



Universidade do Minho  
Departamento de Informática

Perfil Sistemas Inteligentes  
Computação Natural  
Edição 2018/2019

1º Trabalho prático

**Tema** Redes Neurais Artificiais em Exames Clínicos

**Objetivos de aprendizagem**

Com a realização deste trabalho prático, pretende-se que os alunos aprendam os seguintes procedimentos:

- Preparação e análise de datasets;
- Treino e validação de modelos de aprendizagem, especificamente de Redes Neurais Artificiais (ANN);
- Optimização de parâmetros do modelo de aprendizagem;

**Enunciado**

Este enunciado pretende ser o ponto de partida para o desenvolvimento de um modelo preditivo utilizando o ambiente de desenvolvimento Python ou Julia. Para isso, será necessário o desenvolvimento de uma solução para o seguinte problema:

*Classificar se uma massa detectada num exame de mastografia é benigna ou maligna, através da aplicação de um modelo deep learning (ANN).*

Neste projecto, tenciona-se aplicar a matéria leccionada durante a unidade curricular de Computação Natural para a detecção de massas malignas no tecido mamário, através da implementação de algoritmos ANN. Tendo em conta o problema definido, serão utilizados um conjunto de dados clínicos de casos reais (961 instâncias/exames) que apresentam um conjunto de métricas analisadas durante exames de mastografia, com o objectivo de determinar se existem ou não massas benignas/malignas. Como tarefas, cada grupo deverá implementar um conjunto de procedimentos que possibilitem o tratamento/preparação do respectivo dataset, como forma de proceder à criação de modelos deep learning (ANN) que solucionem o problema definido. Neste processo, deverão ter em conta a arquitectura do modelo (número de layers, número de nodos, funções de activação, entre outros), como forma de maximizar a sua performance. Para validação do modelo, deverá ser aplicado o método cross-validation (k=10). Como tarefa extra, a aplicação de Algoritmos Genéticos para uma optimização eficiente do modelo ANN deverá ser aplicado.

Definido o problema, é demonstrado uma breve síntese sobre o dataset:

- Núm. instâncias: 961;
- Núm. atributos: 6 (1 target, 5 features);
- Info. atributos:
  1. Avaliação BI-RADS: 1-5 (ordinal);
  2. Idade: Idade do paciente em anos (integer);
  3. Forma da massa: redonda=1, oval=2, lobular=3, irregular=4 (nominal);
  4. Margem da massa: circunscrito=1, micro-lobulado=2, obscurecido=3, mal-definido=4, espiculado=5 (nominal);
  5. Densidade da massa: alto=1, médio=2, baixo=3, contem gordura=4 (ordinal);
  6. Gravidade: benigno=0, maligno=1 (binominal);

- 
- Dataset c/valor de atributos em falta:
    1. Avaliação BI-RADS: 2;
    2. Idade: 5;
    3. Forma da massa: 31;
    4. Margem da massa: 48;
    5. Densidade da massa: 76;
    6. Gravidade: 0;
  - Distribuição p/classes: benigno: 516; maligno: 445;

O trabalho prático compreende a entrega do código desenvolvido e do relatório, definindo todos os procedimentos aplicados e respetiva justificação da sua utilização, apoiando-se na demonstração dos resultados adquiridos.

Poderão utilizar quaisquer bibliotecas que o grupo entender serem uteis para a resolução do problema (tendo em conta os ambientes de desenvolvimento definidos), onde é sugerida a utilização das bibliotecas leccionadas na u.c.

---

### Entrega

O código resultante da realização do trabalho prático e o respetivo relatório em formato digital .PDF deverão ser enviados por correio eletrónico para [cesar.analide@di.uminho.pt](mailto:cesar.analide@di.uminho.pt) e [fgoncalves@algoritmi.uminho.pt](mailto:fgoncalves@algoritmi.uminho.pt), em ficheiros compactados (formato ZIP). Tanto o assunto da mensagem como o ficheiro deverão ser identificados na forma “[CN: F1GXX]”, em que [XX] designa o número do grupo de trabalho.

A sessão de apresentação do trabalho prático terá lugar no dia 25 de março de 2019, em formato a anunciar oportunamente.

---

### Referências bibliográficas

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Kudlur, M. (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)* (pp. 265-283).

Valarmathi, P., & Robinson, S. (2016). An improved neural network for mammogram classification using genetic optimization. *Journal of Medical Imaging and Health Informatics*, 6(7), 1631-1635.