Capítulo 1

Introdução

Problemas são resolvidos em computação por meio da escrita de um algoritmo ou pseudocódigo, que especifica passo a passo como o problema pode ser resolvido. No entanto, não é fácil escrever um programa de computador que realize com eficiência algumas tarefas que realizamos com facilidade no nosso dia a dia, como reconhecer pessoas pelo rosto ou pela fala. Que características dos rostos ou da fala serão consideradas? O que fazer para diferentes expressões faciais de uma mesma pessoa, alterações na face, como o uso de óculos ou bigode, cortes de cabelo, mudanças na voz por uma gripe ou estado de espírito? No entanto, os seres humanos conseguem realizar essas tarefas com relativa facilidade. Fazem isso por meio de reconhecimento de padrões, quando aprendem o que deve ser observado em um rosto ou na fala para conseguir identificar pessoas após terem tido vários exemplos de rostos ou falas com identificação clara.

Um bom médico consegue, dado o conjunto de sintomas e de resultados de exames clínicos, diagnosticar se um paciente está com problemas cardíacos. Para isso o médico utiliza conhecimento adquirido durante sua formação e experiência proveniente do exercício da profissão. Como escrever um programa de computador que, dados os sintomas e os resultados de exames clínicos, apresente um diagnóstico que seja tão bom quanto o de um médico experiente?

Também pode ser difícil escrever um programa que faça algumas análises de dados de venda de uma loja. Para descobrir quantas pessoas fizeram mais de uma compra em uma loja no ano anterior, programas para gerenciamento de bancos de dados podem ser facilmente utilizados. Entretanto, como escrever um programa que responda a questões mais complicadas como:

- Identificar conjuntos de produtos que são frequentemente vendidos em conjunto.
- Recomendar novos produtos a clientes que costumam comprar produtos semelhantes.
- Agrupar os consumidores da loja em grupos de forma a ter melhores resultados nas operações de marketing.

Apesar da dificuldade de escrever um programa de computador que possa lidar de forma eficiente com essas tarefas, o número de vezes em que tarefas tão complexas como essas precisam ser realizadas diariamente é muito grande. Isso, aliado ao volume de informações que precisam ser consideradas para sua realização, torna difícil ou mesmo impossível a sua realização por seres humanos.

Técnicas de Inteligência Artificial (IA), em particular de Aprendizado de Máquina (AM), têm sido utilizadas com sucesso em um grande número de problemas reais, incluindo os problemas citados anteriormente.

Este capítulo está organizado da seguinte forma. A Seção 1.1 apresenta a relação entre AM e IA, mostrando alguns exemplos da utilização de AM em problemas reais. Na Seção 1.2 é introduzida a relação entre um conjunto de dados e a qualidade da hipótese induzida por um algoritmo de AM. O conceito de viés indutivo, essencial para que o aprendizado ocorra, é discutido na Seção 1.3. A Seção 1.4 descreve as diferentes tarefas de aprendizado, que são agrupadas em aprendizado preditivo e aprendizado descritivo. Finalmente, a Seção 1.5 apresenta a estrutura dos capítulos do livro.

1.1 Inteligência Artificial e Aprendizado de Máquina

Até alguns anos atrás, a área de IA era vista como uma uma área teórica, com aplicações apenas em pequenos problemas curiosos, desafiadores, mas de pouco valor prático. Boa parte dos problemas práticos que precisam de computação era resolvida pela codificação em alguma linguagem de programação dos passos necessários para a sua solução.

A partir da década de 1970, houve uma maior disseminação do uso de técnicas de computação baseadas em IA para a solução de problemas reais. Muitas vezes, esses problemas eram tratados computacionalmente por meio da aquisição de conhecimento de especialistas de um dado domínio, por exemplo, do domínio da medicina, que era então codificado, frequentemente por regras lógicas, em um programa de computador. Esses programas eram conhecidos como Sistemas Especialistas ou Sistemas Baseados em Conhecimento.

O processo de aquisição do conhecimento normalmente envolvia entrevistas com os especialistas para descobrir que regras eles utilizavam quando da tomada de decisão. Esse processo possuía várias limitações, como subjetividade, decorrente do uso pelo especialista de sua intuição na tomada de decisão, e, muitas vezes, pouca cooperação por parte do especialista, por causa do seu receio de ser dispensado após repassar o conhecimento solicitado.

Nas últimas décadas, com a crescente complexidade dos problemas a serem tratados computacionalmente e do volume de dados gerados por diferentes setores, tornou-se clara a necessidade de ferramentas computacionais mais sofisticadas, que fossem mais autônomas, reduzindo a necessidade de intervenção humana e dependência de especialistas. Para isso, essas técnicas deveriam ser capazes de criar por si próprias, a partir da experiência passada, uma hipótese, ou função, capaz de resolver o problema que se deseja tratar. Um exemplo simples é a descoberta de uma hipótese na forma de uma regra ou conjunto de regras para definir que clientes de um supermercado devem receber material de propaganda de um novo produto, utilizando para isso dados de compras passados dos clientes cadastrados na base de dados do supermercado. A esse processo de indução de uma hipótese (ou aproximação de função) a partir da experiência passada dá-se o nome Aprendizado de Máquina (AM).

A capacidade de aprendizado é considerada essencial para um comportamento inteligente. Atividades como memorizar, observar e explorar situações para aprender fatos, melhorar habilidades motoras/cognitivas por meio de práticas e organizar conhecimento

novo em representações apropriadas podem ser consideradas atividades relacionadas ao aprendizado.

Existem várias definições de AM na literatura. Uma delas, apresentada em Mitchell (1997), define AM como:

> *"A capacidade de melhorar o desempenho na realização de alguma tarefa por meio da experiência."*

Em AM, computadores são programados para aprender com a experiência passada. Para tal, empregam um princípio de inferência denominado indução, no qual se obtêm conclusões genéricas a partir de um conjunto particular de exemplos. Assim, algoritmos de AM aprendem a induzir uma função ou hipótese capaz de resolver um problema a partir de dados que representam instâncias do problema a ser resolvido. Esses dados formam um conjunto, simplesmente denominado conjunto de dados (Seção 1.2).

Embora AM seja naturalmente associado à IA, outras áreas de pesquisa são importantes e têm contribuições diretas e significativas no avanço do AM, como Probabilidade e Estatística, Teoria da Computação, Neurociência, Teoria da Informação, para citar algumas. AM é uma das áreas de pesquisa da computação que mais tem crescido nos últimos anos. Diferentes algoritmos de AM, diferentes formas de utilizar os algoritmos existentes e adaptações de algoritmos são continuamente propostos. Além disso, surgem a todo instante novas variações nas características dos problemas reais a serem tratados.

Existem várias aplicações bem-sucedidas de técnicas de AM na solução de problemas reais, entre as quais podem ser citadas:

- Reconhecimento de palavras faladas;
- Predição de taxas de cura de pacientes com diferentes doenças;
- Detecção do uso fraudulento de cartões de crédito;
- Condução de automóveis de forma autônoma em rodovias;
- Ferramentas que jogam gamão e xadrez de forma semelhante a campeões;
- Diagnóstico de câncer por meio da análise de dados de expressão gênica.

Além do grande volume de aplicações que se beneficiam das características da área de AM, outros fatores têm favorecido a expansão dessa área, como o desenvolvimento de algoritmos cada vez mais eficazes e eficientes e a elevada capacidade dos recursos computacionais atualmente disponíveis.

Outras motivações para as pesquisas em AM incluem a possibilidade de aumentar a compreensão de como se dá o aprendizado nos seres vivos. Além disso, algumas tarefas são naturalmente mais bem definidas por meio de exemplos. Os modelos gerados são ainda capazes de lidar com situações não apresentadas durante seu desenvolvimento, sem necessariamente necessitar de uma nova fase de projeto.

1.2 Indução de Hipóteses

Para ilustrar a relação entre AM e indução de hipóteses, imagine um conjunto de dados composto de pacientes de um hospital. Nesse conjunto, que será denominado `hospital`, cada dado (também chamado objeto, exemplo, padrão ou registro) corresponde a um paciente e é uma tupla formada pelos valores de características ou atributos referentes ao paciente, que descrevem seus principais aspectos. Os atributos (também chamados campos ou variáveis) utilizados para cada paciente podem ser, por exemplo, sua identificação, nome, idade, sexo, estado de origem, sintomas e resultados de exames clínicos. Exemplos de sintomas podem ser presença e distribuição de manchas na pele, peso e temperatura do corpo.

Conforme será visto adiante, para algumas tarefas de aprendizado, um dos atributos é considerado um atributo de saída (também chamado atributo alvo ou atributo meta), cujos valores podem ser estimados utilizando os valores dos demais atributos, denominados atributos de entrada (também chamados atributos previsores). O objetivo de um algoritmo de AM utilizado nessas tarefas é aprender, a partir de um subconjunto dos dados, denominado conjunto de treinamento, um modelo ou hipótese capaz de relacionar os valores dos atributos de entrada de um objeto do conjunto de treinamento ao valor de seu atributo de saída.

Um requisito importante para algoritmos de AM é que eles sejam capazes de lidar com dados imperfeitos. Muitos conjuntos de dados apresentam algum tipo de problema, como presença de ruídos, dados inconsistentes, dados ausentes e dados redundantes. Algoritmos de AM devem, idealmente, ser robustos aos problemas presentes nos dados, minimizando sua influência no processo de indução de hipóteses. Entretanto, dependendo de sua extensão, esses problemas podem prejudicar o processo indutivo. Técnicas de pré-processamento são utilizadas com frequência para identificar e minimizar a ocorrência desses problemas.

Voltando ao exemplo dos pacientes, considere a situação em que um algoritmo de AM é utilizado para aprender uma hipótese (por exemplo, uma regra) capaz de diagnosticar pacientes de acordo com os valores associados aos seus atributos de entrada, representados por parte dos demais atributos. Os atributos referentes à identificação e nome do paciente não são considerados entradas relevantes, uma vez que não possuem relação nenhuma com o diagnóstico de uma doença.

O que se deseja, na verdade, é induzir uma hipótese capaz de fazer diagnósticos corretos para novos pacientes diferentes daqueles que foram utilizados para aprender a regra de decisão. Assim, uma vez induzida uma hipótese, é desejável que ela também seja válida para outros objetos do mesmo domínio ou problema que não fazem parte do conjunto de treinamento. A essa propriedade de uma hipótese continuar a ser válida para novos objetos dá-se o nome capacidade de generalização da hipótese. Para ser útil quando aplicada a novos dados, uma hipótese precisa apresentar uma boa capacidade de generalização.

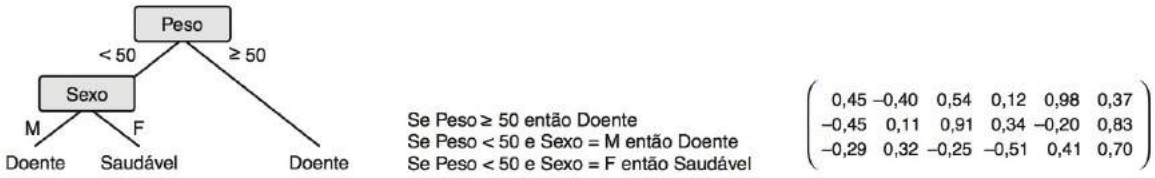
Quando uma hipótese apresenta uma baixa capacidade de generalização, a razão pode ser que ela está superajustada aos dados (*overfitting*). Nesse caso, também é dito que a hipótese memorizou ou se especializou nos dados de treinamento. No caso inverso, o algoritmo de AM pode induzir hipóteses que apresentem uma baixa taxa de acerto mesmo no subconjunto de treinamento, configurando uma condição de subajustamento (*underfitting*). Essa situação pode ocorrer, por exemplo, quando os exemplos de treinamento disponíveis são pouco representativos ou o modelo usado é muito simples e não captura os

padrões existentes nos dados (Monard e Baranauskas, 2003). Na Seção 7.2.1, esses conceitos são ilustrados e discutidos novamente. São feitas então considerações e motivações sobre a escolha de modelos com boa capacidade de generalização.

1.3 Viés Indutivo

Quando um algoritmo de AM está aprendendo a partir de um conjunto de dados de treinamento, ele está procurando uma hipótese, no espaço de possíveis hipóteses, capaz de descrever as relações entre os objetos e que melhor se ajuste aos dados de treinamento.

Cada algoritmo utiliza uma forma ou representação para descrever a hipótese induzida. Por exemplo, redes neurais artificiais representam uma hipótese por um conjunto de valores reais, associados aos pesos das conexões da rede. Árvores de decisão utilizam uma estrutura de árvore em que cada nó interno é representado por uma pergunta referente ao valor de um atributo e cada nó externo está associado a uma classe. A representação utilizada define a preferência ou viés (*bias*) de representação do algoritmo e pode restringir o conjunto de hipóteses que podem ser induzidas pelo algoritmo. A Figura 1.1 ilustra o viés de representação utilizado por técnicas de indução de árvores de decisão, redes neurais artificiais e regras de decisão.



(a) Árvore de decisão (b) Conjunto de regras (c) Redes neurais

Figura 1.1 *Diferentes vieses de representação.*

Além do viés de representação, os algoritmos de AM possuem também um viés de busca. O viés de busca de um algoritmo é a forma como o algoritmo busca a hipótese que melhor se ajusta aos dados de treinamento. Ele define como as hipóteses são pesquisadas no espaço de hipóteses. Por exemplo, o algoritmo ID3, que é utilizado para indução de árvores de decisão, tem como viés de busca a preferência por árvores de decisão com poucos nós, conforme será apresentado no Capítulo 6.

Assim, cada algoritmo de AM possui dois vieses, um viés de representação e um viés de busca. O viés é necessário para restringir as hipóteses a serem visitadas no espaço de busca. Sem viés não haveria aprendizado/generalização. Os modelos seriam especializados para os exemplos individuais. Embora, à primeira vista, o viés pareça ser uma limitação dos algoritmos de AM, segundo Mitchell (1997), sem viés um algoritmo de AM não consegue generalizar o conhecimento adquirido durante seu treinamento para aplicá-lo com sucesso a novos dados.

1.4 Tarefas de Aprendizado

Algoritmos de AM têm sido amplamente utilizados em diversas tarefas, que podem ser organizadas de acordo com diferentes critérios. Um deles diz respeito ao paradigma

de aprendizado a ser adotado para lidar com a tarefa. De acordo com esse critério, as tarefas de aprendizado podem ser divididas em:

- Preditivas e
- Descritivas.

Em tarefas de previsão, a meta é encontrar uma função (também chamada de modelo ou hipótese) a partir dos dados de treinamento que possa ser utilizada para prever um rótulo ou valor que caracterize um novo exemplo, com base nos valores de seus atributos de entrada. Para isso, cada objeto do conjunto de treinamento deve possuir atributos de entrada e de saída.

Os algoritmos ou métodos de AM utilizados nessa tarefa induzem modelos preditivos. Esses algoritmos seguem o paradigma de aprendizado supervisionado. O termo supervisionado vem da simulação da presença de um "supervisor externo", que conhece a saída (rótulo) desejada para cada exemplo (conjunto de valores para os atributos de entrada). Com isso, o supervisor externo pode avaliar a capacidade da hipótese induzida de predizer o valor de saída para novos exemplos.

Em tarefas de descrição, a meta é explorar ou descrever um conjunto de dados. Os algoritmos de AM utilizados nessas tarefas não fazem uso do atributo de saída. Por isso, seguem o paradigma de aprendizado não supervisionado. Uma tarefa descritiva de agrupamento de dados, por exemplo, tem por meta encontrar grupos de objetos semelhantes no conjunto de dados. Outra tarefa descritiva é encontrar regras de associação que relacionam um grupo de atributos a outro grupo de atributos.

A Figura 1.2 apresenta uma hierarquia de aprendizado de acordo com os tipos de tarefas de aprendizado. No topo aparece o aprendizado indutivo, processo pelo qual são realizadas as generalizações a partir dos dados. Tem-se em seguida os tipos de aprendizado supervisionado (preditivo) e não supervisionado (descritivo). As tarefas supervisionadas se distinguem pelo tipo dos rótulos dos dados: discreto, no caso de classificação; e contínuo, no caso de regressão. As tarefas descritivas são genericamente divididas em: agrupamento, em que os dados são agrupados de acordo com sua similaridade; sumarização, cujo objetivo é encontrar uma descrição simples e compacta para um conjunto de dados; e associação, que consiste em encontrar padrões frequentes de associações entre os atributos de um conjunto de dados. Com exceção da sumarização, as demais tarefas serão descritas neste livro.

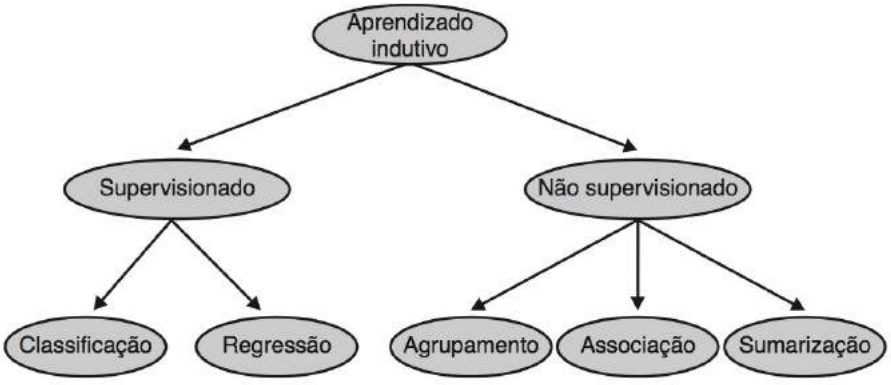


Figura 1.2 *Hierarquia de aprendizado.*

Deve ser observado que, apesar dessa divisão básica de modelos em preditivos e descritivos, um modelo preditivo também provê uma descrição compacta de um conjunto de dados e um modelo descritivo pode prover previsões após ser validado.

Uma tarefa de aprendizado que não se enquadra nas tarefas anteriores é a de aprendizado por reforço. Nessa tarefa, a meta é reforçar ou recompensar uma ação considerada positiva e punir uma ação considerada negativa. Um exemplo de tarefa de reforço é a de ensinar um robô a encontrar a melhor trajetória entre dois pontos. Algoritmos de aprendizado utilizados nessa tarefa, em geral, punem a passagem por trechos pouco promissores e recompensam a passagem por trechos promissores. Por causa do foco adotado para este livro, essa tarefa não será abordada.

1.5 Estrutura do Livro

Este livro tem por objetivo apresentar os principais conceitos e algoritmos de AM e mostrar como AM pode ser utilizado para a solução de problemas reais. Para isso, serão cobertos tanto temas tradicionais como resultados de pesquisas recentes na área.

De forma a agrupar os temas cobertos de uma maneira mais uniforme, os capítulos do livro foram organizados em cinco grandes temas ou módulos:

- **Preparação de dados:** engloba tópicos de descrição dos dados, análise estatística de dados e pré-processamento de dados.
- **Métodos preditivos:** é relacionado a aprendizado supervisionado e que, após definir os conceitos gerais referentes a esse tema, descreve os principais algoritmos de aprendizado preditivo, como hipóteses podem ser combinadas formando comitês, possíveis estratégias para planejar experimentos com esses métodos e as principais métricas empregadas para avaliar seu desempenho.
- **Métodos descritivos:** é o equivalente do módulo anterior para aprendizado não supervisionado. Nesse módulo são abordados os principais conceitos básicos necessários para descrever essa abordagem de AM, os principais algoritmos utilizados, formas como eles podem ser avaliados e combinados. É também discutido como experimentos utilizando esses métodos podem ser planejados e avaliados.
- **Tópicos avançados:** incluem temas de pesquisas recentes na área de AM. Esses temas, que incluem fluxos de dados, meta-aprendizado, estratégias para classificação multiclasse, classificação hierárquica e classificação multirrótulo, são aplicados com sucesso em um grande número de problemas reais.
- **Aplicações:** ilustram alguns exemplos de aplicações reais relacionadas a diferentes domínios em que técnicas de AM têm sido empregadas com sucesso.

Esses tópicos foram cuidadosamente escolhidos para que os leitores tenham uma dosagem equilibrada em abrangência e profundidade dos temas básicos e avançados nas áreas de Inteligência Artificial que utilizam aprendizado para indução de modelos.