UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
DISCIPLINA: REDES NEURAIS - DEEP LEARNING
Iª LISTA DE EXERCÍCIO) - 2025.1
Data de entrega da lista com apresentação: 24/04/2023

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

1-) As métricas de distancia entre vetores são aplicadas nos estudos de aprendizagem de máquina como medidas de similaridade/dissimilaridade entre vetores que representam padrões (atributos). Apresente um estudo sobre as seguintes distancias: Distancia Euclidiana, Distancia de Minkowski, Distancia City Block, Distancia de Mahalanobis, Coeficiente de Correlação de Pearson e Similaridade Cosseno. Apresente neste estudo aplicações onde cada tipo de métricas de distância é mais adequada.

Sugestão de Aplicações: Classificação de padrões, Clustering, Reconhecimento de padrões, Modelos de Linguagem Natural, etc.

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 $\mathbf{g}(\mathbf{w}) = \dfrac{\partial E(\mathbf{w})}{\partial \mathbf{w}}$ e $\mathbf{H}(\mathbf{w})$ é matriz Hessiana, definida por $\mathbf{H}(\mathbf{w}) = \dfrac{\partial^2 E(\mathbf{w})}{\partial \mathbf{w}^2}$.

Demonstre com base na expansão em série de Taylor:

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

$w(n+1) = w(n) - \eta g(w(n))$. Onde η é o coeficiente de aprendizagem ou passo do método

b-) que o método de Newton é dado por

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

c-) que o η^* ótimo é dado por $\mathbf{g}(\mathbf{w})\mathbf{g}^t(\mathbf{w})/\mathbf{g}^t(\mathbf{w})\mathbf{H}(\mathbf{w})\mathbf{g}(\mathbf{w})$, assumindo a matriz $\mathbf{H}$ definida positiva

Considere também as seguintes questoões:

d-) Qual será o valor de η sob as condições acima se vetor gradiente estiver alinhado com a autovetor correspondente ao autovalor máximo (λ_{max}).

e) Em um determinado ponto **w o** gradiente da função E(**w**) é igual zero, isto é, **g(w)=0.** Isto não é suficiente para indicar que este ponto corresponde a um ponto de mínimo. Pode ser um ponto de máximo ou um ponto de sela. Com base nos autovalores da matriz Hessiana, definida como $\mathbf{H}(\mathbf{w}) = \dfrac{\partial^2 E(\mathbf{w})}{\partial \mathbf{w}^2}$, apresente a condição para este ponto ser:

(i) um ponto de mínimo, (ii) um ponto de máximo (iii) um ponto de sela da função.

3-) Apresente um estudo comparando os seguintes algoritmos de otimização: **Gradiente Estocástico (SGM)**, **AdaGrad, RMSProp e Adam.** Estes métodos ou otimizadores são utilizados no processo de aprendizagem de redes neurais/deep learning.

4-) O modelo de neurônio artificial de Mc-Culloch-Pitts faz uso da função de ativação para resposta do neurônio artificial. A função sigmoíde (ou função logística) e a função tangente hiperbólica (ou tangsigmoíde) são normalmente utilizadas nas camadas ocultas das redes neurais perceptrons de múltiplas camadas tradicionais (uma ou duas camadas ocultas - shallow network). A função ReLu (retificador linear) é normalmente utilizadas nas camadas ocultas das redes Deep Learning. Segue abaixo as expressões matemáticas de cada uma:

a-) $\varphi(v) = \dfrac{1}{1+\exp(-av)}$ (sigmoíde); b-) $\varphi(v) = \dfrac{1-\exp(-av)}{1+\exp(-av)} = \tanh(\dfrac{av}{2})$ (tangente hiperbólica ou 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.

b-) As funções de saída das redes neurais dependem do modelo probabilístico que se busca gerar com a rede. Faça uma análise sobre a escolhas destas funções considerando os seguintes problemas:

(i) Classificação de padrões com duas classes.
(ii) Classificação de padrões com múltiplas classes.
(iii) Problema de regressão (aproximação de funções)

c-) Para cada uma das condições do item (b) apresente a função custo a ser considerada no processo de treinamento de uma rede neural com múltiplas camadas em um processo de aprendizagem supervisionada.

4-) O algoritmo backpropagation é o algoritmo base no processo de treinamento de redes neurais do tipo: rede perceptron de múltiplas camadas (MLP), rede convolucional (CNN), rede recorrente (ex: LSTM).

(i) Apresente a dedução do algoritmo para uma rede perceptron de múltiplas camadas com múltiplas variáveis de entradas e múltiplas variáveis de saída.
(ii) Apresente o pseudo código do algoritmo
(iv) Pesquise apresente um estudo da implementação computacional do algoritmo backpropagation para deep learning fazendo uso de tensores e computação gráfica.

5-) Implementações Computacionais de Redes Neurais.

Para cada um dos problemas abaixo apresente a solução fazendo uso de implementações computacionais. Apresente na solução a curva do erro de treinamento e o erro de validação:

5.1-) Defina a estrutura de uma rede perceptron de múltiplas camadas para aproximar as funções abaixo. Gere o conjunto de treinamento e de validação. Apresente na solução a curva da função custo no treinamento em função das iterações. Como se trata de uma regressão a função custo é o erro médio quadrático. Apresente a curva do erro de validação. Apresente também a superfície correspondente a função e a superfície correspondente a função aproximada pela rede:

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

$$f(\mathbf{x}) = 16x_1^2 + x_1x_2 + 8x_2^2 - x_1 - x_2 + \ln(1 + x_1^2 + x_2^2)$$

5.2-) Considere o problema das espirais. Sendo a espiral 1 uma classe e a espiral 2 outra classe. Gere as curvas das espirais usando as seguintes equações:

para espiral 1 $x = \frac{\theta}{4}\cos\theta \qquad y = \frac{\theta}{4}\operatorname{sen}\theta \qquad \theta \ge 0$

para espiral 2 $x = (\frac{\theta}{4} + 0.8)\cos\theta \qquad y = (\frac{\theta}{4} + 0.8)\operatorname{sen}\theta \qquad \theta \ge 0$

Solucione este problema utilizando uma rede perceptron de múltiplas camadas. Gere a partir das equações os dados para treinamento e teste. Determine a matriz de confusão.

5.3-) Considere o problema de classificação de padrões bidimensionais constituído neste caso de 5 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 os centros de quatro semicírculos que se interceptam no interior do quadrado, originando quatro classes. As regiões de não interseção formam a quinta classe. Após gerar aleatoriamente dados que venham formar estas distribuições de dados, selecione um conjunto de treinamento e um conjunto de validação. Defina uma a arquitetura da rede perceptron a ser usada para solução do problema. Treine a rede perceptron para classificar os padrões associados a cada uma das classes. Verifique o desempenho do classificador usando o conjunto de validação, calculando: a matriz de confusão, a acurácia, o recall, a precisão e o F1-score.

5.4-) Considere o problema de predição de uma série temporal definida como $x(n) = v(n) + \beta v(n-1)v(n-2)$, com média zero e variância dada por $\sigma_x^2 = \sigma_v^2 + \beta^2 \sigma_v^2$ onde $v(n)$ é um ruído branco gaussiano, como variância unitária e $\beta = 0.5$. Utilizando uma rede perceptrons de múltiplas camadas (rede do tipo feedforward) estime $\hat{x}(n+1) = f(x(n), x(n-1), x(n-2), x(n-3))$ usando como entradas o valor presente e os três últimos valores da série, isto é, no conjunto de treinamento utilize uma janela deslizante com a amostra n as três amostras anteriores. Isto corresponde as entradas da rede neural. Avalie o desempenho mostrando a curva da série temporal , a curva de predição e a curva do erro de predição e(n+1)=x(n+1)-x^(n+1).

Trabalho:

Apresente um trabalho sobre uma ou mais aplicações de redes neurais do tipo perceptron de múltiplas camadas. Sugestões:

(i) Transformer para área de linguagem natural
(ii) Classificação de padrões
(iii) Regressão
(iv) Filtragem

(v) Uma outra aplicação que seja do seu interesse na sua área de estudo

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 de forma individual ou em grupo de até 3 componentes. Na apresentação os componentes poderão ser submetidos a questionamentos sobre a solução da lista e o desenvolvimento dos trabalhos.