

Universidade Federal da Paraíba Centro de Informática Departamento de Informática	Aprendizado Profundo Período 2024.1 Professor: Tiago Maritan
---	--

**1A LISTA DE EXERCÍCIOS**  
**Data de Entrega: 20/Agosto/2024**

**ORIENTAÇÕES:**

- A lista pode ser resolvida em grupo de até 3 integrantes.
- No dia da entrega da lista de exercícios, o(s) grupo(s) deverão(m) fazer uma apresentação para a turma de cerca de 15 a 20 min, e enviar um link com a sua resolução da lista (contendo códigos-fontes, resultados, etc) no formulário próprio disponibilizado pelo professor no SIGAA;

1) Implemente uma **rede perceptron de múltiplas camadas** e utilize-a para aproximar as funções abaixo.

a)  $f(x) = \exp(-x)$ , onde  $1 \leq x \leq 10$

b)  $f(x) = \log_{10}(x)$ , onde  $1 \leq x \leq 10$

c)  $f(x) = 10x^5 + 5x^4 + 2x^3 - 0.5x^2 + 3x + 2$ , onde  $0 \leq x \leq 5$

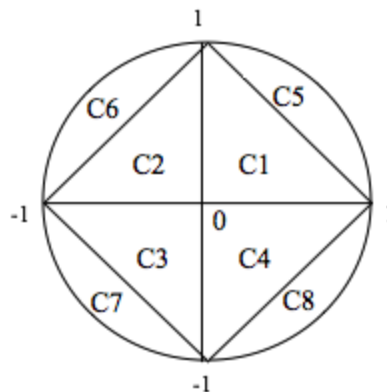
Para cada mapeamento, estabeleça dois conjuntos de dados, um para o treinamento e outro para o teste. Nesse conjunto, cada amostra deve ser representada da seguinte forma:  $x$  é a entrada e  $f(x)$  é a saída desejada - rótulo. Essas amostras devem ser divididas em, pelo menos, dois conjuntos: treinamento e validação. Treine um perceptron de múltiplas camadas para que ele aprenda a aproximar a função a partir do conjunto de treinamento, e vá testando com o conjunto de validação.

Em seguida, compare os resultados com as curvas exatas. Apresente também a curva do erro médio de treinamento com relação ao número de épocas e a curva do erro médio com o conjunto de validação.

2) Considere um problema de classificação de padrões constituído de oito padrões. A distribuição dos padrões forma um círculo centrado na origem de raio unitário e contido no círculo um losango também centrado na origem e com lados iguais à raiz de 2. Os dados das classes  $C_1$ ,  $C_2$ ,  $C_3$ ,  $C_4$  correspondem aos quatro setores do losango e as outras quatro classes correspondem aos setores delimitados pelo círculo e os lados do losango. Após gerar aleatoriamente dados que venham formar estas distribuições de dados, selecione um conjunto de treinamento e um conjunto de validação. Treine **duas redes perceptron (uma rede**

utilizando a regra delta convencional, e outra usando a regra delta com termo do momento), para classificar os padrões associados a cada uma das classes. Verifique o desempenho dos classificadores usando o conjunto de validação e calculando a matriz de confusão.

Dica: Considere que os oitos padrões estavam dispostos geometricamente da seguinte forma:



3) Utilize a **rede neural perceptron de múltiplas camadas** para fazer a predição de um passo, até predição de três passos, da série temporal:  $y(n)=2*\sin(0.5*n)+\cos(1.5*n +(\pi/4))$ , para  $n \geq 0$ . Avalie o desempenho mostrando para cada caso os erros de predição.

Para fazer a predição, estabeleça dois conjuntos de dados, um para o treinamento e outro para o teste. Nesse conjunto, cada amostra/exemplo pode ser construída utilizando como K números anteriores da série como atributos de entrada ( $y(n)$ ,  $y(n-1)$ ,  $y(n-2)$ , ...,  $y(n-K)$ ), e a saída desejada seriam os próximos números da série ( $y(n+1)$ ,  $y(n+2)$  e  $y(n+3)$ ). Por exemplo, para  $K = 10$ . Por exemplo:

- Exemplo de treinamento 1:
  - Entrada:  $y(0)$ ,  $x(1)$ ,  $y(2)$ ...,  $y(9)$
  - Saída desejada:  $y(10)$ ,  $y(11)$  e  $y(12)$
- Exemplo de treinamento 2:
  - Entrada:  $y(1)$ ,  $y(2)$ ,  $y(3)$ ...,  $y(10)$
  - Saída desejada:  $y(11)$ ,  $y(12)$  e  $y(13)$
- ...
- Exemplo de treinamento n:
  - Entrada:  $y(20)$ ,  $y(21)$ ,  $y(22)$ ...,  $y(29)$
  - Saída desejada:  $y(30)$ ,  $y(31)$  e  $y(32)$

4) Implemente uma Rede Perceptron de Múltiplas Camadas que faça a predição se um passageiro sobreviveu ou não ao acidente do Titanic. Embora houvesse algum elemento de sorte envolvido na sobrevivência dos envolvidos, aparentemente alguns grupos de pessoas tinham mais probabilidade de sobreviver do que outros. Neste exercício, você deve tentar criar um modelo preditivo, usando os dados de passageiros (ou seja, nome, idade, sexo, classe socioeconômica, etc.), que procure responder à pergunta: "que tipo de pessoa tem mais probabilidade de sobreviver?" A base de dados do problema pode ser acessada através do seguinte link: <https://www.kaggle.com/c/titanic/data>.

Apresente também a curva do erro médio de treinamento com relação ao número de épocas e a curva do erro médio com o conjunto de validação ou teste. Calcule e apresente também a matriz de confusão do modelo.

Dica: Analisem cuidadosamente a base de dados utilizada. Observem que alguns usuários possuem dados faltantes, inconsistentes, e algumas características (features) são redundantes. Dessa forma, a aplicação de técnicas de pré-processamento vistas nas nossas aulas pode ser importante para resolução do problema.