UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
DISCIPLINA: REDES NEURAIS - DEEP LEARNING
Iª LISTA DE EXERCÍCIO) - 2021.2
Data de entrega da lista com apresentação: 21/09/2021

Objetivos da lista: consolidar os conceitos de: redes neurais, processos de aprendizagem de máquinas e estimular uma revisão dos conceitos de álgebra linear, cálculo de múltiplas variáveis, métodos de otimização e os fundamentos de aprendizagem de máquinas.

1-) Considere o problema de estimativa da matriz de covariância de uma dada distribuição de vetores aleatórios dados por $\mathbf{x_1} = \begin{bmatrix} 7 \\ 3 \\ 9 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 4 \\ 6 \\ 11 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 4 \\ 2 \\ 5 \end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix} 5 \\ 5 \\ 7 \end{bmatrix}$. Estime o vetor média (média amostral) e a matriz de covariância amostral $\mathbf{S} = \{s_{ij}\}$ assumindo que a matriz é simétrica. Observando que a média amostral é dada por $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$ e a variância amostral é dada por $s_{ij} = \frac{\sum_{l=1}^{N}(x_{il} - \hat{\mu}_i)(x_{jl} - \hat{\mu}_j)}{N-1}$. Calcule também a partir da matriz de covariância obtida:

a-) Os autovalores e autovetores da matriz e mostre a forma fatorada da matriz isto é $\mathbf{S} = \mathbf{V}^{-1}\mathbf{D}\mathbf{V}$. Onde $\mathbf{D}$ é a matriz diagonal dos autovalores e $\mathbf{V}$ é a matriz dos autovetores correspondentes.

b-) Calcule fazendo uso da forma fatorada $\mathbf{S}^4$

c-) Calcule a norma Euclidiana da matriz e compare com seu valor estimado pelo traço da matriz.

d-) Calcule a distancia Euclidiana dos vetores aleatórios.

e-) Calcule a distancia de Mahalanobis dos vetores aleatórios e compare com a distancia Euclidiana

2) Considere a função $E(\mathbf{w})$ onde $\mathbf{w} = [w_1, w_2, \ldots, w_n]^t$ é um vetor com múltiplas variáveis. Usando a expansão em série de Taylor a função pode ser expressa como $E(\mathbf{w}(n) + \mathbf{\Delta w}(n)) = E(\mathbf{w}(n)) + \mathbf{g}^t(\mathbf{w}(n))\mathbf{\Delta w}(n) + \frac{1}{2}\mathbf{\Delta w}^t(n)\mathbf{H}(\mathbf{w}(n))\mathbf{\Delta w}(n) + O(\|\mathbf{\Delta w}\|^3)$, onde $\mathbf{g}(\mathbf{w}(n))$ é o vetor gradiente local definido por $\mathbf{g}(\mathbf{w}) = \frac{\partial E(\mathbf{w})}{\partial \mathbf{w}}$ e $\mathbf{H}(\mathbf{w})$ é matriz Hessiana, definida por $\mathbf{H}(\mathbf{w}) = \frac{\partial^2 E(\mathbf{w})}{\partial \mathbf{w}^2}$.

Demonstre:

a-) que o método do gradiente da descida mais íngreme é dado por

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \eta \mathbf{g}(\mathbf{w}(n))$$

b-) que o método de Newton é dado por

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mathbf{H}^{-1}(\mathbf{w}(n))\mathbf{g}(\mathbf{w}(n))$$

e indique que condição a matriz Hessiana $\mathbf{H}(\mathbf{w}) = \dfrac{\partial^2 E(\mathbf{w})}{\partial \mathbf{w}^2}$ precisa atender para o método poder ser aplicado.

3-) Considere a função Rosenbrock com duas variáveis
f (x, y) = (1 -x)^2 + 100 (y - x^2)^2 que é uma função comumente utilizada para avaliar o desempenho de um algoritmo de otimização. Seu ponto de mínimo é (1,1). Utilize os métodos do gradiente da descida mais íngreme assim como o método de Newton para solução numérica do cálculo do mínimo. Avalie o desempenho de cada um dos métodos.

4-)Considere o problema de otimização com restrições:

$$\begin{aligned} &\text{minimize } f(\mathbf{x}) = 2x_1^2 - 2x_1x_2 + 4x_2^2 + 5 \\ &\text{sujeito a:} \\ &g_1(x_1, x_2) = x_1 + x_2 = 0 \end{aligned}$$

formule o problema usando o funcional de Lagrange e calcule os valores (x_1^*, x_2^*) para mínimo da função.

5-) Apresente um estudo sobre o método do gradiente estocástico (SGM).

6-) O modelo de neurônio artificial de Mc-Culloch-Pitts faz uso da função de ativação para resposta do neurônio artificial. As funções sigmoíde (ou função logística) e tangente hiperbólica (tangsigmoíde) são normalmente utilizadas nas redes neurais perceptrons de múltiplas camadas tradicionais. A função ReLu (retificador linear) é normalmente utilizadas em redes Deep Learning. Segue abaixo as expressões matemáticas de cada uma:
a-) $\varphi(v) = \dfrac{1}{1+\exp(-av)}$ (sigmoíde); b-) $\varphi(v) = \dfrac{1-\exp(-av)}{1+\exp(-av)} = \tanh(\dfrac{av}{2})$ (tangsigmoíde) ;
c-) $\varphi(v) = \max(0, v)$ (Re-Lu).

i) Faça uma análise comparativa de cada uma destas funções apresentando de forma gráfica a variação da função e da sua derivada com relação a v (potencial de ativação)
ii) Mostre que $\varphi'(v) = \dfrac{d\varphi(v)}{dv} = a\varphi(v)[1-\varphi(v)]$ para função sigmoíde.
iii) Mostre que $\varphi'(v) = \dfrac{d\varphi(v)}{dv} = \dfrac{a}{2}[1-\varphi^2(v)]$ para função tangsigmoíde.

7-) Dada a função E(n) que corresponde a função custo com base no erro ao quadrado, calculada na saída do neurônio j de uma rede neural perceptron de múltiplas camadas

$\mathrm{E}(\mathrm{n}) = \frac{1}{2}\mathrm{e}_{\mathrm{j}}^{2}(n)$ *com*

$e_j(n) = d_j(n) - y_j(n)$,

$y_j(n) = \varphi(v_j(n)); \quad v_j(n) = \sum_{i=0}^{m} w_{ji}(n) y_i(n)$

i) Demonstre que $\dfrac{\partial \mathrm{E(n)}}{\partial \mathrm{w_{ji}(n)}} = -e_j(n)\varphi'(v_j(n))y_i(n)$

ii) Considerando as funções de ativação da questão anterior apresente a expressão $\dfrac{\partial \mathrm{E(n)}}{\partial \mathrm{w_{ji}(n)}} = -e_j(n)\varphi'(v_j(n))y_i(n)$

8-) O método da máxima verossimilhança (Max-Likelihood) aplicado na determinação de parâmetros desconhecidos de um modelo de distribuição de probabilidades. O método é bastante utilizado para o desenvolvimento de algoritmos de aprendizagem de máquinas em particular em redes neurais. O problema considerado nesta questão envolve o modelo probabilístico de um comitê de redes especialistas e uma rede que indica qual das redes especialista é a mais adequada para processar uma determinada entrada. O problema consiste em determinar os parâmetros **w** das redes especialistas e **a** da rede que decide qual a rede especialista deve processar um determinada entrada. O modelo probabilístico é denominado de modelo de misturas de gaussianas, dado por

$$f(\mathbf{d} \mid \mathbf{x}; \mathbf{w}, \mathbf{a}) = \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{K}} g_i f(\mathbf{d} \mid \mathbf{x}, \mathbf{i}; \mathbf{w}, \mathbf{a}) = \frac{1}{(2\pi)^{q/2}} \sum_{i=1}^{K} g_i \exp(-\frac{1}{2}\left\|\mathbf{d} - \mathbf{y}_i\right\|^2)$$

$$y_i^{(m)} = \mathbf{x}^{\mathbf{t}}\mathbf{w}_{\mathbf{i}}^{(\mathbf{m})}, \quad \begin{cases} i = 1,2,\ldots,K \\ m = 1,2\ldots,q \end{cases}$$

$$g_i = \frac{\exp(u_i)}{\sum_{j=1}^{K}\exp(u_j)} \qquad 0 \le g_1 \le 1 \qquad \sum_{i=1}^{K} g_i = 1$$

$$u_i = \mathbf{x}^{\mathbf{t}}\mathbf{a}$$

sendo **d** vetor de resposta desejada, $\mathbf{y_i}$ vetor de saída de cada rede especialista, $\mathbf{w_i}$ o vetor de pesos de cada rede especialista, **a** o vetor de pesos da rede decisora, q a dimensão do vetor de entrada e do vetor de pesos e K o número de redes especialistas . A função $g_i = \dfrac{\exp(u_i)}{\sum_{j=1}^{K}\exp(u_j)}$ é conhecida como função sofmax e corresponde neste problema a probabilidade a priori da rede i ser escolhida. A função h_i definida abaixo corresponde a probabilidade da posteriori da rede ser escolhida.

Definindo função log de verossimilhança como

$$l(\mathbf{w},\mathbf{a}) = \ln \sum_{i=1}^{K} g_i \exp(-\frac{1}{2}\|\mathbf{d}-\mathbf{y}_i\|^2)$$

o problema consiste em com relação aos parâmetros **w** e **a. F**azendo uso do método do gradiente da descida mais íngreme para determinação max l(**w,a).**

mostre que

$$\mathbf{w}_\mathbf{i}^{(\mathbf{m})}(n+1) = \mathbf{w}_\mathbf{i}^{(\mathbf{m})}(n) + \eta h_i(n) e_i^{(m)}(n)\mathbf{x}$$
$$\mathbf{a}_\mathbf{i}(n+1) = \mathbf{a}_\mathbf{i}(n) + \mu[h_i(n) - g_i(n)]$$
$$e_i^{(m)} = d^{(m)} - y_i^{(m)}(n)$$
$$h_i(n) = \frac{g_i(n)\exp(-\frac{1}{2}\|\mathbf{d}-\mathbf{y}_\mathbf{i}(n)\|^2)}{\sum_{j=1}^{K} g_j(n)\exp(-\frac{1}{2}\|\mathbf{d}-\mathbf{y}_j(n)\|^2)} \qquad 1 \le h_i(n) \le 1 \qquad \sum_{i=1}^{K} h_i(n) = 1$$

9-) Implementações Computacionais de Redes Neurais.

Para cada uma das aplicações abaixo apresente a curva do erro de treinamento e de validação:

9.1-) Defina a estrutura de uma rede perceptron de múltiplas camadas para aproximar as funções abaixo. Defina o conjunto de treinamento e de validação. Compare para os itens b e c a aproximação da função obtida pela rede neural com as curvas exatas. Apresente para cada caso a curva do erro médio de treinamento com relação ao número de épocas e a curva do erro médio com o conjunto de validação.

a) a função lógica $f(x_1, x_2, x_3) = x_1 \oplus x_2 \oplus x_3$

b) a função real

$$f(x_1, x_2) = \left(\frac{\cos(2\pi x_1)}{1-(4x_1)^2} sen(\pi x_1)/\pi x_1\right)\left(\frac{\cos(2\pi x_2)}{1-(4x_2)^2} sen(\pi x_2)/\pi x_2\right) \qquad -4\pi \le x_1 \le 4\pi \qquad -4\pi \le x_2 \le 4\pi$$

9.2-) Utilize a rede neural perceptron de múltiplas camadas para fazer a predição de um passo (x^(n+1)) da série temporal $x(n) = 1 + \cos(n + \cos^2(n))$. Avalie o desempenho mostrando a curva a série temporal, a curva de predição o e curva do erro de predição.

9.3-) Considere o problema de classificação de padrões bidimensionais constituído neste caso de 2 padrões. A distribuição dos padrões tem como base um quadrado centrado na origem interceptando os eixos nos pontos +1 e -1 de cada eixo. Os pontos +1 e -1 de cada eixo são centros de quatro semicírculos que se interceptam no interior do quadrado originando uma classe e a outra classe corresponde as regiões de não interseção. Após gerar aleatoriamente os dados que venham formar estas distribuições de dados, selecione um conjunto de treinamento e um conjunto de validação. Solucione este problema considerando uma rede perceptron de múltiplas camada

10 - Trabalho::

Apresente um estudo com uma análise comparativa dos algoritmos: **AdaGrad, RMSProp e Adam, que tem como base o** método do gradiente estocástico (SGM) e são utilizados no processo de aprendizagem de redes neurais deep learning.

Data de entrega: 21/09/2021

A entrega e apresentação dos trabalhos correspondem a um processo de avaliação. Portanto a presença é obrigatória.

O trabalho e a lista podem ser feitos em grupo de até três componentes.

Na apresentação os componentes poderão ser submetidos a questionamentos sobre a solução da lista e o desenvolvimento dos trabalhos.