

NOTAS SOBRE OS FUNDAMENTOS MATEMÁTICOS DA INTELIGÊNCIA ARTIFICIAL

NOTES ON THE MATHEMATICAL FOUNDATIONS OF ARTIFICIAL INTELLIGENCE

Rodrigo Siqueira-Batista^{1,2}, Eugênio Silva^{3,4}

¹LAP/LMECS – UFV, ²Escola de Medicina – FADIP, ³CCT – UNIFESO, ⁴UComp – UEZO ¹rsbatista@ufv.br, ³eugeniosilva@unifeso.edu.br

Resumo

A área de Inteligência Artificial (IA) representa uma das principais fronteiras do conhecimento, na atualidade, por sua aplicabilidade em diferentes setores da vida humana. O correto entendimento de seus conceitos, e o consequente bom uso de suas técnicas, dependem do adequado estudo da Matemática, uma vez que o nascimento daquele campo se articula, profundamente, com desenvolvimentos dessa disciplina no século XX. Na área da IA, merecem destaque as Redes Neurais Artificiais (RNA) e as Máquinas de Vetor de Suporte (MVS), cuja característica marcante é a versatilidade, o que permite que sejam empregadas na construção de modelos de solução para problemas de interesse de uma ampla variedade de campos da atividade humana. Embora promissoras, a obtenção de modelos bem-sucedidos depende diretamente da correta parametrização dos algoritmos por trás dessas técnicas, o que, por sua vez, torna necessária uma boa compreensão dos seus fundamentos matemáticos. Desta feita, o objetivo do presente texto é indicar o *ferramental* matemático que deve ser compreendido para que se possa tirar o melhor proveito das RNA e das MVS.

Palavras-chave: Inteligência Artificial; Matemática; Redes Neurais Artificiais; Máquinas de Vetor de Suporte.

Abstract

Artificial Intelligence (AI) represents one of the main frontiers of knowledge, today, due its applicability in different sectors of human life. The correct understanding of IA and then the efficient usage of its techniques, depends on the proper study of Mathematics, since the birth of that field is deeply articulated with developments of this discipline in the twentieth century. In AI, Artificial Neural Networks (ANN) and Support Vector Machines (SVM) have a special importance due to its versatility, once they can be applied in the construction of solution models for problems of a wide range of human interests. Although promising, obtaining successful models depends directly on the correct parameterization of the algorithms behind these techniques, which, in turn, depends on a good understanding of their mathematical foundations. Therefore, the objective of this text is to indicate the mathematical tools that must be understood in order to get the best benefit from ANN and SVM.

Keywords: Artificial Intelligence; Mathematics; Artificial Neural Networks; Support Vector Machines.

Introdução

A Inteligência Artificial (IA) é um dos temas mais instigantes da atualidade. O termo – mote de obras cinematográficas, tais como *Ela*, *Eva*, *Ex machina* e *Westworld* (recente série lançada pela HBO) – é frequentemente apropriado pela mídia leiga, com citações em sites, jornais e revistas, tornando-o bastante familiar à sociedade contemporânea. Do ponto de vista da ciência e da tecnologia, tem-se experimentado um marcante crescimento da participação da IA em distintos campos do conhecimento: nas *ciências exatas* – Astronomia, Física e Matemática –, nas

ciências da saúde e da vida – Biologia Molecular, Genética e Medicina –, nas *ciências humanas e sociais* – Bioética, Direito e Economia – e nas Artes – Cinema e Música –, somente para citar algumas esferas que têm se beneficiado dos atuais avanços da área (BITTAR et al., 2018; CAETANO et al., 2005; CHANG, 2018; GUNKEL, 2017; WUERGES; BORBA, 2010; ALTHOFF et al., 2018). Com efeito, pode-se dizer que a IA “*é relevante para qualquer tarefa intelectual; é verdadeiramente um campo universal*” (RUSSELL; NORVIG, 2013, p. 3).

A expressão “Inteligência Artificial” articula duas palavras – “inteligência” e

“artificial” – as quais são de difícil definição (NILSSON, 1998). Pode-se entender *inteligência* “como uma capacidade de certos organismos para adaptar-se a situações novas, utilizando o conhecimento adquirido no curso de processos anteriores de adaptação. Deste ponto de vista, a inteligência é considerada como uma capacidade de aprendizagem e de aplicação da aprendizagem” (FERRATER MORA, 2001, p. 1873; trad. livre dos autores). O *artificial* pode ser compreendido como “*tudo aquilo que é feito pelo ser humano, ou seja, um artefato*” (VON ZUBEN, 2011). Tais definições não são consensuais e, por conseguinte, estão sujeitas a distintas críticas; sem embargo, compondo-as, pode-se chegar a certa compreensão da IA como “*A arte de criar máquinas [artefatos] que executam funções que exigem inteligência [capacidade de aprendizagem, por exemplo] quando realizado por pessoas*” (KURZWEIL, 1990, p. 14; trad. livre dos autores). Outra definição bastante interessante é a proposta por Rich e Knight (1991): “*O estudo de como fazer computadores realizarem coisas nas quais, no momento, as pessoas são melhores*”, p. 3; trad. livre dos autores).

A IA está presente, hoje, nos mais díspares artefatos e técnicas, incluindo as ferramentas de tradução, o planejamento logístico, o reconhecimento de face e de voz, os jogos e a robótica, entre outros (CASTRO; FERRARI, 2016; ARAUJO JUNIOR et al., 2017; NASCIMENTO et al., 2016). A incorporação da IA aos diferentes campos da atividade humana evoca adicional responsabilidade em termos do conhecimento das suas potencialidades e dos seus limites, bem como dos seus fundamentos. Nesse sentido, vale lembrar que a IA não é “*mágica ou ficção científica, mas ciência, engenharia e matemática*” (RUSSELL; NORVIG, 2013, p. 27), ressaltando-se a importância desse último campo para os estudos na área e para a ampliação de seu entendimento.

Diante disso, o escopo do presente artigo consiste em apresentar os aspectos matemáticos

que fundamentam duas das principais técnicas empregadas no âmbito da aprendizagem de máquina, a saber: Redes Neurais Artificiais (RNA) e Máquinas de Vetor de Suporte (MVS). Essa escolha se justifica pela ampla variedade de problemas em que são aplicadas com considerável sucesso. O bom entendimento de seus fundamentos é condição *sine qua non* para o uso correto e proveitoso de ambas as técnicas. É importante destacar que está fora do escopo do artigo o detalhamento dos fundamentos matemáticos aqui apontados. Por outro lado, é apresentada literatura, pertinente, na qual esse conteúdo está disponível.

Histórico

De uma perspectiva histórica, a IA é conceito recentemente incluído na ciência hodierna, remontando sua primeira aparição ao século XX. As origens podem ser recontadas a partir de diferentes espaços e tempos, mas, de um modo geral, acompanhando o clássico texto de Russell e Norvig (2013), reconhece-se a proposição, em 1943, do neurônio artificial – por Warren McCulloch e Walter Pitts (1943), autores que utilizaram (i) fundamentos de neurofisiologia, (ii) a lógica proposicional de Bertrand Russell e Whitehead e (iii) a teoria da computação de Alan Turing – como um dos eventos originários no âmbito da IA (McCULLOCH; PITTS, 1943; TURING, 1936). Em relação a esse último autor, cabe destacar que em um mesmo artigo de 1950 (TURING, 1950), Turing apresentou os conceitos de “aprendizagem de máquina”, de “aprendizagem por reforço” e de “algoritmos genéticos”, além de ter proposto o célebre teste de Turing, o qual foi “*projetado para fornecer uma definição operacional satisfatória de inteligência. O computador passará no teste se um interrogador humano, depois de propor algumas perguntas por escrito, não conseguir descobrir se as respostas escritas vêm de uma pessoa ou de um computador*” (RUSSELL; NORVIG, 2013, p. 4).

A realização de um seminário – sob a organização de John McCarthy, Marvin

Minsky, Claude Shannon e Nathaniel Rochester – dirigido ao debate acadêmico sobre a inteligência, os autômatos e as redes neurais, em Dartmouth, EUA, 1956, pode ser considerado o marco da IA (McCARTHY *et al.*, 1955). Vários aspectos foram debatidos – como possibilidades de uma máquina jogar xadrez, resolver problemas matemáticos e de ser capaz de “pensar” (McCARTHY *et al.*, 1955; NEWELL *et al.*, 1958) – e, segundo consta, o termo IA foi oficialmente utilizado, pela primeira vez, justamente nessa reunião (RUSSELL; NORVIG, 2013, p. 17). Ademais, se reconhece que o referido evento contribuiu, decisivamente, para a delimitação inicial do campo científico de estudo da IA.

Seguiu-se um período de grande euforia, especialmente com o sucesso obtido pelo GPS (*General Problem Solver*), programa desenvolvido por Newell e Simon (1961), para a solução de problemas, e pelo desenvolvimento de outros computadores, dotados de capacidade similar, por diferentes cientistas. Ademais, o desenvolvimento da linguagem Lisp (que se tornou a principal linguagem de programação em IA), por McCarthy, e da proposição, pelo mesmo autor, do programa *Advice Taker* (McCARTHY, 1958), foram significativos impulsos para o crescimento e o desenvolvimento do recém-nascido campo da IA. Outra contribuição importante – que deve ser brevemente mencionada – é o advento das primeiras redes neurais artificiais (WINOGRAD; COWAN, 1963), os *perceptrons* (ROSENBLATT, 1962).

Às conquistas iniciais da IA seguiram-se inúmeros insucessos – muito relacionados às tentativas de solução de problemas difíceis, às questões da chamada “explosão combinatória” e à capacidade limitada de aprendizagem dos *perceptrons* –, os quais produziram importantes dúvidas em relação aos projetos da IA (LIGHTHILL, 1973). Foram necessários mais alguns anos de pesquisa – reconhecendo-se que, nesse período, os subsídios para a investigação científica em IA decresceram significativamente – para que a implementação

dos *sistemas especialistas* (final dos 1970 e início dos 1980) e o “renascimento” das *redes neurais artificiais* pudessem trazer novo “alento” ao campo de estudo. Ademais, recentes desenvolvimentos – tais como os *agentes inteligentes*, os *sistemas multiagentes*, a mineração de dados e a *deep learning*, entre outras estratégias – têm reconfigurado o trabalho na área de IA e ampliado sua utilização no cotidiano. Cabe destaque, nesse contexto, à aprendizagem de máquina, entendida como “*um ramo da inteligência artificial que tem como objetivo desenvolver técnicas capazes de ensinar ao computador a aprender e/ou desempenhar determinada tarefa de forma melhor a partir das próprias experiências*” (SANTOS, 2016, p. 698).

Aprendizagem de máquina

A IA se vale, não raro, de metáforas para melhor representar os conceitos por trás de suas ferramentas e no caso da aprendizagem de máquina não é diferente. No contexto computacional, o termo “aprendizagem” consiste em “*mudanças adaptáveis no sistema, no sentido de que permitem que o sistema, da próxima vez, faça a mesma tarefa ou tarefas tiradas do mesmo grupo de forma mais eficiente e eficaz*” (SIMON, 1983). De outra forma, trata-se de um processo computacional iterativo em que os parâmetros de um modelo matemático são ajustados a fim de minimizar uma medida de erro. Assim, em termos gerais, quanto menor o erro cometido pelo modelo na execução de uma determinada tarefa, melhor será o seu desempenho nessa tarefa.

As técnicas de aprendizagem de máquina são utilizadas, principalmente, na solução de problemas que envolvam fenômenos para os quais não se conhecem modelos analíticos que os representem adequadamente. Para essas questões, em geral, a inexistência de representações analíticas fidedignas faz com que as soluções humanas sejam mais eficientes do que as computacionais e a proposta da IA, segundo Rich e Knight (1991), é fazer com que

os computadores “alcancem” os humanos e, quem sabe, até os ultrapasse.

O tema aprendizagem de máquina se divide em algumas modalidades, nas quais se destacam a aprendizagem supervisionada, a não supervisionada e a aprendizagem por reforço (FACELI et al., 2011). As atenções deste artigo estão concentradas especificamente na modalidade supervisionada, uma vez que é nela que se enquadram as técnicas de RNA e de MVS aqui exploradas. Na aprendizagem supervisionada, às vezes denominada aprendizagem com professor, é necessária a existência de um conjunto de treinamento – formado por exemplos entrada-saída –, que sejam representativos da tarefa que o sistema deve aprender. A entrada corresponde a um estímulo recebido pelo sistema e a saída consiste na ação desejada para aquele estímulo. Dessa forma, o propósito da aprendizagem consiste em ajustar os parâmetros internos do sistema para que este emita respostas tão semelhantes às respostas desejadas, quanto possível. Do ponto de vista matemático, significa encontrar a função que estabelece o melhor mapeamento entre os dados de entrada e de saída. A Figura 1 ilustra uma representação esquemática da aprendizagem supervisionada.

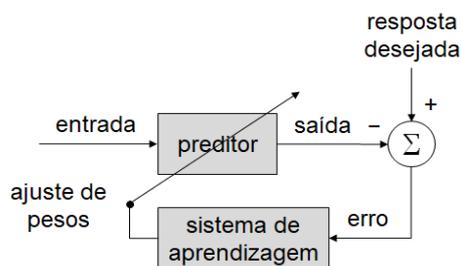


Figura 1. Esquema de funcionamento da aprendizagem supervisionada.

Ilustração elaborada por Eugênio Silva.

Nesse esquema, observa-se que para cada estímulo recebido pelo sistema de aprendizagem é emitida uma resposta, que em seguida é comparada à resposta desejada (supervisor) para aquele estímulo; havendo divergência entre essas respostas, é gerado um sinal de erro que é usado como subsídio para a

atualização dos parâmetros internos do sistema de aprendizagem. O grande objetivo da aprendizagem é alcançar a generalização, ou seja, obter a calibração adequada dos parâmetros do sistema de modo a responder coerentemente a estímulos futuros, que – obviamente – não estavam presentes no processo de aprendizagem, mas que seguem o mesmo padrão daqueles que estavam (NASCIMENTO et al., 2016).

Para que a aprendizagem aconteça, a medida mais básica a ser tomada consiste em representar adequadamente os estímulos que serão apresentados como entrada ao sistema. Invariavelmente, essa representação deve ser simbólica ou numérica, destacando-se que no caso das técnicas de RNA e MVS, a representação é essencialmente numérica. Isso significa que um estímulo é representado por um vetor de valores numéricos, onde cada valor corresponde a uma característica extraída desse estímulo. Em termos geométricos, esse vetor representa um ponto em um espaço n -dimensional, onde n representa o número de características que o compõem.

Em se tratando de aprendizagem supervisionada, há dois tipos de tarefas que essa modalidade de aprendizagem se propõe a resolver: *classificação* e *regressão*. Na primeira, a resposta a um estímulo é sempre um valor numérico discreto, enquanto na segunda a resposta é sempre um valor contínuo. Tanto RNA quanto MVS podem ser usadas como base para a construção de modelos para problemas que envolvem ambas as tarefas e os conceitos matemáticos a serem compreendidos nessas situações são basicamente os mesmos. Em geral, é mais comum a aplicação dessas técnicas na solução de problemas de classificação e, por isso, este é o foco do artigo.

Do ponto de vista geométrico, um problema de classificação consiste em encontrar uma superfície de decisão capaz de separar espacialmente os estímulos de acordo com os rótulos (respostas) que lhes são atribuídos durante o treinamento. A quantidade de rótulos diferentes determina a quantidade de

classes envolvidas no problema. As Figuras 2-a e 2-b representam dois problemas de classificação hipotéticos em que os estímulos são representados por duas características (c_1 e c_2) e são mapeados para uma dentre duas classes possíveis (\bullet e \blacklozenge). No primeiro caso, observa-se que a superfície de decisão é linear, o que caracteriza um problema em que as classes são linearmente separáveis. No segundo, tem-se um problema semelhante, porém com classes que não são linearmente separáveis. Obviamente, esses exemplos podem ser extrapolados para situações em que os estímulos são representados por mais características e, conseqüentemente, os pontos e as superfícies seriam representados em espaços de dimensão superior.

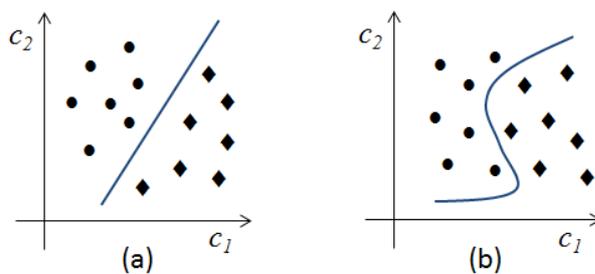


Figura 2. Representação geométrica de problemas de classificação
 (a) linearmente separável e (b) não linearmente separável.
 Ilustrações elaboradas por Eugênio Silva.

Sob essa perspectiva, o aprendizado, seja no contexto de RNA ou de MVS, consiste em parametrizar corretamente essa superfície de decisão (linear ou não) para que a separação entre as classes seja a melhor possível, não só para os estímulos de treinamento, mas também para os estímulos futuros, sempre com vistas à obtenção de uma generalização satisfatória. Nesse ponto fica clara a necessidade de conhecimento de um conceito matemático básico que é a representação de **pontos e superfícies (lineares ou não) em espaços multidimensionais** (FACELI, et al., 2011; STEINBRUCH; WINTERLE, 1987).

Redes neurais artificiais

As Redes Neurais Artificiais (RNA) são sistemas computacionais cuja organização é inspirada no sistema nervoso do *Homo sapiens* (NUNES, 2012). De fato, a unidade das RNA é o neurônio artificial (NA) proposto inicialmente – como já comentado – por McCulloch e Pitts (1943), tal qual explicitado na Figura 3.

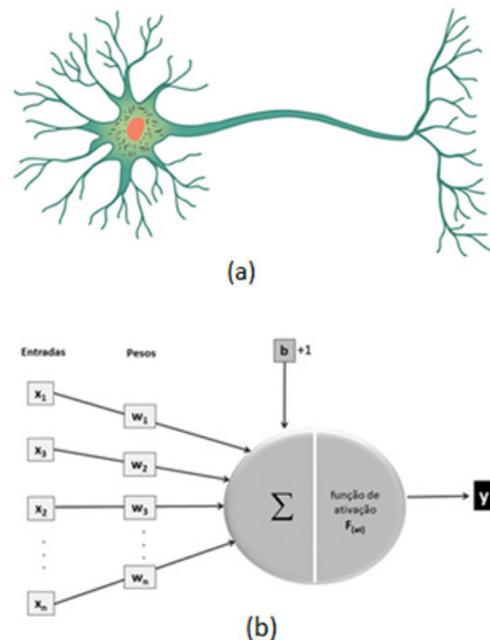


Figura 3. Representação esquemática do neurônio (a) humano e (b) artificial.

Ilustrações elaboradas por Rodrigo Siqueira-Batista¹.

Os neurônios artificiais, também denominados nodos, são modelos – matemático-computacionais (SEUNG; YUSTE, 2014) – que agregam um *combinador linear* e uma *função de ativação*, podendo adquirir dois “estados” – “equivalentes em termos concretos a uma proposição que definia seu estímulo adequado” (RUSSELL; NORVIG, 2013, p. 16) – a saber: “ligado” (“ativo”) ou “desligado” (“inativo”). Na Figura 3-b observa-se que o estímulo de entrada é representado por um vetor X composto por n características que descrevem aquele estímulo. A cada entrada x_i , com $i = 1, 2, \dots, n$, está associado um peso w_i e

¹ A ilustração do neurônio humano foi vetorizada por Ademir Nunes Ribeiro Júnior.

esse conjunto de pesos descreve um vetor W . O produto interno entre X e W associado a um termo de polarização b produz um potencial de ativação u . Em seguida, de acordo com o valor de u , a função de ativação decide se a saída do nodo é excitatória ou inibitória, caso u esteja, respectivamente, acima ou abaixo de um determinado limiar de ativação θ . Assim, a ativação de um NA pode ser descrita conforme as Equações 1 e 2 a seguir:

$$u = \sum_{i=1}^n w_i x_i + b \quad (1) \quad y = \begin{cases} 1 & \text{se } u \geq \theta \\ 0 & \text{se } u < \theta \end{cases} \quad (2)$$

A função de ativação – ao se considerar os valores 0 ou 1 para y – exemplificada na Equação 2 é uma das mais simples, recebendo a denominação de *função degrau* (NUNES, 2012) ou *função escada*, com aplicabilidade em problemas triviais (LU; CHEN, 2005). Há outros tipos de função de ativação – como a *linear*, *sigmoidal*, *tangente hiperbólica*, *gaussiana*, somente para citar algumas (NUNES, 2012) – as quais se prestam à execução de diferentes tarefas.

Os NA podem ser organizados em distintas arquiteturas de RNA, as quais são dotadas de capacidade de aprendizagem, como assinalado por Donald Hebb (1949), autor que propôs uma regra de atualização para modular as intensidades das *sinapses* – conexões – entre NA. Um dos primeiros desenhos – e dos mais simples – de RNA é denominado *perceptron* (ROSENBLATT, 1957), no qual se aplica a regra de Hebb ao neurônio da Figura 3-b para que “aprenda” a resolver um dado problema de classificação. No contexto das RNA, o aprendizado significa ajustar adequadamente, em um processo iterativo, os pesos associados às sinapses do neurônio, a fim de alcançar uma configuração que permita que esse neurônio descreva a superfície de decisão que melhor separa os estímulos de uma classe dos de outra. A associação de um estímulo a uma classe é dada pelo valor y de saída do neurônio, onde $y = 1$ representa uma classe e $y = 0$ representa

outra. A regra de atualização dos pesos de um *perceptron* é dada pela Equação 3.

$$W(t + 1) = W(t) + \lambda \cdot e(t) \cdot X(t) \quad (3)$$

Pela Equação 3 observa-se que o vetor W no instante seguinte $t + 1$ depende do valor W no instante t atual, de uma taxa de aprendizagem λ e também do erro cometido na classificação e do vetor X , ambos no instante atual t . O entendimento do processo de dedução que resulta nessa regra demanda o conhecimento de pelo menos dois conceitos matemáticos, a saber: **produto interno** entre vetores e **desigualdade de Cauchy-Schwarz** (STEINBRUCH; WINTERLE, 1987; SANTANA; QUEIRÓ, 2010; ANTON; RORRES, 2012).

Deve-se comentar que o *perceptron* é uma RNA de camada única, ou seja, dispõe de apenas uma camada de pesos ajustáveis. Com isso, a modelagem de relações mais complexas entre os valores de entrada e aqueles de saída tem limitações com o *perceptron*, restringindo a sua aplicação apenas à solução de problemas linearmente separáveis, nos moldes daquele apresentado na Figura 2-a. Para superar essa limitação, propôs-se uma rede de múltiplas camadas denominada *multilayer perceptron* MLP (Figura 4).

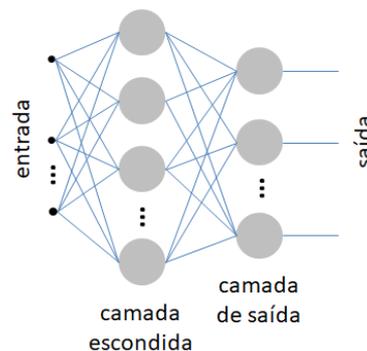


Figura 4. Esquema de Redes Neurais Artificiais “Multilayer Perceptron” (MLP).
 Ilustração elaborada por Rodrigo Siqueira-Batista.

No modelo MLP, a RNA usualmente contém uma ou duas camadas intermediárias – as quais representam *camadas ocultas* – e a *camada de saída*, a qual é composta por uma

quantidade de nodos correspondente à quantidade de classes envolvidas no problema. Dessa forma, a associação de um determinado estímulo a uma das classes consiste em obter uma saída próxima de 1 no nodo correspondente àquela classe e próxima de 0 em todos os outros. As redes MLP utilizam, tanto nos nodos da(s) camada(s) oculta(s) quanto nos da camada de saída, funções de ativação mais complexas, incluindo a função sigmoide (Equação 4) e a função tangente hiperbólica (Equação 5) (TAN; STEINBACH; KUMAR, 2006; BARNARD; VAN NIEKERK, 2018). Em ambas as equações a saída y_i representa a saída do neurônio i , β controla a inclinação da função e u é o potencial de ativação.

$$y_i = \frac{1}{1 + e^{-\beta u}} \quad (4) \quad y_i = \frac{1 - e^{-\beta u}}{1 + e^{-\beta u}} \quad (5)$$

Essa arquitetura de múltiplas camadas juntamente com a utilização de funções de ativação dos tipos citados é o que confere a uma MLP a capacidade de resolver problemas mais complexos, ou seja, aqueles que se caracterizam pela não linearidade da superfície de separação das classes (Figura 2-b). O que torna isso possível é a **combinação linear** de funções não lineares (STEINBRUCH; WINTERLE, 1987; SANTANA; QUEIRÓ, 2010; ANTON; RORRES, 2012), que é efetuada ao longo das camadas da RNA até a sua saída. Aqui, portanto, depara-se com mais um conceito matemático importante para o entendimento de como uma MLP é capaz de aproximar funções complexas.

A regra de aprendizagem de uma RNA do tipo MLP guarda algumas semelhanças com a regra de aprendizagem do *perceptron*, mas o processo dedutivo para se chegar até ela é bastante diferente. O algoritmo de retropropagação (*backpropagation*) do erro, também conhecido como *regra delta generalizada* (HAYKIN, 2001; BRAGA; CARVALHO; LUDERMIR, 2007), atualiza os

pesos das sinapses de um nodo da rede segundo a Equação 6.

$$W(t + 1) = W(t) + \lambda \cdot e(t) \cdot f'(u) \cdot X(t) \quad (6)$$

Observa-se nessa equação que, além dos termos presentes na equação de atualização dos pesos de um *perceptron*, há aqui também a necessidade de calcular $f'(u)$ que corresponde à **derivada** (STEWART, 2017a; GUIDORIZZI, 2018a) da função de ativação em relação ao potencial de ativação. Portanto, mais esse conceito matemático precisa ser compreendido.

A *regra delta generalizada* interpreta a tarefa de atualização dos pesos como um problema de otimização em que o erro – divergências entre as respostas emitidas pela RNA e as respostas desejadas – deve ser minimizado. Como o erro é calculado em função dos pesos, para as inúmeras combinações possíveis de valores para esses pesos é possível traçar uma superfície com as variações do erro de acordo com os pesos. A Figura 5 ilustra um exemplo de uma superfície em que o erro é dado pela variação de apenas dois pesos.

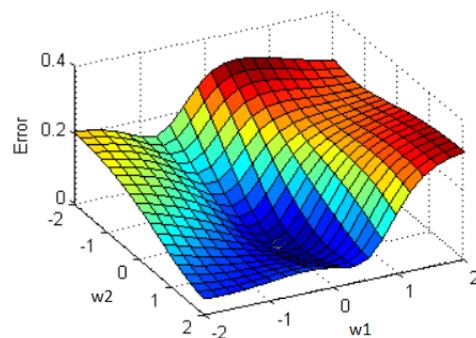


Figura 5. Exemplo de superfície de erro em função dos pesos de uma RNA.

Ilustração elaborada por Eugênio Silva.

O processo de otimização consiste em “caminhar” por essa superfície de erro a fim de tentar encontrar a configuração de pesos correspondente ao seu ponto mais profundo, ou seja, ao seu ponto de erro mínimo. No caso da *regra delta generalizada*, essa busca é guiada por um método iterativo que se baseia no

gradiente (STEWART, 2017b; GUIDORIZZI, 2018b) da função de erro para estabelecer a direção e o sentido da busca. O cálculo do gradiente, por sua vez, demanda a compreensão prévia do conceito de **derivada parcial** (STEWART, 2017b; GUIDORIZZI, 2018b) de funções multivariáveis. Portanto, mais esses dois conceitos figuram como essenciais para o correto entendimento de como transcorre a aprendizagem em RNA do tipo MLP.

A importância dos conceitos apontados aqui não se limita à compreensão do processo de aprendizagem. A configuração adequada dos parâmetros de aprendizagem do algoritmo – quantidade de ciclos (épocas) de treinamento, taxa de aprendizagem, quantidades de camadas ocultas e quantidades de neurônios por camada – também depende do entendimento desses conceitos.

Máquinas de vetor de suporte

As Máquinas de Vetor de Suporte (MVS) são uma técnica de *aprendizagem supervisionada* que provém da Teoria do Aprendizado Estatístico (TAE). Assim como acontece com as RNA e outras técnicas de classificação, no processo de aprendizagem as MVP também se propõem a encontrar a superfície que separa os estímulos pertencentes a uma classe dos de outra. Contudo, a proposta aqui vai além, uma vez que o objetivo das MVP é encontrar a *superfície de decisão ótima* (JARA ESTUPINAN et al., 2016), conforme ilustra a Figura 6.

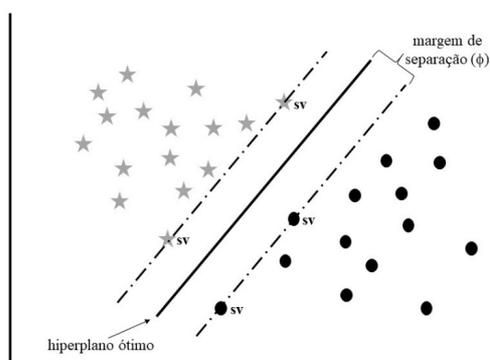


Figura 6. Máquina de vetor de suporte. Observe o *suporte vetorial* (sv) e a margem máxima de separação (ϕ). Ilustração elaborada por Rodrigo Siqueira Batista.

Observando a Figura 6, pode-se notar que o resultado da aprendizagem é “*um separador de margem máxima – um limite de decisão com a maior distância possível a pontos de exemplo*” (RUSSELL; NORVIG, 2013, p. 648). Como no caso das redes MLP, o processo de obtenção desse separador também se traduz em um problema de otimização, que aqui é resolvido por meio do método dos multiplicadores de Lagrange (STEWART, 2017b; GUIDORIZZI, 2018b). Nesse método, a função objetivo a ser minimizada deve ser reescrita como uma **função de Lagrange** (ou **lagrangiana**) (NETO, 2013; TAYLOR, 2013) para o problema de otimização e, portanto, mais esse conceito precisa ser compreendido. Para a minimização da lagrangiana recorre-se mais uma vez ao conceito de **gradiente** e, conseqüentemente, de **derivada parcial**.

A formulação básica de uma MVS permite apenas a construção de *separadores lineares*, o que limita sua utilização a problemas semelhantes aos da Figura 2-a. Para problemas mais complexos – não linearmente separáveis – é preciso aplicar uma **transformação não-linear** (HOFMAN; SCHOLKOPF; SMOLA, 2008; LIU; PRÍNCIPE; HAYKIN, 2010) no espaço original das características que descrevem os estímulos, de forma a projetá-las em um espaço de ordem superior em que as classes se tornem linearmente separáveis (VON ZUBEN, 2013; ESPINOSA-OVIEDO et al., 2017). Essa transformação se dá pela aplicação do chamado *truque do núcleo*, que consiste em aplicar uma **função de núcleo** (HOFMAN; SCHOLKOPF; SMOLA, 2008; LIU; PRÍNCIPE; HAYKIN, 2010) para obter a transformação desejada. Após a transformação, a formulação básica de uma MVS pode ser usada na solução do problema.

Aqui nota-se que alguns dos conceitos necessários ao bom entendimento das MVS são comuns àqueles apontados no contexto das RNA. Como naquele caso, a compreensão das concepções destacadas é de grande importância para a parametrização correta de uma MVS, com destaque para a escolha correta da função

de núcleo quando da necessidade de solução de problemas mais complexos.

Considerações finais

O presente ensaio propôs uma abordagem da fundamentação matemática da IA. Os aspectos brevemente apresentados no texto permitiram que fossem estabelecidas, de modo claro, as relações entre o conhecimento matemático e a emergência da IA. Nesse sentido, reforça-se a relevância do estudo das bases matemáticas aos interessados em apreender os meandros da IA, para uma melhor compreensão do alcance e das limitações dessa contemporânea ferramenta, a qual tem modificado – substantivamente – as vidas das pessoas no século XXI.

Agradecimentos

Os autores são gratos ao Prof. Ademir Nunes Ribeiro Júnior (FADIP), pela ajuda na elaboração do esquema do neurônio humano.

Referências

- ARAÚJO JUNIOR, C. A.; LEITE, H. G.; SOARES, C. P. B.; BINOTI, D. H. B.; SOUZA, A. P.; SANTANA, A. F. S.; TORRE, C. M. M. E. A multi-agent system for forest transport activity planning. *Cerne*, v. 23, n. 3, p. 329-337, 2017.
- ALTHOFF, D.; BAZAME, H. C.; FILGUEIRAS, R.; DIAS, S. H. B. Heuristic methods applied in reference evapotranspiration modeling. *Ciência e Agrotecnologia*, v. 42, n. 3, pp.314-324, 2018.
- ANTON, H.; RORRES, C. *Álgebra Linear com Aplicações*. 10. ed. Porto Alegre: Bookman, 2012.
- BARNARD, M.; VAN NIEKERK, T. I. Neural network fault diagnosis system for a diesel-electric locomotive's closed loop excitation control system. *SAIEE Africa Research Journal*, v. 109, n. 1, p. 23-35, 2018.
- BITTAR, R. D., ALVES, S. M. F.; MELO, F. R. Estimation of physical and chemical soil properties by artificial neural networks. *Revista Caatinga*, v. 31, n. 3, p. 704-712, 2018.
- BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. *Redes Neurais Artificiais - Teoria e Aplicações*. 2. ed. Rio de Janeiro: LTC, 2007.
- CAETANO, M.; MANZOLLI, J.; VON ZUBEN, F. J. *BioMúsica: Aplicações de Inteligência Artificial e Algoritmos Bio-Inspirados em Música*. In: *Proceedings of the 2. II Seminário de Música, Ciência e Tecnologia*, 2005, São Paulo (SP) [online]. 2005 [citado em 10 de outubro de 2018]. Disponível em: <http://www.proceedings.scielo.br/scielo.php?>
- CASTRO, L. N.; FERRARI, D. G. *Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações*. 1. ed. São Paulo: Saraiva, 2016. 351p.
- CHANG, A. *Artificial Intelligence in Pediatric Critical Care Medicine: Are we (finally) ready?* *Pediatric Critical Care Medicine*, v. 19, n. 10, p. 997-998, 2018.
- FERRATER MORA, J. *Diccionario de filosofía*. Tomo II, Barcelona: Editorial Ariel S.A., 2001.
- ESPINOSA-OVIEDO, J. E.; ZULUAGA-MAZO, A.; GOMEZ-MONTOYA, R. A. Kernel methods for improving text search engines transductive inference by using Support Vector Machines. *Tecciencia*, v. 12, n. 22, p.51-60, 2017.
- FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. L. F. *Inteligência Artificial – Uma Abordagem de Aprendizado de Máquina*, Rio de Janeiro: LTC; 2011.
- GUIDORIZZI, H. L. *Um Curso de Cálculo – Volume 1*. 6. ed. Rio de Janeiro: LTC, 2018a.
- GUIDORIZZI, H. L. *Um Curso de Cálculo – Volume 2*. 6. ed. Rio de Janeiro: LTC, 2018b.
- GUNKEL, D. J. *Comunicação e inteligência artificial: novos desafios e oportunidades para a pesquisa em comunicação*. *Galáxia*, n. 34, p. 5-19, 2017.
- HAYKIN, S. *Redes Neurais: Princípios e Prática*. 2. ed. Porto Alegre: Bookman, 2001.

- HEBB, D. O. *The Organization of Behavior*. Wiley: New York, 1949. 335p.
- HOFMAN, T.; SCHOLKOPF, B.; SMOLA, A. J. Kernel Methods in Machine Learning. *The Annals of Statistics*. 2008, Vol. 36, No. 3, 1171-1220.
- JARA ESTUPINAN, J.; GIRAL, D.; MARTINEZ SANTA, F. Implementación de algoritmos basados en máquinas de soporte vectorial (SVM) para sistemas eléctricos: revisión de tema. *Tecnur*, v. 20, n. 48, p.149-170, 2016.
- KURZWEIL, R. *The age of intelligent machines*. Cambridge: MIT Press, 1990, 565 p.
- LIGHTHILL, J. Artificial intelligence: A general survey. In: LIGHTHILL, J.; SUTHERLAND, N. S.; NEEDHAM, R. M.; LONGUET-HIGGINS, H. C.; MICHIE, D. *Artificial Intelligence: A Paper Symposium*. Science Research Council of Great Britain, 1973.
- LIU, W.; PRÍNCIPE, J. C.; HAYKIN, S. *Kernel Adaptive Filtering: A Comprehensive Introduction*. John Wiley & Sons, 2010.
- LU, W.; CHEN, T. Dynamical behaviors of Cohen-Grossberg neural networks with discontinuous activation functions. *Neural Networks*, v. 18, p. 231-242, 2005.
- McCARTHY, J.; MINSKY, M. L.; ROCHESTER, N.; SHANNON, C. E. Proposal for the Dartmouth summer research project on artificial intelligence. *Technical Reports*, Dartmouth College, 1955.
- McCARTHY, J. Programs with common sense. *Proceedings of Symposium on Mechanisation of Thought Processes*, v. 1, p. 77-84, 1958.
- McCULLOCH, W.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, p. 115-137, 1943.
- NASCIMENTO, C. D. L.; SILVA, S. D. S.; SILVA, T. A.; PEREIRA, W. C. A.; COSTA, M. G. F.; COSTA FILHO, C. F. F. Breast tumor classification in ultrasound images using support vector machines and neural networks. *Research on Biomedical Engineering*, v. 32, n. 3, p. 283-292, 2016.
- NETO, J. B. *Mecânica Newtoniana, Lagrangiana e Hamiltoniana*. 2. ed. Livraria da Física, 2013.
- NEWELL, A.; SHAW, J. C.; SIMON, H. A. Chess playing programs and the problem of complexity. *IBM Journal of Research and Development*, v. 4, n. 2, p. 320-335, 1958.
- NEWELL, A.; SIMON, H. A. GPS, a program that simulates human thought. In BILLING, H. *Lernende Automaten*. Olden Bourg, 1961, p. 109-124.
- NILSSON, N. J. *Artificial Intelligence: a new synthesis*. San Francisco: Morgan Kaufmann, 1998.
- NUNES, W. V. Redes neurais artificiais: aspectos introdutórios. In: ESPERIDIÃO ANTONIO, V. *Neurociências: diálogos e interseções*. Rio de Janeiro: Rubio, 2012, p. 255-303.
- RICH E.; KNIGHT, K. *Artificial Intelligence*. 2nd edition. McGraw-Hill, 1991.
- ROSENBLATT, F. *The perceptron, a perceiving and recognizing automaton*. Cornell Aeronautical Laboratory report. New York: Buffalo, 1957.
- ROSENBLATT, F. *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. Washington: Spartan Books, 1962. 616p.
- RUSSELL, S. J.; NORVIG, P. *Inteligência artificial*. Trad. Regina C. Simille. Rio de Janeiro: Elsevier, 2013.
- SANTANA, A. P.; QUEIRÓ, J. F. *Introdução à Álgebra Linear*. Lisboa: Gradiva, 2010.
- SANTOS, A. C. M. *Aprendizado de máquina aplicado ao diagnóstico de Dengue*, 2016, Recife (PE) [online]. XIII Encontro Nacional de Inteligência Artificial e Computacional. 2016 [citado em 10 de outubro de 2018]. Disponível em: <http://www.lbd.dcc.ufmg.br/colecoes/eniac/2016/059.pdf>.
- SIMON, H. A. Why Should Machines Learn? In: Michalski R.S., Carbonell J.G., Mitchell

- T.M. (eds) Machine Learning. Symbolic Computation. Springer, Berlin, Heidelberg, 1983.
- SEUNG, S.; YUSTE, R. Redes neurais. In: KANDEL, E. R.; SCHWARTZ, T. M.; SIEGELBAUM, S. A.; HUDSPETH, A. J. Princípios de Neurociências. Porto Alegre: AMGH, 2014. p. 1378-1395.
- STEINBRUCH, A; WINTERLE, P. Álgebra Linear, 2. ed. São Paulo: McGraw-Hill, 1987.
- STEWART, J. Cálculo – Volume 1. Tradução da 8. ed. São Paulo: Cengage Learning, 2017a.
- STEWART, J. Cálculo – Volume 2. Tradução da 8. ed. São Paulo: Cengage Learning, 2017b.
- TAN, P. N.; STEINBACH, M.; KUMAR, V. Introduction to data mining. Boston: Addison-Wesley; 2006.
- TAYLOR, J. R.; ROQUE, W. L. Mecânica Clássica. Porto Alegre: Bookman, 2013.
- TURING, A. On computable numbers, with an application to the Entscheidungsproblem. Proceedings of the London Mathematical Society, 2d series, 42, 230-265, 1936.
- TURING, A. Computing machinery and intelligence. Mind, v. 59, p. 433-460, 1950.
- WINOGRAD, S.; COWAN, J. D. Reliable Computation in the Presence of Noise. Cambridge: MIT Press, 1963.
- VON ZUBEN, F. J. O mundo natural e o mundo artificial, 2011, Campinas (SP) [online]. 2011 [citado em 10 de outubro de 2018]. Disponível em:<ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia707_1s11/notas_de_aula/topico1_IA707_1s11.pdf>
- VON ZUBEN, F. J. Máquinas de vetores de suporte, 2013, Campinas (SP) [online]. 2013 [citado em 10 de outubro de 2018]. Disponível em:<ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia353_1s13/topico8_1s2013.pdf>
- WUERGES, A. F. E.; BORBA, J. A. Redes neurais, lógica nebulosa e algoritmos genéticos: aplicações e possibilidades em finanças e contabilidade. Journal of Information Systems and Technology Management, v. 7, n. 1, p.163-182, 2010.