Aprendizagem de Máquina

Conteudista: Prof. Me. Manuel Fernández Paradela Ledón
Revisão Técnica: Prof. Me. Douglas Almendro
Revisão Textual: Prof. Me. Luciano Vieira Francisco

Objetivos da Unidade:

- Conhecer os conceitos relacionados à aprendizagem de máquina;
- Identificar os diferentes tipos de aprendizagem.

📄 Material Teórico

Aprendizagem de Máquina

Algumas das grandes empresas multinacionais de informática, tais como *Google*, IBM e *Microsoft*, investem fortemente em pesquisas e desenvolvimento de sistemas relacionados à inteligência artificial e, especificamente, em aprendizagem de máquina, de modo que devemos imaginar que se trata de algo importante – no Material Complementar desta Unidade são recomendadas algumas leituras interessantes.

Assim, nesta oportunidade veremos este assunto amplamente abordado dentro da inteligência artificial: a aprendizagem ou o aprendizado de máquina. Conheceremos os conceitos relacionados à aprendizagem de máquina e identificaremos os diferentes tipos de aprendizagem.

O que Significa Aprendizagem de Máquina?

Quando utilizamos a expressão *machine learning* – em inglês –, que significa aprendizagem ou aprendizado de máquina, utilizamos o ponto de vista da inteligência artificial, ou seja, não consideramos, por exemplo, o enfoque didático-pedagógico da aprendizagem. Assim, aprendizado e aprendizagem podem ser considerados sinônimos, dado que aprendizado diz respeito ao: "Ato, ou processo ou efeito de aprender; aprendizagem" (AURÉLIO, 2009).

Segundo Russell e Norvig (2004), sob o enfoque de agentes inteligentes:

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

Uma conclusão importante, segundo Luger (2004), é que: “*The ability to learn must be part of any system that would claim to possess general intelligence*”. Como outros estudiosos, Luger (2004) afirma que a habilidade de aprender deve ser parte de qualquer sistema que exija possuir inteligência geral.

Podemos considerar que aprendizagem de máquina é a capacidade de melhorar o desempenho na realização de alguma tarefa por meio da experiência, ou seja, do que se aprende – lembrando-nos que aprender significa “[...] tornar-se apto ou capaz de alguma coisa, em consequência de estudo, observação, experiência, advertência etc.” (AURÉLIO, 2009).

Tipos de Aprendizagem de Máquina

Diferentes autores consideram três tipos de aprendizagem de máquina, a saber:

- Supervisionada;
- Não supervisionada;
- Por reforço ou recompensa.

A **aprendizagem supervisionada** pressupõe a presença de um supervisor, que consideraremos como algum critério ou medida de adequação, ou qualquer outro tipo de processo que oriente o treinamento ou aprendizado.

Na aprendizagem supervisionada teremos, frequentemente, a aprendizagem de uma função a partir de exemplos de entradas e saídas, isto é, será produzida uma nova função como resultado da aprendizagem.

Tipicamente, a aprendizagem supervisionada é apresentada como indutiva, com algoritmos que recebem dados de treinamento pré-classificados ou conjuntos de treinamento – exemplos: árvores de decisão, redes neurais supervisionadas ajustando seus pesos etc.

A **aprendizagem não supervisionada** – em inglês, *unsupervised learning* – elimina a figura do supervisor e requer que os aprendizes formem e avaliem seus conceitos por si. Dito de outra forma, métodos de aprendizado não supervisionado aprendem sem qualquer intervenção humana.

Nenhum tipo de etiqueta é dado ao algoritmo de aprendizado não supervisionado, deixando-o sozinho para encontrar uma estrutura nas entradas fornecidas. O aprendizado não supervisionado pode ser um objetivo em si – por exemplo, analisar e descobrir novos padrões dentro dos dados, ou um meio para atingir um fim.

Em uma situação de aprendizagem puramente não supervisionada não existem informações adicionais sobre qual seria um estado desejável e quais seriam as ações corretas – dado que o processo transcorre sem receber dados de treinamento.

Dois exemplos simplificados do dia a dia de aprendizagem não supervisionada poderiam ser:

- Um motorista de táxi aprende e desenvolve conceitos de **melhor percurso**, **dias de tráfego ruim**, **em dia de chuva** etc.;
- Os cientistas criam teorias e aprendem com pouca supervisão.

A **aprendizagem por reforço** – em inglês, *reinforcement learning* – acontece com ações ou respostas do ambiente. Significa aprender a partir do reforço ou recompensa.

Quem utiliza aprendizado com reforço recebe um positivo ao operar corretamente e um negativo ao operar incorretamente. Em robótica, por exemplo, quando o robô pega um objeto corretamente, recebe uma recompensa ou reforço positivo – a robótica utiliza aprendizagem por reforço e supervisionada. Exemplos de aprendizagem por reforço seriam:

- A utilização em jogos, controle de robôs;
- As crianças aprendem com os elogios pelos seus progressos ao caminhar ou falar adequadamente;
- Com a gorjeta pelo atendimento adequado, o garçom aprende quando o seu comportamento e educação são adequados;
- Com as notas, o aluno aprende sobre os métodos de estudo que lhe são mais efetivos.

Aprendizagem Indutiva como um Caso de Aprendizagem Supervisionada

Mencionamos que na aprendizagem supervisionada será necessário algum critério ou medida de adequação para aprender. Ademais, na aprendizagem indutiva encontraremos essa medida de adequação com base em situações/resultados/exemplos.

A palavra **indutiva** tem relação com o verbo **induzir**, este que significa causar, inspirar e/ou produzir, de modo que na inferência indutiva pura:

- Temos um grupo de atributos **x** que produzem uma saída **f(x)**;
- Esta função **f** poderia ser de uma ou várias variáveis – ou atributos, parâmetros etc. –, como veremos adiante, no exemplo de árvores de decisão;

- Os atributos (x) e o resultado (f(x)) poderiam ser valores reais ou de outros tipos de dados; geralmente, o resultado será booleano – verdadeiro ou falso;

- Cada par (x, f(x)) será chamado de **exemplo**;

- Dada uma coleção de **exemplos** de f, **retornaremos outra função induzida h(x)** – chamada de **hipótese** –, que se aproxime de **f**, parecida com **f**;

- Uma boa **hipótese h** deverá generalizar bem, prever correta e eficientemente **exemplos** – situações –, inclusive ainda não vistos.

Aprendizagem Utilizando Árvores de Decisão

Uma "árvore de decisão" toma como entrada um conjunto de atributos e retorna como saída uma decisão, comumente de valor lógico – verdadeiro ou falso, sim ou não –, dependendo da entrada.

A árvore de decisão do exemplo presente na Figura 1 poderia ser utilizada para prever se uma música de uma banda de *rock* será um *top hit* internacional. Obviamente, é um exemplo simplista, quase que uma brincadeira, mas que nos permite mostrar como funciona este tipo de árvore.

Assim, observe nas Figuras 1 e 2 que as conclusões significam: **sim** – a música de uma banda de *rock* poderá ser um *top hit* internacional – e **não** – a música não terá sucesso.

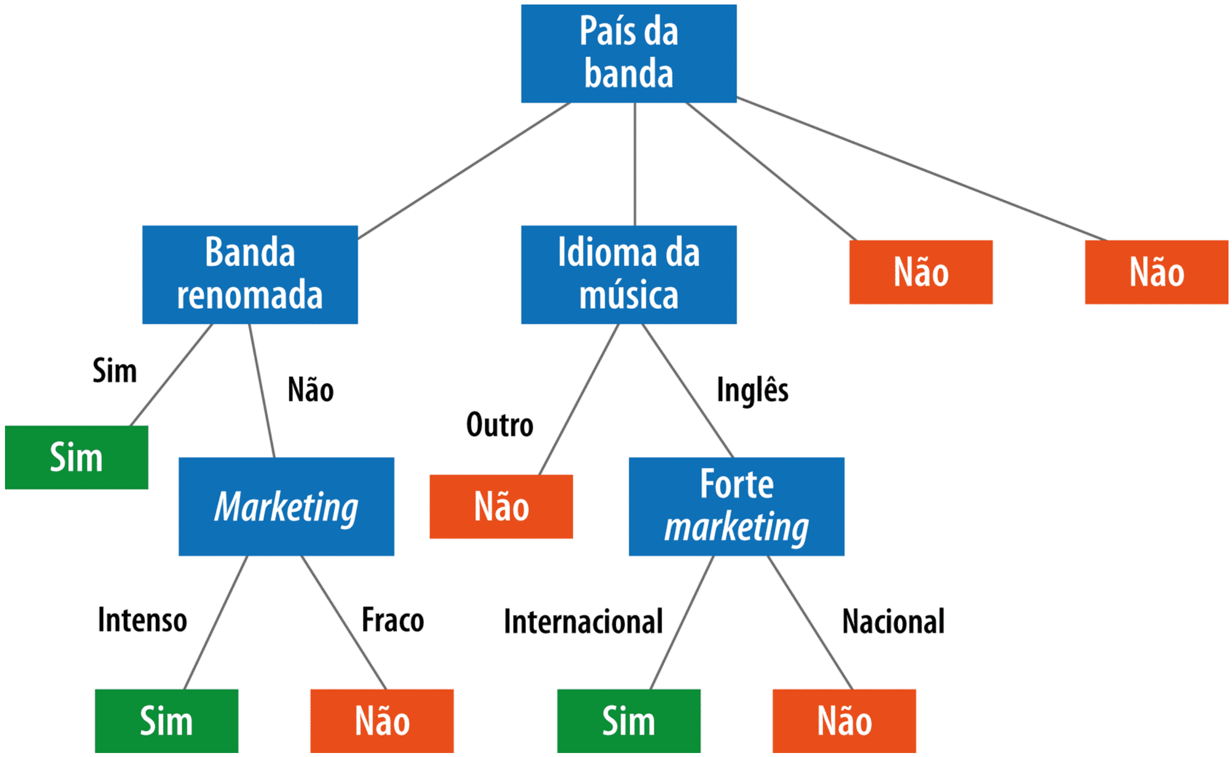

Figura 1 – Árvore de decisão original para prever se uma música de uma banda de *rock* será *top hit* internacional

Já o Quadro 1 evidencia a parte de uma representação equivalente da árvore de decisão da Figura 1. Logo, uma "árvore de decisão booleana", onde X1, X2, X3... Xn são "exemplos", podendo-se considerar como vetores de dados de entrada – atributos: país da banda, idioma da música etc. – e um valor booleano de saída – a última coluna do Quadro, onde o resultado é **sim** ou **não**. O grupo completo de exemplos é chamado de **conjunto de treinamento**.

Quadro 1 – Conjunto de "exemplos" para a árvore de decisão da Figura 1

Exemplo	País da banda	Banda renomada	Idioma da música	*Marketing*	Forte *marketing*	Meta: *hit* internacional
X1	Estados Unidos	Sim	Inglês	Sim	-	Sim
X2	Europa	-	Inglês	-	Internacional	Sim
X3	Europa	-	Outro	-	Nacional	Não
X4	África	-	-	-	-	Não
X5	Ásia	-	-	-	-	Não
X6	Estados Unidos	Não	Inglês	Não	-	Não
...	...	...	...	...	...	...

O mais interessante no algoritmo de aprendizagem em árvore de decisão (RUSSELL; NORVIG, 2004, p. 637) será testar primeiro os atributos mais importantes – colunas –, que permitem chegar à classificação correta com um pequeno número de testes.

A árvore de decisão “induzida” da Figura 2 é mais simples e permite chegar a uma decisão efetuando menos testes. Um algoritmo que permita gerar uma nova árvore de decisão induzida mais eficiente ou de forma mais simples “aprenderá” e gerará novo conhecimento, inéditas formas de raciocínio. Por este motivo, estamos diante de uma situação de **aprendizagem**.

Figura 2 – Árvore de decisão induzida para prever se uma música de uma banda de *rock* será *top hit* internacional

- Uma "árvore de decisão booleana" – X1... Xn no Quadro 1, "exemplos" – pode ser considerada como um vetor de dados de entrada – país da banda, idioma da música etc. – e um valor booleano de saída – a última coluna do Quadro, com o resultado **sim** ou **não**;

- O conjunto completo de "exemplos" é chamado de conjunto de treinamento;

- O algoritmo de aprendizagem em árvore de decisão gerará, produzirá "nova árvore de decisão induzida" – como vimos na Figura 2 –, mais simples que a árvore original, com menor profundidade e mais eficiente;

- Foi induzida uma nova árvore de decisão a partir de "exemplos". Por este motivo, o método é chamado de indução de árvores de decisão. A partir de uma função inicial **f(...)** foi obtida uma nova função **h(...)**, o que podemos considerar como uma situação de aprendizagem;

- A árvore de decisão é o exemplo mais utilizado de aprendizagens indutiva e supervisionada;

- A seleção de testes de atributos é um tema importante dentro das árvores de decisão e, assim, significativamente complexo – tal assunto poderá ser aprofundado no livro de Russell e Norvig (2004).

Aprendizagem Estatística como um Caso de Aprendizagem Supervisionada

A aprendizagem estatística é um modo de formular a atividade de aprendizagem utilizando inferência probabilística – raciocínio com incerteza –, a partir de observações. Segundo Russell e Norvig (2004): "A aprendizagem bayesiana simplesmente calcula a probabilidade de cada hipótese, considerando-se os dados, e faz previsões de acordo com ela".

Como observamos, qualquer raciocínio probabilístico ou "bayesiano" está relacionado a cálculos da teoria das probabilidades. Por isto, começaremos lembrando de algumas definições de probabilidades.

Probabilidades

Se você conhece e se lembra dos princípios da teoria das probabilidades, poderá pular esta e as duas próximas seções.

Começaremos com um exemplo simples: em um sorteio de “amigo secreto” na universidade, devemos escolher uma entre oito pessoas, sendo que são quatro homens e quatro mulheres. Qual será, então, a probabilidade de sortear uma mulher?

Ninguém duvidaria em responder: 50% ou 0,5. Mas como foi calculada tal probabilidade?

As probabilidades são utilizadas em experimentos aleatórios, com situações imprevisíveis, onde um resultado possível, que chamaremos de **evento**, é um subconjunto do conjunto de todas as possibilidades que chamaremos de **espaço amostral** ou **universo**.

Exemplo 1:

Retomando o caso do “amigo secreto”, representaremos este evento com dois conjuntos: o primeiro será o espaço amostral – que é o universo, as oito pessoas:

U = {Jorge, Ana, Pedro, Luzia, Luis, Vanda, José, Rosa}

O evento de sortear uma mulher é um conjunto que possui quatro elementos:

A = {Ana, Luzia, Vanda, Rosa}

O espaço amostral é:

U = {Jorge, Ana, Pedro, Luzia, Luis, Vanda, José, Rosa}

O evento A – conjunto de quatro elementos, quatro mulheres – é:

A = {Ana, Luzia, Vanda, Rosa}

A probabilidade **p(A)** de que o evento **A** aconteça será:

p(A) = n(A) / n(U) = 4 / 8 = 0,5 (ou 50%)

Sendo:

n(A) = número de elementos do evento **A**.

n(U) = número de elementos no espaço amostral **U**.

Importante!

A e U são conjuntos equiprováveis e finitos, ou seja, todos os elementos têm a mesma chance de acontecer.

Exemplo 2:

Temos um dado clássico de seis lados e queremos calcular a probabilidade de “sair” um número maior que quatro.

Figura 3
Fonte: Getty Images

U = {1, 2, 3, 4, 5, 6} universo, conjunto de números do dado.

A = {5, 6} evento **A**: sair um número > 6.

A probabilidade **P(A)** do evento “sair um número maior que 4” será:

P(A) = n(A) / n(U) = 2 / 6 (0,33 ou 33%)

Reflita

Qual será a probabilidade de “sair um número par” com um único dado?

Qual será a probabilidade de acertar uma questão da prova que possui cinco alternativas? E se tal questão tiver seis alternativas?

Probabilidades – Características e Axiomas

- A probabilidade de um evento impossível **A** será zero. Observe, então, que n(A) = 0, logo, n(A) / n(U) = 0 / n(U) = 0;

- A probabilidade de um **evento certo** – fato – é de valor um. Observe que n(A) / n(U) = 1. Por exemplo, se o espaço amostral for de oito mulheres, a probabilidade de sortear uma mulher é de 1 (100%);

- **Resumindo:** uma probabilidade sempre será um valor entre 0 e 1;

- A soma das probabilidades de um evento **A** e de seu complemento **A'** será a unidade (valor 1). Exemplo: para, em um evento, sortear uma mulher e um homem, em um universo de quatro homens e quatro mulheres, teremos: 0,5 + 0,5 = 1,0.

- **Axioma:** P(A Ú B) = P(A) + P(B) - P(A Ù B).

Exemplo 3:

Queremos calcular a probabilidade de “sortear um número ímpar ou menor ou igual que quatro” a partir de um grupo de oito valores: 1, 2, 3, 4, 5, 6, 7, 8. Vejamos:

U = {1, 2, 3, 4, 5, 6, 7, 8} espaço amostral de oito valores.

A = {1, 3, 5, 7} evento “sortear número ímpar”.

B = {1, 2, 3, 4} evento “sortear um número <= 4”.

A Ù B = {1, 3} interseção – elementos comuns – dos conjuntos anteriores **A** e **B**.

Temos que:

P(A) = n(A) / n(U) = 4 / 8 = 0,5

P(B) = n(B) / n(U) = 4 / 8 = 0,5

p(A Ù B) = n(A Ù B) / n(U) = 2 / 8 = 0,25

Então, usando o axioma P(A Ú B) = P(A) + P(B) - P(A Ù B), temos:

P(A Ú B) = P(A) + P(B) - P(A Ù B) = 0,5 + 0,5 - 0,25 = 0,75

Ou seja, a probabilidade de “sortear um número ímpar ou menor ou igual que 4” a partir desse grupo de oito valores é de 0,75 (75%).

Exemplo 4:

Trata-se de calcular a probabilidade de “sortear um número que seja ímpar e par”, dentro de um conjunto de oito valores: 1, 2, 3, 4, 5, 6, 7, 8.

Uma vez que conhecemos o seguinte axioma: P(A Ú B) = P(A) + P(B) - P(A Ù B), do qual podemos trocar termos, ficaria desta maneira: P(A Ù B) = P(A) + P(B) - P(A Ú B).

U = {1, 2, 3, 4, 5, 6, 7, 8}   espaço amostral de oito valores.

A = {1, 3, 5, 7}   evento “sortear número ímpar”.

B = {2, 4, 6, 8}   evento “sortear número par”.

A Ú B = {1, 2, 3, 4, 5, 6, 7, 8}   união dos conjuntos **A** e **B**.

Temos que:

P(A) = n(A) / n(U) = 4 / 8 = 0,5

P(B) = n(B) / n(U) = 4 / 8 = 0,5

P(A Ú B) = n(A Ú B) / n(U) = 8 / 8 = 1,0 p(A Ù B) = p(A) + p(B) - P(A Ú B) = 0,5 + 0,5 - 1,0 = 0

Ou seja, a probabilidade de “sortear um número que seja ímpar e par” dentro do conjunto de oito valores (1, 2, 3, 4, 5, 6, 7, 8) é zero ou 0%. Resultado óbvio, dado que é impossível sortear um número que seja ímpar e par ao mesmo tempo.

Probabilidade Incondicional ou a *Priori*

É o grau de crença em ausência de quaisquer outras informações. Tal probabilidade não depende de outras condições ou probabilidades.

Exemplo: probabilidade de que um paciente tenha cárie, segundo o dentista: P(cárie = verdadeiro) = P(cárie) = 0,5.

Outro exemplo: probabilidade de “sair” coroa no lançamento de uma moeda no universo: {cara, coroa} será:

P(coroa) = 0,5

Porque:

n(P) / n(U) = 1 / 2 = 0,5

Mais um exemplo: probabilidade de que Ana seja aprovada, segundo o professor:

P(Ana) = 0,2

Probabilidade Condicional ou Posterior

A probabilidade de a, dado que tudo o que sabemos é **b**:

P(a | b)   notação utilizada.

É uma probabilidade condicionada a outra probabilidade prévia.

Existe uma fórmula importante, conhecida como **regra do produto**, vejamos:

P(a | b) = P(a Ù b) / P(b)   **regra do produto**   (se P(b) > 0)

Exemplo: P(cárie | dordedente) = 0,8

Logo, trata-se da probabilidade de um paciente ter cárie conhecendo a probabilidade – prévia – de sentir dor de dente.

Exemplo 5:

Considerando que temos um típico baralho de 52 cartas, onde existem 13 copas, qual é a probabilidade de tirar duas cartas consecutivas – sem reposição – e que estas duas sejam copas? Observe que depois de tirar a primeira carta – supondo que seja copa – ficarão 51 cartas e, entre as quais, 12 copas.

Eventos: Copa1 (tirar a primeira copa) e Copa2 (tirar a segunda copa depois de tirar a primeira).

A expressão P(Copas2 | Copas1) é uma probabilidade condicional ou posterior.

Logo, calcularemos P(Copas2 Ù Copas1), ou seja, “probabilidade de tirar uma segunda copa e tirar uma primeira”.

Utilizando a regra do produto, temos que:

P(Copas2 | Copas1) = P(Copas2 Ù Copas1) / P(Copas1)

P (Copas2 Ù Copas1) = P(Copas2 | Copas1) x P(Copas1)

P (Copas2 Ù Copas1) = 12 / 51 x 13 / 52 = 0,0588 = 5,88%

Interpretando os cálculos, encontramos a “probabilidade de tirar uma segunda copa e tirar uma primeira” P(Copas2 Ù Copas1), conhecendo a probabilidade condicional de que tiremos uma segunda carta de naipe copa, dado que tiramos uma primeira P(Copas2 | Copas1). Utilizamos, então, a regra do produto e chegamos à conclusão de que existem 5,88% de probabilidade de tirar uma primeira copa do baralho de 52 cartas e tirar uma segunda copa das 51 cartas que ficaram.

Probabilidades – O Teorema de Bayes

Existe outra fórmula importante, conhecida como lei, regra ou teorema de Bayes, que relaciona probabilidades e que será útil para efetuar cálculos e raciocínios baseados em probabilidades, vejamos:

$$P(b|a) = \frac{P(b|a)\,P(b)}{P(a)} \quad \text{teorema de Bayes.}$$

A regra ou teorema de Bayes é significativamente útil porque, às vezes, vários termos são conhecidos e necessitamos encontrar uma probabilidade incógnita. Assim, o teorema de Bayes recebe este nome devido ao matemático inglês Thomas Bayes (1701-1761), quem originalmente forneceu tal equação para prever novas evidências, relacionando probabilidades *a priori* com probabilidades condicionadas. De modo que a inferência probabilística em inteligência artificial – para a aprendizagem estatística – utiliza como base essa regra.

Exemplo 6 – de Russell e Norvig (2004, p. 466-467):

Temos dois eventos:

S = rigidez no pescoço.

M = doença meningite.

Os médicos sabem que:

P(S|M) = 0,5 (a rigidez no pescoço condicionada a ter meningite; é uma probabilidade condicional).

P(M) = 1 / 50.000 (ou seja, uma em cada 50.000 pessoas tem meningite; é uma probabilidade *a priori*).

P(S) = 1 / 20 (uma em cada vinte pessoas tem rigidez no pescoço; é outra probabilidade *a priori*).

Assim, segundo a regra de Bayes:

P(M | S) = P(S | M) P(M) / P(S)

Com os valores conhecidos – vide probabilidades anteriores –, calculamos:

P(M | S) = 0,0002

De modo que ter meningite está condicionado a ter rigidez no pescoço com probabilidade de 0,0002 (0,02 %).

Este é um exemplo de utilização da regra de Bayes para um caso de diagnóstico médico, mas poderíamos usar tal lei em outras situações e em diferentes áreas do conhecimento.

Aprendizagem Estatística – Aprendizagem Bayesiana

Conforme comentado, a aprendizagem estatística é um modo de formular a atividade de aprendizagem utilizando inferência probabilística, a qual considera o raciocínio com incerteza a partir de observações. Segundo Russell e Norvig (2004): "A aprendizagem bayesiana simplesmente calcula a probabilidade de cada hipótese, considerando-se os dados, e faz previsões de acordo com ela".

Dentro do raciocínio com probabilidades – que podemos chamar de "inferência bayesiana" –, a situação mais comum é encontrar a hipótese mais provável **h**, **h ϵ H**, sendo fornecidos os dados de treinamento **D** – observações. Vejamos a hipótese com o máximo *a posteriori* – de maior probabilidade.

Fórmula: hipótese h com o Máximo A Posteriori (MAP):

$$
\begin{aligned}
h_{MAP} &= \underset{h \in H}{\arg\max}\, P(h \mid D) \\
&= \underset{h \in H}{\arg\max}\, \frac{P(D \mid h)\, P(h)}{P(D)} \\
&= \underset{h \in H}{\arg\max}\, P(D \mid h)\, P(h)
\end{aligned}
$$

Figura 4 – Fórmula: hipótese h com o Máximo *A Posteriori* (MAP)

A fórmula anterior utiliza o teorema de Bayes e desconsidera, finalmente, o termo **P(D)**, uma vez que este é uma constante independente de **h**, que afetaria igualmente todas as hipóteses.

Assim, na fórmula da Figura 4 temos que:

P(h) é a probabilidade *a priori* – incondicional – da hipótese **h**.

P(D | h) é a probabilidade condicionada de **D** dado **h**.

P(D) é a probabilidade *a priori* dos dados de treinamento **D**.

Observemos que, para várias hipóteses h1, h2, h3..., calcularemos a probabilidade com a fórmula anterior para cada hipótese. Logo, a **hipótese que tiver maior valor será a melhor previsão**, ou seja, a hipótese que, entre todas as hipóteses, tiver o maior valor será a mais segura, a que melhor se aproximará de uma conclusão ou previsão correta.

Exemplo 7 – baseado no caso trazido por Koerich (2008):

Considere um problema de diagnóstico médico onde existem duas hipóteses:

- O paciente tem câncer: câncer;
- O paciente não tem câncer: ¬câncer.

Os dados disponíveis são de um exame de laboratório com dois resultados possíveis:

r+ positivo.

r- negativo.

Temos o conhecimento prévio de que na população inteira somente 0,008 tem essa doença – probabilidade de 0,008, sendo este o termo P(h) do caso anterior.

O teste retorna um resultado positivo correto em 98% dos casos nos quais a doença está realmente presente – probabilidade de 0,98.

Ademais, o teste retorna um resultado negativo correto em 97% dos casos nos quais a doença não está presente (0,97, logo, de dar positivo se não houver câncer, 0,03).

Calculando a hipótese com maior probabilidade a posteriori, temos que:

P(r+ | câncer) P(câncer) = 0,98 . 0,008 = 0,0078 primeira hipótese.

P(r+ | ¬câncer) P(¬câncer) = 0,03 . 0,992 = 0,0298 segunda hipótese maior.

Assim, hMAP = ¬câncer, de modo que esta será a melhor previsão – a hipótese com MAP.

As vantagens do aprendizado estatístico de máquina, como é o caso da aprendizagem bayesiana aqui estudada, giram em torno de:

- Poder utilizar probabilidades conhecidas dentro do domínio da situação ou problema analisado;
- Utilizarmos algoritmos para inferir ou raciocinar e chegar a novas instâncias ou inéditos conhecimentos com bons resultados gerados, o que reforça a ideia de que houve aprendizagem.

Por outro lado, a principal desvantagem da utilização de aprendizado estatístico de máquina está na necessidade de processamento estatístico, este que é a base de tal aprendizagem, de modo que:

- Muitas probabilidades poderiam ser calculadas dentro do algoritmo de inferência;
- Isto pode ocasionar alto custo computacional.

Alguns estudos – que estão fora do alcance deste material introdutório – abordam algoritmos e técnicas para melhorar o desempenho – *performance* – dentro da aprendizagem bayesiana, como é o caso do chamado classificador naive Bayes – Naïve Bayes.

Material Complementar

Indicações para saber mais sobre os assuntos abordados nesta Unidade:

Sites

Produtos de IA e *Machine Learning*

Clique no botão para conferir o conteúdo.

ACESSE

IBM Watson – Brasil

Clique no botão para conferir o conteúdo.

ACESSE

Leitura

TensorFlow: Aprendizado de Máquina mais Inteligente para Todos

Clique no botão para conferir o conteúdo.

ACESSE

Como Selecionar Algoritmos do *Azure Machine Learning*

Clique no botão para conferir o conteúdo.

ACESSE

Esta é a Aposta do *Google* para Popularizar a Aprendizagem de Máquina

Clique no botão para conferir o conteúdo.

ACESSE

Aprendizado Bayesiano Aplicado ao Processamento de Línguas Naturais

Clique no botão para conferir o conteúdo.

ACESSE

Referências

AURÉLIO. **Novo dicionário eletrônico Aurélio**. Versão 6.0. Curitiba, PR: Positivo, 2009.

HAUSER, L. *Artificial intelligence*. ***Internet Encyclopedia of Philosophy***, [20--]. Disponível em: <http://www.iep.utm.edu/art-inte>. Acesso em: 06/09/2017.

KNAPP, S. *Artificial intelligence*: *past*, *present*, *and future*. ***Vox of Dartmouth***, jul. 24, 2006. Disponível em: <http://www.dartmouth.edu/~vox/0607/0724/ai50.html>. Acesso em: 06/07/2017.

KOERICH, A. L. **Aprendizagem de máquina:** aprendizagem bayesiana. 2008. Disponível em: <http://www.ppgia.pucpr.br/~alekoe/AM/2008/4-AprendizagemBayesiana-ApreMaq-2008.pdf>. Acesso em: 08/10/2017.

LUGER, G. F. **Inteligência artificial:** estruturas e estratégias para a resolução de problemas complexos. Trad. Paulo Martins Engel. 4. ed. Porto Alegre, RS: Bookman, 2004.

RICH, E.; KNIGHT, K. **Inteligência artificial**. 2. ed. São Paulo: Makron Books, 1994.

ROSA, J. L. G. **Fundamentos da inteligência artificial**. Rio de Janeiro: LTC, 2011.

RUSSELL, S. J.; NORVIG, P. **Inteligência artificial:** referência completa para cursos de Computação. 2. ed. Rio de Janeiro: Elsevier, 2004.

WINSTON, P. H. **Inteligência artificial**. Rio de Janeiro: LTC, 1988.