

LICENCIATURA EM ENGENHARIA INFORMÁTICA E SISTEMAS  
INTELIGÊNCIA COMPUTACIONAL 20/21

PROJETO - FASE I

---

## 1. Âmbito e Objectivo

O Projeto (Fase I, Seminário e Fase II) combina a realização de um trabalho de investigação com o desenho e implementação de uma aplicação para classificação ou regressão. As diversas fases envolvem a escolha de um problema (caso real), definição de uma arquitetura para aprendizagem automática (Fase I), estudo de um algoritmo para otimização da arquitetura (seminário), implementação e validação da solução final (Fase II).

Esta primeira fase do trabalho envolve:

- Escolha de um problema (dataset);
- Análise do Problema e recolha de dados;
- Escolha do modelo (definição da arquitetura de uma rede neuronal: número de camadas, neurónios por camada, funções de ativação, algoritmo de treino, coeficientes de aprendizagem, etc);
- Treino do modelo;
- Avaliação do desempenho.

As fases de Seleção de características, Otimização do modelo (escolha de hiper-parâmetros) e produção (implementação de um protótipo) será objeto de estudo na Fase II do Projeto.

Esta primeira fase poderá ser realizada em Matlab ou Python.

## 2. Caso de Estudo

A metodologia deverá ser aplicada a um problema de classificação ou regressão, com os seguintes requisitos mínimos:

- Um mínimo de 3 classes para problemas de classificação;
- Um valor mínimo de 5000 amostras para treino;
- Um mínimo de 10 features (ou atributos);

Como possíveis escolhas e respeitando os requisitos definidos apresentam-se três problemas:

- Fashion Mnist

Consiste num problema com 60000 exemplos para treino e 10000 exemplos para teste. Cada imagem é representada por uma matriz 28x28 “grayscale”. Existem 10 classes a identificar pelo modelo a implementar.

<https://www.kaggle.com/zalando-research/fashionmnist>

- CIFAR 10

Consiste num “dataset” com 60000 imagens 32x32 correspondentes a 10 classes, com 6000 imagens por classe.

<https://www.cs.toronto.edu/~kriz/cifar.html>

- BlogFeedback Dataset

Consiste num Dataset com 60021 instâncias e 281 atributos, correspondentes a features extraída de blog posts – problema de regressão.

<https://archive.ics.uci.edu/ml/datasets/BlogFeedback>

Poderá ser escolhido um destes problemas ou outro a propor que satisfaça os requisitos indicados. Os datasets devem estar descritos num repositório público, tal como:

- UCI, <https://archive.ics.uci.edu/ml/index.php>
- Kaggle, <https://www.kaggle.com/>
- OpenML, <https://www.openml.org/home>
- PMLB, <https://epistasislab.github.io/pmlb/>

Para o problema selecionado, deve descrever o objetivo do classificador e responder às questões: Quais as características das amostras? Como são representadas? Quantos exemplos existem por classe? Qual o modelo de classificador mais adequado?

### 3. Desenvolvimento do Modelo – Classificador

Deverá ser desenvolvido um modelo baseado numa rede neuronal capaz de resolver a tarefa de classificação ou regressão.

O desempenho do modelo (classificador) deve ser avaliado para o conjunto de teste, apresentando a matriz de confusão e resultado para as principais métricas: “accuracy”, sensibilidade, especificidade, f-measure e AUC. Para os modelos de regressão deverão ser avaliados o RMSE e R2.

Devem ser testadas várias configurações de redes neuronais. Pode ser considerada uma rede MLP, CNN ou outras.

Caso o modelo seja implementado em Python deve recorrer-se às bibliotecas:

- <https://scikit-learn.org/stable/>
- <https://www.tensorflow.org/>
- <https://keras.io/>

Deve-se analisar e avaliar diferentes soluções e estabelecer uma análise comparativa e crítica.

#### 4. Relatório do Projeto

O relatório deve seguir a seguinte estrutura, com o máximo de 8 páginas:

- Cap. 1: *Descrição do caso de estudo o objetivos do problema;*
- Cap. 2: *Descrição da **Implementação** dos algoritmos;*
- Cap. 3: **Análise de Resultados;**
- Cap. 4: **Conclusões.**
- **Referências**

#### 5. Avaliação

O trabalho é individual ou em grupos de 2 alunos.

A documentação final a submeter no Moodle consiste em:

- Relatório;
- Código;
- Slides de apresentação.

**Submissão de tema – 07 Novembro 2021**

**Submissão de relatório final – 17 Novembro 2021**

**Apresentação e defesa:** em dia e hora a marcar

**Cotação: 2 valores**