



Universidade de Brasília

DEPARTAMENTO DE ESTATÍSTICA

16 de junho de 2023

Lista 3: Otimização em RNA.

Prof. Guilherme Rodrigues

Redes Neurais Profundas

Tópicos especiais em Estatística 1

- (A) As questões deverão ser respondidas em um único relatório *PDF* ou *html*, produzido usando as funcionalidades do *Rmarkdown* ou outra ferramenta equivalente.
- (B) O aluno poderá consultar materiais relevantes disponíveis na internet, tais como livros, *blogs* e artigos.
- (C) O trabalho é individual. Suspeitas de plágio e compartilhamento de soluções serão tratadas com rigor.
- (D) Os códigos *R* utilizados devem ser disponibilizados na íntegra, seja no corpo do texto ou como anexo.
- (E) O aluno deverá enviar o trabalho até a data especificada na plataforma Aprender 3.
- (F) O trabalho será avaliado considerando o nível de qualidade do relatório, o que inclui a precisão das respostas, a pertinência das soluções encontradas, a formatação adotada, dentre outros aspectos correlatos.
- (G) Escreva seu código com esmero, evitando operações redundantes, comentando os resultados e usando as melhores práticas em programação.

Considere a função

$$f(x_1, x_2) = x_1^4 + x_2^4 + x_1^2 x_2 + x_1 x_2^2 - 20x_1^2 - 15x_2^2$$

para responder os itens a seguir.

a) Apresente um gráfico com as curvas de nível de $f(x_1, x_2)$. Quantos pontos críticos a função parecer? Dica para usuários do R: use a função `geom_contour_filled()`.

b) Encontre (algericamente) o gradiente de f em relação ao vetor $x = (x_1, x_2)$. Isso é,

$$\nabla_x f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right).$$

c) Crie uma função computacional que implemente o método do gradiente para minimizar a função em estudo. Permita ao usuário definir a taxa de aprendizado, o número de passos e o ponto de partida.

d) Use a função criada no item c) para encontrar o valor obtido pelo método do gradiente partindo-se do ponto inicial $(x_1^{(0)}, x_2^{(0)}) = (0, 5)$. Use taxa de aprendizado igual a 0.01 e execute 100 passos.

e) Repita o item d), agora com as seguintes taxas de aprendizado: 1, 0.1, 0.01, 0.001, 0.0001. Qual dessas opções lhe parece mais apropriada nesse caso? Justifique sua resposta.

f) Fixe a semente do gerador de números aleatórios no valor 123 (se estiver usando o R, basta executar o código `set.seed(123)`). Repita novamente o item d), agora partindo de 20 pontos escolhidos aleatoriamente (uniformemente) no quadrado $-5 < x_1, x_2 < 5$. Refaça o gráfico do item a) e adicione uma linha representando o caminho percorrido por cada uma das 20 otimizações. Qual foi o percentual de vezes em que o algoritmo encontrou o mínimo global da função (despresando um eventual desvio de menor importância)?

g) Repita o item d), substituindo o método do gradiente pelo método do gradiente com momento (veja a Seção 8.3.2 do livro *Deep Learning*). Use taxa de aprendizado $\epsilon = 0.01$, parâmetro de momento $\alpha = 0.9$ e velocidade inicial $v = 0$.

h) Repita o item d), substituindo o método do gradiente pelo método RMSProp (veja a Seção 8.5.2 do livro *Deep Learning*). Use taxa de aprendizado $\epsilon = 0.001$, taxa de decaimento $\rho = 0.9$ e constante $\delta = 10^{-6}$.

i) Repita o item d), substituindo o método do gradiente pelo método ADAM (veja a Seção 8.5.3 do livro *Deep Learning*). Use taxa de aprendizado $\epsilon = 0.001$ e taxas de decaimento $\rho_1 = 0.9$ e $\rho_2 = 0.999$.

j) Apresente graficamente, em uma única figura, os caminhos percorridos pelas otimizações executadas nos itens d), g), h) e i).