

 <b>INSTITUTO FEDERAL</b> Ceará	<b>INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ</b> Campus Canindé - Graduação Tecnológica em Análise e Desenvolvimento de Sistemas		<b>Nota:</b>
	<b>Disciplina:</b> Inteligência Artificial		
	<b>Professor:</b> Raphael Torres Santos Carvalho		<b>Data:</b> 21/06/2022
<b>1ª AVALIAÇÃO DA N2 - TRABALHO</b>			

## 1. INTRODUÇÃO

Inteligência Computacional Aplicada é uma área de conhecimento da computação que vem ganhando bastante destaque nos últimos anos em virtude de ser composto de diversas ferramentas que permitem resolver problemas que não são possíveis de serem solucionados pela programação tradicional.

Classificadores neurais são ferramentas que realizam reconhecimento de padrões por meio da classificação baseadas nos neurônios artificiais do tipo perceptron. São exemplos de algoritmos de classificação neural o classificador perceptron simples (PS), o classificador perceptron simples com multiclases e o classificador perceptron multicamadas (MLP).

Com esta avaliação da N2, os alunos serão capazes de desenvolver scripts na linguagem R para resolução de problemas utilizando algoritmos de classificação não-neural.

## 2. REQUISITOS E CRITÉRIOS DE AVALIAÇÃO

Para cada questão apresentada na próxima seção, o aluno deverá elaborar os script (códigos) na linguagem R e realizar os procedimentos solicitados.

Ao final, o aluno deverá preparar um relatório contendo o código elaborado na linguagem R, os gráficos solicitados e responder às perguntas solicitadas em cada questão.

Cada questão será avaliada de forma **percentual, atribuindo-se os percentuais de 0%, 25%, 50%, 75% e 100% sobre a pontuação indicada para a questão/item.**

Este trabalho tem uma questão extra ao final que vale 2,0 pontos extras, de modo que a nota total poderá ser 12 pontos.

O relatório e todos os códigos scripts em R elaborados deverão ser enviados pela atividade postada na Sala de Aula do Google da disciplina dentro dos prazos previstos neste documento. **Caso seja observada a cópia – no todo ou em parte do teor do trabalho, seja de outros alunos ou da Internet, a(s) equipe(s) envolvida(s) terá(ão) sua nota anulada.**

## 3. ENTREGA DO TRABALHO

O trabalho é **INDIVIDUAL** e cada discente deverá preparar os artefatos solicitados em cada questão e enviá-los, devidamente compactados no formato ZIP ou RAR, na atividade “1ª Avaliação da N2” postada na sala de aula do Google da disciplina. Qualquer falha na compactação dos arquivos, bem como atraso nas entregas fora é de responsabilidade do aluno.

A data limite para entrega do trabalho **será dia 02/07/2022**. Não será aceito a entrega de nenhum arquivo ou do trabalho após o dia 02/07/2022, resultando em **nota zero** no item ou em toda a avaliação.

## 4. QUESTÕES

### QUESTÃO 1 (4,0 pontos)

Nesta questão, o aluno deverá utilizar novamente o conjunto de dados denominado **“iris”**, que contém 3 classes com 50 amostras de cada referente aos tipos da planta Íris (iris setosa, iris virginica e iris versicolor)

Assim, utilizando o script em R do classificador perceptron multicamadas (MLP) apresentado em sala para o conjunto de dados Íris, faça uma alteração para testar o classificador utilizando diferentes combinações dos atributos de entrada (comprimento da pétala, largura da pétala, comprimento da sépala e largura da sépala), ou seja, realize o treinamento e teste com as amostras utilizando menos do que 4 características. Em seguida, faça um comparativo entre as combinações de características e seus resultados de taxa de acerto.

Para isso, deve-se dividir o conjunto de dados em dois conjuntos (treinamento com 80% e teste com 20%) selecionados aleatoriamente, e devem ser realizadas 10 repetições entre treinamento e teste de modo a determinar as seguintes métricas de avaliação: Taxas de acerto, máxima, mínima e média, e a Matriz de Confusão do melhor resultado.

No relatório final, apresente a configuração dos parâmetros utilizada no treinamento do classificador, explicando o motivo dos parâmetros utilizados no treinamento.

*Maiores detalhes sobre o dataset “Iris”:* <https://archive.ics.uci.edu/ml/datasets/Iris>

### QUESTÃO 2 (3,0 pontos)

Nesta questão, o aluno deverá utilizar novamente o conjunto de dados denominado **“parkinsons.csv”**, com medições biomédicas de voz de 31 pessoas, 23 com doença de Parkinson (DP). Cada coluna na tabela é uma medida de voz específica, e cada linha corresponde a uma das 195 gravações de voz desses indivíduos (coluna "nome"). Estes dados discriminam pessoas saudáveis daquelas com DP, de acordo com a coluna "status" que é definida como 0 para saudáveis e 1 para DP.

Assim, implemente o script em R do classificador perceptron multicamadas (MLP) para multiclases, dividindo o conjunto de dados em dois conjuntos (treinamento com 80% e teste com 20%) selecionados aleatoriamente, e devem ser realizadas 10 repetições entre treinamento e teste de modo a determinar as seguintes métricas de avaliação: Taxas de acerto, máxima, mínima e média, e a Matriz de Confusão do melhor resultado.

Por fim, compare os resultados obtidos com os demais classificadores estudados na disciplina: Perceptron Simples, KNN e DMC. No relatório final, apresente a configuração dos parâmetros utilizada no treinamento do classificador, explicando o motivo dos parâmetros utilizados no treinamento.

*Maiores detalhes sobre o dataset “parkinsons”:* <https://archive.ics.uci.edu/ml/datasets/Parkinsons>

### QUESTÃO 3 (3,0 pontos)

Nesta questão, o aluno deverá utilizar novamente o conjunto de dados denominado **“dermatology.csv”**, que é composto por dados de doenças eritemato-escamosas que são um problema real na dermatologia. As doenças deste grupo são psoríase (classe 1), dermatite seborréica (classe 2), líquen plano (classe 3), pitiríase rósea (classe 4), dermatite crônica (classe 5) e pitiríase rubra pilar (classe 6).

Este conjunto de dados contém 366 amostras de pacientes com 34 atributos, sendo 12 atributos obtidos na consulta clínica e outros 22 atributos histopatológicos, que são determinados por uma análise das amostras ao microscópio. Todas as outras características (clínicas e histopatológicas) receberam um grau na faixa de 0 a 3, em que 0 indica que a característica não estava presente, 3 indica a maior quantidade possível e 1, 2 indicam os valores intermediários relativos. Apenas o atributo “histórico familiar” tem valor 1, se paciente tem alguma dessas doenças na família, e 0 caso contrário. E o atributo “idade” que indica a idade do paciente.

Assim, implemente o script em R do classificador perceptron multicamadas (MLP) para multiclases, dividindo o conjunto de dados em dois conjuntos (treinamento com 80% e teste com 20%) selecionados aleatoriamente, e devem ser realizadas 10 repetições entre treinamento e teste de modo a determinar as seguintes métricas de avaliação: Taxas de acerto, máxima, mínima e média, e a Matriz de Confusão do melhor resultado

Por fim, compare os resultados obtidos com os demais classificadores estudados na disciplina: Perceptron Simples, KNN e DMC. No relatório final, apresente a configuração dos parâmetros utilizada no treinamento do classificador, explicando o motivo dos parâmetros utilizados no treinamento.

Maiores detalhes sobre o dataset "dermatology": <https://archive.ics.uci.edu/ml/datasets/Dermatology>

### QUESTÃO EXTRA (2,0 pontos)

Implementar as portas lógicas XOR (OU-EXCLUSIVO) e XNOR (COINCIDÊNCIA) usando o classificador perceptron multicamadas (MLP). Deve ser mostrado um gráfico com a reta que separa os pontos classificados como 0 ou 1. Considere a tabela lógica abaixo para as funções.

x1	x2	XOR	XNOR
0	0	0	1
0	1	1	0
1	0	1	0
1	1	0	1