

TRABALHO PRÁTICO

1. Objetivos

Pretende-se que os alunos explorem e aprofundem os conceitos de raciocínio baseado em casos e de redes neurais (RNs).

O trabalho deve ser implementado usando o Matlab, nomeadamente a toolbox Deep Learning (redes neurais).

A descrição do trabalho e das tarefas a desenvolver encontram-se na secção 3.

2. Regras para a realização do trabalho

- Os grupos de trabalho são de dois alunos. Apenas serão aceites trabalhos individuais após uma justificação válida que deve ser apresentada ao docente das aulas práticas.
- A escolha do *dataset* deve ser feita pelo grupo, no entanto, apenas um dos alunos do grupo deve preencher o referendo que será disponibilizado no Moodle em dia a anunciar brevemente.
- A **data única de entrega** do trabalho é até às 23.59 do dia **12 de maio de 2024**.
- Ficheiro a submeter no Moodle:
 - Deve ser em formato ZIP
 - Nome do ficheiro: deve ter os nomes, números, e turma dos alunos dos elementos do grupo (ex: **AnaMelo22222_P1_PedroCosta33333_P3.ZIP**)
 - Deve conter: código Matlab de todas as tarefas implementadas, ficheiro Excel com os resultados globais recolhidos nas diferentes tarefas e PDF do relatório
- As defesas serão nas aulas das semanas **13 a 23 de maio**.
- Para a defesa, cada grupo terá de fazer a inscrição de um (**só 1**) dos seus elementos no Moodle na turma que frequentam e onde fizeram a submissão do trabalho. Os *slots* para inscrição nas defesas serão oportunamente disponibilizados.
- A defesa do trabalho é **obrigatória** e com a presença de todos os membros do grupo. Não fazendo a defesa, o trabalho será cotado com zero valores.
- O trabalho prático tem a cotação de 8 valores (numa escala de 0 a 20).

3. Descrição do trabalho

Para a realização do trabalho prático, deverá executar as tarefas descritas de seguida.

As ambiguidades ou situações menos detalhadas devem ser resolvidas pelos alunos, fundamentando no relatório as escolhas e decisões tomadas.

3.1. Escolha do Dataset

No moodle encontra-se um referendo onde cada grupo (apenas um dos elementos) deve escolher o dataset que pretende usar para a realização do trabalho prático. Existem 3 datasets diferentes para realizar tarefas de classificação. A descrição de cada dataset encontra-se no ficheiro **descrição_datasets.pdf**.

- Hepatitis
- Stroke
- Cirrosis

Haverá vagas limitadas para a escolha de cada dataset.

3.2. Datasets

Para cada dataset existem 3 ficheiros.

- START – dataset com 10 exemplos, sem necessidade de preparação. Deve ser usado para a realização das tarefas descritas em **3.4 a)**
- TRAIN – dataset sem pré-processamento. Precisa de ser analisado, preparado de forma a ficar com as mesmas características do *dataset* do ficheiro START. Deve ser usado para a realização das tarefas descritas em **3.3 e 3.4 b)**
- TEST - dataset com 10 exemplos, sem necessidade de preparação. Deve ser usado para a realização das tarefas descritas em **3.4 c)**

3.3. [30%] Preparação do dataset

O dataset do ficheiro TRAIN deve analisado e preparado para que possa ser usado no estudo e análise de desempenho de redes neuronais feedforward da secção seguinte.

- a) Todos os atributos devem ser convertidos em valores numéricos ou booleanos. Veja os tipos de dados dos mesmos atributos do dataset do ficheiro START e faça as alterações necessárias. Mais informações no ficheiro **descrição_datasets.pdf**
Pode executar esta tarefa usando o Excel, ou implementando um script em Matlab que faça a conversão.
- b) Identifique qual o atributo que corresponderá à saída desejada (*target*) da rede neuronal. Esta coluna possui alguns valores em falta identificados com **NA**.
- c) Implemente um sistema de **raciocínio baseado em casos** para preencher os atributos com valores em falta (NA). Implemente apenas a fase de RETRIEVE para encontrar o caso mais semelhante àquele onde pretende o preenchimento e use os valores desse caso para preencher os valores em falta.

Nota: caso não consiga executar esta tarefa, e para conseguir prosseguir com as tarefas da secção 3.4, remova as linhas do dataset onde existem valores em falta.

No relatório não esqueça de:

- Explicar como fez o ajuste dos tipos de dados pedido em b)
- Justificar as funções de similaridade local e global, bem como os pesos de cada atributo usados em c)

3.4. Estudo e análise de redes neuronais

Usar o *dataset* escolhido para estudar, treinar e analisar redes neuronais *feedforward* para as tarefas de classificação de cada *dataset*.

a) [20%] Usar o ficheiro **START**:

- Utilize uma rede neuronal *feedforward* com uma camada de 10 neurónios, a parametrização das funções de treino e ativação deve ser a *default*.
- Use **todos os exemplos** no treino (sem segmentação).
- Divida o dataset START nas variáveis de *entrada* e de *target*. O *target* corresponde à saída da rede, isto é, a categoria de classificação de cada dataset.
- Registe o tempo de execução e as métricas de desempenho (precisão total, erro). Faça algumas alterações apenas às funções de treino e de ativação.
- No relatório, registe os dados obtidos e apresente as conclusões observadas.

b) [30%] Usar o ficheiro **TRAIN** devidamente preparado de acordo com a secção 3.3:

- Divida o dataset TRAIN nas variáveis de entrada e de target. O target corresponde à saída da rede, isto é, a categoria de classificação de cada dataset
- Estude a influência de *diferentes arquiteturas* de redes feedforward (diferentes números de camadas e de neurónios em cada camada)
- Estude a influência de *diferentes funções de treino*
- Estude a influência de *diferentes funções de ativação*
- Estude influência de *diferentes segmentações* para os dados de *train*, *validation*, *test*.
- Grave as **três** redes neuronais com melhores desempenhos.

No relatório, faça um resumo dos principais resultados e conclusões obtidas.

NOTA: sugere-se a adaptação do ficheiro Excel dado nas aulas práticas, para registar resultados de precisão global e precisão de teste das diferentes configurações e obter as conclusões.

c) [20%] Usar o ficheiro **TESTE**:

Nesta alínea pretende-se que analise a capacidade de generalização e aprendizagem das melhores redes gravadas em b)

- Divida o dataset do ficheiro TEST nas variáveis de entrada e de target. O target corresponde à saída da rede, isto é, a categoria de classificação de cada dataset. Estes exemplos nunca foram usados no treino da rede, permitem ver como a rede se comporta com casos novos.
- Utilize este dataset e, **sem treinar**, use as redes obtidas em b)
- Compare a saída das redes com o *target* desejado e registe as métricas de acerto.

No relatório mostre os resultados obtidos e registe as conclusões mais relevantes.

d) [bónus de 1 valor extra]

Desenvolva uma aplicação gráfica simples em Matlab que ofereça as seguintes funcionalidades:

- Criar uma rede neuronal, permitindo a configuração do número de camadas, números de neurónios por camada, funções de ativação, funções de treino, rácios de segmentação
- Treinar a rede neuronal criada após a escolha de um dataset
- Carregar uma rede neuronal previamente treinada e gravada em disco (obtidas na alínea b) e usar a RN para um dataset, mostrando os desempenhos obtidos
- Permitir introduzir os valores dos atributos de apenas um caso e verificar qual a classificação dada pela rede. Para testar e verificar se a rede acerta ou falha, pode usar os casos do ficheiro TEST

IMPORTANTE: No relatório, indique qual o *dataset* foi usado e descreva as tarefas desenvolvidas em 3.3 e 3.4, os resultados e conclusões obtidas. Uma má qualidade do relatório pode descontar até 50% na classificação total obtida nos pontos anteriores.