Ficha de Unidade Curricular A3ES
Elementos de Aprendizagem Estatística
Mestrado em Matemática Aplicada para a Indústria
2024-25

---

1. Caracterização da Unidade Curricular

1.1 Designação
[4289] Elementos de Aprendizagem Estatística / Elements of Statistical Learning

1.2 Sigla da área científica em que se insere
MAT

1.3 Duração
Unidade Curricular Semestral

1.4 Horas de trabalho
162h 00m

1.5 Horas de contacto
Total: 67h 30m, 72h 30m das quais T: 45h 00m, h 00m | TP: 22h 30m, 45h 00m | P: 22h 30m, h 00m | O: 5h 00m, h 00m

1.6 ECTS
6

1.7 Observações
Unidade Curricular Obrigatória, Unidade Curricular Opcional
Unidade Curricular comum ao(s) curso(s) de PGMCD

---

2. Docente responsável [2039] Carlos José Brás Geraldes

---

3. Docentes e respetivas cargas letivas na unidade curricular Não existem docentes definidos para esta unidade curricular

4. Objetivos de aprendizagem (conhecimentos, aptidões e competências a desenvolver pelos estudantes)

Esta UC introduz os algoritmos e métodos estatísticos para mineração de dados. A sua natureza interdisciplinar combina tópicos de estatística, bases de dados e ciência da computação¿. Os objetivos são:

1. Identificar as fases de um projeto de *Data Science* . Conhecer os conceitos e técnicas estatísticas de aprendizagem automática (AA), exemplificando as suas aplicações e funcionalidades

2. Identificar os vários tipos de dados de input e conhecer os métodos para a sua preparação e pré-processamento, bem como as principais técnicas para transformação e redução da dimensionalidade dos dados

3. Conhecer os fundamentos teóricos matemáticos dos métodos de AA apresentados, saber manuseá-los, bem como identificar e interpretar várias formas e tipos de resultados

4. Avaliar os resultados obtidos com as técnicas de AA, usando e interpretando as medidas de desempenho

5. Usar softwares apropriados para resolver vários desafios

6. Realizar na integra um projeto de *Data Science* usando metodologias adequadas

4. Intended learning outcomes (knowledge, skills and competences to be developed by the students)

This CU introduces statistical data mining algorithms and methodologies. It´s interdisciplinary nature combines topics of statistics, databases, and computer science. The intended outcomes are:

1.To describe the several stages of a Data Science project. To know the concepts and statistical techniques of machine learning (ML), exemplifying its applications and functionalities

2.To identify the several types of input data and to know the methodologies for its preparation and preprocessing, as well as the main techniques for transformation and reduction of data dimensionality

3.To know the mathematical theoretical foundations associated to the ML methods, to know how to handle them, as well as to identify and interpret the several forms and types of results

4.To evaluate the results obtained with the ML techniques, using, and interpreting the performance measures

5. Use appropriate software for solving several challenges

6.To complete a Data Science project using appropriate methodologies

5. Conteúdos programáticos

1. Introdução. Fases de um projeto de Data Science. Conceitos e ferramentas de mineração de dados e aprendizagem automática. Input: tipos de dados e preparação de dados. Output: tipos de representação
2. Classificação. Método do Vizinho Mais Próximo (KNN)
3. Classificadores de Bayes. Classificador *Naive Bayes*
4. Estimação da Densidade
5. Desempenho do Classificador
6. Árvores de Decisão para a Classificação
7. Classificador de Máquinas de Vetores de Suporte
8. Redes Neuronais para Classificação
9. Análise de Componentes Principais
10. Análise de Clusters
11. Análise de Componentes Independentes
12. Classificação com Análise Discriminante
13. Avaliação de Desempenho dos Classificadores

5. Syllabus

1. Introduction. Phases of a Data Science project. Data mining and machine learning concepts and tools. Input: data types and data preparation. Output: representation types
2. Classification. Nearest Neighbor Method (KNN)
3. Bayes classifiers. Naive Bayes Classifier
4. Density estimation methods
5. Classifier Performance
6. Decision Trees for Classification
7. Support Vector Machine Classifier
8. Neural Networks for Classification
9. Principal Component Analysis
10. Cluster Analysis
11. Independent Component Analysis
12. Classification with Discrimina nt Analysis
13. Classifier Performance Assessment

6. Demonstração da coerência dos conteúdos programáticos com os objetivos de aprendizagem da unidade curricular

Os pontos 1 e 2 dos conteúdos programáticos pretende alcançar o ponto 1 dos objetivos

Os pontos 2, 3, 4, 6, 7, 8, 9, 10, 11 e 12 dos conteúdos programáticos apresentam os conceitos necessários para atingir o ponto 3 dos objetivos

Os pontos 5 e 13 dos conteúdos programáticos pretendem alcançar o ponto 4 dos objetivos

A totalidade dos pontos dos conteúdos programáticos permitem atingir e consolidar os objetivos 5 e 6

6. Evidence of the syllabus coherence with the curricular unit´s intended learning outcomes

Topics 1 and 2 of the program content aim to achieve objective 1.

Topics 2, 3, 4, 6, 7, 8, 9, 10, 11, and 12 of the program content present the necessary concepts to achieve objective 3.

Topics 5 and 13 of the program content aim to achieve objective 4.

The entirety of the program content points allow for achieving and consolidating objectives 5 and 6.

7. Metodologias de ensino (avaliação incluída)

Aulas teórico-práticas. A parte teórica baseia-se no formalismo matemático e na compreensão dos algoritmos subjacentes à implementação dos métodos de AA. Os alunos serão familiarizados com as técnicas de aprendizagem estatística, que incluem a inferência e a interpretação dos modelos de classificação. Na parte prática são implementados computacionalmente os algoritmos de AA e os procedimentos abordados na teórica, com base em casos reais ou simulados. Se reunidas as condições necessárias, esta UC poderá ser parcialmente lecionada à distância de forma síncrona (1/3 das horas de contacto semanais). Avaliação distribuída com exame final, tendo 2 partes pedagogicamente fundamentais: a teórica é constituída por um exame (nota mínima de 9,50 valores) e a prática integra um projeto de grupo e uma prova oral obrigatória. A nota final, NF, é dada por: NF=0,5 NT+0,5 NP, onde NT é a nota da teórica e NP a nota da prática, sendo esta a média aritmética das notas do projeto e da prova oral.

7. Teaching methodologies (including assessment)

Theoretical-practical classes. The theoretical part is based on mathematical formalism and the understanding of the algorithms underlying the implementation of AA methods. Students will be familiarized with statistical learning techniques, which include inference and interpretation of classification models. In the practical part, the AA algorithms and the procedures covered in the theoretical one are computationally implemented, based on real or simulated cases. If the necessary conditions are met, this CU can be partially taught remotelt in a sychronous manner (1/3 of the weekly contact hours). The assessment is distributed with a final exam, with 2 pedagogically fundamental parts: the theoretical one consists of an exam and the practical includes 2 parts: a group project and a mandatory oral test. The final grade, NF, is given by: NF=0.5 NT+0.5 NP, where NT is the theoretical grade and NP the practical grade, this being the arithmetic average of the project and oral exam grades.

8. Demonstração da coerência das metodologias de ensino com os objetivos de aprendizagem da unidade curricular

A aprendizagem estatística é focada fundamentalmente na formalização de todos os aspetos do processo de mineração de dados, principalmente naqueles que permitem a interpretação e avaliação dos resultados obtidos pelo modelo de classificação. É importante que o aluno adquira competências na utilização do raciocínio estatístico para resolver problemas de modelação de complexidade mais elevada.

As metodologias de ensino são consistentes com os objetivos da unidade curricular, uma vez que na parte teórica é ensinado o formalismo estatístico necessário para o desenvolvimento da capacidade crítica na escolha e construção de um modelo de classificação. Adicionalmente, na parte prática, o aluno pode aplicar as aptidões já mencionadas para resolver problemas reais ou simulados (próximos da realidade).

A implementação computacional dos processos envolvidos, com recurso à utilização de um *software* livre, irá possibilitar ao aluno o desenvolvimento das suas próprias ferramentas bem como um melhor entendimento da resolução prática de problemas.

A avaliação da aprendizagem com base num exame permitirá aferir os conhecimentos e competências individuais adquiridas pelo aluno. O projeto de grupo permitirá avaliar a capacidade cooperativa na resolução dos problemas.

8. Evidence of the teaching methodologies coherence with the curricular unit´s intended learning outcomes

Statistical learning is fundamentally focused on formalizing all aspects of the data mining process, particularly those that allow for the interpretation and evaluation of the results obtained by the classification model. It is important for students to acquire skills in using statistical reasoning to solve more complex modeling problems.

The teaching methodologies are consistent with the objectives of the course unit, as the theoretical part teaches the necessary statistical formalism for developing critical capacity in the selection and construction of a classification model. Additionally, in the practical part, students can apply the aforementioned skills to solve real or simulated problems (close to reality).

The computational implementation of the involved processes, using open-source software, will enable students to develop their own tools as well as gain a better understanding of the practical resolution of problems.

The learning assessment based on an exam will allow for the evaluation of the individual knowledge and skills acquired by the student. The group project will assess the cooperative ability to solve problems.

9. Bibliografia de consulta/existência obrigatória

1. Bishop, C. M., Pattern Recognition and Machine Learning, Springer (2006).
2. Duda, R. O., Hart, P. E. and Stork, D. G., Pattern Classification, Wiley (2001).
3. Hastie, T., Tibshirani, R. and Friedman, J., The elements of statistical learning. Springer (2017).
4. Lantz, B., Machine Learning with R, Packt (2013).
5. Muller, A. and Guido, S., Introduction to Machine Learning with Python, O?Reilly (2017)
6. Murphy, K., Machine Learning: A Probabilistic Perspective, MIT Press (2012).
7. Witten, I. H., Frank, E. and Hall M. A., Data mining: practical machine learning tools and techniques. Morgan Kaufmann (2011).

10. Data de aprovação em CTC 2024-07-17

11. Data de aprovação em CP 2024-06-26