

# Como Ler Artigos que usam Machine Learning

Beatriz Guedes da Silva, Carlos Eugênio Ferreira Neto

## I. MACHINE LEARNING NA MEDICINA

A medicina certamente foi uma das áreas mais influenciadas pelo avanço da computação. A área de *Machine Learning*, também conhecida como Aprendizado de Máquina, é uma das mais exploradas para auxiliar os profissionais da saúde em seus diagnósticos e tratamentos. Mas, afinal, o que seria o *Machine Learning*, quais são suas vantagens para a sociedade em geral e por que os profissionais da área da saúde devem ao menos conhecer os termos técnicos para avaliar soluções que utilizam esta ferramenta?

O *Machine Learning* consiste em modelos computacionais projetados para aprender a partir de uma certa base de dados sobre um determinado assunto e, assim, verificar onde um dado que não foi utilizado para treino se encaixa de acordo com os rótulos fornecidos ou até mesmo criar os rótulos a partir de semelhanças na sua base de dados.

Nos últimos anos, muitas novas ferramentas de diagnóstico clínico foram desenvolvidas utilizando métodos complexos de *Machine Learning* [1]. O artigo que baseia este texto não foi elaborado para estudantes de computação, o que pode parecer contraditório inicialmente, mas conscientizar os profissionais da saúde para o fato de que sim, deve haver um entendimento dos mesmos para com a tecnologia é necessário para melhores resultados no que diz respeito à qualidade de vida e tratamento de seus pacientes.

O *Machine Learning* pode ser aplicado em todas as etapas de um exame de sintomas, desde o pré-diagnóstico até o pós-diagnóstico, melhorando a precisão do diagnóstico com modelos treinados para analisar os sintomas de doenças específicas e também pode ajudar a corrigir diagnósticos errôneos que podem trazer aflição e outros problemas ao paciente. É necessário entender que o modelo de *Machine Learning* não resolverá sozinho os problemas da medicina, um trabalho conjunto entre homem e máquina para aumentar a eficiência dos resultados é indicado.

O profissional da saúde também deve entender a interface de determinado modelo para não haver confusões ao utilizá-lo, considerando que o setor da saúde possui maior responsabilidade em caso de falhas. Em resumo, é importante que os profissionais da área da saúde entendam como funcionam os algoritmos de *Machine Learning* para compreender como as previsões do modelo podem auxiliar no diagnóstico do paciente e acelerar o tratamento de doenças, potencializando as chances de cura ou melhora da qualidade de vida do paciente.

Quanto ao *Machine Learning*, é importante entender que o mesmo possui dois tipos de aprendizado: aprendizado supervisionado e aprendizado não-supervisionado. O aprendizado supervisionado utiliza dados já rotulados para que o modelo aprenda, ou seja, é exibido um dataset à máquina

contendo todos os rótulos que ela deve aprender com base nas características da imagem, texto, áudio, etc. Para melhor compreensão, é como se a máquina recebesse várias fotos de uma parte do corpo de vários pacientes dizendo se o mesmo tem certa doença ou está saudável, assim, a máquina só pode produzir essas duas saídas para qualquer entrada. Enquanto o aprendizado não supervisionado consiste em agrupar vários dados sem rótulos e a missão do modelo é agrupá-los segundo suas características. Um exemplo seria expor o modelo a várias imagens de raio-x e o mesmo verifica quais imagens têm semelhanças entre si, assim, um especialista analisa os resultados, corrige os erros e rotula os dados separados pelo modelo de acordo com alguma doença ou se a parte examinada está saudável.

O processo de aprendizado da máquina requer muitos cuidados, pois complicações significativas podem ocorrer durante a exposição de dados ao modelo como *underfitting* e *overfitting*. O *underfitting* ocorre, por exemplo, quando o modelo encontra um ponto médio entre o resultado dos dados e sempre gera aquela saída para qualquer entrada. Em contrapartida, o *overfitting* é um excesso de ajuste do modelo aos dados fornecidos. Isso quer dizer que o modelo sempre fornecerá a saída correta para determinada entrada para os dados usados como treino, porém, não generaliza bem para novos conjuntos de dados, gerando resultados controversos com a amostra de dados usada para validação e teste da máquina.

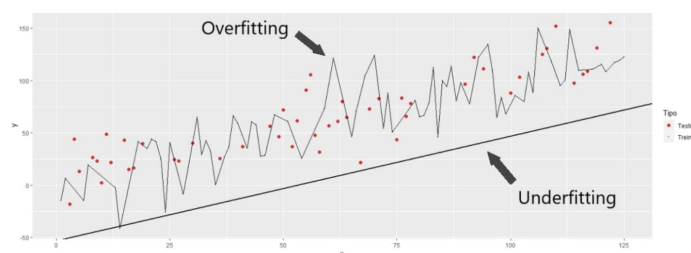


Figura 1. Gráfico que exhibe *overfitting* e *underfitting* em um modelo. [2]

O *underfitting* é facilmente detectado, mas para perceber o *overfitting* é necessário comparar o desempenho do modelo nos conjuntos de ajuste e validação. É necessário que ambas as curvas tenham o mesmo comportamento no gráfico. Se a linha do conjunto de validação não seguir o comportamento da linha de ajuste, é certo que há algum erro no modelo de *Machine Learning*. É necessário contatar os especialistas para os mesmos analisarem se a eficácia do modelo está aceitável. Como já citado anteriormente, a relação humano-máquina é essencial para a otimização de resultados.

Para prevenir o *overfitting*, uma técnica utilizada é a regularização, que suaviza o ajuste do modelo de *Machine Learning* de forma a evitar o *overfitting* a qualquer conjunto de

dados fornecido. Um dos métodos de regularização mais eficiente se chama *early stopping*, que interrompe o treinamento do modelo ao perceber que o mesmo não está mais tendo avanços na aprendizagem. Para isso, é necessário estipular um número de épocas que o modelo pode passar sem melhorar seu desempenho. Se o número limite de épocas for atingido e o número de vezes que a máquina é exposta aos dados não tiver atingido o limite, devemos parar imediatamente o aprendizado, de forma a prevenir o *overfitting*.

Diante do exposto, é fundamental que profissionais da saúde saibam lidar com essa tecnologia, pois o que o *Machine Learning* tem a oferecer tanto à medicina quanto a inúmeras áreas é um universo de infinitas possibilidades. Para tal, os especialistas da área devem estar alinhados com as técnicas de *Machine Learning* para verificar se as respostas fornecidas pelo modelo estão corretas. Também é indispensável não se ater a apenas um profissional para estudar o modelo. Quanto mais especialistas, menos riscos de decisões erradas ou tomadas por viés de qualquer espécie. Assim, haverá uma melhora cada vez mais evidente da qualidade de vida da sociedade.

#### REFERÊNCIAS

- [1] Y. Liu, P.-H. C. Chen, J. Krause, and L. Peng, “How to read articles that use machine learning: Users’ guides to the medical literature,” *JAMA*, vol. 322, no. 18, pp. 1806–1816, Nov. 2019.
- [2] Didática.Tech. (2023) Underfitting e overfitting. 2024-04-26. [Online]. Available: <https://didatica.tech/underfitting-e-overfitting/>