo de treinamento.
- **VERMELHO:** esta cor, pode-se visualizar os resultados do treinamento. Esta visualização permite interação com o webcam e apresentação de objetos a serem classificados. As imagens capturadas pelo webcam serão processadas em tempo real e classificadas segundo as amostras previamente aprendidas. Também existe uma opção de exportar o modelo pronto, de forma que possa ser utilizado em outro ambiente de programação e, assim, expandir a sua aplicação.

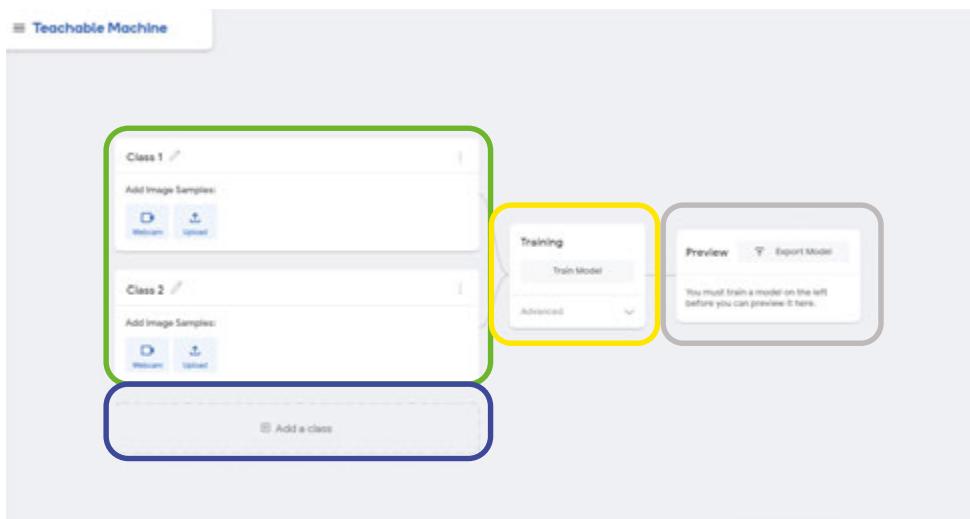


Figura 8 - Exemplo: Passo 5 / Fonte: os autores.

Descrição da Imagem: a imagem apresenta quatro retângulos de destaque. No primeiro, de cor verde, encontra-se: dois campos idênticos, divididos em "Class 1" e o símbolo de uma caneta; abaixo há um campo escrito "Add image samples"; abaixo há dois ícones, um de webcam e outro de Upload; esse campo se repete, mas agora com o nome "Class 2". No segundo, amarelo, encontra-se "Trainning" e abaixo um campo escrito "Train Modal". No terceiro, cinza, está escrito "Preview". No último retângulo, de cor azul, localizado abaixo desses três, encontra-se apenas um campo escrito "add a class".

Salientamos que, a princípio, vamos definir apenas duas classes de forma a identificar apenas dois objetos. Além disso, vamos manter os valores apresentados como padrão.

PASSO 6 - Assumimos que você compreendeu os recursos da *Teachable Machine*, até aqui, apresentados. Vamos, então, planejar a estrutura da IA para atender à necessidade da nossa aplicação. Com um exemplo simples, nós vamos treinar a IA para classificar um caderno e um copo. Teremos o cuidado de prever classificação para um objeto apresentado que não tenha sido classificado. Para tal, devemos ter uma classe que identifique uma “imagem vazia”.

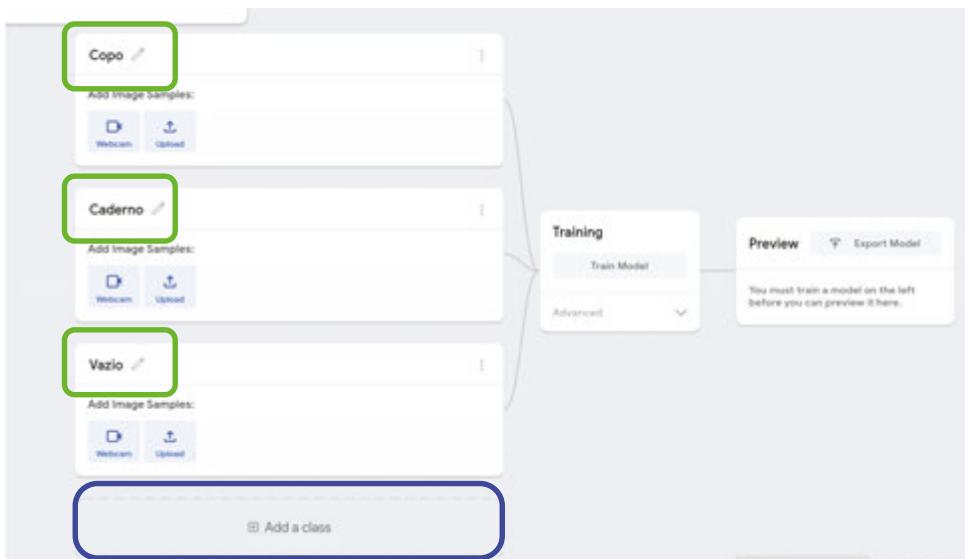


Figura 9 - Exemplo: Passo 6 / Fonte: os autores.

Descrição da Imagem: na imagem, há um retângulo vermelho dando destaque a uma caixa de texto, em que se encontra “New Image Project” e abaixo “Standard image model”, do lado direito encontra-se escrito “Embedded image model”. Há outras frases escritas, mas elas são ilegíveis.

Observe a imagem anterior: para adicionar os objetos, você deve escolher a opção **Add a class (adicionar uma classe)**, salientada em azul. Após adicionar, é possível alterar o nome da classe por meio do **símbolo de lápis** (ao lado no nome de cada objeto demarcado pela cor verde).

Sinta-se à vontade para escolher outros objetos que estejam a sua disposição, podendo, inclusive, adicionar mais classes. Atenção! Apenas os nomes das classes devem ser significativos aos objetos que serão utilizados na classificação. Neste momento, podemos dizer que a arquitetura da IA está finalizada!

PASSO 7 – Agora, especificaremos amostras para treinar a IA. Para cada classe definida, devemos inserir amostras que serão utilizadas no treinamento. Lembre-se que escolhemos a opção de modelo de imagens. Portanto, as amostras serão inseridas, a partir de upload de imagens já existentes ou a partir de imagens capturadas pelo webcam do seu computador.

Antes de iniciar o processo, e se for necessário, altere a configuração no computador para permitir que a plataforma possa usar o webcam.

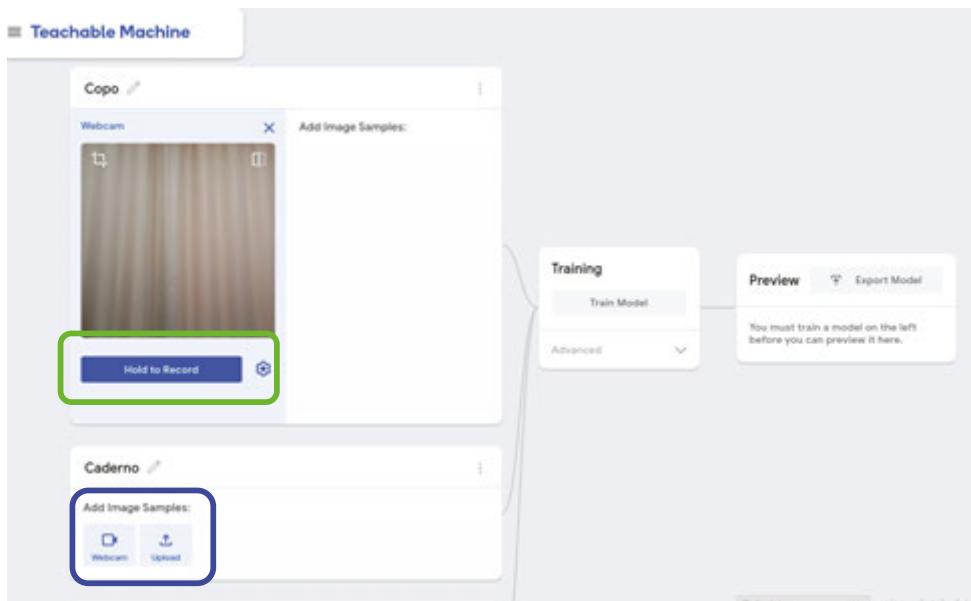


Figura 10 - Exemplo: Passo 7 / Fonte: os autores.

Descrição da Imagem: a imagem anterior se repete, mas, agora, abaixo, no campo intitulado “Copo”, há uma imagem de cor marrom e um botão azul clicável escrito “Hold to Record”. Em volta dele, há um retângulo de cor verde para dar destaque.

A imagem anterior salienta as opções para se capturar amostras por meio do webcam. Procure realizar os seguintes passos:

- Posicione o objeto em frente ao webcam (segurando com as mãos ou com apoio de um suporte) e aperte a opção **Webcam** destacada em azul. Note que será exibida a imagem de seu webcam, para verificar se o objeto está bem posicionado.

- b) Selecione a opção ***Hold to record*** (em português, **iniciar registro**) destacada em verde para que a plataforma inicie a obtenção das amostras de seu webcam. Enquanto a opção de captura estiver pressionada, a plataforma irá capturar uma sequência de imagens.

Para ampliar os resultados e permitir identificar diferentes características do objeto, é importante que, durante a captura de amostras, a posição do objeto seja alterada. Isto é, procure aproximar, afastar, inclinar, girar e movimentar o objeto. Quanto mais amostras forem registradas, melhor será o desempenho da IA. É necessário repetir os passos A e B para cada uma das classes. Faça de modo que as quantidades de amostras das classes sejam equilibradas.

PASSO 8 - Após finalizar a coleta das amostras para todas as classes, o treinamento pode ser iniciado. Para tal, selecione a opção ***Training*** (em português, **treinando**), conforme demarcado na imagem, a seguir, com a cor roxa.

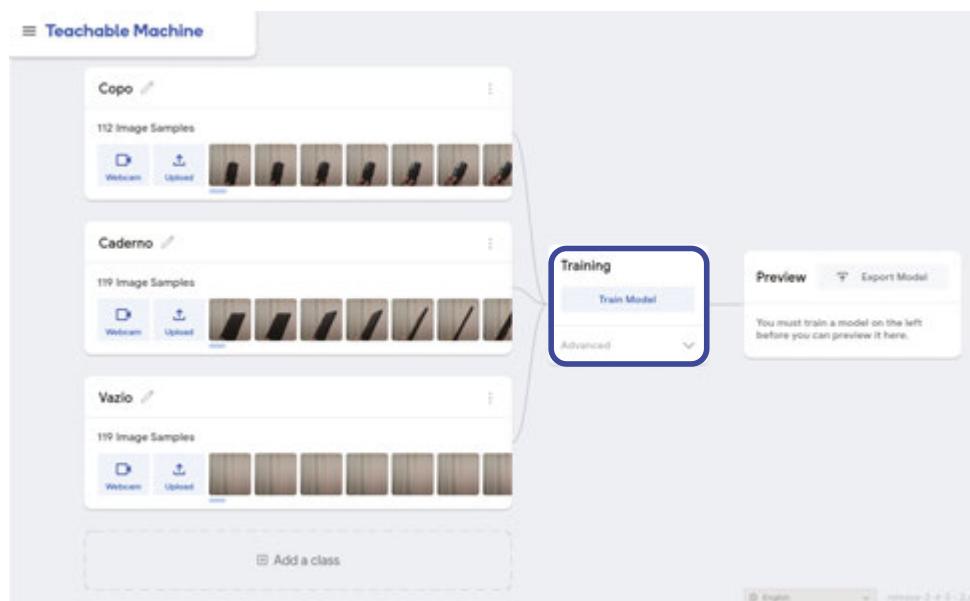


Figura 11 - Exemplo: Passo 8 / Fonte: os autores.

Descrição da Imagem: é a mesma imagem disponibilizada anteriormente, mas, agora, em cada campo “Copo”, “Caderno” e “Vazio” há várias imagens em tom de marrom, em posições variáveis. No Campo “Training”, há um retângulo roxo para dar destaque.

O processo de treinamento é, então, iniciado, sendo possível acompanhar o avanço do treinamento no mesmo bloco. Note que, durante o treinamento, a plataforma indica que a aba do navegador não seja alterada ou fechada. Esse processo de treinamento leva pouco tempo (na ordem de segundos), depende da quantidade de amostras e da complexidade da arquitetura construída.

PASSO 9 - Ao término do treinamento, a IA terá uma base de conhecimento sobre as classes adicionadas. O modelo fica disponível para ser aplicado na janela de *preview* (em português, **antevisão**).

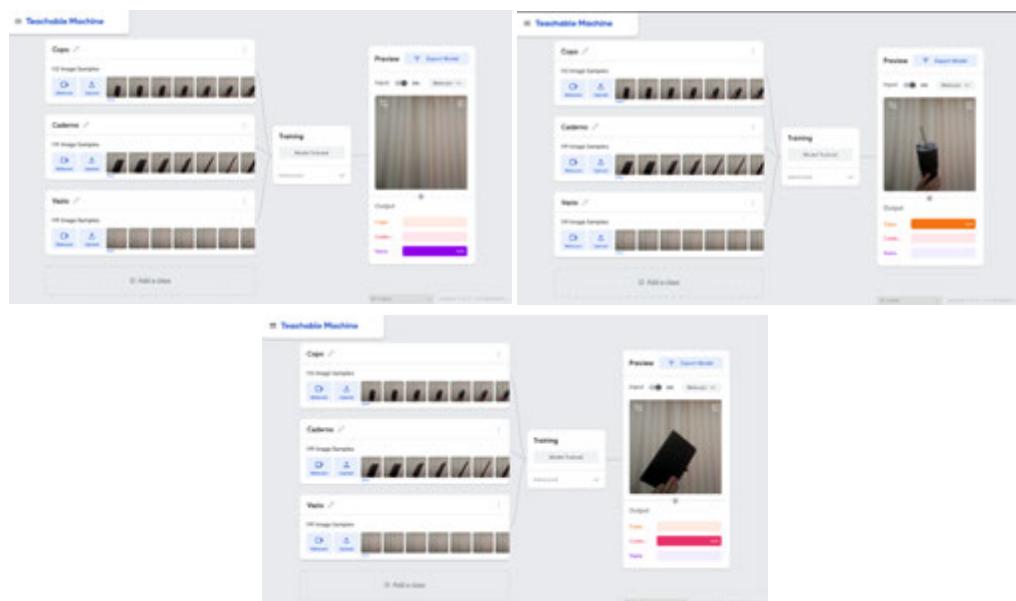


Figura 12 - Exemplo: Passo 9 / Fonte: os autores.

Descrição da Imagem: a figura apresenta a imagem descrita anteriormente replicada três vezes. Ao lado de cada uma, há uma caixa apresentando a imagem selecionada da vez.

É possível notar, na imagem anterior, que, em todos os objetos, a IA consegue classificar com 100% de certeza. A classificação ocorre em tempo real, note isto, ao movimentar o seu objeto. A probabilidade de classificação sempre será utilizada para definir toda a imagem, mesmo que, em algumas vezes, não seja atingido 100% de certeza, a classe que atingir maior probabilidade é utilizada para definir a classe do objeto presente na imagem. Dessa forma, quando a probabilidade de certeza é baixa, provavelmente a IA conseguiu observar características que podem pertencer a mais de um objeto, não oferecendo grande certeza na classificação.

MAPA MENTAL



Mapa Mental sobre aprendizado

Um bom método para fixação de conteúdos é o mapa mental. Um mapa mental é uma forma prática de deixar mais visual as informações e os conceitos que estudamos. A técnica nos ajuda a entender melhor as relações entre ideias, fazendo assimilação do conteúdo de um jeito mais interessante. Desta forma, os estudos se tornam mais proveitosos! Com os mapas mentais, você pode estruturar — de maneira rápida, simples e dinâmica — tudo o que for preciso para uma posterior revisão. O melhor de tudo é que essa prática possibilita gerar novas ideias, auxiliando o fluxo de pensamentos.

Na prática, o mapa mental permitirá que conteúdos estudados possam ser interligados, se possível, gerando uma malha que integra esses conteúdos. Para isto, basta utilizar elipses para representar as palavras-chave de referência de cada assunto, interligando subtópicos a um tópico principal.

Nesta atividade, você deve construir um mapa mental que contenha os principais tópicos estudados, procurando, assim, sintetizar os principais conceitos e suas relações. Pretende-se que você reforce o conhecimento apresentado nesta unidade. Portanto, o tópico principal será “Machine Learning”.

Esse é o momento para usar e abusar da criatividade. Acredite: ela está dentro de nós e só precisa ser estimulada. Você pode combinar exercícios de criatividade, durante o processo, para se aperfeiçoar, mas, a partir da construção de um mapa mental do seu próprio jeito, as coisas podem fluir bastante. Pegue uma folha em branco e comece à escrever ou desenhar alguma ideia ou conceito que deseja aprender ou desenvolver. É interessante, nesse processo, utilizar a folha na horizontal, para ter mais possibilidades de criação. Como o pensamento, comece puxando setas do conceito inicial e vá adicionando tudo o que vier à mente sobre ele. Dentro desses tópicos, é muito relevante adicionar subtópicos, como datas, curiosidades sobre o assunto, filmes e outras coisas que podem ajudar a memorizar o que está sendo estudado ou planejado. Desenhos e figuras são muito interessantes nessa etapa. Não se esqueça de variar o uso das cores, para que tudo chame atenção e tenha um impacto visual significativo. É bom experimentar ao máximo durante a construção do mapa. Segue uma imagem que exemplifica um Mapa Mental sobre o tema principal “Argumentação”.



Figura 13 - Mapa mental / Fonte: Benjamin (2021, on-line).

MEU ESPAÇO





Deep Learning

Esp. Janaina Aparecida de Freitas

OPORTUNIDADES DE APRENDIZAGEM

O objetivo desta unidade é apresentar conceitos, métodos e algoritmos baseados em *Deep learning*, os quais permitem aplicar processos de análise estatística e realizar previsões. Para tanto, serão compreendidos os conceitos, o uso e as aplicações com as redes neurais.

Olá, Aluno(a)! Esperamos que você já compreenda a área que envolve o aprendizado de máquina. Vamos, agora, refletir sobre o aprendizado profundo? Você sabe que, hoje, podemos treinar computadores para automatizar atividades que nós, seres humanos, realizamos? Essas atividades podem se relacionar ao reconhecimento de voz, à identificação de objetos e/ou imagens ou a previsões de tempo e de outros comportamentos que envolvem informações com registro histórico. Em muitas dessas atividades, nós trabalhamos com um grande volume de dados que são armazenados e organizados para serem executados, seguindo regras ou equações predefinidas.

Já pensou como podemos automatizar essas atividades? Podemos fazê-lo, configurando parâmetros de entrada do programa de computador, e, na sequência, treinar este computador para que ele aprenda sozinho durante a realização das atividades. Mas como esse computador pode fazer isso? Por meio de parâmetros predefinidos e do subsequente reconhecimento dos padrões. Geralmente, este tipo de treinamento — para realizar tarefas ou agir como seres humanos — envolve muito poder computacional, devido à natureza dos algoritmos, do aumento da complexidade na medida que são inseridas novas regras e grande volume de dados, para o treinamento deste computador.



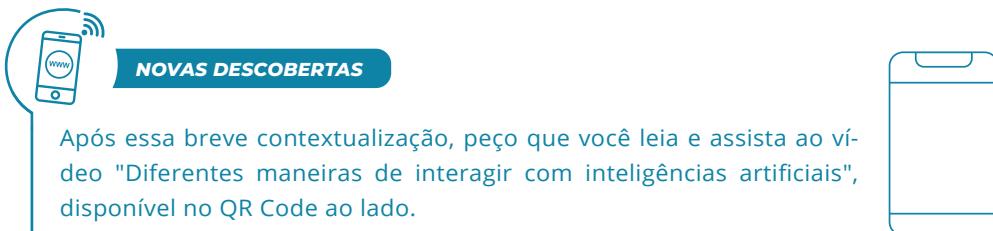


Conforme já apresentado nas unidades anteriores, a Inteligência Artificial (IA) contempla todas as teorias e tecnologias, abrangendo as duas categorias de *Machine Learning* (ML) e o *Deep Learning* (DL). Além disso, o DL é uma subárea do ML. O DL se deve à fascinação que temos por IA, com técnicas que ajudam a aprimorar a capacidade das máquinas em aprender tarefas, como: reconhecer, classificar, descrever e detectar parâmetros, ou seja, compreender.

A partir do DL, podemos, por exemplo, classificar imagens, realizar o reconhecimento de fala, detectar diversos objetos disponíveis em ambientes (fechados ou abertos) e descrever conteúdos advindos de imagens, entre outros. Pense em tudo o que você tem a sua volta: mouse, teclado, tela sensível ao toque (em inglês, *touch screen*), linguagem natural, gestos.

Agora, reflita sobre a quantidade de dados gerados pela Internet das Coisas (em inglês, *Internet of things - IoT*), dados textuais disponíveis nas mídias sociais, informações advindas de cadastros online, receitas médicas digitais, transcrições de dados investigativos que estão armazenadas na nuvem. Todas essas informações permitem que seja construída a personalização de clientes para que as empresas possam ser mais efetivas nas suas operações analíticas, como localização, redes, atividades, interesses e, com isso, ter uma maior compreensão do público-alvo.

Você deve estar pensando que treinar um computador para realizar tarefas, como um ser humano, é projeto realizado por pesquisadores ou obra de ficção científica. Contudo o DL já é disponibilizado por muitas empresas conhecidas em suas aplicações e/ou aparelhos. Você quer um exemplo? O Xbox, Skype e o Google realizam atividades de reconhecimento de voz. As crianças já utilizam o recurso de busca pela internet, apenas falando com o aplicativo da Google! Que outros aplicativos e/ou recursos tecnológicos permitem isso?



Após essa breve contextualização, peço que você leia e assista ao vídeo "Diferentes maneiras de interagir com inteligências artificiais", disponível no QR Code ao lado.

Neste texto, são apresentados diversos filmes de ficção científica que nos desafiam a pensar sobre como os seres humanos podem interagir com as máquinas. Procure fazer um comparativo sobre o que já utilizamos e o que é proposto nesses filmes. Não se esqueça de refletir sobre quão próximos (ou ainda distantes) estamos desses recursos tecnológicos.

DIÁRIO DE BORDO

Esta unidade baseia-se na área de estudo denominada por **Redes Neurais Artificiais (RNA)**, que, usualmente, é apenas chamada por “Redes Neurais”. A motivação para esta área é que o cérebro humano tem a capacidade de processar muitas informações de uma maneira diferente das máquinas. Conforme Haykin (2007, p. 27), “o cérebro humano é um computador (sistema de processamento de imagens) altamente complexo, não-linear e paralelo. Ele tem a capacidade de organizar seus constituintes estruturais, conhecidos por neurônios, de forma a realizar processamentos rápidos”.

Para melhor compreender, refletiremos sobre a visão humana, uma tarefa de processamento de informação. Essa tarefa permite a representação do ambiente ao qual devemos interagir. Conforme Haykin (2007), o cérebro humano realiza tarefas de reconhecimento de imagens de forma rotineira em, aproximadamente, 100-220 ms. Este mesmo autor salienta que um computador convencional pode levar dias para executar tarefas de menor complexidade.

Vamos, agora, refletir sobre o sonar de um morcego, um sistema ativo de localização por eco. Este sistema fornece informações sobre a distância do morcego até seu alvo, a velocidade relativa e o tamanho deste alvo, além de várias outras características. É necessário muito processamento neural, no cérebro pequeno do morcego, sobre as informações obtidas pelo sistema de localização dos morcegos.

Neste momento, você pode se perguntar: como o cérebro humano ou o cérebro do morcego consegue realizar todo esse processamento? Quando se nasce, o cérebro tem uma estrutura e uma habilidade enorme de se desenvolver por meio da “experiência”, que se aprimora com o passar dos tempos.

De acordo com Haykin (2007, p. 28), “um neurônio em desenvolvimento é sinônimo de um cérebro plástico: a plasticidade permite que o sistema nervoso em desenvolvimento se adapte ao seu meio ambiente”. Essa plasticidade, também, é construída com neurônios artificiais, em que:



[...] uma rede neural é uma máquina que é projetada para modelar a maneira como o cérebro realiza uma tarefa particular ou função de interesse; a rede é normalmente implementada utilizando-se componentes eletrônicos ou é simulada por programação em um computador digital. Para alcançarem bom desempenho, as redes neurais empregam uma interligação maciça de células computacionais simples denominadas “neurônios” ou “unidades de processamento” (HAYKIN, 2007, p. 28).

Uma definição para RNA é que são uma estrutura complexa, interligadas por elementos de processamento simples (denominados por neurônios) e que podem realizar muitas operações complexas, tais como: cálculos em paralelo para a representação de conhecimento ou para processamento de dados. Conforme a Figura 1, a RNA assemelha-se ao cérebro nos seguintes aspectos:

1. O conhecimento pode ser adquirido pela rede, a partir do ambiente, através de um processo de aprendizagem.
2. As forças de conexão dos neurônios (pesos sinápticos) são utilizadas para armazenar o conhecimento adquirido no processo de aprendizagem.

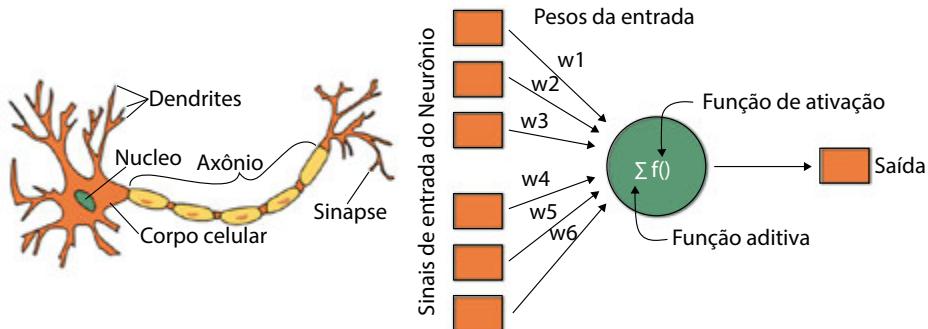


Figura 1 - Neurônio biológico vs Modelo de Neurônio [artificial]
Fonte: adaptada de Uledagsozluk Galexy ([2021], on-line).

Descrição da Imagem: duas ilustrações representam, ao lado esquerdo, neurônio biológico e, ao lado direito, um neurônio artificial. A primeira imagem apresenta o desenho de um neurônio com os dendritos, núcleo, corpo celular, axônio e impulso nervoso. Na segunda imagem, temos um neurônio artificial, representado por um círculo ao meio, sendo este a função de ativação e a função aditiva. À esquerda, desse círculo, temos seis retângulos, que simbolizam os sinais de entrada e os pesos de entrada; à esquerda do círculo, temos a saída.

Antes de nos aprofundarmos no conhecimento sobre redes neurais, precisamos compreender melhor os **modelos de um neurônio**. Para Haykin (2007, p. 36), um **neurônio** é “uma unidade de processamento de informação que é fundamental para a operação de uma rede neural”, conforme apresentado na Figura 2.

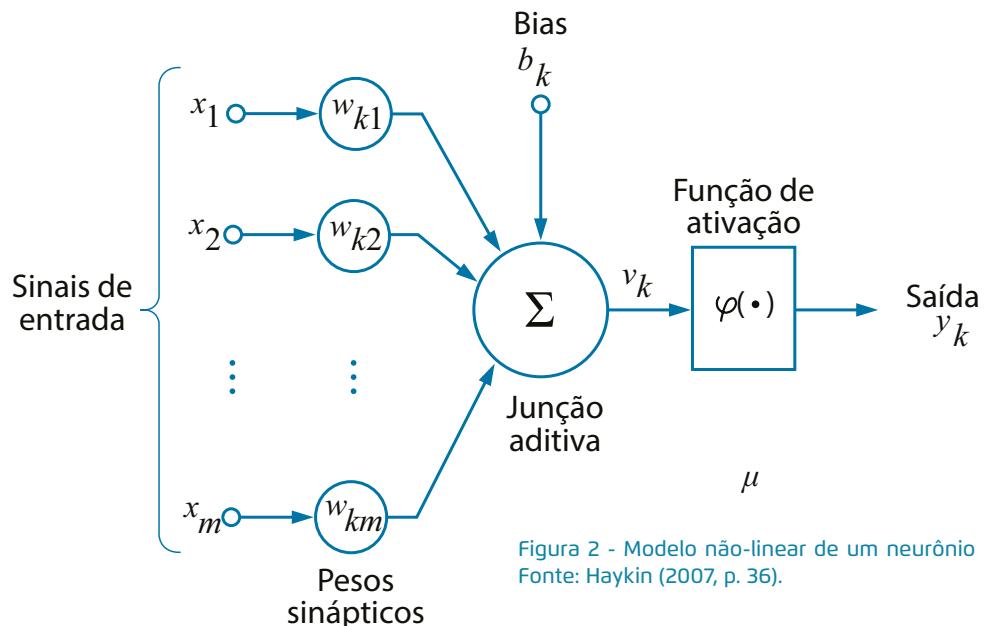


Figura 2 - Modelo não-linear de um neurônio
Fonte: Haykin (2007, p. 36).

Descrição da Imagem: : a figura ilustra o modelo não-linear de um neurônio, em que, à esquerda, identificamos os "Sinais de entrada" das sinapses: X1, X2 e Xm, conectadas ao peso wk1, wk2 e wkm que estão em círculos com setas que levam para um círculo ao centro, com o símbolo da notação de somatório identificado como "Junção aditiva" e, acima deste, há o texto "Bias - bk" conectada à função por uma seta. Conectada à função de somador, temos uma seta identificada como "vk" que conecta a um quadrado com o símbolo "" identificado como "Função de ativação", seguido de uma seta que conecta à "Saída - yk".

No modelo de um neurônio, apresentado na Figura 2, vê-se as **sinapses** (também chamados de elos de conexão) que são caracterizadas por um peso ou força própria, em que:

x_j um sinal na entrada da sinapse **j**
conectada ao neurônio **k** é
multiplicado pelo peso sináptico **w_{jk}**

É importante notar que: o índice **k** se refere ao neurônio em questão; o índice **j** se refere ao terminal de entrada da sinapse, a qual o peso se refere, e, finalmente, o peso sináptico de um neurônio pode estar em um intervalo com valores negativos ou positivos.

Ainda, sobre o modelo de um neurônio da Figura 3, o **somador** Σ faz a soma dos sinais de entrada, ponderando as sinapses do neurônio. As operações descritas constituem um **combinador linear**. Na sequência, a **função de ativação** é utilizada para restringir a amplitude da saída de um neurônio a um valor finito. Finalmente, **Bias** são aplicadas, externamente, e tem o propósito de aumentar ou diminuir (dependendo do valor ser positivo ou negativo) a entrada líquida da função de ativação.

Uma boa solução para vários problemas e aplicações do mundo envolve uma boa e detalhada representação do conhecimento. Isso, também, ocorre com as redes neurais. De acordo com Haykin (2007, p.49), “as formas possíveis de representação, desde as entradas até os parâmetros internos da rede, são muito diversificadas, o que tende a tornar o desenvolvimento de uma solução satisfatória, utilizando uma rede neural, um desafio real de projeto”.

Para que tal ocorra, nós precisamos de uma representação do conhecimento adequada. Compreenda que **conhecimento** se refere às informações armazenadas, ou modelos que são utilizados pela máquina, para: interpretar, processar, prever e responder às questões de forma apropriada ao mundo exterior. A representação do conhecimento deve, portanto, ser direcionada a um objetivo e satisfazer as seguintes características:

A informação é realmente tornada explícita.

A informação é codificada, fisicamente, para o uso subsequente.



NOVAS DESCOBERTAS

Este livro (em português, *A Organização do Comportamento*) indica que as criações de redes neurais acontecem quando a conectividade do cérebro modifica, à medida que o organismo aprende tarefas funcionais diferentes, em outras palavras, esta seria a capacidade de aprendizado em redes neurais. O conteúdo apresentado, neste livro, desempenha um papel significativo no estímulo à investigação dos fundamentos neurais do comportamento e continua a ser inspirador, porque fornece uma estrutura geral para relacionar o comportamento à organização sináptica, por meio da dinâmica das ANN.



O computador pode aprender uma tarefa por tentativa e erro, realizando análise de muitos exemplos de treinamento. As RNA podem aprender a resolver diferentes tipos de problemas — desde a identificação de gatos em fotos até a direção de um carro que dirige sozinho — e são utilizadas para construir soluções de problemas complexos. Um **Algoritmo de Aprendizagem** baseado em RNA é, portanto, um procedimento utilizado para realizar todo o processo de aprendizagem e modificar os pesos sinápticos da rede para alcançar o objetivo desejado, de forma ordenada. A modificação dos pesos sinápticos, de acordo com Haykin (2007), é o método tradicional para os projetos que usam redes neurais.

As RNA extraem seu poder computacional tanto de uma estrutura paralela distribuída quanto da habilidade de aprender e generalizar. Quando falamos “generalizar”, é porque a rede neural produz saídas adequadas para as entradas que, ainda, não foram aprendidas via processo de treinamento. Com todo esse poder computacional, as RNA podem resolver problemas complexos e considerados intratáveis. Na sequência, a Tabela 1 apresenta uma síntese das propriedades e capacidades que as RNA devem possuir.

Propriedades e Capacidades	Descrição
Não-linearidade	<p>Neurônio artificial pode ser linear ou não-linear.</p> <p>Rede neural: constituída de neurônios não-lineares, é ela mesma não-linear.</p> <p>A não-linearidade é um tipo especial de rede neural, por ela ser distribuída por toda a rede.</p>
Mapeamento de entrada e saída	<p>Envolve os pesos sinápticos de uma rede neural pela aplicação de um conjunto de amostras de treinamento ou exemplos de tarefas.</p> <p>Cada amostra ou exemplo é um sinal de entrada único e sua correspondente resposta esperada.</p> <p>O treinamento da rede é repetido para muitos exemplos do conjunto até a rede alcançar um estado estável.</p> <p>A rede aprende a partir dos exemplos a construir o mapeamento de entrada e saída para um determinado problema.</p>
Adaptabilidade	<p>Redes neurais têm a capacidade inata de adaptar seus pesos sinápticos conforme o ambiente se modifica.</p> <p>Rede neural treinada pode operar em um ambiente específico e pode ser treinada, novamente, para lidar com as modificações que podem ocorrer no ambiente.</p>
Resposta a evidências	<p>Rede neural pode ser projetada para fornecer informações sobre qual padrão particular selecionar, em caso de classificação de padrões.</p> <p>Rede neural pode ser projetada para se ter a confiança ou crença na decisão tomada, em caso de classificação de padrões. Isso permite descartar padrões ambíguos e melhorar o desempenho e a qualidade.</p>
Informação contextual	<p>Na rede neural, o conhecimento é representado pela própria estrutura e estado da rede.</p> <p>Cada neurônio da rede é afetado pela atividade de todos os neurônios da rede e, com isso, a informação contextual é tratada, naturalmente, pela rede.</p>

Tolerância a falhas	<p>Rede neural implementada na forma física tem o potencial de ser tolerante a falhas ou capaz de realizar computação robusta (desempenho degradado suavemente, em condições adversas).</p> <p>Se um neurônio ou conexões são danificados, a recuperação de um padrão armazenado é prejudicada em qualidade, devido à natureza distribuída da informação armazenada na rede.</p> <p>Rede neural exibe uma degradação suave de desempenho, em vez de uma falha catastrófica.</p>
Implementação em VLSI	<p>Natureza paralela da rede neural a faz ser, potencialmente, rápida na computação de certas tarefas, como para implementação, utilizando tecnologia de integração em escala muito ampla.</p> <p>A tecnologia VLSI fornece um meio de capturar o comportamento, realmente, complexo de uma forma hierárquica.</p>
Uniformidade de Análise e Projeto	<p>Redes neurais desfrutam de universalidade como processadores de informação.</p> <p>Ou seja, a mesma notação é utilizada em todos os domínios, envolvendo as redes neurais.</p>
Analogia neuro-biológica	Redes neurais são motivadas pela analogia do cérebro humano, que tem o processamento paralelo tolerante a falhas rápido e poderoso.

Tabela 1 - Propriedades e capacidades das redes neurais / Fonte: adaptada de Haykin (2007).

**NOVAS DESCOBERTAS**

Estudos recentes no MIT apresentam a Rede Neural "Líquida" (RNL) que é capaz de aprender a estrutura de causa e efeito da tarefa de navegação para a qual é treinada. As RNLs podem mudar suas equações subjacentes para se adaptar, continuamente, às novas entradas. Por exemplo, uma RNL com 19 neurônios de controle é capaz de controlar, autonomamente, um veículo. Estas redes compreendem a tarefa diretamente de dados visuais, são mais eficazes do que outra rede neuronal ao navegar em ambiente complexo, por exemplo, um local com árvores densas ou com condições climáticas alternadas. No futuro, pretende-se aprimorar a confiabilidade de forma a realizar tarefas de alto risco, como: dirigir um veículo autônomo em uma rodovia movimentada.

Fonte: Zewe (2021, on-line, tradução nossa).

Neste momento, podemos refletir sobre qual é a diferença fundamental entre um processamento de informação clássico (por padrões) e um projeto de RNA? No processamento de informação clássico, primeiro, é formulado um modelo matemático das observações do ambiente. Após, o modelo deve ser validado com dados reais para, na sequência, o projeto ser estruturado com base no modelo construído. Já, no projeto de uma RNA, o embasamento é direto nos dados do mundo real e, com isso, a rede neural fornece um modelo implícito do ambiente (ou mundo) mais uma função de processamento da informação de interesse.

Note que a RNA precisa de um **conjunto de dados de treinamento**, em que cada dado de treinamento corresponde a um par de sinal de entrada-saída, ou seja, um sinal de entrada e sua correspondente resposta esperada. Portanto, uma das tarefas importantes para a RNA é aprender sobre o “modelo do ambiente” ou, conforme (HAYKIN, 2007), “mundo” ao qual a RNA está inserida. Esta rede deve manter esse modelo do ambiente consistente para que os objetivos sejam atendidos. Dois tipos de informações sobre o conhecimento desse modelo do ambiente são apresentados na Tabela 2.

Tipo de Informação	Descrição
Estado conhecido do mundo	Representado sobre o que é e o que era conhecido (fatos). Chamado de informação prévia.
Observações do mundo	<p>Obtidas por meio de sensores projetados no ambiente, em que a RNA vai operar.</p> <p>Normalmente, são inherentemente ruidosas, podendo apresentar erros devido ao ruído do sensor.</p> <p>As informações obtidas fornecem um conjunto de informações de onde são retirados os exemplos usados nos treinos da RNA.</p>

Tabela 2 - Tipos de informação do conhecimento do mundo / Fonte: adaptada de Haykin (2007).

Importante ressaltar, também, que os exemplos utilizados para treinar a RNA representam o conhecimento acerca do ambiente de interesse, podendo ser:

- **Rotulados:** cada exemplo que representa um sinal de entrada é associado a uma resposta desejada correspondente.
- **Não-rotulados:** consistem de exemplos de ocorrências diferentes dos sinais de entrada.

Para compreender melhor a dinâmica de uma RNA, trabalharemos o exemplo:

Reconhecer um dígito manuscrito em que cada sinal de entrada consiste de uma imagem com pixels pretos e brancos, e a imagem contém 10 dígitos, estes estão separados. Qual é a resposta desejada? A resposta deve apresentar a identificação de cada dígito em particular, e a imagem deste dígito deve ser apresentada como um sinal de saída.

Para que isso ocorra, deve-se ter um conjunto de dados de treinamento que possuem uma variedade de dígitos manuscritos, estes representam diferentes estilos de escrita do mundo real. Após ser determinado o conjunto de dados de treinamento (ou exemplos), o **projeto para uma RNA** ocorre da seguinte maneira:

1. **Aprendizagem:** a arquitetura é selecionada para a RNA com camadas de entrada com nós de fonte igual em números (pixel de uma imagem de entrada) e uma camada de saída (10 neurônios, um para cada dígito). Após, um algoritmo apropriado é usado para treinar a rede com um subconjunto de exemplos.
2. **Generalização:** testes são realizados com dados que não foram apresentados, anteriormente, para verificar o desempenho de reconhecimento da RNA treinada. Para tal, apresenta-se uma entrada que é a imagem de um dígito, mas não se informa a identidade deste. A partir disso, o desempenho da RNA é estimado, verificando se o reconhecimento do dígito corresponde à identidade real deste dígito.



EXPLORANDO IDEIAS

Ficou curioso como isso pode ser feito? Procure realizar a atividade proposta no tutorial "Como construir uma rede neural para reconhecer dígitos manuscritos com o TensorFlow" ([clique aqui](#) para acessar).

Importante ressaltar que os exemplos usados para treinar uma RNA podem ser **positivos** ou **negativos**. O problema a ser resolvido determina quais serão usados como exemplos positivos ou negativos. Por exemplo, em um problema de detecção passiva de sonar, em que os dados de treinamento de entrada devem conter o(s) alvo(s) de interesse. O submarino é um exemplo positivo, e exemplos negativos são os elementos que compõem a vida marinha, os quais podem gerar alarmes falsos. Os exemplos negativos são, portanto, incluídos nos dados de treinamento para evitar que se confundam na identificação.

Relembre que a representação do conhecimento em uma RNA é definida por valores assumidos por parâmetros livres (denominados em pesos sinápticos e bias) da rede neural. A chave para o desempenho desta rede será a forma da representação do conhecimento. A Figura 3 apresenta uma RNA que funciona como um classificador binário simples que permite mapear as entradas x em função de um valor de saída y , criando uma matriz $f(x) = 1$ se $w \cdot x + u > 0$, senão o valor y é 0.

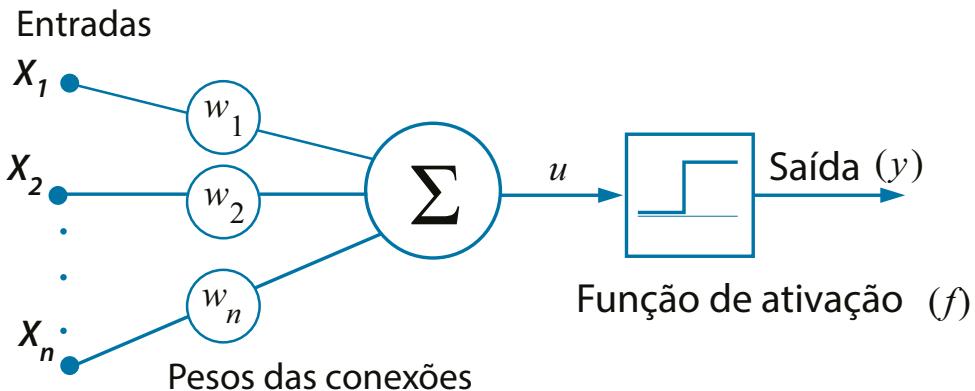


Figura 3 - Interconexão entre os neurônios e seus valores de entrada e saída
Fonte: Silva et al. (2019, p.116).

Descrição da Imagem: representação da Interconexão entre os neurônios e seus valores de entrada e saída. Onde temos os sinais de entradas X_1 , X_2 e X_n ligados a círculos com pesos W_1 , W_2 e W_n , que estão conectados a outro círculo central com a função de somador, ligada a uma seta com o símbolo u como texto acima da seta, conectado a um quadrado com a função de ativação, ligadas a uma seta com sinais de saída y .

Nesta altura, você deve se questionar: “Como podemos incorporar a informação no projeto de uma rede neural?”. Para incorporar informação prévia ao seu projeto, Haykin (2007, p. 54) diz que “não há atualmente regras bem definidas para fazer isto; em vez disso, temos alguns procedimentos ad-hoc que sabemos que produzem resultados úteis”. Portanto, vamos lhe apresentar duas técnicas:

- **Campos Receptivos:** restringe a arquitetura de rede pelo uso de conexões locais.
- **Compartilhamento de Pesos:** restringe a escolha de pesos sinápticos.

Essas técnicas podem ser combinadas, e a combinação delas ajudará a reduzir, significativamente, o número de parâmetros livres. Além disso, devemos considerar que as redes neurais possuem duas estruturas principais:

- **RNA de camada única** é a estrutura fundamental com todas as entradas conectadas, diretamente, com uma única saída. Esta estrutura de rede de camada única é considerada a forma mais simples para a estruturação de uma rede neural artificial, pois se constitui de uma camada neural e de um neurônio, conforme apresentado na Figura 4.

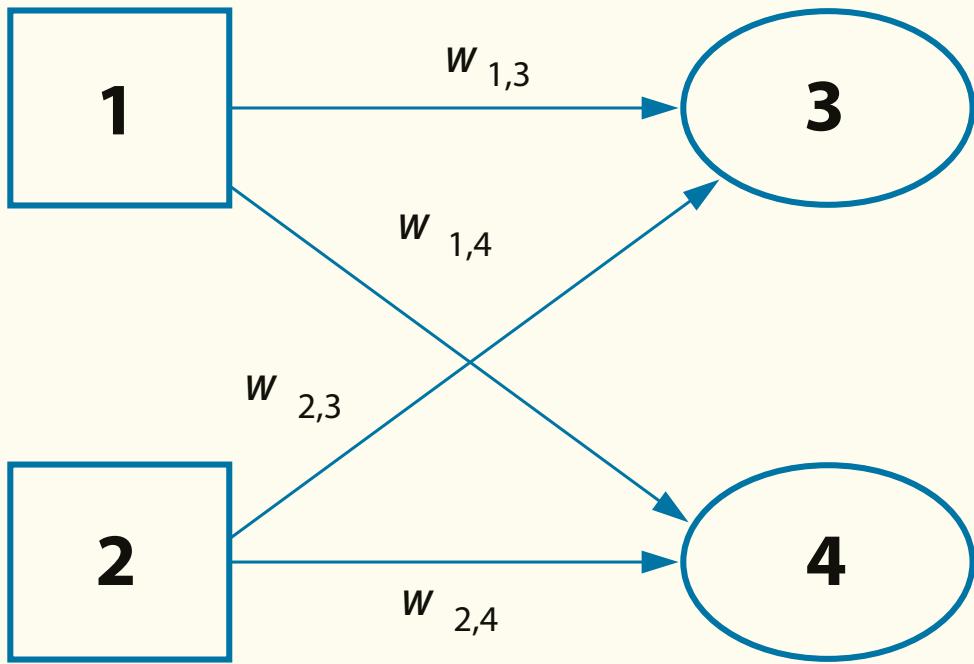


Figura 4 - Rede de Camada única simples de duas entradas com duas interligações
 Fonte: Silva et al. (2019, p.115).

Descrição da Imagem: representação de rede neural de camada única. Há dois quadrados, um abaixo do outro, numerados como 1 e 2 e, ao lado, há dois círculos, um abaixo do outro, numerados como 3 e 4. O Quadrado 1 está ligado ao círculo 3 por pesos $W_{1,3}$ e ao círculo 4 por pesos $W_{1,4}$. O Quadrado 2 está ligado ao círculo 3 por pesos $W_{2,3}$ e ao círculo 4 por pesos $W_{2,4}$.

RNA multicamada é uma rede composta de neurônios ligados entre si por sinapses com pesos, similar à representação de um neurônio em uma camada, com mais de uma camada, conforme apresentado na Figura 5.

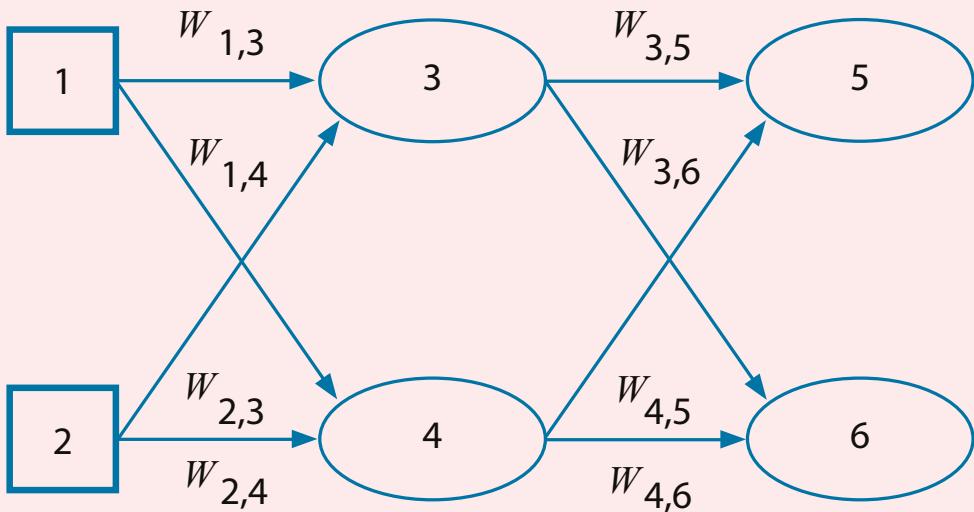


Figura 5 - Exemplo de rede neural artificial de multicamadas / Fonte: Silva et al. (2019, p.116).

Descrição da Imagem: representação de exemplo de rede neural artificial de multicamadas. Há dois quadrados, um abaixo do outro, numerados como 1 e 2 e, ao lado, há dois círculos, um abaixo do outro, numerados como 3 e 4, e, ao lado, há mais dois círculos, um abaixo do outro, numerados como 5 e 6. O Quadrado 1 está ligado ao círculo 3 por pesos $W_{1,3}$ e ao círculo 4 por pesos $W_{1,4}$. O Quadrado 2 está ligado ao círculo 3 por pesos $W_{2,3}$ e ao círculo 4 por pesos $W_{2,4}$. O círculo 3 está ligado ao círculo 5 por pesos $W_{3,5}$ e ao círculo 6 com pesos $W_{3,6}$. O círculo 4 está ligado ao círculo 5 com pesos $W_{4,5}$ e ao círculo 6 com pesos $W_{4,6}$.

A vantagem na construção de uma RNA multicamadas é que ela tem uma entrada de estímulos e uma camada intermediária de construção da resposta para encaminhar a saída, formando grafos que simulam neurônios em multicamada, conforme apresentado na Figura 6.

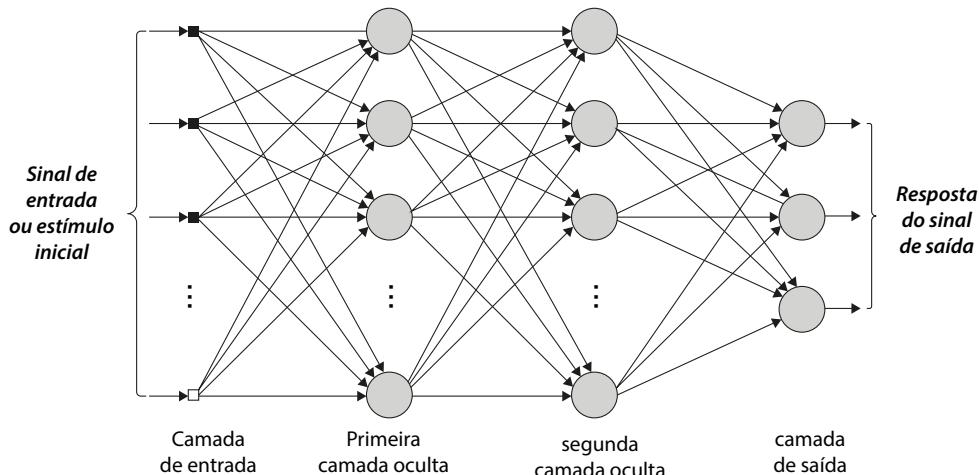


Figura 6 - Construção da rede neural artificial de multicamadas / Fonte: Silva et al. (2019, p.117).

Descrição da Imagem: representação da construção da rede neural artificial de multicamadas. Temos os sinais de entrada ou estímulo inicial que fazem parte da camada de entrada ligadas por setas e círculos que pertencem a primeira camada oculta, que estão ligadas por setas e círculos que pertencem a segunda camada oculta, que estão ligadas a círculos que pertencem a camada de saída que mostram a resposta do sinal de saída.

Conforme Silva et al. (2019, p. 141), “apesar de sempre existir mais do que uma camada de neurônios, a rede de camada única possui apenas neurônios de entrada e saída e a rede de multicamadas tem ao menos uma camada oculta que não representa a entrada e nem a saída”, ou seja, uma camada intermediária de neurônios. As redes multicamadas são aproximadoras universais que solucionam problemas de diversos tipos, desde que sejam dadas as quantidades necessárias de entrada e saída.



O IBM Watson, que foi desenvolvido por iniciativa da International Business Machines, é uma plataforma de Inteligência Artificial que combina as novas metodologias de aprendizado de máquina com o uso de supercomputadores, podendo ser aplicado em projetos de diferentes áreas. O Watson tornou-se muito conhecido pela sua capacidade de análise linguística, por ser o primeiro computador a vencer concorrentes humanos em um programa de perguntas e respostas norte-americano, programa este considerado muito competitivo.

Fonte: Silva et al. (2019, p.158).

Neste momento, acreditamos que você deve se questionar: "Em quais áreas as RNA podem ser aplicadas?".

Nas mais diversas áreas e de forma crescente, pois tanto a Inteligência Artificial quanto os recursos tecnológicos estão em constante evolução. Além disso, as RNA beneficiam-se da grande quantidade de dados gerados pelos mais diversos programas de software e aplicativos.

A **Internet das Coisas** (em inglês, *Internet of Things - IoT*) é muito utilizada em conjunto com a RNA. Note que tem aumentado, consideravelmente, a quantidade de sensores nos mais diversos setores (comercial, industrial e educacional) e, até, no meio ambiente. Isto permite o fornecimento de grande quantidade de dados. Além disso, há muitos avanços tecnológicos, por exemplo, a nanotecnologia aliada à IA, em que dispositivos pequenos e com custo reduzido têm realizado tarefas complexas. O conceito de RNA, de acordo com Silva *et al.* (2019), é muito útil para a IoT, porque:



[...] permite que uma rede de pequenos dispositivos seja capaz de interagir de forma descentralizada, processando as informações de cada elemento na nuvem, assim, os dados são processados pelo conjunto de operações efetuadas por cada dispositivo, se aproximando muito do conceito de neurônios em uma rede. Porém, junto dos neurônios da IoT não há apenas o fluxo de dados, como também a sua obtenção. Portanto, é função dos dispositivos obterem e processarem dados, bem como, em alguns casos, servirem de meio físico de atuação (SILVA *et al.*, 2019, p. 156).

A RNA aplicada à **área da medicina** é muito utilizada para o processamento de imagens, pela sua capacidade de aprender sozinha e de reconhecer padrões nas imagens (até mesmo, as que não têm, ainda, reconhecimento da ciência) com uma velocidade de aprendizado que permite melhorar o desempenho da observação de diversas amostras, realizando em um curto intervalo de tempo. Perceba que temos possibilidade de ter um grande banco de dados, com todo tipo de imagens, geradas por ressonância magnética, tomografia e DNA. As RNA podem ser treinadas para percorrer esse banco de dados, buscando por respostas para diagnósticos ou testando novas metodologias de aprendizado de máquina.

Na **agronomia**, RNA são utilizadas para o imageamento, via satélite e em diferentes frequências, para obter informações sobre temperatura, umidade do ar, profundidade e altitudes estimadas. Estas informações são usadas para o reconhecimento geológico e climático da região. A análise dessas informações é complexa, e as RNA podem reconhecer padrões de comportamento e, assim, permitir que o agricultor tenha uma melhor tomada de decisão para os momentos de plantio, colheita e irrigação.

Agora, falaremos sobre as redes neurais convolucionais, estas são aplicadas com bastante êxito para **análise e processamento de imagens**. Note que gerar imagens se tornou corriqueiro para o usuário final, pois, por exemplo, os celulares tiram fotos de alta qualidade acrescida à possibilidade de ter diversos softwares que ajudam na melhoria da iluminação, nitidez, efeitos, entre outros; sem falar de recursos tecnológicos que dão apoio à área médica. Realizar, contudo, a análise computacional dessas imagens é algo bastante dispendioso, pois, conforme a qualidade da imagem aumenta, necessita-se melhor e maior capacidade de processamento e memória computacional. Com isso, técnicas na área de IA são desenvolvidas para processar imagens, por exemplo, as RNA. Porém é, ainda, um recurso computacional dispendioso. Para o reconhecimento de imagens, modelos RNA foram adaptados para se ter um conjunto de treinamento chamado ao *overfitting*. Neste caso, redes neurais modificadas são criadas e podem ser gerenciadas, mesmo que sejam imagens “pesadas”. Essas redes são chamadas de **redes neurais convolucionais**.

A área de processamento de imagens é vasta e abrange um amplo campo do conhecimento. Por exemplo, não existe um limite determinado entre **processamento de imagens e visão computacional**. Muitos pesquisadores consideram o processamento de imagens como métodos, técnicas e algoritmos que estejam relacionados a um processo, em que as entradas e as saídas são imagens. Mas, em alguns casos, o processamento realizado tem imagens que são as entradas, mas as saídas apresentam outro conteúdo (texto, por exemplo).



A robótica tenta simular a visão humana por meio da visão computacional, utilizando, inclusive, o aprendizado de máquina para conseguir fazer deduções e mostrar reações que simulam o comportamento humano ou atuam de forma inteligente. Essa área está relacionada a um campo do conhecimento chamado inteligência artificial (SILVA *et al.*, 2019, p. 169).

Existem softwares que apoiam áreas da medicina, biologia, aerofotogrametria, agricultura, meteorologia, indústria cinematográfica, televisiva, indústria de jogos, publicidade e propaganda, imagens de satélites, restauração, arqueologia, astronomia etc. Além disto, aplicações que possibilitam o processamento digital de imagens é vasto, tais como: PhotoShop, PhotoScan, Acute3D, Paint, Canva, PicMonkey, Seashores, Gimp, Photoscape, Polarr, Illustrator, Pixlr, entre outros.

Referente ao processamento computacional desses softwares e aplicações, existem três níveis de processamento computacional contínuo, sem necessariamente haver um divisor explícito entre estes níveis:

Níveis	Processamento computacional
Nível 1	Envolve os processos com operações básicas, como redução de ruídos, aumento e redução de contrastes e melhorias de imagem de um modo geral. Neste nível, tanto a entrada como a saída são imagens.
Nível 2	Envolve processos de separação das imagens em segmentos, objetos ou regiões, com sua redução a uma forma que seja adequada ao processamento computacional e reconhecimento (classificação) de segmentos, objetos ou regiões individuais. Neste nível, as entradas são imagens, mas as saídas incluem atributos e características que são extraídas das imagens. Exemplo: linhas, contornos, bordas etc.
Nível 3	Envolve ações que dão sentido ao conjunto reconhecido, como na análise e classificação de imagens ou na geração de comportamentos inteligentes associados à visão humana.

Tabela 3 - Níveis de processamento computacional / Fonte: Silva et al. (2019, p. 117).


EXPLORANDO IDEIAS

Suponha que você queira efetuar o reconhecimento de impressão digital. O processo de aquisição da imagem e o pré-processamento desta será realizado por um scanner, este tipo de processamento está no **primeiro nível** da Tabela 3. Na sequência, a extração de características significativas da imagem (impressão digital) e a preparação desta imagem para realizar processamento computacional estão no **segundo nível** da Tabela 3, pois são técnicas de comparação e reconhecimento a partir de imagens catalogadas em um banco de dados. Finalmente, após o reconhecimento da impressão digital, as ações realizadas devem demonstrar comportamento inteligente, estas técnicas pertencem ao **terceiro nível** apresentado na Tabela 3.

Fonte: Silva *et al.* (2019, p. 158).

Além do processamento, você deve, também, compreender os formatos das imagens:

- **Imagen de Rastreio** (ou *Raster*): representação em uma matriz de duas dimensões, em que cada elemento desta dimensão é um elemento de imagem. São as mais comuns de encontrar na internet (GIF, JPEG, BMP, PNG etc.). Contudo, conforme o aumento da dimensão da imagem, a qualidade desta tende a piorar.
- **Imagen Vetorial**: usada na área de computação gráfica e é composta de entidades como polígonos, retas, pontos, entre outros elementos. A qualidade da imagem não piora quando as suas dimensões são aumentadas, pois sua representação se baseia em entidades e equações matemáticas. Além disso, o tamanho do arquivo gerado é menor do que os arquivos de rastreio. Esta representação é aplicada em desenhos assistidos por computador, desenhos industriais e cinematográficos, como também em projetos de arquitetura, engenharia e de visualização 3D.

Você não está curioso como ocorre o processo de conversão da imagem real para uma imagem digital? A Figura 7 apresenta os passos necessários. Inicialmente, a imagem deve ser capturada por um dispositivo de aquisição, que pode ser câmera digital, scanner, sensor, ou máquinas de ultrassom, de raio x, de ressonância magnética, entre outros. Cada dispositivo permite regular os parâmetros e características da imagem, podendo esta ser processada por programas de computador. A fase seguinte é denominada de “conversor análogo/digital”, em que ocorre o processo de **separar as amostras pontuais**, igualmente espaçadas, da imagem contínua; após, as **amostras são quanti-**

zadas para que possam, na sequência, serem codificadas. Após este processo de conversão, é gerado o arquivo para armazenamento.

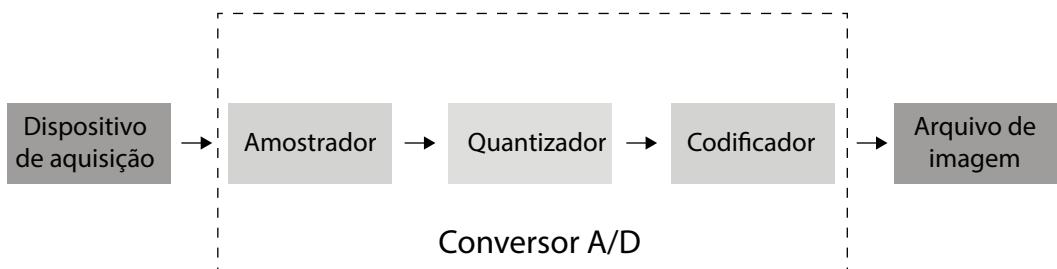


Figura 7 - Sistema de aquisição de imagem simplificado / Fonte: Silva et al. (2019, p.174).

Descrição da Imagem: representação do sistema de aquisição de imagem simplificado. Temos um retângulo com o texto Dispositivo de Aquisição ligado a um retângulo maior pontilhado chamado de Conversor A/D. Dentro deste retângulo pontilhado temos o retângulo com o texto Amostrador ligado ao retângulo com o texto Quantizador, que está ligado do retângulo com o texto Codificador, que está ligado, fora do retângulo pontilhado, ao retângulo com o texto Arquivo de Imagem.

Importante salientar que a **extração de características de imagens** desempenha um papel fundamental no processamento digital de imagens. As características de uma imagem permitem computar, de forma eficiente, os valores que descrevem a imagem (ou parte dela) e são representados por vetores de características. Essas características são definidas como qualquer elemento capaz de ser expresso, quantitativamente, e são variáveis. São calculadas em 2D, mas, comumente, estão relacionadas a objetos 3D. A extração de características de imagens pode ser agrupada conforme:

- **Globais:** são descritores obtidos com base na localização, intensidade e relações espaciais.
- **Locais:** são calculadas conforme o contorno do objeto ou de uma região da imagem, segmentos de contornos, curvaturas, detecção de concavidades ou cantos etc.
- **Relacionais:** são medidas de orientação relativa, distâncias entre as características ou posições de regiões, contornos fechados, concavidades ou outras características locais.

Quando se utiliza o reconhecimento ou classificação, os padrões reconhecidos podem se dividir a partir dessas características. Importante ressaltar que, para caracterizar um objeto que se quer reconhecer ou classificar, **descritores de características** (apresentados na Figura 8) permitem especificar o aspecto desejado da imagem.

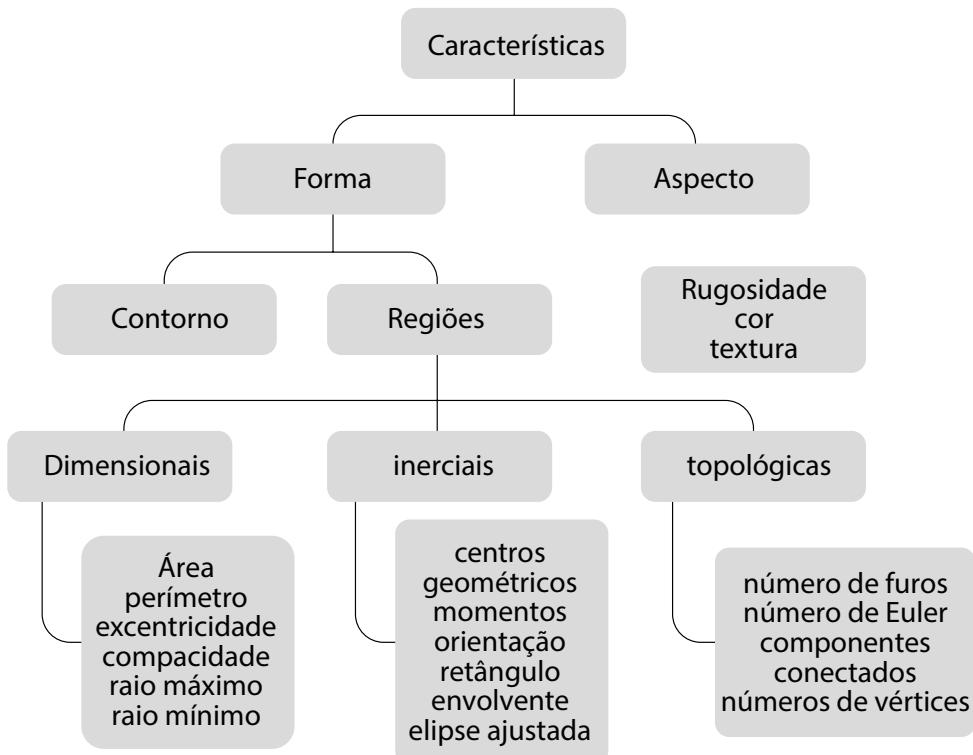


Figura 8 - Descritores das características de imagens / Fonte: Silva et al. (2019, p. 178).

Descrição da Imagem: representação dos Descritores das características de imagens. Temos um organograma que inicia com um retângulo com o texto Características que está ligado logo abaixo a dois retângulos com os textos Forma e Aspecto. O retângulo Forma está ligado, logo abaixo, a dois retângulos com os textos Contorno e Regiões. O retângulo Aspecto está ligado, logo abaixo, a um retângulo com o texto Rugosidade, Cor, Textura. O retângulo Regiões está ligado, logo abaixo, com três retângulos com os textos Dimensionais, Inerciais e Topológicas. O retângulo Dimensionais está ligado, logo abaixo, ao retângulo com o texto Área, Perímetro, Excentricidade, Compacidade, Raio Máximo e Raio Mínimo. O retângulo Inerciais está ligado, logo abaixo, ao retângulo com os textos Centros Geométricos, Momentos Orientação, Retângulo envolvente e Elipse ajustada. O retângulo Topológicas está ligado, logo abaixo, ao retângulo com os textos Números de furos, Números de Euler, Componentes conectados e Número de Vértices.

São várias características que dependem das tarefas, estas podem ser realizadas, ou do tipo de imagem a ser tratada, por exemplo:

- **Reconhecer impressões digitais:** consideram-se os descritores, tais como contorno, arcos e bifurcações.
- **Reconhecimento facial:** consideram-se as dimensões, tais como distância entre os olhos, nariz, comprimento da boca, entre outros.

Penso que você já deve ter compreendido que é um processo complexo realizar as etapas e detalhes até agora descritos, devido ao grande volume de armazenamento (memória) que o processamento de imagens exige, da mesma maneira que os algoritmos para realizar análise e reconhecimento das imagens.

De acordo com Silva *et al.* (2019, p.179), extrair características da imagem é “fundamental, porque simplifica o conjunto de dados em um conjunto mais simples de se trabalhar, que permita a realização da tarefa desejada e represente a imagem original com mais precisão”.

Portanto, percebe-se que o processamento de imagens está ligado a várias áreas do conhecimento humano, e que o seu uso nas tarefas do cotidiano está aumentando cada vez mais. Seu potencial é muito utilizado na Inteligência Artificial (apoio tomada de decisão, entre outros), como nas áreas de Realidade aumentada e Realidade virtual.



As redes convolucionais têm vantagens significativas sobre as demais técnicas quando processa sinais naturais, cuja diferença computacional necessária é expressiva se comparada às redes neurais densas. Elas, ainda, conseguem aprimorar o processamento, subdividindo a análise em partes menores, por meio de filtros que se adaptam para encontrar características específicas, sons e imagens. Assim, elas são divididas em camadas, cada qual com uma função claramente definida, para somente então entregar a classificação da imagem nos neurônios de saída com o auxílio de uma rede totalmente conectada às densas.

(F. M. Silva, M. L. Lens, P. H. C. Freitas, S. C. B. Santos)

Note que o objetivo de qualquer RNA é a utilização de modelos de neurônios artificiais para obter respostas automáticas e que possam ser consideradas racionais, já que se inspiram no comportamento humano. A ligação entre os neurônios pode seguir, apenas, uma direção ou ambas as direções, conforme apresentado na Figura 9.

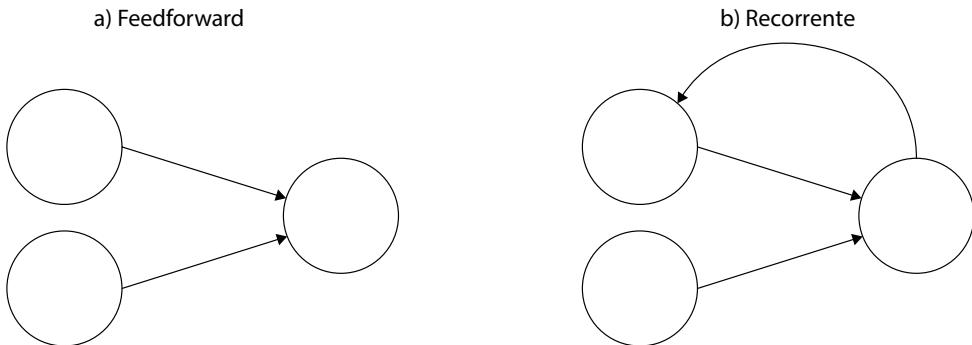


Figura 9 (a) - Ligação entre os neurônios: Feedforward; Figura 9 (b) - Ligação entre os neurônios: Recorrente
Fonte: Silva et al. (2019, p.183).

Descrição da Imagem: representação da Ligação entre os neurônios. Temos dois círculos do lado esquerdo, um embaixo do outro, ligados por uma seta reta em um círculo do lado direito. E temos outros dois círculos do lado esquerdo, um embaixo do outro, ligados por uma seta reta em um círculo do lado direito e uma seta curva que volta a um dos círculos do lado esquerdo.

A Figura 9 (a) apresenta a **RNA Feedforward**, em que a ligação entre os neurônios ocorre em apenas uma direção, ou seja, temos dois neurônios de entrada e um de saída. Já a Figura 9 (b) apresenta a **RNA Recorrente**, e a ligação entre os neurônios ocorre nas duas direções; pode-se, também, ver uma conexão de retorno (*feedback*). Conforme Silva et al. (2019, p. 169):



[...] as redes rasas e densas formam a primeira divisão no grupo das redes neurais, cuja diferença é simples: a rasa possui poucas camadas (três ou menos) e a densa tem muitas camadas. Ambas resolvem problemas complexos, porém, a maior quantidade de camadas foi superior tanto no treinamento como na eficiência de resolução comparada a uma rasa com mesma quantidade de conexões e neurônios. Além disso, as redes rasas apresentavam boa eficiência para problemas simples e bem definidos, mas eram insuficientes na resolução de problemas reais, especialmente aqueles que envolviam sinais naturais.

Agora, podemos apresentar as RNA Densas e Convolucionais. A **RNA Densa** reflete quando se começou a observar como o cérebro humano lidava com as informações, separando o processo em múltiplas camadas. Novas técnicas de RNA. Densas surgem, e que cada neurônio de determinada camada recebe a totalidade.

de saídas dos neurônios da camada, imediatamente, anterior. Com isso, espera-se obter maior capacidade de raciocínio e adaptação. Para cada conexão, tem-se um valor de peso associado, independentemente dos demais.

Atenção! Essa liberdade pode ser uma vantagem, mas também pode se tornar um problema e, às vezes, ser um limitador operacional. O processamento de máquina é finito, os tamanhos de redes viáveis são limitados e este limite ocorre pelo tempo de aprendizado dessas redes. Para Silva *et al.* (2019, p. 185), desde que os “valores de peso e *bias* entre os neurônios estejam definidos, não haverá grandes dificuldades para realizar a operação em alta velocidade, mas no treinamento, uma série de cálculos de parametrização deve ser feito, recalculando o valor de peso de todas conexões”. Para compreender melhor, observe o exemplo de RNA Densa apresentado na Figura 10.

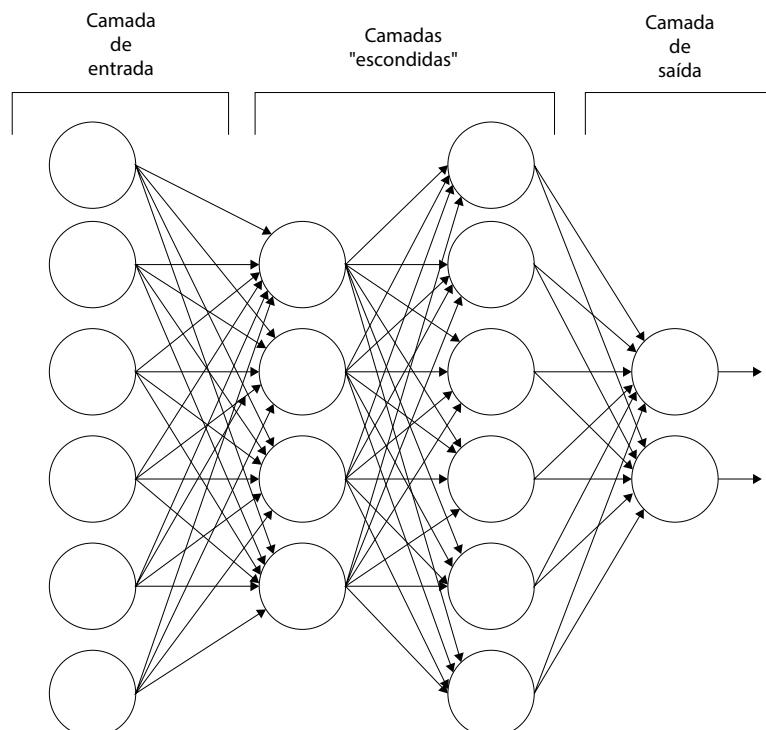


Figura 10 - Exemplo de Rede Neural densa / Fonte: Silva *et al.* (2019, p.184).

Descrição da Imagem: representação de um exemplo da rede neural densa. Temos o texto “Camada de entrada” seguido por vários círculos, um abaixo do outro, com várias setas ligadas a outros círculos, um abaixo do outro, que pertencem a “Camada ‘escondidas’” e que estão ligadas a dois círculos, um abaixo do outro, que pertencem à “Camada de saída”.

**EXPLORANDO IDEIAS**

O processamento de imagens é uma situação que demanda volume considerável de entradas e inúmeros treinamentos para obter resultados satisfatórios, sendo que o treinamento consiste no ajuste do peso de cada conexão. Para uma rede convencional, os valores de peso são independentes; e os pixels, tratados de forma individual. Já para uma rede convolucional, o uso do Kernel para calcular todas as entradas de um neurônio garante que, naquele campo receptivo observado pelo Kernel, a cada iteração haverá uma contribuição entre os pixels próximos, tornando o aprendizado da rede mais conectado às características espaciais da imagem.

Fonte: Silva et al. (2019, p.158).

As **RNA Convolucionais** fracionam o processamento em conjuntos (grupos) menores, no qual não há uma conexão rígida de todos os neurônios de uma camada com as camadas seguintes. Contudo os pesos são dependentes de cada *Kernel* — que procura buscar a mesma característica em diferentes regiões da imagem —, pois haverá pesos suficientes para cobrir uma região e replicar para as demais. A inspiração dessa rede baseia-se no sistema visual humano, este procura dividir o processamento em filtros menores, diferentemente das redes convencionais que não tem uma ligação total entre todas as camadas.

**PENSANDO JUNTOS**

Fique atento. Apesar de redes densas terem inúmeras camadas com pesos e resultados de treinamento diferentes, as redes convolucionais são mais eficientes por conseguirem fragmentar camadas em várias características, menos abstratas e mais especializadas na resolução do problema para a qual foram treinadas.

(F. M. Silva, M. L. Lens, P. H. C. Freitas, S. C. B. Santos)

Atenção! As RNA Convolucionais são divididas em quatro camadas:

Camada	Conceito
Convolução	Camada em que a imagem passa por diferentes operações de convolução pela busca de características específicas, por exemplo: reconhecimento de partes de um caractere ou membros de um animal. Resultado: redução do tamanho da imagem, as bordas são eliminadas no processo e os pixels que possuem pouco vínculo com as características de busca. Gera-se uma nova matriz de pixels, chamada de mapa de atributos (<i>features map</i>).
Retificação	Camada que transforma os pixels do mapa de atributos, utilizando a função não linear. Podem ser usadas várias funções diferentes, dependendo das necessidades. Evita que o aprendizado se torne demasiadamente lento por dispersar o sinal, após a sua passagem entre as camadas.
Subamostragem (em inglês, <i>pooling</i>)	Camada que reduz o tamanho dos mapas de atributos, utilizando tamanho fixo, que substitui por um pixel correspondente ao valor máximo do grupo (máximo morfológico). Diminui bastante a quantidade de dados e, consequentemente, de neurônios e conexões necessários, sem que as informações sejam perdidas.
Totalmente conectada	Camada em que se espera, após a convolução que boa parte dos pixels seja nulo e as características em regiões próximas estejam devidamente expressas por seus máximos.

Tabela 4 - Divisão das redes convolucionais em categorias
Fonte: adaptada de Silva et al. (2019, p.188).

As RNA Convolucionais se tornam especialistas, quando se criam camadas com características bem definidas, para resolver os problemas específicos para as quais foram treinadas. Essas redes servem de apoio para entender como os padrões são criados e auxiliam a Pesquisa Cognitiva.



NOVAS DESCOBERTAS

“RNA melhoraram vídeo de 1896, e agora parece que ele foi filmado com um smartphone”

Artigo que mostra como as RNA foram utilizadas para melhorar um curta-metragem de 1896, intitulado *L'Arrivée d'un train en gare de La Ciotat* (em português, *A chegada de um comboio à estação de La Ciotat*), que mostra lentamente um trem carregando um comboio e chegando a uma estação. Foram utilizadas técnicas de *Deep learning*, como demonstra Denis Shiryaev, que puderam melhorar consideravelmente o vídeo em preto e branco. Agora, parece que ele foi filmado com um smartphone!



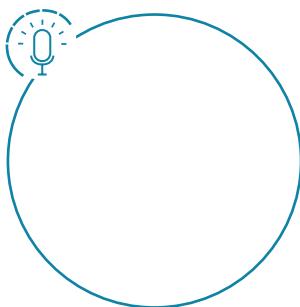
Uma região, ou uma quantidade de informação de qualquer tamanho, é uma característica relevante para o processamento da imagem. Assim, um grupo de *pixels*, borda, reta ou outro padrão de reconhecimento que auxilie a RNA a classificar e distinguir imagens. Pense como as redes sociais, tal como o Facebook, conseguem nomear, automaticamente, cada uma das pessoas que aparecem em uma foto?

Isso é possível, mas são necessários muitos procedimentos de detecção de características, por exemplo, o contorno dos olhos e a posição adequada destes em relação ao restante da imagem, características consideradas as mais gerais. Já para características mais específicas — tais como dígito manuscrito, círculos, retas verticais e horizontais — (juntos) ajudam a RNA a classificar, de forma adequada, a imagem selecionada.



EXPLORANDO IDEIAS

Perceptron é um método de aprendizagem supervisionado, proposto por Frank Rosenblatt, em 1943, e posteriormente analisado e refinado por Minsky e Papert, em 1969. A apresentação da rede de Perceptrons demonstrou que os neurônios deveriam ser organizados em camadas: a camada de entrada possui conexões fixas (simulando a retina humana); depois, recebe impulsos devido a essas conexões (atribuindo-se pesos) e envia para uma camada de saída (respostas). Veja o material que ensina como criar uma rede neural do zero: "Rede Neural Perceptron Adaline" ([clique aqui](#) para acessar este conteúdo). A rede Adaline (*Adaptive Linear Element*) foi proposta por Widrow e Hoff, em 1960, e tem a estrutura do Perceptron, diferenciando-se apenas no algoritmo de treinamento. Procure ler com atenção e, se possível, realize todas as atividades propostas.



Caros Aluno(a), nesta conversa, quero fazer uma comparação entre os seres humanos. Os nossos órgãos dos sentidos são **sensores** (tais como a boca, os olhos e os ouvidos); já os braços, as pernas, as mãos são **atuadores**. Na robótica, os sensores que captam as entradas são as câmeras, sensores térmicos e infravermelhos, entre outros; e os atuadores são os motores com braços mecânicos.

Para aplicar o conhecimento visto em um ambiente profissional ou real, devemos pensar no aumento do poder de processamento disponível e como os avanços tecnológicos na IA permitem coletar dados, mas também entender os detalhes por trás desses dados. Imagine todo esse poder utilizado para compreender a quantidade de dados existentes nas redes sociais, no *e-commerce* ou nas recomendações personalizadas. A partir disso, convido você a passar pela seguinte experiência.

Abra o seu navegador e realize a busca por um filme, ou por um produto do seu interesse. Se você tem Netflix, busque por filmes e, depois, saia do aplicativo. Se a busca foi realizada no seu navegador, abra agora uma rede social de sua preferência. Você verificará que aparecem recomendações de filmes ou de produtos, similares ao que você pesquisou. Se a busca foi feita no Netflix, ao abri-lo novamente, você verá recomendações de filmes (ou séries) semelhantes ao que você pesquisou ou que já tenha assistido.

Para que isso aconteça, são aplicados algoritmos de *Deep Learning*, que mostram qual o próximo conteúdo que são de seu interesse, baseados no seu perfil, nas suas buscas ou com características similares aos que você já tenha

consumido. Este mesmo tipo de experiência pode ser realizada com o *Spotify* e outros tipos de serviços, tais como compras em lojas de *E-commerce*. Após uma compra, seus dados são cruzados e o serviço procura despertar o seu interesse para comprar produtos comprados por consumidores similares a você.



AGORA É COM VOCÊ

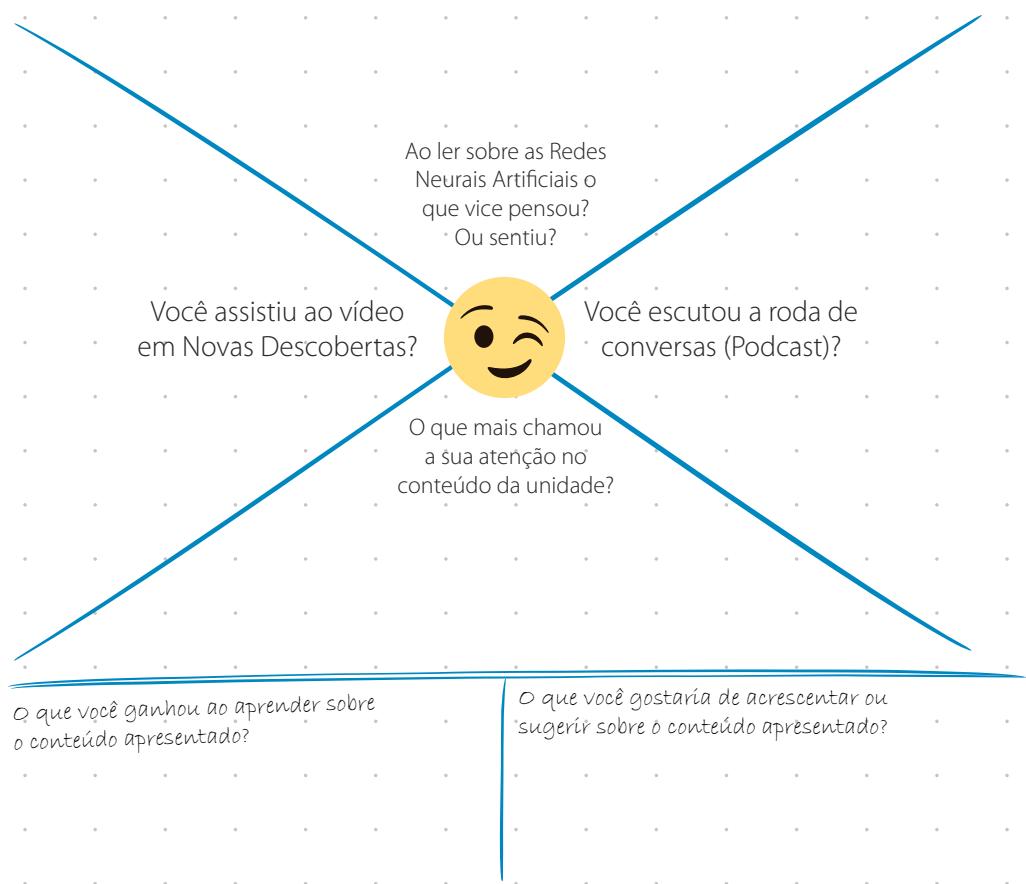


Figura 11 - Mapa da empatia / Fonte: a autora.

MEU ESPAÇO



Aprofundando em Modelos, Métodos e Algoritmos

Dra. Iara Carnevale de Almeida

OPORTUNIDADES DE APRENDIZAGEM

O objetivo desta unidade é se aprofundar em conceitos sobre modelos, métodos e algoritmos de classificação presentes na Inteligência Artificial, tanto para *Machine learning* quanto para *Deep learning*.

Olá, aluno(a)! Assumimos que você compreendeu melhor os princípios do *Machine Learning* (em português, Aprendizado de Máquina — AM) e do *Deep Learning* — DL (em português, Aprendizado Profundo). Você, também, já deve ter percebido como esses aprendizados estão presente em diversas soluções atuais do mercado, não é verdade? Cada vez mais, podemos observar soluções que, há bem pouco tempo, pareciam ser futuristas, mas, agora, estão presentes em nossas vidas. Muito se deve ao avanço da área de Inteligência Artificial — IA (em inglês, *Artificial Intelligence*), visto que esta área engloba AM e, por consequência, o DL. Mas como construir as soluções apresentadas nas unidades anteriores? Quais são os modelos, métodos e algoritmos que devemos conhecer para estas construções? Será que existem ferramentas prontas?

Relembre que discutimos, nas unidades anteriores, as soluções baseadas em IA que estão presentes no nosso dia a dia. Uma dessas soluções consiste em perguntar para um telefone celular “Qual a previsão do tempo para a minha cidade amanhã?”. Tanto a Siri (se for um Iphone) quanto o Google Assistente (em dispositivos móveis com Android) irão reconhecer sua voz, descobrir a sua cidade, a partir dos dados de sua geolocalização ou de suas preferências de localização, procurar pela previsão do tempo e responder-lhe, com uma voz nada robótica. Todo esse processo é muito rápido!

Recorde-se, também, que aplicações baseadas em IA estão disponíveis em sistemas computacionais e utilizadas por organizações em diferentes setores: comercial, industrial e educacional. Acreditamos, contudo, que você tenha dúvidas sobre quais são os modelos, métodos e algoritmos mais utilizados. Para tal, conto que você leia com atenção esta unidade e, principalmente, procure realizar as atividades propostas durante a leitura deste material.

Vamos, então, compreender melhor alguns dos modelos, métodos e algoritmos baseados em *Machine Learning*?

Para que você compreenda melhor a complexidade dessa unidade, além da quantidade e variedade de modelos, métodos e algoritmos baseados em *Machine Learning* e *Deep Learning*, procure fazer uma busca no Google Acadêmico, utilizando as palavras-chave seguintes indicadas (procure respeitar o uso de aspas, AND e OR), e registre o número de resultados que obtém.

Primeiro em português,

- Modelos AND (“aprendizado de máquina” OR “aprendizado profundo”).
- Métodos AND (“aprendizado de máquina” OR “aprendizado profundo”).
- Algoritmos AND (“aprendizado de máquina” OR “aprendizado profundo”).

Depois em inglês,

- Models AND (“machine learning” OR “deep learning”).
- Methods AND (“machine learning” OR “deep learning”).
- Algorithms AND (“machine learning” OR “deep learning”).

O número de artigos foi vasto, não? É porque esta é uma temática que está em franca expansão, tanto na pesquisa (busca por novos modelos, métodos e algoritmos) quanto na aplicação desses em organizações, nos mais diversos setores. Por isso, não se surpreenda se encontrar novo material além do apresentado por aqui!

Em novembro de 2021, a IBM publicou em seu site um artigo que trata da **IA Explicável**, indicando que ela aumenta a capacidade de interpretação da IA. Salienta-se que ela avalia e minimiza os riscos da IA e que permite uma implementação da IA com confiança e convicção. Clique [aqui](#) para ter acesso ao artigo.

Diversas temáticas sobre IA são debatidas; nesta reflexão, o foco é a confiança na IA. Acesse ao artigo disponibilizado, relembre os conteúdos apresentados, até aqui, e reflita sobre a confiança que precisamos ter na aplicação de modelos, métodos e algoritmos de IA. Aproveite seu diário de bordo para fazer suas anotações e registrar os pontos que você considerar relevantes.

DIÁRIO DE BORDO

Antes de dissertarmos sobre algoritmos e métodos, precisamos compreender melhor sobre os **tipos de modelo de aprendizagem**: preditivos e descriptivos. A Figura 1 apresenta a hierarquia dos modelos de aprendizagem preditivos e descriptivos. Essa divisão não é tão rígida, é apenas por questões didáticas, e, por isso, muitas vezes, esses modelos acabam por se misturar no projeto, pois o modelo preditivo, também, pode prover descrição dos dados, e o modelo descriptivo pode prover previsões após validado.



Figura 1 - Tipos de aprendizagem / Fonte: José (2018, on-line).

Descrição da Imagem: os tipos de aprendizagem ilustrados, por meio de um esquema hierárquico, em caixas de texto. No topo, a primeira caixa contém “aprendizagem” que se desmembra para outras duas caixas, em um nível abaixo. À esquerda, contém “supervisionada: modelos preditivos” e, à direita, o termo “não supervisionada: modelos descriptivos”. A caixa da esquerda se desmembra para outras duas caixas denominadas “classificação” e “regressão”, e a caixa da direita se desmembra em outras três denominadas: “agrupamento”, “associação” e “sumarização”.

Modelos preditivos (aprendizado supervisionado) possuem tarefas para encontrar uma função, modelo ou hipótese utilizados para realizar uma previsão. Por exemplo, posso querer prever um valor de um imóvel ou, por exemplo, o estado para um novo paciente daqui a 5 meses, depois da cirurgia, para indicar se ele estará doente ou saudável. Nestas previsões, temos uma entrada geralmente representada por X, e uma saída (geralmente representada por Y). Este modelo pretende:

- **Classificação:** os rótulos são discretos como os diagnósticos, se uma pessoa está doente ou saudável, ou se essa pessoa é boa ou má pagadora. Também, pode-se indicar se uma fruta é banana, maçã ou laranja.

- **Regressão:** os rótulos não são discretos, eles são contínuos, por exemplo: peso, altura etc.

Os **modelos descritivos** (aprendizado não supervisionado) possuem tarefas que são, basicamente, para explorar ou descrever um conjunto de dados, conforme uma das opções:

- **Agrupamento:** conforme a similaridade dos dados.
- **Associação:** explora-se as regularidades dos dados, procurando determinar as associações entre os atributos.
- **Sumarização:** objetiva a descrição compacta dos dados similares.

Importante ressaltar que existem diferentes tipos de análise de negócios: descritiva, diagnóstica, preditiva e prescritiva, conforme apresentado na Figura 2. Nós estudamos, portanto, duas das quatro possibilidades.

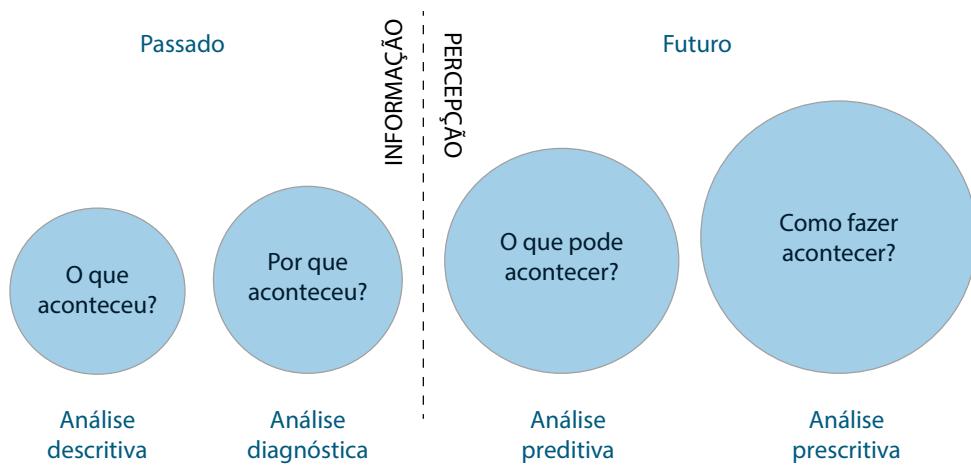


Figura 2 - Tipos de Análise de negócios / Fonte: Ferri (2019, on-line).

Descrição da Imagem: a ilustração representa os tipos de Análise de negócios. Têm-se quatro círculos em tamanho crescente da esquerda para direita. Os dois primeiros referem-se ao passado e informação, e os dois últimos ao futuro e percepção. No primeiro círculo está escrito "O que aconteceu?", referindo-se à "análise descritiva". No segundo círculo está escrito "por que aconteceu?", referindo-se à "análise diagnóstica". No terceiro círculo está escrito "o que pode acontecer?", referindo-se à "análise preditiva" e, por último, no quarto círculo está escrito "Como fazer acontecer?", referindo-se à análise prescritiva.

Ressalta-se que a **análise descritiva** é a classe mais simples de análise, pois trabalha com os dados brutos de forma a identificar o que aconteceu. A **análise diagnóstica** procura explicar porque aconteceu. A **análise preditiva** procura prever o que vai acontecer no futuro. A **análise prescritiva** pode ajudar as empresas a alterar o futuro, a partir de recomendações inteligentes para gerar ou acelerar resultados desejados. A principal diferença entre preditivo e prescritivo é que o primeiro prevê possíveis resultados futuros, enquanto o segundo ajuda a elaborar recomendações específicas para resultados desejados.

Para compreender melhor os algoritmos apresentados, a seguir, precisamos aprofundar o conteúdo referente à regressão linear.

Na **regressão linear**, você pode modelar a relação entre variáveis numéricas de forma que a variável dependente Y e as variáveis exploratórias X fazem parte do processo. A variável Y é o objetivo que se pretende fazer previsão, e as variáveis exploratórias X são os atributos previsores. Na sequência, alguns exemplos para melhor compreensão:

- O valor da temperatura, da umidade e da pressão do ar são variáveis exploratórias X, para que se possa fazer a previsão da velocidade do vento (variável Y).
- Os gastos no cartão de crédito e o histórico do cliente são as variáveis exploratórias X, para que se possa fazer a previsão do limite do cartão (variável Y).
- A idade de uma pessoa é variável exploratória X, para se conseguir fazer a previsão do custo do plano de saúde para esta pessoa (variável Y).

A **regressão linear simples** ocorre, quando temos apenas uma variável exploratória que utiliza a idade para prever esse custo. Vamos compreender melhor por meio do exemplo da especificação do seguro de saúde? Este é um exemplo clássico de desafio para a regressão linear simples resolver, porque esse tipo de regressão trabalha com apenas uma variável exploratória: a idade das pessoas.

Exemplificando, assuma que temos a idade da pessoa e o custo mensal para o plano de saúde, conforme a idade da pessoa: R\$800,00 (18 anos), R\$1.600,00 (32 anos), R\$2.400,00 (42 anos), R\$3.200,00 (54 anos) e R\$4.000,00 (66 anos). Ou seja, uma pessoa com, aproximadamente, 17 anos tem o custo do seu plano

de saúde por volta de R\$800,00; enquanto que uma pessoa com mais de 66 anos terá R\$4.000,00 de custo com o seu plano de saúde. Assuma, também, que esses dados já estão cadastrados na base de dados da empresa de planos de saúde. O objetivo é, com base na idade da pessoa, realizar a previsão do custo do seu plano de saúde. Note que as intersecções das variáveis X (idade) com Y (custo) demonstra que existe uma relação linear entre essas variáveis: quanto maior a idade, maior será o custo do seu plano de saúde.

A **regressão linear múltipla** acontece quando não podemos explicar a variável de resposta com apenas uma variável preditora, sendo o caso de incorporarmos várias variáveis independentes para chegarmos a solução. Por exemplo, estabelecer a relação entre os números da criminalidade com o índice de desemprego e escolaridade dos seus moradores, ou seja, se houver mais desemprego e menor escolaridade, haverá mais crimes.

Portanto, na **regressão linear simples** existe apenas uma variável de saída y e uma de entrada x de forma que obtemos $y = f(x)$. Em uma **regressão linear múltipla** existe apenas uma variável de saída y com várias de entrada x_i ($i = 1, \dots, n$) de forma que obtemos $y = f(x_1, x_2, \dots, x_n)$.



EXPLORANDO IDEIAS

Para compreender melhor sobre regressão linear, sugerimos a leitura do material "Regressão simples", disponibilizado por estatísticos. O objetivo deste material é facilitar a compreensão sobre métodos e verificações de dados disponíveis na ferramenta **Minitab Statistical Software**.



A **análise de regressão** pretende verificar a possibilidade de ajuste de um modelo que demonstra a relação, após explorar a natureza dessa relação. A análise de regressão linear tem por finalidade determinar o modelo que demonstre essa relação (equação de regressão), enquanto que a análise de correlação linear busca apresentar o grau de relacionamento entre as duas variáveis. A análise de regressão constrói um modelo matemático para representar dois atributos (variáveis) x e y :

- $y = f(x)$ onde $f(\cdot)$ é a função que relaciona x e y .
- x é a variável independente da equação.
- $y = f(x)$ é a variável dependente das variações de x .

Usamos esse modelo para prever o valor de Y para um determinado valor de X, fazendo, assim, previsões sobre o comportamento futuro de algum evento da realidade. Denominamos X como sendo a variável independente ou variável explicativa (manipulável) e Y como sendo a variável dependente ou variável resposta (observa o efeito).

Para representar visualmente os dados, utilizamos o **gráfico de dispersão** (em inglês, *Scatterplot*), do tipo cartesiano e com os pares de informação (referente a cada observação) x e y, formado por uma “nuvem de pontos” que define um eixo ou direção e caracteriza o padrão de relacionamento entre as variáveis x e y.

Para compreender melhor os gráficos de dispersão, veja o gráfico apresentado na Figura 3. Ele representa a relação entre Produto Interno Bruto per capita (PIB per capita) e a expectativa de vida das pessoas. Mais especificamente, cada ponto do gráfico apresenta o valor do logaritmo do PIB per capita de um país em relação à expectativa de vida desse país. Observa-se que quanto maior o PIB per capita, maior a expectativa de vida, ou seja, existe uma **correlação positiva** entre essas variáveis (à medida que uma variável aumenta, a outra variável também aumenta).

PIB per capita X expectativa de vida

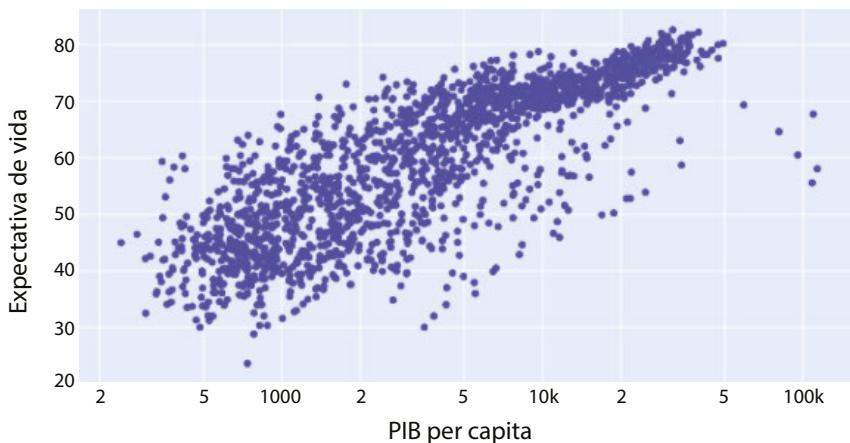


Figura 3 - Gráfico de Dispersão / Fonte: Scudiloh (2020, on-line).

Descrição da Imagem: Tem-se, na vertical, expectativa de vida, com os valores, de baixo para cima, 20, 30, 40, 50, 60, 70 e 80 e, na horizontal, PIB per capita, com os valores 1000, 10.000 e 100.000. Dentro do gráfico, há inúmeros pontos em azul, em que não se consegue ter a definição de um valor predominante, mas se percebe o crescimento da esquerda para a direita e de baixo para cima.

O parágrafo anterior apresenta o termo “correlação positiva”. Você deve ter se questionado: o que isso significa? Temos que compreender, porém, que uma variável causa efeito em outra, se existe uma relação. Além disso, podemos determinar a intensidade dessa relação, denominada **correlação**, que pode ser positiva, negativa ou neutra, linear ou não linear. A Figura 4 apresenta gráficos que representam esses tipos.

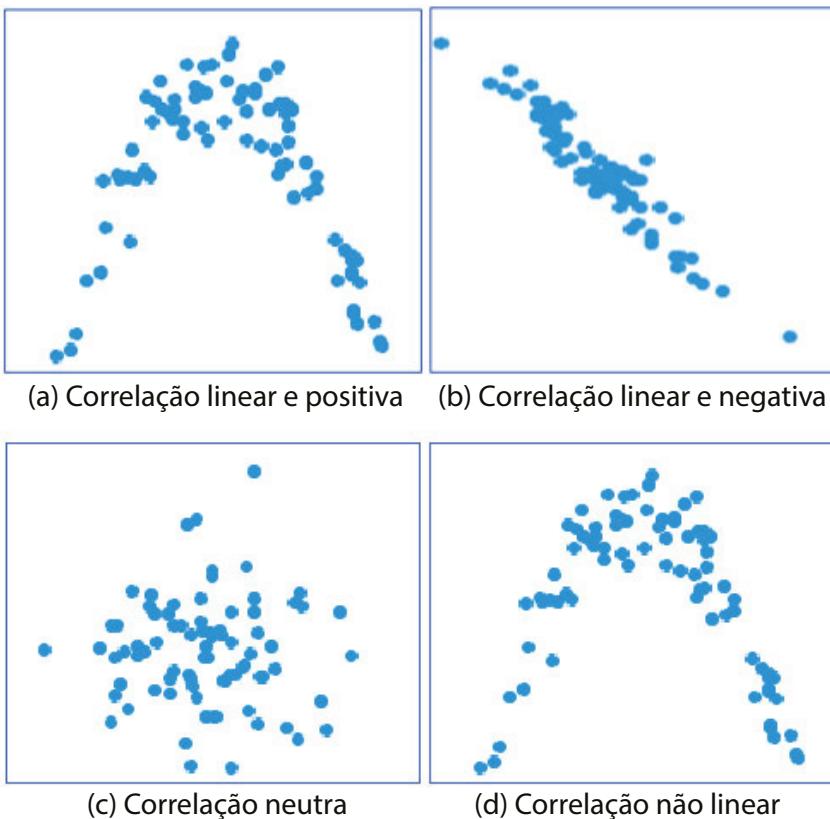
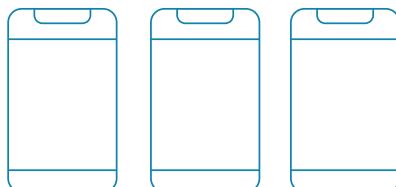


Figura 4 - Exemplos de correlação / Fonte: Scudiloh (2020, on-line).

Descrição da Imagem: exemplo com quatro gráficos, o primeiro gráfico apresenta a correlação linear e positiva, em que os valores, representados por pontos, iniciam na parte inferior e à esquerda do gráfico e crescem acima e à direita do mesmo. O segundo gráfico exemplifica a correlação linear e negativa, em que os pontos iniciam na parte superior esquerda e caem abaixo e à direita. O terceiro gráfico apresenta a correlação neutra, em que os pontos estão espalhados de forma irregular pelo gráfico. O quarto e último gráfico apresenta a correlação não linear, em que os pontos iniciam na parte de baixo, à esquerda, e crescem até o centro, acima, caindo à direita.

Importante salientar que a regressão linear nem sempre pode ser utilizada para que se encontre a melhor função dentro de um determinado conjunto de dados. Logo, existem outras formas de trabalhar com regressão, que não serão abordadas nesta unidade. Sugerimos, por isso, que você leia o material “Scatter plot: Um Guia Completo para Gráficos de Dispersão”.

Conforme Perelmutter (2020), os avanços das técnicas de Inteligência Artificial permitem que sistemas computacionais aprendam a inferir e a extrapolar com base em situações que são apresentadas e ensinadas, ao longo do tempo. Esse mesmo autor indica que o desenvolvimento de técnicas para a análise de grande quantidade de dados alavanca, de forma exponencial, a capacidade das máquinas para processar informações e gerar recomendações.



Existem plataformas que disponibilizam meios e recursos para automatizar os métodos e algoritmos que estamos estudando. Um exemplo é a plataforma Minitab que possibilita a integração, automação e governança de dados para insights abrangentes. O site oficial desta plataforma está disponível no segundo QR Code. Outras informações sobre esta plataforma podem ser encontradas no arquivo, em PDF, disponibilizado pela própria plataforma.

A partir desse ponto da unidade, apresentamos alguns dos algoritmos mais citados na atualidade, que são: árvores de classificação e regressão, k-vizinho mais próximo, redes neurais artificiais, máquinas de vetores de suporte e, finalmente, regressão logística. Precisamos, porém, inicialmente, relembrar alguns conceitos apresentados na Unidade 2, que são:

- **Algoritmo** é o conjunto de instruções dado a um computador para a execução de uma determinada tarefa. Quanto melhor ficam os computadores, melhores, mais eficientes e inteligentes ficam os algoritmos, realizando tarefas que, antes, eram realizadas exclusivamente pelos seres humanos.
- **Infraestrutura** é o conjunto de equipamentos físicos e os meios de comunicação que são disponibilizados para os usuários.
- **Ferramentas** são os algoritmos e programas disponibilizados para os usuários.

Ao utilizar a infraestrutura e ferramentas de terceiros, o tempo e o custo de implantação de soluções de IA podem ser reduzidos! Portanto, vamos também procurar apresentar ferramentas que são baseadas nesses algoritmos.

O algoritmo **Árvores de Classificação e Regressão** (em inglês, *Classification and Regression Trees - CART*) destina-se a produzir não uma árvore, mas uma sequência de árvores podadas aninhadas, em que cada uma delas é uma candidata a ser a árvore ótima. A árvore de decisão CART permite o particionamento recursivo binário capaz de processar atributos contínuos e nominais, vistos como alvos e preditores.

A árvore de “tamanho certo” é identificada, quando se avalia o desempenho preditivo de cada árvore, na sequência da poda com dados de teste independentes. O desempenho da árvore é medido com dados de testes independentes (ou por meio de validação cruzada) e a seleção de uma árvore ocorre após avaliação baseada nos dados desses testes. Se o teste ou a validação cruzada não foram realizados, o algoritmo CART permanece agnóstico em relação à qual das árvores é a melhor.

A seguir, um algoritmo simplificado, esboçando o mecanismo para construir a maior árvore possível, sem fazer referência aos valores ausentes, atribuições de classe ou outros detalhes essenciais da CART. Os dados são tratados em sua forma bruta; nenhum tratamento é necessário ou recomendado. Começa-se pelo nó raiz, os dados são divididos em dois ramos filhos. Na sequência, cada ramo filho, por sua vez, é dividido em ramos netos. As árvores vão crescendo até o seu tamanho máximo, não há uma regra de parada. Essencialmente, o processo de crescimento da árvore para quando não há mais divisões possíveis devido a falta de dados. A árvore resultante é, então, podada de volta à sua raiz, por meio de um método de poda de complexidade de custo. O ramo a ser podado é o que menos contribui para o desempenho geral da árvore para o treinamento dos dados.

INÍCIO: Atribuir todos os dados de treinamento ao nó raiz

Defina o nó raiz como um nó terminal

DIVISÃO:

NovasDivisões = 0

PARA cada nó terminal na árvore:

-Se o tamanho da amostra do nó terminal for muito pequeno ou todas as instâncias no nó pertencerem a mesma classe de destino, vá para OBTER_PRÓXIMA.

-Encontre um atributo que melhor separe o nó em dois nós filhos, aplicando uma regra de divisão permitida.

-NovasDivisões + 1

OBTER_PRÓXIMA:

PRÓXIMA

A seguir, o esboço simplificado do algoritmo para realizar a poda, ignorando antecedentes e custos. O procedimento começa determinando a maior árvore construída (T_{max}), depois remove todas as divisões e gera dois nós terminais que não melhoraram a precisão da árvore com dados de treinamento. Esse é o ponto de partida para fazer a poda, ou seja, a remoção iterativa dos elos mais fracos da árvore que menos contribuem para o desempenho da árvore sobre dados de teste. No algoritmo apresentado a seguir, a poda é restrita aos pais de dois nós terminais.

DEFINA: $r(t)$ = taxa de classificação incorreta de dados de treinamento no nó t

$p(t)$ = fração dos dados de treinamento no nó t

$R(t) = r(t) * p(t)$

t_{left} = filho esquerdo do nó t

t_{right} = filho direito do nó t

$|T|$ = número de nós terminais na árvore T

INÍCIO: T_{max} = maior árvore construída

`Current_Tree = T_{max}`

Para todos os pais t de dois nós terminais

Remova todas as divisões para as quais $R(t) = R(t_{left}) + R(t_{right})$

`Current_Tree = T_{max} após poda`

PODA: Se $|Current_Tree| = 1$ então vá para FEITO

Para todos os pais t de dois nós terminais

Remova todo nó t tal que $R(t) - R(t_{left}) - R(t_{right})$ é mínimo

`Current_Tree = Current_Tree depois da poda`

FEITO:

Software que implementa o algoritmo CART está disponível na Salford Systems, disponível em www.salford-systems.com, de forma que versões para avaliação sem custo podem ser baixadas mediante solicitação.



Sugiro que assista ao vídeo sobre a ferramenta CART, disponível no QR Code. Este vídeo procura salientar que essa ferramenta descobre maneiras de dividir os dados em segmentos menores e, a seguir, seleciona as divisões de melhor desempenho, recursivamente, até que uma coleção ideal seja encontrada.



O algoritmo **k-vizinho mais próximo** (em inglês, *k-Nearest Neighbors* — kNN) procura encontrar um grupo de objetos k no conjunto de treinamento que estão mais próximos do objeto de teste, baseando-se na atribuição de um rótulo na predominância de uma classe particular nesta vizinhança. Isso ressalta o problema de que, em muitos conjuntos de dados, é improvável que um objeto corresponda exatamente a outro, bem como o fato de que informações conflitantes sobre a classe de um objeto podem ser fornecidas pelos objetos mais próximos a ele. Existem vários elementos-chave dessa abordagem:

1. O conjunto de objetos rotulados a serem usados para avaliar uma classe de objeto de teste.
2. Uma distância ou métrica de similaridade que pode ser usada para calcular a proximidade dos objetos.
3. O valor de k , o número de vizinhos mais próximos.
4. O método usado para determinar a classe do objeto de destino, com base nas classes e distâncias dos k vizinhos mais próximos.

Em sua forma mais simples, o algoritmo kNN pode envolver a atribuição de um objeto à classe de seu vizinho mais próximo ou da maioria de seus vizinhos mais próximos, mas uma variedade de melhorias é possível e discutida a seguir.

Entrada: D (o conjunto de objetos de treinamento), z (objeto teste, um vetor com valores de atributo), e L (conjunto de classes utilizadas para denominar os objetos).

Saída: $CZ \in L$, a classe de Z

Para cada objeto $y \in D$ faça

Calcular $d(z,y)$, a distância entre z e y ;

Fim

Selecionar $N \subseteq D$, o conjunto (vizinho) de k o mais próximo dos objetos de treinamento para z :

$$c_z = \operatorname{argmax}_{v \in L} \sum_{y \in N} I(v = \operatorname{class}(c_y));$$

onde $I(\cdot)$ é uma função indicador que retorna o valor 1 se é o seu argumento; 0, caso contrário.

Dado um conjunto de treinamento D e um objeto de teste z, que é um vetor de valores de atributos e tem um rótulo de classe desconhecido, o algoritmo calcula a distância (ou similaridade) entre z e todos os objetos de treinamento para determinar sua lista de vizinhos mais próximos. Isso, então, atribui uma classe a z, pegando a classe da maioria dos objetos vizinhos. Laços são quebrados de maneira não especificada, por exemplo, aleatoriamente ou tirando o máximo de classe frequente no conjunto de treinamento.

A complexidade de armazenamento do algoritmo é $O(n)$, em que n é o número de treinamento objetos. A complexidade do tempo também é $O(n)$, uma vez que a distância precisa ser calculada entre o alvo e cada objeto de treinamento. Não há, porém, tempo para a construção do modelo de classificação, por exemplo, uma árvore de decisão ou avaliação hiperplano. Assim, kNN é diferente da maioria das outras técnicas de classificação, que têm estágios de construção de modelos moderados, com valores diversificados.

A Figura 5 mostra um objeto de teste sem rótulo “x”, e objetos de treinamento que pertencem a uma classe “+” ou “-”.

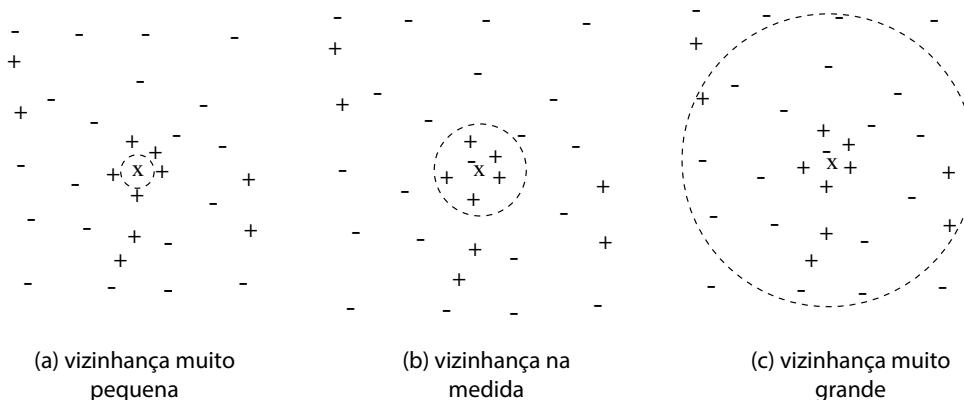


Figura 5 - Classificação kNN com k pequeno, médio e grande / Fonte: a autora.

Descrição da Imagem: Ilustração composta por três quadros, todos com os mesmos símbolos negativo (-) e positivo (+), diversos e dispersos pelo quadro. Ao meio dos três quadros têm um símbolo x. O primeiro quadro possui um pequeno círculo com o símbolo x dentro e sinais de positivo (+) ao redor do círculo, abaixo lê-se “vizinhança muito pequena”. O segundo quadro apresenta um círculo médio, com o símbolo x ao centro mais alguns sinais de positivo (+) e um negativo (-) ao redor deste símbolo x, abaixo a legenda “vizinhança na medida”. O terceiro e último quadro apresenta um grande círculo com o símbolo x, os mesmos sinais de positivo (+) e um negativo (-) ao redor do símbolo x, seguido pela legenda “vizinhança muito grande”.

Note que, se k for muito pequeno, então o resultado pode ser sensível a pontos de ruído. Por outro lado, se k for muito grande, então a vizinhança pode incluir muitos pontos de outras classes. Uma estimativa de um melhor valor para k pode ser obtida por validação cruzada. Se as amostras são suficientes, valores maiores do que k são mais resistentes ao ruído.

Conforme já apresentado na Unidade 4, **Redes Neurais Artificiais** (em inglês, *Artificial Neural Networks — ANN*) da IA disponibilizam algoritmos que permitem simular a rede neural do cérebro humano. Relembando, uma ANN é composta por um conjunto de neurônios (unidade computacional básica e conectados em rede). A Figura 6 demonstra a diversidade de ANN que um computador pode construir.

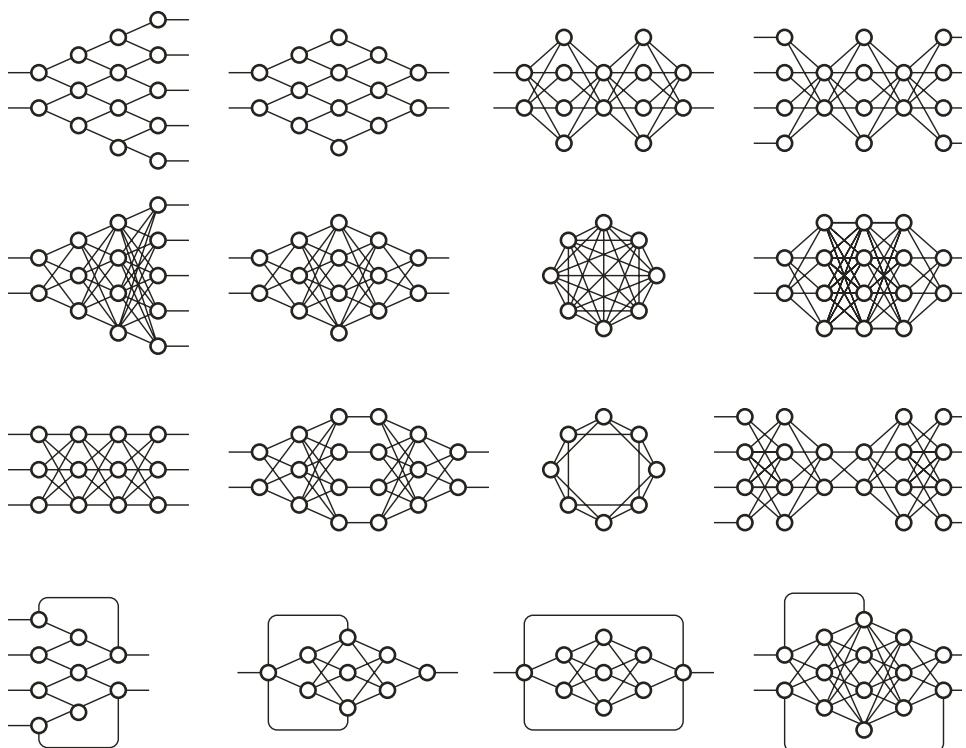
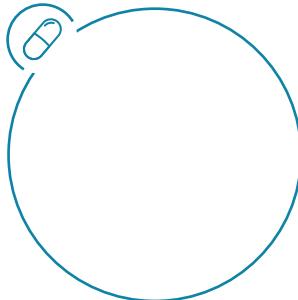


Figura 6 - Exemplos de Redes Neurais Artificiais

Descrição da Imagem: Ilustração que exemplifica diferentes configurações de redes neurais artificiais. Cada uma das 16 configurações é formada por pequenos círculos conectados por linhas simples. Em algumas destas configurações, as conexões entre os pontos formam imagens mais uniformes; em outras configurações existem mais conexões entre os pontos e, por consequência, as imagens são mais complexas.

Para compreender os **algoritmos de ANN**, apresentaremos o Perceptron, um método clássico de aprendizagem supervisionado, proposto por Frank Rosenblatt, em 1943, e posteriormente analisado e refinado por Minsky e Papert, em 1969.



Para compreender melhor a proposta do Algoritmo Perceptron, você deve relembrar alguns conceitos de Álgebra, tais como: vetor, representação de vetores, produto escalar de dois vetores, ângulo entre dois vetores e configurando o problema. Assista ao vídeo e, depois, continue a leitura deste texto! Acredito que você vai compreender melhor!

A apresentação da rede de Perceptrons, conforme a Figura 7, demonstra que os neurônios devem ser organizados em camadas: a camada de entrada possui conexões fixas (simulando a retina humana); depois, recebe impulsos devido a essas conexões (atribuindo-se pesos), sendo enviados para uma camada de saída (respostas).

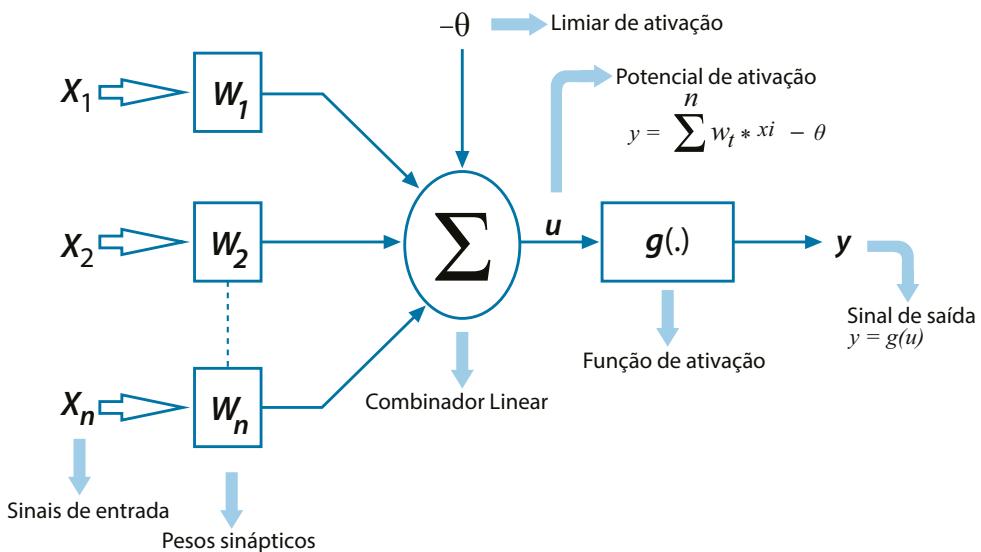


Figura 7 - Rede Perceptrons / Fonte: Embarcados ([2021], on-line).

A seguir, um exemplo do algoritmo Perceptron que pretende encontrar um vetor w , este pode classificar os dados de entrada positivos e os dados de entradas negativos.

```

P ← entradas com label 1;
N ← entradas com label 0;

INICIALIZAR w RANDOMICAMENTE;

ENQUANTO! converge FAÇA

    Escolha aleatória x ∈ P ∪ N ;

    SE x ∈ P and w.x < 0 ENTÃO

        w = w + x;

    SE x ∈ N and w.x ≥ 0 ENTÃO

        w = w - x;

    FIM

// o algoritmo converge quando todas as entradas são classificadas corretamente.

```

O vetor \mathbf{w} é inicializado com um valor aleatório. Em seguida, itera-se sobre todos os exemplos nos dados ($P \cup N$), ou seja, exemplos positivos e negativos. Reflita: se uma entrada x pertence a P , o que seria idealmente o produto escalar $\mathbf{w} \cdot \mathbf{x}$? Deve ser um valor maior ou igual a 0, porque essa é a única coisa que este algoritmo Perceptron pretende. Note que, se x pertencer a N , o produto escalar DEVE ser menor que 0. Observe as condições SE no ENQUANTO:

- a) Quando x pertence a P e seu produto escalar $w \cdot x < 0$
- b) Quando x pertence a N e seu produto escalar $w \cdot x \geq 0$

Apenas para esses dois casos, o w será alterado aleatoriamente. Isso ocorre porque os casos (A) e (B) estão violando a própria regra de um *perceptron*. Caso contrário, nada será alterado. Para que você compreenda melhor o processo de desenvolvimento de aplicações, utilizando ANN. Esse processo é organizado nas seguintes etapas:

1. Coleta de dados relacionados à tarefa (ou problema a resolver).
2. Separação em conjuntos (ou seja, a classificação dos dados).
3. Configuração.
4. Treinamento e teste da ANN.
5. Integração desta em um sistema computacional.

Sugiro que consulte e leia com atenção a página “Cinco etapas para criar e implantar redes neurais profundas de aprendizado” ([clique aqui para acessá-la](#)), pois apresenta com detalhe como são estas cinco etapas. Peremulter (2020) indica que sistemas baseados em DL utilizam, quase sempre, redes neurais artificiais (inspiradas no funcionamento do cérebro) e são capazes de extrair características necessárias para a criação de regras e padrões de grande quantidade de dados e de respostas. Ressalta-se que os sistemas baseados em DL são, **teoricamente**, capazes de melhorar a precisão de tarefas que exijam reconhecimento de padrões e classificação; contudo são muito dependentes/sensíveis da qualidade e dos tipos de dados que utilizam.



NOVAS DESCOBERTAS

Estudos recentes no MIT apresentam a Rede Neural "Líquida" (RNL) que é capaz de aprender a estrutura de causa e efeito da tarefa de navegação para a qual está sendo treinada. As RNL podem mudar suas equações subjacentes para se adaptar continuamente às novas entradas. Por exemplo, uma RNL com 19 neurônios de controle é capaz de controlar autonomamente um veículo. Estas redes compreendem a tarefa diretamente de dados visuais, são mais eficazes do que outra rede neuronal ao navegar em ambiente complexo, por exemplo, um local com árvores densas ou com condições climáticas alternadas. No futuro, pretende-se aprimorar a confiabilidade de forma a realizar tarefas de alto risco, por exemplo, dirigir um veículo autônomo em uma rodovia movimentada.



Fonte: Zewe (2021, on-line).



NOVAS DESCOBERTAS

Veja o material que ensina como criar uma rede neural do zero: "Rede Neural Perceptron Adaline". A rede Adaline (Adaptive Linear Element) foi proposta por Widrow e Hoff, em 1960, e tem a mesma estrutura do Perceptron, diferenciando apenas no algoritmo de treinamento. Procure ler com atenção e, se possível, realize todas as atividades propostas.



O algoritmo **Máquinas de Vetores de Suporte** (em inglês, *Support Vector Machines* - SVM) é um dos métodos robustos e precisos para a mineração de dados. O SVM tem uma base teórica sólida, enraizada na teoria do aprendizado estatístico, requerendo apenas alguns exemplos para o treinamento e, muitas

vezes, são insensíveis ao número de dimensões. Uma das desvantagens dos algoritmos SVMs é, na fase de treinamento, uma onerosa complexidade computacional, não permitindo aplicar em grandes conjuntos de dados. Este problema, no entanto, é resolvido com sucesso. Aplica-se a abordagem de quebrar um grande problema de otimização em uma série de problemas menores, em que cada problema envolve apenas algumas variáveis cuidadosamente escolhidas para que a otimização possa ser feita de forma eficiente. Esse processo itera até que todos os problemas de otimização sejam resolvidos com sucesso.

No kernel dos algoritmos do SVM, as funções principais devem satisfazer o teorema de Mercer. Para compreensão desse teorema, é preciso uma boa base matemática. Se você quiser se aprofundar, sugiro que conheça os três tipos de funções — sigmóide, polinomial e funções de base radial —, as quais, às vezes, podem limitar a aplicabilidade do truque do kernel: um método que permite simplificar o processamento de forma a evitar computar diretamente o espaço de recursos complexos, não apenas no cálculo de produtos internos, mas também para o projeto do classificador. Existem dois softwares famosos que implementam o algoritmo SVM:

- **LibSVM**: Apresenta códigos-fonte C++ e Java que podem ser melhorados, revisados e aplicados em diferentes sistemas operacionais. Além disso, disponibiliza poucos parâmetros ajustáveis para o algoritmo SVM e fornece muitos parâmetros padrões que permitem resolver diversos problemas reais.
- **SVMlight**: foi implementado em linguagem C, ele adota uma técnica de seleção de conjunto eficiente com base na descida mais íngreme possível com as políticas computacionais de avaliações: “Redução” e “Cache”; inclui ainda os programas: SVM learn, este permite aprender amostras de treinamento e treinar o correspondente classificador, e SVM classify, este permite classificar amostras de teste. Além disso, existem muitas Toolbox completas de aprendizado de máquina, incluindo algoritmos SVM, como Torch (em C ++), Spider (em MATLAB) e Weka (em Java), disponíveis [aqui](#).

Sugerimos que você acesse a esses dois sites, procure ver as orientações de instalação e uso e, finalmente, procure usá-los! Modelos de previsão de risco clínico são aplicados em muitos domínios médicos, pois eles visam a prever um resultado clinicamente relevante, usando informações no nível da pessoa. Uma abordagem

tradicional para desenvolver esses modelos envolve o uso de modelos de regressão, por exemplo, **Regressão logística** (em inglês, *Logistic Regression - LR*) para prever a presença da doença (ou seja, um diagnóstico) ou resultados da doença (ou seja, prognóstico). Outras aplicações reconhecidas que envolvem LR são:

- Fazer a previsão de risco na área tributária, em que se pode calcular a probabilidade do contribuinte ficar inadimplente (ou adimplente), após o parcelamento de seus tributos.
- Classificar grupo de empresas solvente ou insolvente.
- Determinar quais são as características que levam as empresas a adotarem o balanced scorecard.

A LR permite estimar a probabilidade associada à ocorrência de um determinado evento, em face do conjunto de variáveis explanatórias. Recomenda-se aplicar quando a variável dependente é de natureza dicotómica ou binária (ou seja, apenas dois valores possíveis) e as variáveis independentes, podendo ser categóricas ou não.

A LR busca estimar a probabilidade da variável dependente e assumir um valor em função dos valores conhecidos de outras variáveis, os resultados desta análise estão contidos no intervalo dos valores de 0 até 1. Ou seja, a probabilidade de ocorrência de um evento pode ser estimada da seguinte forma: assuma que uma variável dependente Y tem apenas dois valores possíveis — 0 ou 1 — e que existe um conjunto P de variáveis independentes x_i . O modelo de regressão logística pode ser escrito da seguinte forma:

$$P(Y=1) = \frac{1}{1+e^{-g(x)}} \quad \text{onde } g(x) = B_0 + B_1X_1 + \dots + B_P X_P$$

Os coeficientes B_0, B_1, \dots, B_p são estimados a partir do conjunto dados, pelo método da «máxima verossimilhança» que procura encontrar a combinação de coeficientes que maximiza a probabilidade da amostra ter sido observada. Considerando a combinação de coeficientes B_0, \dots, B_P , procurando variar os valores de X.

A curva logística apresenta um comportamento probabilístico na forma da letra S, característica da regressão logística, conforme pode ser visto na Figura 8. Portanto, $g(x) \rightarrow +\infty$ então $P(Y=1) \rightarrow 1$; caso contrário, $g(x) \rightarrow -\infty$ então $P(Y=1) \rightarrow 0$.

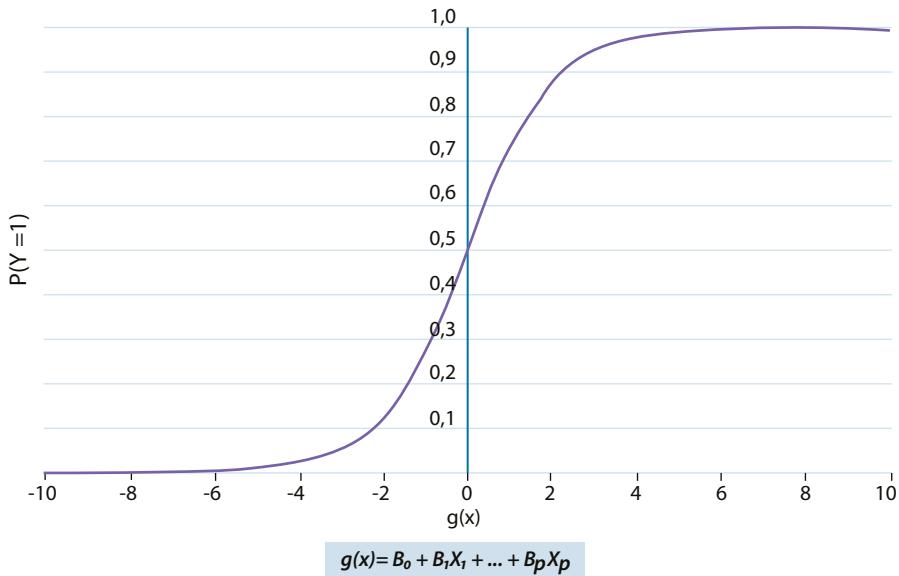


Figura 8 - Regressão Logística / Fonte: a autora.

Descrição da Imagem: gráfico representando a regressão logística em que, na vertical, tem-se $P(Y=1)$, com os valores de 0,1 a 1,0 e, na horizontal, $g(x)$, com os valores de -10 até 10. Temos a curva em S que se inicia em -10 e finaliza em 1,0.

Finalmente, algumas vantagens de usar regressão logística: facilidade para lidar com variáveis independentes categóricas; fornece resultados em termos de probabilidade; facilidade de classificação de indivíduos em categorias; requer pequeno número de suposições e alto grau de confiabilidade.

Acreditamos que você deve estar fascinado com as soluções apresentadas, principalmente pelas aplicações disponíveis para uso.

Você se recorda que, na Unidade 3, foi apresentado o algoritmo K-means? Na sequência desta unidade, citamos algumas das variações deste algoritmo, denominadas por: expectativa-maximização, diagramas de Voronoi e K-means++.

O algoritmo K-means pode ser compatibilizado com o algoritmo denominado de **algoritmo de expectativa-maximização** (em inglês, *Expectation–maximization — EM*), pois este é aplicável em atividades relacionadas com a clusterização. O EM visa a estimar parâmetros para modelos estatísticos e até na reconstrução de imagens de tomografias médicas computadorizadas.

Além disso, usar **diagramas de Voronoi** (em inglês, *Mosaicos de Dirichlet*) é uma forma intuitiva de referenciar o algoritmo K-means. Esse diagrama compõe um determinado espaço em subconjuntos, compondo uma espécie de mosaico que delimita os subconjuntos, sendo bastante adequado para a compreensão da aplicação do algoritmo K-means, pois é capaz de representar os clusters gerados a partir de um conjunto completo de elementos.

Para a geração do diagrama de Voronoi, como é visto na Figura 9, os pontos utilizados para a construção do diagrama se baseiam nos centróides de cada iteração e cada área de cor distinta representa um cluster de dados. Basicamente, o algoritmo repete iterações que geram novos centróides até que determinado estado seja atingido e, no caso da geração do diagrama, em cada iteração, os pontos são gerados a certa distância da posição obtida na iteração anterior, e o processo encerra, quando a distância desse movimento for menor que um valor limite estabelecido de tolerância. Como a geração da primeira amostra é aleatória de certa forma, é comum que o algoritmo seja repetido várias vezes, de forma a se observar o comportamento do algoritmo e resultados obtidos a cada execução e, ao final, os melhores resultados avaliados.

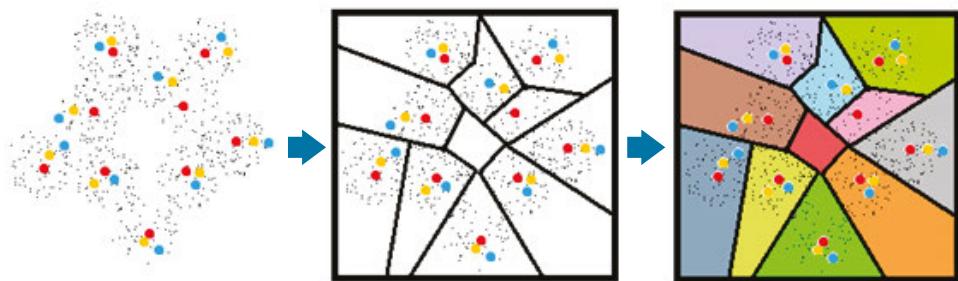


Figura 9 - Exemplo ilustrativo de Diagrama Voronoi aplicado ao método K-means / Fonte: a autora.

Descrição da Imagem: a figura traz os passos para a geração de um diagrama, que inicia a partir do estado final da Figura 3, da Unidade 3, com um conjunto de pontos em que alguns estão destacados em azul, outros em amarelo e outros em vermelho. Depois, no passo seguinte, um quadrado é desenhado, ao redor do conjunto, e linhas cortam a figura, separando o conjunto de pontos, formando um mosaico de peças, a partir da posição dos trios de pontos coloridos próximos que devem ficar em áreas separadas do mosaico. Por fim, as áreas são coloridas com cores diversas, diferentes umas das outras.

O **algoritmo K-means++** é mais uma variante do algoritmo K-means que busca gerar centróides para a amostra inicial com uma certa distância mínima de forma que se tenha uma certa isotropia para que melhores resultados sejam provavelmente obtidos. Analisando a Figura 9, observa-se uma preocupação em manter os centróides iniciais para a amostra de aplicação do algoritmo a uma distância, relativamente, simétrica entre si, para que, nas iterações subsequentes, os próximos centrinos sejam gerados até que a condição de parada das iterações seja satisfeita. Já, no exemplo da Figura 10, é perceptível que os centróides gerados na iteração inicial tenham certa simetria em termos de distância entre si para que se tenha a impressão de que centróides bem distribuídos tenham sido escolhidos para que, nas próximas iterações, possam ser obtidos bons centróides e, ao final, uma boa organização de clusters.



Figura 10 - Exemplo ilustrativo de inicialização de um algoritmo K-means++ / Fonte: a autora.

Descrição da Imagem: a figura contém um conjunto de pontos espalhados e, nestes pontos, estão alguns deles destacados em laranja, indicando a primeira escolha de pontos do algoritmo.

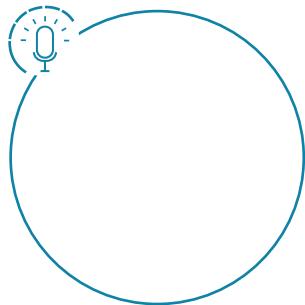


Figura 11 - Exemplo de inicialização de um algoritmo K-means++ / Fonte: a autora.

Descrição da Imagem: na figura, há quatro grupos de pontos separados, formando um conjunto de pontos laranja, outro azul, outro verde e outro lilás. Estes conjuntos possuem um ponto destacado em um tom de azul escuro para indicar os pontos, inicialmente, escolhidos.

O algoritmo K-means++ pode ser, inclusive, utilizado para gerar amostras iniciais que podem ser chamadas de sementes para outros algoritmos, de forma a obter amostras controladas e que, provavelmente, gerarão melhores resultados nos algoritmos que utilizarem essas amostras. Finalizamos esta unidade com dicas, adaptadas de Taulli (2020), que podem ajudar você na implementação de soluções de IA:

1. Identifique e defina, com o máximo de precisão, o problema a resolver.
2. Selecione ferramentas e plataformas adequadas.
3. Crie o modelo de IA em conformidade com as ferramentas e plataformas utilizadas.
4. Implante o modelo.
5. Monitore os resultados, e caso não sejam satisfatórios, faça ajustes no modelo.
6. Fique atento(a) à qualidade e aos tipos de dados que serão utilizados.



Caro(a) aluno(a), não perca nosso último podcast da disciplina! Escute no QR Code ao lado.

Caro(a) aluno(a), finalizamos esta unidade. Espero que você tenha percebido a complexidade de construir algoritmos e implementá-los. Espero, também, que você tenha percebido que já existem plataformas (gratuitas e pagas) que estão disponíveis para construção de uma solução IA. Evite inventar a roda! Mas, sim, ter conhecimento do que já temos disponível e usar a seu favor. Além disso, ressaltamos que, para que você tenha plena compreensão sobre esta temática, será muito importante que você procure se aprofundar e, principalmente, construir soluções práticas para adquirir experiência.

AGORA É COM VOCÊ



Caro(a) aluno(a), espero que você tenha compreendido o conteúdo apresentado nesta unidade. Sabemos que esse assunto é complexo e que apresenta muitas informações e conceitos. Sugerimos que você complemente o Mapa Conceitual elaborado na Unidade 2 — Métodos e Algoritmos — para perceber e, depois, relembrar os diversos conceitos apresentados e, principalmente, determinar a relação entre todos os conceitos apresentados nestas duas unidades. O Mapa conceitual é muito utilizado para sintetizar os conceitos e torná-los representativos visualmente, permitindo que você exerçite o seu pensamento crítico! Uma sugestão de ferramenta gratuita é o www.goconqr.com.

REFERÊNCIAS



UNIDADE 1

ALMEIDA, A.; CARVALHO, F.; MENINO, F. Introdução ao Machine Learning: de alunos para alunos. **Dataat**, [S. I.], [2021]. Disponível em: <https://dataat.github.io/introducao-ao-machine-learning/introdu%C3%A7%C3%A3o.html>. Acesso em: 24 nov. 2021.

BUDKEWICZ, M. Aprendizado Supervisionado com exemplos (Supervised Learning). **Horadecodar**, [s. I.], 2018. Disponível em: <https://medium.com/horadecodar/aprendizado-supervisionado-com-exemplos-supervised-learning-f9856fed2445>. Acesso em: 23 nov. 2021.

CARVALHO, J. N. de. **Seleção e classificação inteligente de mangas por análise de imagens**. 2015. 272 f. Tese (Doutorado em Engenharia de Processos) – Programa de Pós-Graduação em Engenharia de Processos, Centro de Ciência e Tecnologia, Universidade Federal de Campina Grande, Paraíba, 2015.

GÉRON, A. Hands-On Machine Learning with Scikit-Learn and TensorFlow. [S. I.]: O'Reilly Media, 2017.

GINSBERG, M. L. **Essentials of artificial intelligence**. San Mateo: Morgan Kaufmann Publishers, 1993.

INTELIGÊNCIA ARTIFICIAL. In: MICHAELIS Brasileiro da Língua Portuguesa. São Paulo: Melhoramentos, [2021]. Disponível em: <https://michaelis.uol.com.br busca?id=PqO2A>. Acessada em: 23 nov. 2021.

MITCHELL, T. **Machine Learning**. New York: McGraw-Hill, 1997.

NORMAN, R. Inteligência Artificial, Machine Learning e o futuro dos profissionais de marketing. **Think with google**, [S. I.], 2018. Disponível em: <https://www.think-withgoogle.com/intl/pt-br/futuro-do-marketing/novas-tecnologias/inteligencia-artificial-machine-learning-e-o-futuro-dos-profissionais-de-marketing/>. Acesso em: 24 nov. 2021.

O QUE É Inteligência artificial? Como funciona, exemplos e aplicações. **Totvs**, [s. I.], 2019. Disponível em: <https://www.totvs.com/blog/inovacoes/o-que-e-inteligencia-artificial/>. Acesso em: 23 nov. 2021.

PIMENTA, I. [Guia Completo] Inteligência Artificial: o que é, conceito e métodos de IA. **Take Blipblog**, [s. I.], 2021. Disponível em: <https://www.take.net/blog/tecnologia/inteligencia-artificial>. Acesso em: 23 nov. 2021.

RICH, E; KNIGHT, K. **Artificial Intelligence**. New York, NY: McGraw Hill 1991.

RUSSEL, S.; NORVIG, P. **Inteligência Artificial**. 3. ed. [S. I.]: Gen Ltc, 2013.

REFERÊNCIAS



- RUSSEL, S.; NORVIG, P.; **Artificial Intelligence**: A Modern Approach. [S. I.]: Prentice-Hall, 1994.
- SAMUEL, A. L. Some Studies in Machine Learning Using the Game of Checkers. **Journal of Research and Development**, [S. I.], v. 3, n., p. 210-229, 1959.
- SILVA, F. M. da Silva; LENN, M. L.; FREITAS, P. H. C.; SANTOS, S. C. B. **Inteligência Artificial**. Porto Alegre: SAGAH, 2019.
- TAULLI, T. **Introdução à Inteligência Artificial**: uma Abordagem Não Técnica. [S. I.]: Novatec, 2020.
- TURING, A. M. Computing machinery and intelligence. **Mind**, [S. I.], v. 59, e. 236, p. 433-60, 1950.

UNIDADE 2

- CASTRO L. N.; FERRARI, D. G. **Introdução à Mineração de Dados**: Conceitos Básicos, Algoritmos e Aplicações. São Paulo: Saraiva, 2016.
- MITCHELL, T. M. Does Machine Learning Really Work? **AI Magazine**, [S. I.], v. 18, n. 3, p. 11-20, 1997. Disponível em: <https://doi.org/10.1609/aimag.v18i3.1303>. Acesso em: 26 nov. 2021.
- PEREMLUTER, G. **Futuro Presente**: o mundo movido à tecnologia. São Paulo: Companhia Editora Nacional, 2020.
- SEJNOWSKI; T. J. **A Revolução do Aprendizado Profundo**. Tradução de Carolina Gaio. Rio de Janeiro: Alta Books, 2019.
- SELEÇÃO um gráfico de dispersão. **Suporte ao Minitab® 19**, [S. I.], [2021]. Disponível em: <https://support.minitab.com/pt-br/minitab/19/help-and-how-to/graphs/scatterplot/create-the-graph/select-a-scatterplot/>. Acesso em: 26 nov. 2021.
- SILVA, J.C. Algoritmos de Aprendizagem de Máquina: qual deles escolher? **Machina Sapiens**, [S. I.], 2018. Disponível em: <https://medium.com/machina-sapiens/algoritmos-de-aprendizagem-de-m%C3%A1quina-qual-deles-escolher--67040ad68737>. Acesso em: 26 nov. 2021.
- SPERLING, E. Deep Learning Spreads. **Semiconductor Engineering**, [S. I.], 2018. Disponível em: <https://semiengineering.com/deep-learning-spreads/>. Acesso em nov. 2021.
- TAN, P. N.; STEINBACH, M.; KUMAR, V. **Introdução ao DATAMINING Mineração de dados**. Rio de Janeiro: Editora Ciência Moderna, 2009.
- TAN, P.N. et al. **Introduction to Data Mining**. 2. ed. Alemanha: Pearson, 2019.
- TAULLI, T. **Introdução à Inteligência Artificial**: uma Abordagem Não Técnica. [S. I.]:

REFERÊNCIAS



Novatec, 2020.

UNIDADE 3

BENJAMIN, M. Mapa mental: aprenda a técnica e construa anotações criativas. **Alura**, [s. l.], 2021. Disponível em: https://www.alura.com.br/artigos/mapa-mental-aprenda-a-tecnica?gclid=CjwKCAiAs92MBhA-XEiwAXTi259WX6yW08M1Fq4TjepSiRxiSkqfa9Xvj-mFv4CTV2SMODpQUXdDyghoCqk8QAvD_BwE. Acesso em: 15 dez. 2021,

DAVENPORT, T. **Reengenharia de processos**. São Paulo: Campus, 1994.

ESTATÍSTICA de como: estatística para o resto de nós. **Statistics How To**, [s. l.], [2021]. Disponível em: <https://www.statisticshowto.com/hierarchical-clustering/>. Acesso em: 14 dez. 2021.

HRUSCHKA, E. R.; EBECKEN, N. F. F. **A Genetic algorithm for cluster analysis**. [s. l.]: IEEE Transactions on Evolutionary Computation, 2001.

MITCHELL, T. M. Does Machine Learning Really Work? **AI Magazine**, [s. l.], v. 18, n. 3, p. 11, 1997. Disponível em: <https://ojs.aaai.org/index.php/aimagazine/article/view/1303>. Acesso em: 14 dez. 2021.

UNIDADE 4

HAYKIN, S. **Redes Neurais**: princípios e práticas. [s. l.]: Bookman, 2007.

SILVA, F. M. da S.; LENZ, M. L.; FREITAS, P. H. C.; SANTOS, S. C. B. **Inteligência Artificial**. Porto Alegre: SAGAH, 2019.

ÜLEDÄGSOZLUK GALEXY. **[Sem título]**. [2021]. 1 figura, color. Disponível em: <https://galeri.uludagsozlu.com/g/bilincin-matematiksel-bir-yap%C4%B1-olma-ih-timali/>. Acesso em: 29 nov. 2021.

Zewe, A. These neural networks know what they're doing: a certain type of artificial intelligence agent can learn the cause-and-effect basis of a navigation task during training. **MIT News**, Cambridge, 2021. Disponível em: <https://news.mit.edu/2021/cause-effect-neural-networks-1014>. Acesso em: 30 nov. 2021.

REFERÊNCIAS



UNIDADE 5

EMBARCADOS. [Sem título]. [2021]. 1 figura. Disponível em: <https://www.embarcados.com.br/wp-content/uploads/2016/09/Perceptron-01.png>. Acesso em: 15 dez. 2021.

FERRI, E. Análise Preditiva ou Prescritiva? Sua empresa precisa de ambos. **Medium**, [s. l.], 2019. Disponível em: <https://medium.com/@edselferri/an%C3%A1lise-preditiva-ou-prescritiva-s%C3%AA-a-empresa-precisa-de-amb%C3%B5s-63b49caf09cf>. Acesso em: 15 dez. 2021.

JOSÉ, I. Tipos de aprendizagem. **Medium**, [s. l.], 2018. Disponível em: <https://medium.com/brasil-ai/tipos-de-aprendizagem-1c1339f73bdf>. Acesso em: 15 dez. 2021.

PERELMUTER, G. **Futuro Presente**: o mundo movido à tecnologia. São Paulo: Companhia Editora Nacional, 2020.

SCUDILIOH, J. Scatter plot: Um Guia Completo para Gráficos de Dispersão. **Flai**, [s. l.], 2020. Disponível em: <https://www.flai.com.br/juscudilio/scatter-plot-um-guia-completo-para-graficos-de-dispersao/>. Acesso em: 15 dez. 2021.

TAULLI, T. **Introdução à Inteligência Artificial**: uma Abordagem Não Técnica. [S. l.]: Novatec, 2020.

ZEWE, A. These neural networks know what they're doing: A certain type of artificial intelligence agent can learn the cause-and-effect basis of a navigation task during training. **MIT News Office**, [s. l.], 2021. Disponível em: <https://news.mit.edu/2021/cause-effect-neural-networks-1014>. Aceso em: 17 dez. 2021.

CONFIRA SUAS RESPOSTAS



UNIDADE 1

1. ML é um sistema que pode alterar seu comportamento autonomamente, tendo como base a sua própria experiência. Ser um sistema que tem a capacidade de aprender é a principal diferença entre o ML e a IA.
2. D.
3. C.
4. B.
5. A.
6. Banco de dados autônomo, combate a fraudes em sistemas de pagamento eletrônico, tradução de textos, recomendação de conteúdos, entre outros.
7. D.

UNIDADE 2

Digo, sempre, para os meus alunos que construir um mapa conceitual é muito particular, pois depende da sua interpretação e/ou compreensão sobre o conteúdo estudado/lido. Não classifico um mapa conceitual como “certo ou errado”, mas, sim, como “completo ou superficial”. Para esta atividade avaliativa, sugiro, como palavras-chave iniciais para o seu mapa conceitual, que você aplique as palavras que estão marcadas em negrito no texto desta unidade. Após, procure definir as “frases de ligação” entre estes conceitos.

CONFIRA SUAS RESPOSTAS



UNIDADE 3

Construir um mapa mental é muito particular, pois depende da interpretação e/ou compreensão sobre o conteúdo estudado/lido. Não classificamos um mapa mental como "certo ou errado", mas, sim, como "completo ou superficial". Para esta atividade avaliativa, sugerimos que apresente os principais conceitos evidenciados pelas palavras-chave em negrito, procurando relacionar estes conceitos.

UNIDADE 4

UNIDADE 5

Digo, sempre, para os meus alunos que construir um mapa conceitual é muito particular, pois depende da interpretação e/ou compreensão sua sobre o conteúdo estudado/lido. Não classifico um mapa conceitual como "certo ou errado", mas, sim, como "completo ou superficial". Para essa atividade avaliativa, sugiro, como palavras chave iniciais para o seu mapa conceitual, que você aplique as palavras que estão marcadas em negrito no texto desta unidade. Após, procure definir as "frases de ligação" entre estes conceitos.

MEU ESPAÇO



MEU ESPAÇO



MEU ESPAÇO



MEU ESPAÇO

