

ISSN: 2319-0124

AVALIAÇÃO DE DIFERENTES ESTRUTURAS DE CAMADAS OCULTAS EM ALGORITMOS DE DEEP LEARNING USANDO ANOVA

João Pedro do N. SILVA¹; Diego SAQUI²

RESUMO

Avaliar a eficiência de um processo de classificação vem se tornando cada vez mais complexo. O número de variáveis envolvidas na equação aumenta a cada dia, e diferentes estruturas de código surgem, dificultando o processo de avaliação. Porém, existem métodos matemáticos e estatísticos que são capazes de gerar uma avaliação desse modelo, a partir de amostras populacionais. Esses processos são capazes de classificar diversas informações dentro de um grupo e definir se diferentes algoritmos possuem a mesma eficiência, ou se um deles é mais eficaz. Neste estudo é detalhado um caso de teste entre 4 diferentes arquiteturas de um classificador baseado em *Deep Learning* usado para o mesmo propósito, e é analisado qual dentre eles apresenta melhores resultados.

Palavras-chave:

Deep Learning; Avaliação; Método; Computação.

1. INTRODUÇÃO

O desenvolvimento da tecnologia está em constante evolução, automatizando cada dia mais processos, quase não precisando de interação humana, o que resulta em diferentes programas, alguns dos quais são algoritmos de Inteligência Artificial (IA). Estes são códigos que podem aprender algo sem a necessidade de ensinar ao algoritmo todos os diferentes resultados, porém muitos deles apresentam um baixo nível de desempenho. Um algoritmo de IA que apresenta uma acurácia baixa e gera um resultado errôneo pode ser muito prejudicial ao processo final, como por exemplo, em tarefas para diagnosticar que uma pessoa tem câncer, como mostra o trabalho de Harada (2021), onde ele detalha o uso de algoritmos de IA para assistência no diagnóstico médico e os riscos que um resultado errôneo pode causar. Com a crescente diversidade de códigos de IA, está ficando difícil avaliar adequadamente sua eficiência devido aos muitos tipos variáveis envolvidos no processo, resultando em muitos algoritmos com baixo desempenho, como indica o artigo de Sichman (2021). Algoritmos com baixo desempenho são códigos que não atingiram seu máximo potencial, e com a menor mudança em seu código, a precisão pode aumentar drasticamente. Como descrito por Campagner (2020), diferentes estudos sobre como classificar e avaliar a performance

¹ Orientado, IFSULDEMINAS – *Campus* Muzambinho. E-mail: jpns1118@gmail.com.

² Orientador, IFSULDEMINAS – *Campus* Muzambinho. E-mail: diego.saqui@muz.ifsuldeminas.edu.br.

da IA estão sendo realizados em todo o mundo. Com o intuito de definir um método de classificar e avaliar algoritmos, neste estudo é apresentada uma estratégia conhecida como ANOVA, ou análise de variância, que pode ser usada para avaliar a melhor arquitetura entre diversos algoritmos de *Deep Learning* e explicar se é possível ou não determinar qual é a melhor modelo de IA.

2. MATERIAL E MÉTODOS

Neste projeto é utilizado o *Deep Learning* que, segundo EMMERT-STREIB et al. (2021) é um tipo de classificador de IA/*Machine Learning* que utiliza técnicas que exploram camadas de dados não lineares para processamento de dados de seleções de padrões e transformação, além do reconhecimento de padrões e classificações. Ele coleta e classifica as informações de um *dataset* em diferentes grupos, analisando os padrões e verificando a probabilidade de serem parte do mesmo e, ao final, o resultado é definido pelo grupo que apresenta maior probabilidade. O *software* usado para criar o código foi o “Google Colaboratory”³, que usa a linguagem de programação Python 3.7.

O método estatístico escolhido foi ANOVA, ou Análise de Variância, que analisa várias amostras de dados pertencentes a diferentes grupos em um nível de intervalo x , onde x é qualquer número entre 0 e 1. Essa probabilidade x determinará se os resultados irão gerar uma resposta que satisfaça a hipótese inicial proposta e não a rejeitamos, ou se mostra uma diferença expressiva e a rejeitamos através de métodos de estatísticas ANOVA.

A população amostral utilizada neste projeto são valores de acurácia geradas por um algoritmo de *Deep Learning* de classificação do Keras, o *fashion_mnist*, onde 4 tipos diferentes de arquitetura geraram 4 modelos treinados diferentes, cada um com 50 gerações do mesmo código. Como na arquitetura do modelo foi utilizado um fator aleatório, os algoritmos de aprendizado profundo nem sempre geram a mesma eficiência do modelo, sendo necessário gerar 50 gerações de cada, para comparar os resultados de cada um deles separadamente. Cada modelo tem um percentual de acurácia diferente, que é um número entre 0 e 100.

| | | | | | | | | | | | | | | | | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 91 | 92 | 91 | 92 | 91 | 91 | 92 | 89 | 92 | 90 | 91 | 91 | 87 | 93 | 90 | 90 | 91 | 92 | 93 | 91 | 93 | 93 | 90 | 93 | 93 |
| 91 | 89 | 91 | 88 | 87 | 91 | 92 | 91 | 89 | 92 | 91 | 93 | 91 | 91 | 90 | 91 | 88 | 92 | 91 | 93 | 90 | 94 | 88 | 91 | 91 |
| 91 | 92 | 87 | 92 | 90 | 91 | 90 | 93 | 94 | 92 | 92 | 88 | 93 | 91 | 91 | 92 | 91 | 90 | 88 | 92 | 90 | 89 | 95 | 93 | 91 |
| 89 | 89 | 91 | 89 | 90 | 91 | 89 | 92 | 93 | 90 | 89 | 89 | 92 | 90 | 91 | 88 | 93 | 89 | 92 | 93 | 92 | 90 | 92 | 94 | 89 |
| 90 | 92 | 91 | 93 | 93 | 92 | 92 | 90 | 90 | 93 | 89 | 91 | 93 | 90 | 88 | 91 | 92 | 91 | 91 | 90 | 92 | 91 | 91 | 91 | 90 |
| 88 | 89 | 91 | 90 | 93 | 92 | 90 | 90 | 91 | 92 | 90 | 92 | 88 | 89 | 93 | 89 | 92 | 90 | 91 | 90 | 90 | 92 | 88 | 92 | 91 |
| 93 | 90 | 90 | 89 | 90 | 89 | 90 | 88 | 89 | 93 | 90 | 88 | 91 | 92 | 91 | 94 | 94 | 91 | 92 | 92 | 89 | 91 | 92 | 94 | 90 |
| 91 | 90 | 88 | 92 | 91 | 90 | 90 | 93 | 90 | 93 | 90 | 89 | 91 | 94 | 91 | 91 | 89 | 84 | 90 | 89 | 89 | 93 | 91 | 91 | 91 |

Tabela 1 - Amostra de dados - acurácia da estrutura

³ https://colab.research.google.com/drive/1aiGwKWRFG5x8ZEeJQCxgBirQk_-sEOA?usp=sharing

Esses valores de precisão são os dados de amostra usados na análise ANOVA. A amostra obtida é exibida na Tabela 1.

3. RESULTADOS E DISCUSSÕES

Usando o método ANOVA e assumindo a Hipótese nula $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$, e H_1 : pelo menos um $\mu_i \neq \mu_j$, geramos os dados a um nível de significância de 0,05, um nível de significância aceitável para este tipo de código. Os resultados da análise são mostrados na Tabela 2, onde contém informações sobre todas as variáveis utilizadas no método ANOVA.

| ANOVA - Fator Único | | | | | | |
|---------------------|----------|------|-------------|---------------|------------|-----------|
| Alpha | 0,05 | | | | | |
| Grupos | Contagem | Soma | Média | Variância | | |
| Linha 1 | 50 | 4559 | 91,18 | 1,86489795918 | | |
| Linha 2 | 50 | 4529 | 90,58 | 2,53428571429 | | |
| Linha 3 | 50 | 4550 | 91 | 3,34693877551 | | |
| Linha 4 | 50 | 4527 | 90,54 | 3,31469387755 | | |
| Origem da Variação | SS | df | MS | F | P-Value | F-Crítico |
| Entre Grupos | 12,895 | 3 | 4,965 | 1,79552751024 | 0,14933498 | 2,6506765 |
| Dentre Grupos | 541,98 | 196 | 2,765204082 | | | |
| Total | 556,875 | 199 | | | | |

Tabela 2 - Análise ANOVA

Ao analisar os dados obtidos nos resultados, podemos ver que a variância não é um valor alto em relação à diferença de valores, sendo de 0 a 100, indicando que há uma grande chance de não rejeitamos a H_0 . Isso acontece porque podemos ter uma ideia geral dos resultados comparando os dados de maior variância com os menores. Se o maior valor é mais que duas vezes o valor do menor, há uma grande chance de rejeitarmos a H_0 . Neste caso, o menor é 1,86 e o maior é 3,31. Para suspeitar da rejeição de H_0 , o maior valor deve ser pelo menos 3,72. Para uma análise mais avançada, após calcular todas as variáveis, a variável que temos que verificar é o P-Value. Neste caso, o P-Value é 0,149, sendo maior que o valor alfa, o que nos leva a conclusão de não rejeitar H_0 . Já que o P-Value é quase três vezes o valor alfa, temos uma alta correlação entre eles, mostrando que precisaria de um valor alfa muito menor para rejeitá-lo.

No cenário em que rejeitamos a hipótese H_0 , poderíamos obter mais informações usando o método de Tukey, como mostra a eq. 1, o qual analisando todos os diferentes tipos de classificação

modelos, podemos descobrir qual deles não suporta H_0 : a avaliação de precisão é igual, e por consequência podemos determinar se o que não é igual apresenta uma melhor métrica de resultados ou uma pior, demonstrando valores métricos dos modelos e definindo qual deles deve ser utilizado.

$$q_{tukey} = \frac{|x_1 - x_2|}{\sqrt{MSE/n_{group}}} \quad (1)$$

4. CONCLUSÕES

No caso, no projeto desenvolvido não podemos determinar totalmente qual dos 4 construídos modelos é o melhor, porque a um nível de significância de 5% não podemos rejeitar H_0 , então todos eles são considerados como tendo a mesma eficiência, podendo utilizar qualquer um deles e o resultado será considerado semelhante. Se H_0 fosse rejeitado seria aplicado o método de Tukey, assim determinando qual deles tem a melhor eficácia para o seu propósito.

AGRADECIMENTOS

Agradecemos ao LABSOFT e ao IFSULDEMINAS - Edital 33/2021 - mobilidade acadêmica Canadá - College of the Rockies, pela oportunidade de intercâmbio e estrutura concedidos para a realização desta pesquisa.

REFERÊNCIAS

CAMPAGNER, Andrea; SCONFENZA, Luca; CABITZA, Federico. H-accuracy, an alternative metric to assess classification models in medicine. In: Digital Personalized Health and Medicine. IOS Press, 2020, Volume 270, p. 242-246.

EMMERT-STREIB, Frank; YANG, Zhen; FENG, Han; TRIPATHI, Shailesh; MATTHIAS, Dehmer. An introductory review of deep learning for prediction models with big data. *Frontiers in Artificial Intelligence*, Volume 3, p. 4, 2020.

HARADA, Yukinori; KATSUKURA, Shinichi; KAWAMURA, Ren; SHIMIZU, Taro. Efficacy of artificial-intelligence-driven differential-diagnosis list on the diagnostic accuracy of physicians: An open-label randomized controlled study. *International journal of environmental research and public health*, Volume 18, n. 4, p. 2086, 2021.

SICHMAN, Jaime Simão. Inteligência Artificial e sociedade: avanços e riscos. *Estudos Avançados*, Volume 35, p. 37-50, 2021.