

Relatório do Laboratório 7:

Redes Neurais

Isabelle Ferreira de Oliveira
CT-213 - Engenharia da Computação 2020
Instituto Tecnológico de Aeronáutica (ITA)
São José dos Campos, Brasil
isabelle.ferreira3000@gmail.com

Resumo—Esse relatório documenta a implementação de uma rede neural de duas camadas para realizar a segmentação de cores para o futebol de robôs. Para isso, foi necessário configurar essa rede neural para realizar classificação multi-classe, implementando os algoritmos de Forward Propagation (inferência) e Back Propagation (treinamento) para essa rede.

Index Terms—Redes neurais, segmentação de cores, Forward Propagation, Back Propagation

I. INTRODUÇÃO

Redes neurais são sistemas de computação que podem reconhecer padrões escondidos, agrupar dados e classificá-los, além de, com o tempo, aprender e melhorar continuamente [3].

Uma rede neural é formada por camadas de neurônios artificiais (baseado no sistema nervoso humano e neurônios naturais), e eles são os responsáveis por realizar as contas que levam de uma entrada a uma saída esperada (para o caso de aprendizado supervisionado). A Figura 1 apresenta um exemplo de rede neural com duas camadas (a camada de entrada geralmente não é considerada nesse tipo de contagem). É possível notar no exemplo que a camada 1 tem 3 neurônios e a camada 2 tem 1 neurônio.

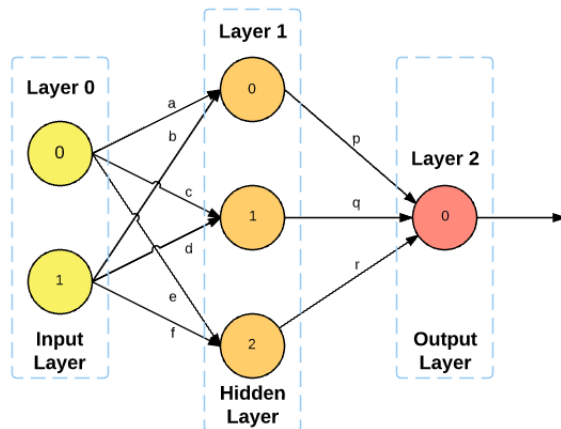


Figura 1. Exemplo de rede neural, com duas camadas (1 camada de entrada, 1 camada escondida e 1 camada de saída), como a trabalhada nesse laboratório. Essa imagem de exemplo foi apresentada no site [2]

Agora analisando um neurônio em particular, como o apresentado na Figura 2, a conta realizada para o aprendizado é que

se encontra no interior do círculo, ou seja: $g(\sum w_j x_j + b) = y$, no qual w são os pesos; b , os bias; x , as entradas; e g , a função de ativação. O objetivo do treinamento é ajustar w e b para aproximar alguma função $y(x)$ de acordo com os valores fornecidos como *outputs* esperados.

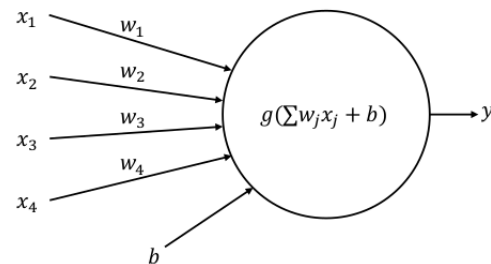


Figura 2. Exemplo de neurônio. Essa imagem de exemplo foi apresentada no roteiro [1]

O processo de inferência é realizado por meio do cálculo da equação apresentada no círculo do neurônio da Figura 2, em um algoritmo chamado *Forward Propagation*. Já o processo de aprendizado é realizado pelo algoritmo de *Back Propagation*, que calcula os pesos e *biases* a partir das saídas esperadas e das decidas de gradientes calculados para esses pesos e *bias*. Já esses gradientes são calculados por meio das saídas esperadas, as entradas iniciais e os resultados intermediários calculados por meio do *Forward Propagation*.

Os pseudo-códigos geral dos algoritmos *Forward Propagation*, *Back Propagation* e o cálculo das decidas de gradientes para pesos e *biases* podem ser visto a seguir. Em seguida, será apresentado como esses algoritmos foram implementados no contexto do laboratório.

```
# Forward Propagation
def forward_propagation(input):
    z[1] = weights[1]input + biases[1]
    a[1] = g[1](z[1])
    z[2] = weights[2]a[1] + biases[2]
    a[2] = g[2](z[2])
    return z, a
```

```
# Gradients' Computation
```

```

def compute_gradient_back_propagation(inputs,
    expected_outputs):
    for i in range(len(inputs)):
        input = inputs[i]
        expected_output = expected_outputs[i]

        z, a = forward_propagation(input)

        dz[2] = a[2] - expected_output
        weights_gradient[2] += dz[2] * a[1].T
        biases_gradient[2] += dz[2]

        dz[1] = weights[2].T * dz[2] * g[1]'(z[1])
        weights_gradient[1] += dz[1] * input.T
        biases_gradient[1] += dz[1]

    return weights_gradient, biases_gradient

```

```

# Back Propagation
def back_propagation(inputs,
    expected_outputs):
    weights_gradient, biases_gradient =
        compute_gradient_back_propagation(inputs,
            expected_outputs)

    weights[1] -= alpha * weights_gradient[1]
    biases[1] -= alpha * biases_gradient[1]
    weights[2] -= alpha * weights_gradient[2]
    biases[2] -= alpha * biases_gradient[2]

```

No pseudocódigo acima, J é a função para medir a qualidade das soluções candidatas; $m0$ e $C0$ são a média e a matriz de covariância iniciais da população que será gerada, respectivamente; m e C são, de forma análoga, a média e a matriz de covariância atuais da população que será gerada, respectivamente; e μ é o tamanho da população considerada como "melhores soluções até então".

II. IMPLEMENTAÇÃO DO ALGORITMO

Para a implementação da rede neural, era necessário preencher a função *forward_propagation()*, *compute_gradient_back_propagation()* e *back_propagation()* da classe *NeuralNetwork*. Essas funções a se completar estavam no código base fornecido [1].

Recebendo os valores de *input*, a função *forward_propagation()* foi implementada conforme apresentado no pseudo-código da Introdução, com o detalhe que a função de ativação g de todos os neurônios utilizada foi a *sigmoid*. Além disso, tanto a função *sigmoid* quanto sua derivada já eram fornecidos pelo código base.

Já na implementação de *back_propagation()*, também foi implementado conforme o sugerido no pseudo-código da Introdução, ou seja, atualizando os valores de pesos e *biases* subtraindo de seu valor um fator de aprendizado α vezes os gradientes de pesos e *biases*, respectivamente. Esses valores de gradientes são aqueles calculados em *compute_gradient_back_propagation()*, supondo que já foi corretamente implementado. Foi utilizado para α o mesmo que o sugerido pelo código base.

Além disso, era necessário comparar os desempenhos desse algoritmo SES com o algoritmo CMA-ES, já fornecido no código base, aplicando-os na otimização de quatro funções, a saber: esfera transladada, e as funções de Ackley, Schaffer 2 e Rastrigin 2D.

A análise de vários pontos do algoritmo descrito acima terá uma breve descrição em alto nível da sua implementação a seguir.

Primeiramente, foi necessário ordenar a população atual tendo em vista os valores de *fitness* associados a ela recebido por parâmetro. Isso foi feito utilizando a função *argsort()* da biblioteca *Numpy*, conforme sugerido pelo roteiro do laboratório.

Após isso, as μ melhores amostras dessa população foram separadas para ajudar no cálculo de sua matriz de covariância e, posteriormente, realizar o cálculo da nova média das melhores amostras dessa geração em questão. Essa matriz de covariância foi feita a partir da multiplicação de uma determinada matriz auxiliar pela sua transposta, sendo essa matriz auxiliar a diferença entre essas melhores amostras e a média encontrada na população anterior.

Dessa forma, tendo em posse a nova matriz de covariância e a nova média das melhores amostras da população anterior, é possível gerar uma nova população, evoluindo para a próxima geração de possíveis candidatas a solução.

Para testar o funcionamento dessa implementação, foram alterados os valores das variáveis *algorithm* (entre "ses" e "cmaes") e *function* (entre as funções *translated_sphere*, *ackley*, *schaffer2d*, *rastrigin*) no arquivo *test_evolution_strategy.py* do código base, gerando imagens dos resultados da otimização de cada uma dessas funções usando a estratégia evolutiva escolhida.

Por fim, a fim de realizar o *benchmark* através de simulações de Monte Carlo para comparar os desempenhos dessa estratégia evolutiva simples e do CMA-ES, foram alterados novamente os valores das variáveis *algorithm* e *function*, dessa vez no arquivo *benchmark_evolution_strategy.py* do código base, gerando dessa vez os gráficos comparativos de rendimento para diferentes situações de SES e CMA-ES. Esses gráficos foram apresentados nas Figuras ?? a ??.

III. RESULTADOS E CONCLUSÕES

A. Teste das Estratégias Evolutivas

A otimização para testar a implementação foi executada para as quatro funções evolutivas já citadas na seção anterior, cuja equação matemática se encontra no roteiro do laboratório [?]. Os resultados dessas execuções foram satisfatórios e saíram conforme o esperado, comprovando o correto funcionamento da implementação e a validade da utilização das estratégias evolutivas na otimização de funções. Esses resultados foram apresentados nas Figuras de 9 a ??.

Vale reparar que os resultados de ambas implementações foram bastante equivalentes, com exceção do caso da função Rastrigin, no qual cada algoritmo convergiu em mínimos locais diferentes.

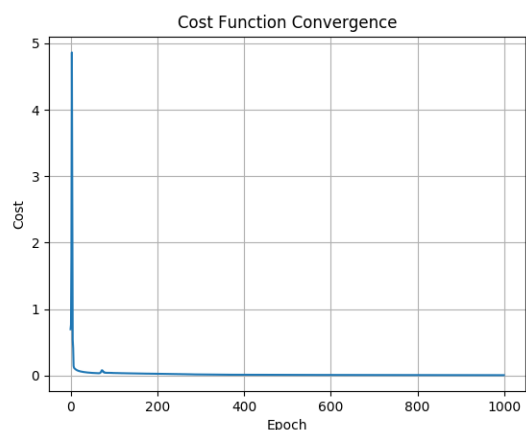


Figura 3. Otimização da função de Esfera Transladada usando a estratégia evolutiva SES. O resultado encontrado é o ponto vermelho.

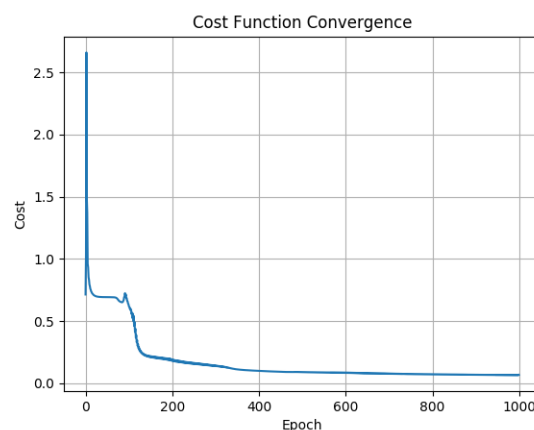


Figura 6. Otimização da função de Esfera Transladada usando a estratégia evolutiva SES. O resultado encontrado é o ponto vermelho.

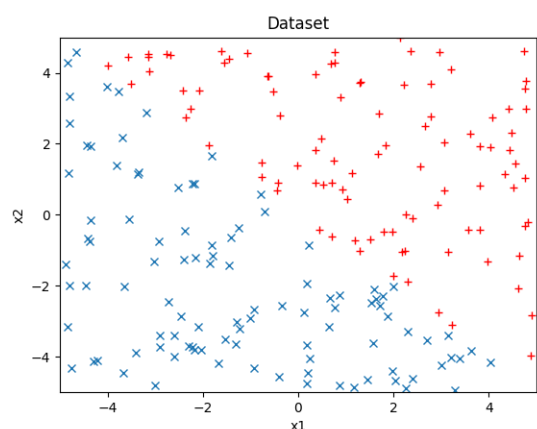


Figura 4. Otimização da função de Esfera Transladada usando a estratégia evolutiva CMA-ES. O resultado encontrado é o ponto vermelho.

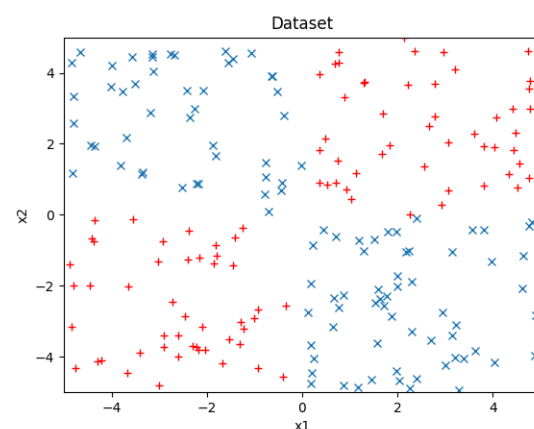


Figura 7. Otimização da função de Esfera Transladada usando a estratégia evolutiva CMA-ES. O resultado encontrado é o ponto vermelho.

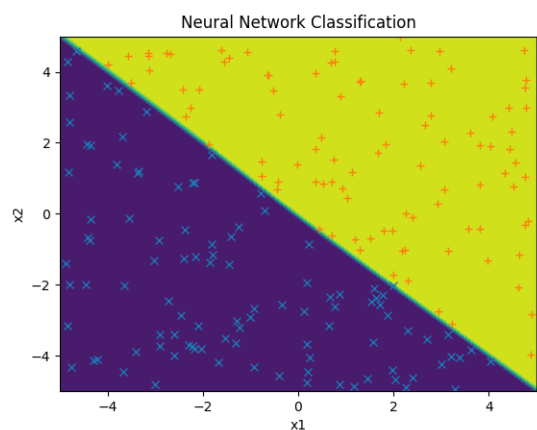


Figura 5. Otimização da função de Ackley usando a estratégia evolutiva SES. O resultado encontrado é o ponto vermelho.

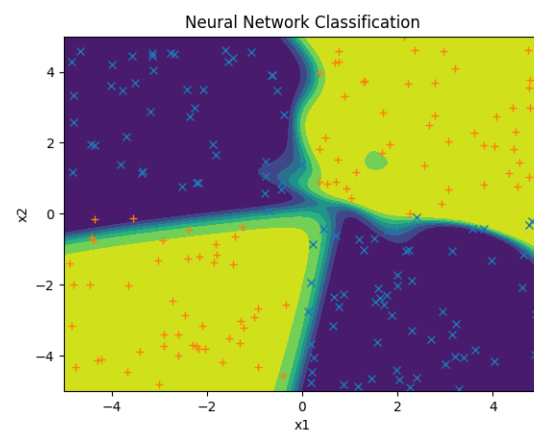


Figura 8. Otimização da função de Ackley usando a estratégia evolutiva SES. O resultado encontrado é o ponto vermelho.

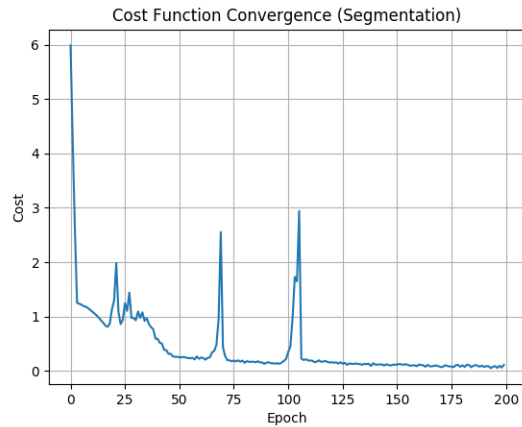


Figura 9. Otimização da função de Esfera Transladada usando a estratégia evolutiva SES. O resultado encontrado é o ponto vermelho.

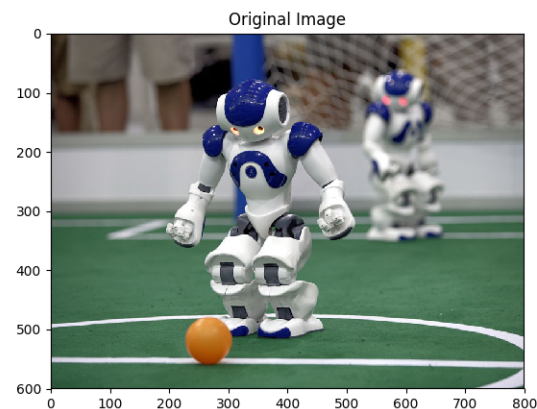


Figura 10. Otimização da função de Esfera Transladada usando a estratégia evolutiva CMA-ES. O resultado encontrado é o ponto vermelho.

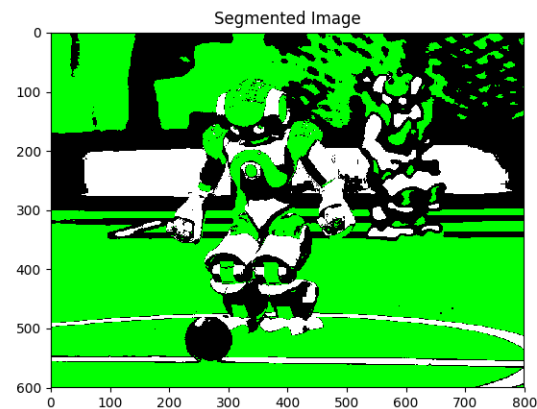


Figura 11. Otimização da função de Ackley usando a estratégia evolutiva SES. O resultado encontrado é o ponto vermelho.

B. Benchmark das Estratégias Evolutivas

A otimização para realizar o *benchmark* do desempenho de cada um dos métodos evolutivos estudados (SES e CMA-ES) para diferentes funções e parâmetros foi executada para as quatro funções evolutivas já citadas. Os resultados dessas execuções demonstraram comportamentos coerentes e foram apresentados nas Figuras de 9 a ??.

É possível salientar que, embora sempre tenha havido convergências após um número significativo de iterações, nem todas as convergências chegaram a mínimos locais, como por exemplo para a função da Esfera Transladada, que só possui um mínimo local (que também é global), mas cada situação testada chegou a valores diferentes de *fitness*.

Além disso, notou-se que, para estratégias evolutivas mais simples, é necessário um número cada vez maior de elementos na população para que os resultados se tornem cada vez mais otimizados. Já para o algoritmo CMA-ES precisou de cerca de 1/4 de população para alcançar resultados similares aos do SES. Isso aconteceu para todos os casos com exceção à função Schaffer N° 2, que estratégias mais simples e com menores populações se saíram melhor em desempenho do que CMA-ES.

Tendo em vista o que foi apresentado, pode-se notar, por fim, que esses algoritmos realmente se demonstraram eficazes em encontrar parâmetros otimizados para uma determinada função.

REFERÊNCIAS

- [1] M. Maximo, "Roteiro: Laboratório 5 - Estratégias Evolutivas". Instituto Tecnológico de Aeronáutica, Departamento de Computação. CT-213, 2019.
- [2] Towards Data Science, "Neural Net from scratch". Acessado em <https://towardsdatascience.com/neural-net-from-scratch-using-numpy-71a31f6e3675>.
- [3] SAS, "Redes Neurais: O que são e qual a sua importância?". Acessado em https://www.sas.com/pt_br/insights/analytics/neural-networks.html.