

	INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ Campus Canindé - Graduação Tecnológica em Análise e Desenvolvimento de Sistemas	Nota:
	Disciplina: Inteligência Artificial	
	Professor: Raphael Torres Santos Carvalho Data: 10/05/2022	Etapa: 1

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 não-neurais são ferramentas que realizam reconhecimento de padrões por meio da classificação utilizando como métrica a distância entre vetores de atributos. São exemplos de algoritmos de classificação não-neural o classificador vizinho mais próximo (NN), o classificador k-vizinhos mais próximos (KNN) e o classificador distância mínima ao centróide (DMC).

Com esta avaliação da N1, 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.**

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 N1” 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 18/05/2022**. Não será aceito a entrega de nenhum arquivo ou do trabalho após o dia 18/05/2022, resultando em **nota zero** no item ou em toda a avaliação.

4. QUESTÕES

QUESTÃO 1 (5,0 pontos)

Nesta questão, o aluno deverá utilizar o conjunto de dados denominado **“parkinsons.csv”**, que é composto por uma série de 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"). O principal objetivo dos dados é discriminar pessoas saudáveis daquelas com DP, de acordo com a coluna "status" que é definida como 0 para saudáveis e 1 para DP.

Assim, implemente os scripts em R dos classificadores NN, KNN e DMC em R e compare o desempenho de cada um dos classificadores. Para cada classificador, o conjunto de dados deverá ser dividido 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, cada script deve apresentar por meio de um gráfico de dispersão dos dados de treinamento, como os dados entre as classes saudável (=0) e com doença de Parkinson (=1) estão definidos.

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

QUESTÃO 2 (5,0 pontos)

Nesta questão, o aluno deverá utilizar 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). Todas essas doenças compartilham as características clínicas de eritema e descamação, com muito poucas diferenças. Normalmente, uma biópsia é necessária para o diagnóstico, mas infelizmente essas doenças também compartilham muitas características histopatológicas. Outra dificuldade para o diagnóstico diferencial é que uma doença pode apresentar as características de outra doença no estágio inicial e pode apresentar as características características nos estágios seguintes.

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 os scripts em R dos classificadores NN, KNN e DMC em R e compare o desempenho de cada um dos classificadores. Para cada classificador, o conjunto de dados deverá ser dividido 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, cada script deve apresentar por meio de um gráfico de dispersão dos dados de treinamento, como os dados entre as classes 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)

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