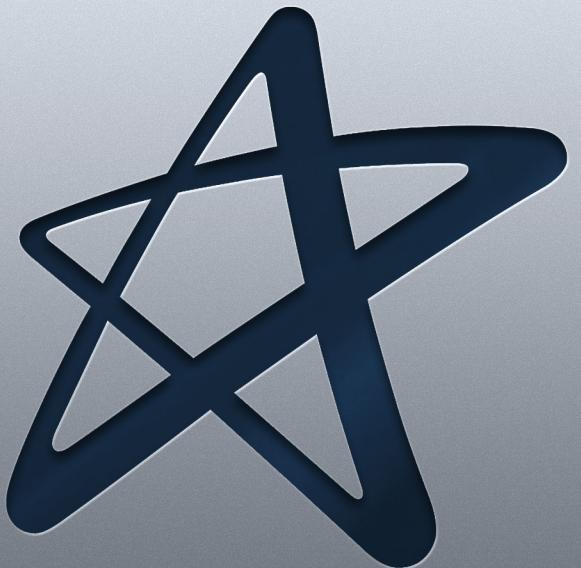


Machine Learning



Cruzeiro do Sul Virtual
Educação a distância

Material Teórico



Fundamentos de *Machine Learning*

Responsável pelo Conteúdo:

Prof. Me. Orlando da Silva Junior

Revisão Textual:

Prof.^a Dr.^a Selma Aparecida Cesarin



- Introdução à Aprendizagem de Máquina;
- Problemas e Sistemas de Aprendizado;
- Modelos de *Machine Learning*;
- Tarefas de *Machine Learning*;
- Aplicações.



OBJETIVOS DE APRENDIZADO

- Entender o funcionamento geral dos Sistemas de Aprendizagem de Máquina;
- Conhecer as tarefas e aplicações de *Machine Learning*.



Orientações de estudo

Para que o conteúdo desta Disciplina seja bem aproveitado e haja maior aplicabilidade na sua formação acadêmica e atuação profissional, siga algumas recomendações básicas:



Assim:

- ✓ Organize seus estudos de maneira que passem a fazer parte da sua rotina. Por exemplo, você poderá determinar um dia e horário fixos como seu “momento do estudo”;
- ✓ Procure se alimentar e se hidratar quando for estudar; lembre-se de que uma alimentação saudável pode proporcionar melhor aproveitamento do estudo;
- ✓ No material de cada Unidade, há leituras indicadas e, entre elas, artigos científicos, livros, vídeos e sites para aprofundar os conhecimentos adquiridos ao longo da Unidade. Além disso, você também encontrará sugestões de conteúdo extra no item **Material Complementar**, que ampliarão sua interpretação e auxiliarão no pleno entendimento dos temas abordados;
- ✓ Após o contato com o conteúdo proposto, participe dos debates mediados em fóruns de discussão, pois irão auxiliar a verificar o quanto você absorveu de conhecimento, além de propiciar o contato com seus colegas e tutores, o que se apresenta como rico espaço de troca de ideias e de aprendizagem.

Introdução à Aprendizagem de Máquina

A Computação tem sido utilizada para resolver diversos problemas do mundo. Parte desses problemas é resolvida com a escrita de um código que o computador executa, conhecido como **algoritmo**. Porém, nem sempre é fácil escrever esse algoritmo, sobretudo em problemas que o ser humano resolve muito bem.



Algoritmo: sequência de passos computacionais que toma um ou mais valores como entrada e produz um ou mais valores como saída.

Você é capaz de reconhecer rapidamente uma pessoa ao ver o rosto dela em uma fotografia. Mas como permitir que um computador faça o mesmo? Quais características ele deveria considerar? Como ele deve se comportar com as diferentes expressões faciais? E em relação aos ornamentos, como brincos e óculos, o que ele deve fazer?

Embora você não faça essas perguntas ao reconhecer alguém, o computador certamente precisará de algum método para aprender a identificar as pessoas, tanto para diferenciar um ser humano de outras espécies quanto para distinguir entre os membros da própria espécie humana.



Você saberia dizer quais passos um computador deveria realizar para poder aprender a realizar tarefas como a de reconhecer pessoas pelo rosto?

Inteligência Artificial e *Machine Learning*

A Disciplina que você está estudando neste momento se propõe a auxiliá-lo(a) na construção de computadores que podem aprender. Para isso, durante todo o seu processo de aprendizagem, nesta disciplina, você estudará técnicas de Inteligência Artificial, particularmente de Aprendizado de Máquina, em um grande número de problemas, como o exemplo citado.

Em 1950, o cientista Alan Turing determinou que a capacidade de aprendizado é uma das características essenciais para definir um comportamento inteligente por parte das máquinas.

Entre as atividades de aprendizado, um computador deve ser capaz de memorizar, observar e explorar situações para entender fatos, aperfeiçoar habilidades motoras e cognitivas, e representar o conhecimento adquirido.



Em 1950, o cientista da computação Alan Turing publicou o artigo *Computing Machinery and Intelligence*. Nesse artigo, o cientista articulou uma visão completa sobre IA, apresentando o conhecido teste de Turing e uma ideia para Aprendizagem de Máquina, Aprendizagem por Reforço e Algoritmos Genéticos. Todas essas ideias foram sendo estudadas ao longo das últimas décadas por diversos pesquisadores ao redor do mundo e se consolidaram como áreas de estudo independentes da IA, ainda que relacionadas à ela. O artigo de Turing tornou-se a base fundamental para o esforço inicial de vários cientistas em encontrar caminhos que permitam a compreensão da inteligência das máquinas e também da inteligência humana.

Até poucos anos atrás, a área de Inteligência Artificial utilizava técnicas que eram aplicadas exclusivamente em problemas de pouco valor prático, como desafios simples e jogos.

Nos últimos anos, parte dessas técnicas passaram a ser utilizadas também na Indústria, permitindo que produtos de *Data Science* e Sistemas Inteligentes fizessem parte da vida comum das pessoas.

A partir da década de 1970, os Sistemas baseados no conhecimento de especialistas de domínios particulares começaram a fazer parte da indústria por solucionarem problemas reais. Por exemplo, um Sistema que tivesse como meta auxiliar um profissional de saúde no diagnóstico médico: esse sistema era construído a partir de entrevistas realizadas com profissionais da Área para entender como eles tomavam decisões. Na maioria da vezes, extraíam-se regras dessas entrevistas para posteriormente codificá-las no computador.

Apesar de esse método ter acompanhado a Área de Inteligência Artificial durante anos, ele carregava uma série de limitações, como a pouca colaboração do profissional, decorrente do medo de ser dispensado. Esse tipo de problema fez com que a área se expandisse a novos desafios, que são enfrentados até os dias de hoje.

Mais recentemente, o grande volume de dados gerado pelo **Big Data** corporativo exigiu que os Sistemas Inteligentes se tornassem mais autônomos e menos dependentes de especialistas de domínio.

A Inteligência Artificial passou a prover técnicas que formulam regras a partir dos dados, reduzindo a intervenção humana e aumentando a capacidade de aprendizagem automática.

Dessa forma, um Sistema de apoio ao diagnóstico médico passa a formular hipóteses sobre os pacientes cadastrados no Sistema, e não mais a partir da opinião de um ou mais profissionais da saúde, exclusivamente.



Big Data: no mundo corporativo, indica os Sistemas que geram dados em três dimensões: volume, velocidade e variedade.



O professor Orlando Junior, pesquisador em Inteligência Artificial, tem um Artigo em seu *website* pessoal que explica em detalhes o que é Inteligência Artificial. Vale a pena conferir: **O guia definitivo do que é (e o que não é!) Inteligência Artificial.**

Disponível em: <https://bit.ly/2HhtF06>

Problemas de *Machine Learning*

O processo de formulação de hipóteses usando dados históricos é conhecido como Aprendizagem de Máquina (ou “Aprendizado de Máquina” ou “*Machine Learning*”).

Segundo Mitchell (1997), a Aprendizagem de Máquina é a capacidade de um Programa de Computador desempenhar melhor uma classe de tarefas a partir da experiência. O desempenho deve ser mensurado e a experiência e a classe de tarefas devem ser especificadas.

Para compreender melhor a ideia por trás de *Machine Learning*, imagine a situação a seguir.

Você é um especialista em Inteligência Artificial e está desenvolvendo um sistema de apoio ao diagnóstico médico para uma Rede Hospitalar de Goiás. O objetivo desse sistema é responder aos profissionais de saúde a possível doença do paciente, a partir de um conjunto de sintomas informados.

Para ter certeza de que o seu Sistema funciona de acordo com as expectativas da Aprendizagem de Máquina, você deverá medir o desempenho do Sistema.

Uma maneira de você fazer isso é verificando a quantidade de diagnósticos computados corretamente. Você poderá compará-los à opinião médica e mensurar a taxa de erro do seu Sistema inteligente.

Se a Rede Hospitalar tiver disponível uma base de dados com os diagnósticos já realizados pelos profissionais de Saúde da Rede, você poderá usar essa informação para facilitar a mensuração do desempenho do Sistema.

Em vez de realizar o diagnóstico duas vezes (uma do Sistema e outra do profissional), você fará isso apenas uma vez: considerando que a base de dados já informa a classificação da doença pelo profissional, você só precisará verificar se os resultados encontrados pelo seu programa ou algoritmo são iguais aos resultados dos especialistas.

Dessa forma, é importante observar que uma base de dados com informações suficientes para o problema de aprendizagem colabora para a mensuração do desempenho do algoritmo.

Além disso, a base de dados pode facilitar no processo de aprendizagem, encontrando a hipótese mais rapidamente. A cada nova informação adicionada à aprendizagem, uma nova experiência é incluída no algoritmo.



Em 1997, o professor e pesquisador Tom Mitchell publicou o artigo *Does machine learning really work?* (em português, *Machine Learning* realmente funciona?) para descrever a ascensão da Inteligência Artificial e o uso das Técnicas de Aprendizagem de Máquina na Indústria. O Artigo descreve os nichos e as áreas de aplicação de *Machine Learning* e fornece, ainda, uma visão de futuro para a Área. A partir dessa leitura, você poderá comparar as direções de futuro do professor Mitchell, em 1997, à atual realidade vivenciada por você.

Disponível em: <https://bit.ly/3cHMaNA>

Problemas e Sistemas de Aprendizado

Na seção anterior, você aprendeu a identificar um problema de Aprendizagem de Máquina. Continue a leitura nesta seção para saber como um problema de Aprendizagem deve ser formulado. Mais à frente, você também verá como funcionam os sistemas de Aprendizagem de Máquina.

Formulação de Problemas

Vamos começar abrangendo a definição de *Machine Learning* apresentada anteriormente, também de Mitchell (1997): “Um Programa de Computador aprende a partir da experiência E em relação a alguma classe de tarefas T e medida de desempenho P se o seu desempenho nas tarefas T, medido por P, melhora com a experiência E” (tradução livre do autor).

O primeiro passo para formular um problema de *Machine Learning* é determinar os elementos T, P e E do Problema de Aprendizagem.

Como exemplo, considere um Programa de Computador que aprende a jogar damas. Esse programa pode melhorar seu próprio desempenho à medida que melhorar sua habilidade em ganhar partidas de jogos de damas. Uma forma de adquirir experiência para alcançar essa habilidade é treinar jogando contra si mesmo.

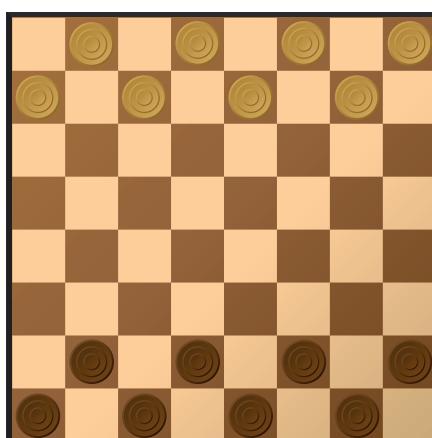


Figura 1 – Ilustração de um tabuleiro de jogo de damas
Fonte: Getty Images

Os elementos T, P e E do problema de aprender damas desse exemplo são:

- **Tarefa T:** jogar damas;
- **Desempenho P:** porcentagem de partidas vencidas contra adversários;
- **Experiência E:** jogar partidas contra si mesmo.

Como segundo exemplo, considere o problema de reconhecer textos manuscritos.

Nesse problema, imagine que você recebeu um conjunto de imagens digitais de redações escritas à mão que foram digitalizados para o computador.



Figura 2 – Ilustração de uma carta escrita à mão

Fonte: Getty Images

Nesse exemplo, os elementos T, P e E são:

- **Tarefa T:** Reconhecimento de texto;
- **Desempenho P:** porcentagem de acertos no reconhecimento de texto;
- **Experiência E:** reconhecer textos digitalizados.

A partir desses exemplos você poderá formular outros problemas de Aprendizagem de Máquina que encontrar ao longo da sua jornada profissional.

Sistemas de Aprendizado de Máquina

Após a formulação do problema, você deverá realizar o Projeto do Sistema de Aprendizagem. Esse Projeto será importante para que você desenvolva uma solução que resolva o problema de aprendizagem.

Vamos conhecer alguns elementos desse Sistema?

Russell e Norvig (2003) descrevem um modelo geral de um Agente de Aprendizagem. Em Inteligência Artificial, um **Agente** é um mecanismo que percebe e atua sobre um ambiente. Um agente que aprende é capaz de operar em ambientes inicialmente desconhecidos e se tornar mais competente do que seu conhecimento inicial sozinho permitiria.

A
Z

Em Inteligência Artificial, um **agente** é um mecanismo capaz de perceber seu ambiente por meio de sensores e de agir sobre esse ambiente por meio de atuadores. O agente pode ser tanto um robô eletromecânico quanto um *software*.

A Figura 3 ilustra o modelo genérico para um Agente de aprendizagem. Esse modelo é formado por quatro componentes conceituais:

- **Elemento de Aprendizado:** responsável pela execução do aperfeiçoamento;
- **Elemento de Desempenho:** responsável pela seleção das ações externas, recebendo percepções e decidindo sobre ações;
- **Crítico:** informa como o agente está funcionando e determina de que maneira o elemento de desempenho deve ser modificado para funcionar melhor no futuro;
- **Gerador de problemas:** responsável por sugerir ações que levarão o agente a novas experiências.

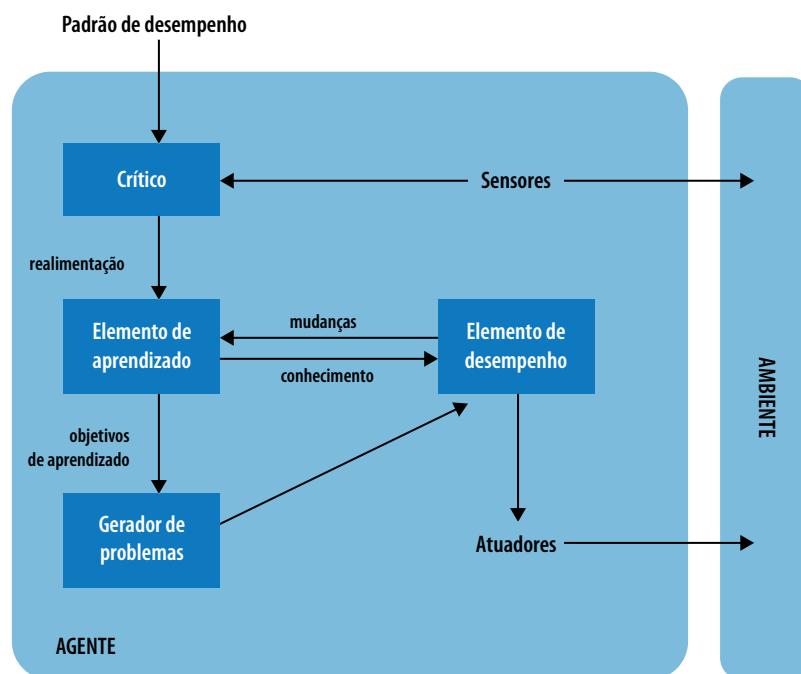


Figura 3 – Modelo genérico de agente com aprendizagem

Fonte: Adaptado de RUSSEL; NORVIG, 2003

O componente principal do agente com aprendizagem é o elemento de aprendizado.

Sem esse componente, o agente ainda é um mecanismo inteligente, mas incapaz de aprender por conta própria. Integrado aos demais elementos, o elemento de aprendizado permite que o agente seja ensinado, aprenda e aprimore o aprendizado à medida que novas experiências sejam apresentadas ao agente.

A aprendizagem acontece à medida que o agente observa suas interações com o mundo e com seus próprios processos de tomada de decisão.

Esse processo aproxima os componentes das informações de realimentação, permitindo que o desempenho global do agente seja melhorado.

Ao passo que o agente observa suas interações com o mundo e com seus próprios processos de tomada de decisão, a aprendizagem passa a acontecer.

A aprendizagem modifica os mecanismos de decisão do agente para melhorar o desempenho. A ideia por trás da aprendizagem é que as percepções devem ser usadas não apenas para agir, mas também para melhorar a habilidade do agente para agir no futuro. Ela pode variar desde a memorização trivial da experiência até a criação de teorias científicas inteiras.



Trocando Ideias...

Você acabou de aprender que uma das habilidades essenciais dos Sistemas Inteligentes é a capacidade de aprendizagem. Essa habilidade também é muito importante para os seres humanos. Você consegue imaginar outras habilidades humanas que também podem ajudar a descrever a inteligência dos computadores?

Modelos de *Machine Learning*

Agora que você já sabe como formular problemas e Sistemas de Aprendizagem, estude, a seguir, como aperfeiçoar o Sistema para os diferentes tipos de problemas de *Machine Learning* que você enfrentará durante a sua jornada profissional.

Indução de Modelos

O propósito dos métodos de *Machine Learning* é induzir modelos (ou hipóteses) que representem a distribuição dos dados observados. Queremos fazer isso para compreender como os dados se comportam e o modelo pode ser utilizado em situações futuras.

Nesse sentido, o método deve aprender, a partir de um subconjunto dos dados – conhecido como conjunto de treinamento – um modelo capaz de relacionar os valores dos atributos de entrada de um exemplo ao valor de seu atributo de saída. Existem diferentes métodos e você conhecerá alguns deles neste Curso.

Embora os métodos sejam importantes, um detalhe especial ao qual você deve ficar atento na construção dos seus modelos é a metodologia da abordagem experimental. Uma vez que o modelo é induzido para os dados de treinamento, ele deve ser válido também para os dados fora desse subconjunto. Essa característica é chamada de capacidade de generalização.

Quando o modelo apresenta baixa capacidade de generalização, os dados podem estar superajustados ao modelo.

Esse fenômeno de superajustamento é conhecido na comunidade de *Machine Learning* como *overfitting* e, na prática, significa que o modelo memorizou os dados de treinamento e não foi capaz de aprender a partir deles, para ser usado em situações futuras.

Outra situação que pode acontecer é o *underfitting*. Nesse caso, o método de aprendizagem apresenta baixa taxa de acerto nos dados de treinamento, fazendo com que o modelo fique subajustado aos dados.

A Figura 4 ilustra os três casos da indução de modelos: o **modelo com *underfitting***, o **modelo ideal** e o **modelo com *overfitting***.

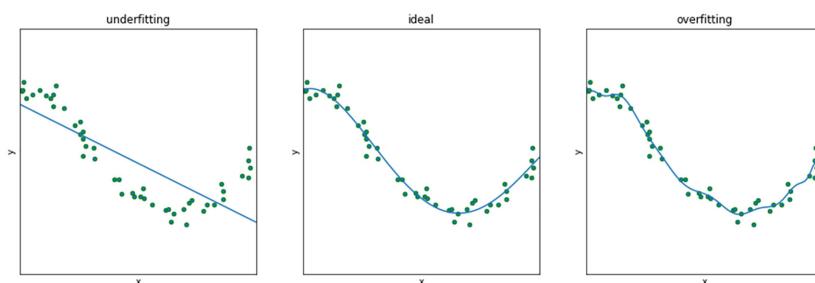


Figura 4 – *Overfitting*, *underfitting* e modelo ideal

Fonte: Acervo do conteudista

Outra questão importante a respeito dos modelos é o viés (ou *bias*, em inglês).

Quando um método de *Machine Learning* busca por modelos que representem os dados, ele selecionará os modelos de acordo com uma representação. Por exemplo, enquanto Redes Neurais representam modelos a partir de números reais, as árvores de decisão utilizam grafos cílicos direcionados. Esse viés de representação determina como os modelos serão induzidos pelo método, o que pode restringir o espaço de busca de modelos.

Além do viés de representação, os métodos também são afetados pelo viés de busca.

Nesse caso, o viés representa a forma como o algoritmo busca o modelo que melhor se ajusta ao subconjunto de treinamento. Sem ambos os vieses, os modelos seriam incapazes de aprender os dados históricos e generalizar para novos dados.

Paradigmas

Para estudarmos mais a fundo os problemas, tarefas e aplicações de *Machine Learning*, é importante conhecermos antes como esses desafios são organizados pela comunidade científica. Uma das principais formas de organizar é por meio dos Paradigmas de Aprendizagem, que permitem a classificação das tarefas em preditivas e descritivas.

Nas tarefas preditivas, o objetivo é encontrar um modelo que possa prever o rótulo ou o valor de um novo exemplo (ou registro) com base nos valores de seus atributos de entrada a partir dos dados de treinamento.

Como exemplo de tarefa preditiva, imagine que você deseja saber o diagnóstico de uma doença específica de um determinado paciente a partir de um conjunto de sintomas que outros pacientes já relataram para essa mesma doença.

Na Tabela 1, as linhas representam os pacientes já diagnosticados e as colunas indicam as características observadas para esses pacientes. A coluna Diagnóstico indica o atributo de saída, e a demais colunas indicam os atributos de entrada.

Tabela 1 – Conjunto de dados de pacientes para o diagnóstico de doenças

| Nome | Idade | Sexo | Temperatura | Dores | Diagnóstico |
|--------|-------|------|-------------|-------|-------------|
| Maria | 54 | F | 39.0 | Sim | Doente |
| João | 33 | M | 38.7 | Não | Saudável |
| José | 29 | M | 35.4 | Sim | Saudável |
| Carlos | 48 | M | 36.0 | Não | Doente |
| Ana | 21 | F | 36.5 | Sim | Doente |

Para tarefas descritivas, o objetivo é encontrar um modelo que descreva um conjunto de dados. Nesse caso, o conjunto não possui um atributo de saída.

Como exemplo, imagine que você deseja oferecer 10 tipos de ofertas especiais aos 30 mil clientes da sua Empresa.

Para descobrir qual das ofertas oferecer a cada cliente, você decide reuni-los em 10 grupos distintos. Nesse caso, como você quer descobrir a resposta – o grupo o qual o cliente irá pertencer –, não existirá um atributo de saída.

Os métodos de *Machine Learning* usados para induzir modelos preditivos seguem o paradigma de Aprendizado Supervisionado porque eles conhecem a saída esperada para cada exemplo. Os modelos descritivos seguem o paradigma de aprendizado não supervisionado.



O professor Pedro Domingo publicou, em 2012, um Artigo resumindo 12 lições importantes para quem deseja saber mais sobre *Machine Learning*.
DOMINGOS, P. *A few useful things to know about machine learning*. *Communications of the ACM*, EUA, v. 55, n. 10, p. 78-87, 2012.

Tarefas de *Machine Learning*

Existem várias tarefas de aprendizado que podem ser estudadas e aplicadas com métodos de *Machine Learning*.

A Figura 5 ilustra a relação entre elas e os Paradigmas de Aprendizagem.

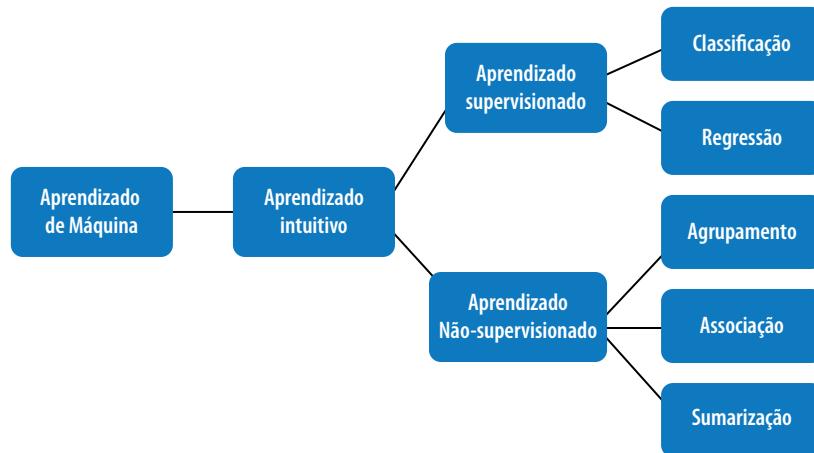


Figura 5

Classificação e regressão são as principais tarefas de aprendizado supervisionado, que se distinguem pelo valor do atributo de saída. Se o rótulo é discreto, estamos falando de classificação; mas se o rótulo for numérico contínuo, estaremos falando de regressão.

Entre as tarefas não supervisionadas, destacam-se:

- **Agrupamento:** tem como meta agrupar os dados de acordo com sua similaridade;
- **Associação:** visa a encontrar padrões frequentes de associações entre os atributos;
- **Sumarização:** tem como meta encontrar uma descrição resumida dos dados.

Uma outra abordagem de tarefas é apresentada em Provost; Fawcett (2016, p. 20-3), que detalham nove tarefas para se trabalhar em Ciência de Dados.

A Tabela 2 sintetiza essas tarefas e descreve o propósito de suas aplicações.

Tabela 2 – Síntese das principais tarefas analíticas

| Tarefa | Propósito |
|--|--|
| Classificação | Predizer (ou prever), para cada indivíduo de uma população, a que conjunto de classes este indivíduo pertence. |
| Regressão | Estimar ou predizer um valor numérico para cada indivíduo da população. |
| Combinação por similaridade | Identificar indivíduos semelhantes com base nos dados conhecidos sobre eles. |
| Agrupamento | Reunir indivíduos de uma população por meio de similaridade, sem, no entanto, estar motivado por algum propósito. |
| Mineração de itens frequentes | Buscar por associações entre entidades com base em transações que as envolvem. |
| Perfilamento (<i>data profiling</i>) | Caracterizar o comportamento típico de um indivíduo, grupo ou população. |
| Previsão de vínculo | Tenta prever ligações entre itens de dados, geralmente sugerindo que um vínculo deveria existir e, possivelmente, também estimando a força do vínculo. |
| Redução de dados | Procura substituir um grande conjunto de dados por um outro conjunto com menos informações, mantendo a mesma qualidade de informações relevantes. |
| Modelagem causal | Auxiliar na análise ajudando a compreender que acontecimentos ou ações realmente influenciam outras pessoas. |

Aplicações

A Área de *Machine Learning* é uma das que mais tem crescido na Pesquisa em Computação nos últimos anos. A cada dia, muitos cientistas têm proposto novos algoritmos, técnicas e ferramentas para resolver os mais variados tipos de problemas.

A Indústria não tem ficado para trás e muitas áreas e segmentos de negócio têm se beneficiado dessas aplicações, especialmente:

- Agricultura e Pecuária;
- Ciência e Bioinformática;
- Biologia, Ecologia e Meio Ambiente;
- Energia;
- Finanças;
- Saúde;
- Telecomunicações.

Entre as várias aplicações de técnicas de *Machine Learning* na solução de problemas reais, podemos destacar:

- Reconhecimento de fala;
- Estimativa do preço de produtos de Mercado;
- Construção de portfólio de investimentos;
- Condução de veículos autônomos;
- Detecção de fraudes em cartões de crédito;
- Aspiradores robóticos que limpam residências;
- Jogos de tabuleiro, como damas e xadrez;
- Diagnóstico de câncer por meio da análise de expressão gênica.



A *Boston Dynamics* é uma Empresa focada na construção de Sistemas Robóticos. Em seu canal no *Youtube*, ela disponibiliza uma série de vídeos com seus robôs de alta mobilidade. Vale a pena conferir. Acesse: <https://bit.ly/3ife1ap>

Para que você conheça o uso prático de *Machine Learning*, leia, a seguir, alguns casos de aplicações em diferentes áreas.

Avaliação da qualidade da carne de carneiro

Em Cortez *et al.* (2006), é avaliada a qualidade da carne de carneiro usando técnicas de *Machine Learning*.

O estudo apresenta uma comparação entre os métodos SVM, redes neurais e regressão múltipla para fazer essa avaliação.

Os autores utilizam um conjunto de dados com 81 animais e 12 atributos de entrada. O sabor da carne (definido por 12 especialistas) e a força de cisalhamento correspondem aos atributos de saída. A solução final corresponde ao uso simplificado e econômico de um método não invasivo com 6 atributos de entrada que pode auxiliar a tomada de decisão em até 24 horas.

Detecção de falhas na rede elétrica

Oleskovicz *et al.* (1998) apresentam uma aplicação para detectar e localizar falhas em linhas de transmissão de energia. O objetivo principal do trabalho é implementar um modelo de proteção alternativo às linhas de transmissão usando Redes Neurais artificiais.

A abordagem utiliza os valores pós-falta das tensões e correntes trifásicas em cinco janelas de dados.

Diagnóstico de pacientes diabéticos

O trabalho de Patil *et al.* (2010) descreve uma abordagem híbrida de *Machine Learning* para diagnosticar pacientes com Diabetes tipo 2.

Sendo uma das patologias mais comuns nos dias de hoje, o correto e antecipado diagnóstico pode auxiliar no tratamento e conduzir o paciente a uma melhor qualidade de vida.

Na abordagem, os autores propõem a solução híbrida combinando os métodos de agrupamento k-médias e indução de árvores de decisão C4.5.

O primeiro algoritmo valida as classes associadas aos exemplos e o segundo algoritmo é aplicado aos novos casos.

Monitoramento da qualidade da água

A Ecologia é um terreno fértil para muitas aplicações de técnicas de *Machine Learning*. Além de permitir que o Ecossistema se beneficie dessas aplicações, os seres humanos e a pesquisa científica da área também podem aproveitar os benefícios que elas podem trazer.

O trabalho de Policastro *et al.* (2004) relata a construção de um Sistema em que Técnicas de Aprendizagem de Máquina são usadas para monitorar a qualidade da água em rios.

A proposta considera três métodos de aprendizagem (SVMs, redes neurais e árvores de decisão), que são utilizados em Comitês para adaptar os casos recuperados de uma base de dados. Os atributos de entrada do Sistema são substâncias químicas e a população de algas dos rios.

Sistemas como esse ajudam a identificar o impacto da intervenção humana nos locais investigados e a diminuir a incidência de casos de processos que sejam tóxicos ao ambiente.

Predição de estruturas de proteínas

Entre os Projetos mais comuns em Bioinformática, a análise de proteínas por meio de técnicas de *Machine Learning* tem se mostrado promissora e capaz de colaborar na solução de vários problemas da Área, como a identificação de genes.

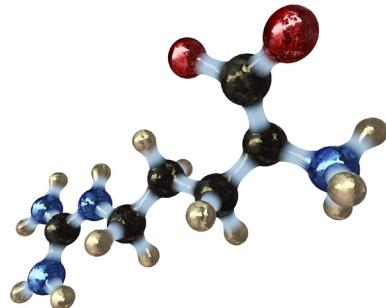


Figura 6 – Proteína de aminoácido de arginina

Fonte: Getty Images

Em Cheng *et al.* (2008), diferentes métodos supervisionados e não supervisionados de *Machine Learning* são estudados na predição de estruturas proteicas.

O trabalho resume 120 outros trabalhos correlatos que enfocaram *Machine Learning* nesse tipo de aplicação.

Material Complementar

Indicações para saber mais sobre os assuntos abordados nesta Unidade:



Livros

O Algoritmo Mestre

DOMINGOS, P. **O Algoritmo Mestre**. São Paulo: Novatec, 2017.

Inteligência Artificial: Uma abordagem de aprendizado de máquina

FACELI, K. *et al.* **Inteligência Artificial: Uma abordagem de aprendizado de máquina**. Rio de Janeiro: LTC, 2011.



Leituras

Conceitos sobre aprendizado de máquina

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. Sistemas inteligentes-Fundamentos e aplicações, São Paulo, v. 1, n. 1, p. 32, 2003.

<https://bit.ly/2EKEYUB>

CS229 Lecture notes

NG, A. *CS229 Lecture notes*, v. 1, n. 1, p. 1-3, 2000.

<https://bit.ly/34gLYRY>

Referências

- CHENG, J.; TEGGE, A. N.; BALDI, P. *Machine learning methods for protein structure prediction*. *IEEE reviews in biomedical engineering*, EUA, v. 1, p. 41-49, 2008.
- CORTEZ, P. *et al.* *Lamb meat quality assessment by support vector machines*. *Neural Processing Letters*, Portugal, v. 24, n. 1, p. 41-51, 2006.
- MITCHELL, T. M. *Machine learning*. New York: McGraw-Hill, 1997.
- OLESKOVICZ, M.; COURY, D. V.; DE CARVALHO, CPLF. *Artificial neural network applied to power system protection*. In: *Proceedings 5th Brazilian Symposium on Neural Networks* (Cat. No. 98EX209). São Paulo, 1998. p. 247-252.
- PATIL, B. M.; JOSHI, R. C.; TOSHNIWAL, D. *Hybrid prediction model for type-2 diabetic patients*. *Expert systems with applications*, India, v. 37, n. 12, p. 8102-8108, 2010.
- POLICASTRO, C. A.; CARVALHO, A. CPLF.; DELBEM, Alexandre CB. *A hybrid case based reasoning approach for monitoring water quality*. In: *International Conference on Industrial, Engineering and other Applications of Applied Intelligent Systems*. Springer, Berlin, Heidelberg, 2004. p. 492-501.
- PROVOST, F.; FAWCETT, T. *Data Science para negócios*. Tradução de Marina Bosatto. Rio de Janeiro: Alta Books, 2016.
- RUSSEL, S. J.; NORVIG, P. **Inteligência Artificial**: uma abordagem moderna. 2.ed. Rio de Janeiro: Editora Elsevier, 2004.



Cruzeiro do Sul
Educacional