



**Ciências
ULisboa**

Prospecção e Descoberta da Informação

Código: 425133

ECTS: 6

Ano Letivo: 2015/16

Carga horária: T: 2:00 h; TP: 1:30 h; OT: 2:00 h;

Departamento: Informática

Área Científica: Informática;

Objetivos da Unidade Curricular

A capacidade actualmente existente para coligir e armazenar dados ultrapassa largamente a nossa capacidade para os analisar, resumir e extrair daí manualmente conhecimento útil, previamente desconhecido mas implicitamente contido nos dados analisados. A Prospecção e Descoberta de Informação ("Data Mining and Knowledge Discovery") é a disciplina científica que permite extrair, de forma automática, informação útil de grandes colecções de dados. Nesta cadeira pretende-se dotar o aluno de conhecimentos específicos nessa área, através do estudo das várias etapas da prospecção de informação, com ênfase no estudo de métodos e algoritmos de aprendizagem automática, originados nas áreas de Inteligência Artificial, Estatística e Análise de Dados, e sua aplicação à descoberta de modelos e padrões interessantes em colecções de dados.

Pré-requisitos

- Programação I (26701)

Conteúdos

Introdução e conceitos gerais: Tarefas e problemas de Prospecção e Descoberta de Informação
Indução de árvores de decisão
Exploração e visualização de dados
Aprendizagem baseada em instâncias
Regressão linear e não linear
Avaliação de Desempenho
Regras de associação
Agrupamento (Clustering)
Máquinas de Vectores de Suporte
Text Mining
Sistemas de Recomendação

Descrição detalhada dos conteúdos programáticos

Componente Teórica

Introdução e conceitos gerais: Tarefas e problemas de Prospecção e Descoberta de Informação
Indução de árvores de decisão
Exploração e visualização de dados

Aprendizagem baseada em instâncias
Regressão linear e não linear
Avaliação de Desempenho
Regras de associação
Agrupamento (Clustering)
Máquinas de Vectores de Suporte
Text Mining
Sistemas de Recomendação

Componente Teórica-Prática

Resolução de exercícios sobre a matéria teórica.

Aulas em laboratório de demonstração do uso das bibliotecas de algoritmos de prospecção de dados associadas à linguagem R; aplicação de algoritmos a pequenos conjuntos de dados.

Bibliografia

Recomendada

- Principles of Data Mining, David Hand, Heikki Mannila e Padhraic Smyth, MIT Press, 2001.
- Modern Applied Statistics with S - 4th Edition, W. N. Venables, B. D. Ripley, Springer, 2002.

Outros elementos de estudo

Data Mining - Practical Machine Learning Tools and Techniques with Java Implementations, second edition, Ian H Witten e Eibe Frank, Morgan Kaufmann, 2005.

Introduction to Machine Learning, Ethem Alpaydin, MIT Press, 2004.

A linguagem R - Programação para a Análise de Dados, Luis Torgo, Escolar Editora, 2009

Pattern Classification, second edition, Richard Duda, Peter Hart and David Stork, Wiley, 2001.

The Elements of Statistical Learning - Data Mining, Inference and Prediction, T. Hastie, R. Tibshirani e J. Friedman, Springer-Verlag, 2001.

Data Preparation for Data Mining, Dorian Pyle, Morgan Kaufmann, 1999.

Métodos de Avaliação

Resolução de exercícios e um projecto final

Língua de ensino

A cadeira é leccionada em Inglês caso estejam inscritos alunos que não dominem a língua portuguesa. Caso contrário é leccionada em português.