



**Ciências
ULisboa**

Aprendizagem Automática

Código: 425216

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

Estudo de diferentes paradigmas e métodos de aprendizagem automática.

Após uma introdução geral aos conceitos fundamentais na área e apresentação de métodos clássicos de aprendizagem supervisionada e não supervisionada, abordam-se em particular métodos para aprendizagem de sequências de ações, como a aprendizagem por reforço.

Pretende-se que os alunos sejam capazes de investigar sobre um tema muito recente relevante mas que não é dado e que façam uma exposição crítica desse tópico ou técnica de aprendizagem, fomentando competências de investigação. Pretende-se que o aluno adquira a capacidade de desenvolver novas aplicações, envolvendo a programação de algoritmos ou a sua combinação e extensão.

Pré-requisitos

- Introdução às Probabilidades e Estatística (22701)
- Desenvolvimento Centrado em Objetos (425189)

Conteúdos

- Introdução aos diferentes paradigmas da aprendizagem
- Aprendizagem baseada em reforço
- Indução de árvores de decisão
- Indução de regras
- Aprendizagem baseada em instâncias
- Aprendizagem simplista de Bayes
- A avaliação da qualidade dos classificadores
- Agrupamento

Descrição detalhada dos conteúdos programáticos

Componente Teórica

- **Introdução aos diferentes paradigmas de aprendizagem** : Aprendizagem supervisionada e não supervisionada. Aprendizagem baseada em reforço.

- **Aprendizagem baseada em reforço:** Processos de decisão de Markov; O problema de aprendizagem baseada no reforço formulado como um problema de optimização num espaço de estados de Markov; O planeamento das melhores acções através da aproximação da função valor. O algoritmo de interação da função valor. Ilustração da técnica de iteração valor. Os algoritmos de aprendizagem baseados na diferenças temporais: o algoritmo-Q e o Sarsa. Exemplos da aplicação desses algoritmos; Exemplos de aplicações bem sucedidas: a pilotagem de um helicóptero e o jogador de gamão

- **Indução de árvores de decisão:** Conceito de árvore de decisão. O algoritmo geral de indução de árvores de decisão para atributos discretos. O conceito de entropia. O algoritmo ID3. O conceito de sobreajustamento em árvores de decisão. A podagem: pré e pos-podagem. Formas de lidar com atributos contínuos: discretização a priori e a discretização em tempo real. Fronteira de decisão para as árvores de decisão.

- **Aprendizagem baseada em regras:** O conceito de regra (proposicional). A transformação de uma árvore de decisão num conjunto de regras. Um sistema de regras visto como uma lista ordenada de regras, em que a precedência ou a resolução de conflitos é importante. A pre e pós-podagem de regras. Algoritmos directos de geração de regras usando a cobertura sequencial. O algoritmo PRISM: ilustração com exemplos. Discretização a priori ou em tempo real, como forma de lidar com atributos contínuos. Regras relacionais. O algoritmo FOIL, um algoritmo de cobertura para a aprendizagem de regras relacionais. Ilustração da execução do FOIL com exemplos.

- **Aprendizagem baseada em instâncias:** O conceito de lazy-learning. Aprendizagem baseada nos vizinhos mais próximos: o algoritmo K-nn e uma sua variante incremental. Funções de distância para atributos contínuos e discretos. A fronteira de decisão como um diagrama de Voronoi. Formas de contornar o sobre-ajustamento. A classificação ponderada pela distância aos vizinhos mais próximos e necessidade de discretização dos argumentos. Como procurar o melhor n usando a validação cruzada estratificada? A selecção e a ponderação dos pesos dos atributos como formas de contornar a maldição da dimensionalidade. O algoritmo RELIEF para definição dos pesos dos atributos. As árvores-KD para guardar de forma mais eficiente os dados de treino.

- **Aprendizagem Bayesiana:** A regra de Bayes e a definição de variável condicionalmente independente; A assunção da independência condicional do método simplista de Bayes; Versão contínua e discreta; A variação usando estimadores de Laplace; O uso de logaritmos para lidar com probabilidades muito reduzidas, próximas de zero.

- **Avaliação de desempenho dos classificadores:** O treino e o teste: divisão em teste e treino (1/3-2/3); validação cruzada, "leave-on-out", bootstrap e as suas versões repetidas; A taxa de sucesso e de erro; a matriz de confusão; precisão e recall, a medida-F e as curvas de roc; medidas dependentes do custo; como lidar com dados desequilibrados e como seleccionar os classificadores

- **Agrupamento (Clustering):** O algoritmo não optimal de partição mais clássico, o K-Means; Diferentes formas de escolha dos centros iniciais; Algoritmos de agrupamento hierárquico aglomerativos e divisivos; Descrição geral do método aglomerativo de agrupamento; diferentes formas de cálculo de distâncias entre grupos: vizinhos mais próximos, mais distantes e centróide. Representação da hierarquia de agrupamentos através de um dendograma; exemplos de aplicação.

Componente Teórica-Prática

Resolução de exercícios sobre a matéria das aulas teóricas.

Uso da biblioteca de algoritmos Bayesianos de aprendizagem Weka.

Uso da linguagem NetLogo para o projecto de Aprendizagem por Reforço

Bibliografia

Recomendada

- Reinforcement Learning: an Introduction, R. S. Sutton and A. Barto, MIT Press, 1998.
- Data Mining - Practical Machine Learning Tools and Techniques with Java Implementations, Second edition, Ian H Witten e Eibe Frank, Morgan Kaufmann, 2005.
- Machine Learning, Tom M. Mitchell, McGraw-Hill, 1997

Outros elementos de estudo

- notas de apoio às aulas teóricas.
- Introduction to Machine Learning, Ethem Alpaydyn, MIT Press, 2004.

Métodos de Avaliação

- Projecto de Aprendizagem por Reforço (Netlogo): 20%
- Projecto de classificação de dados usando o Weka: 20%
- Um projecto de desenvolvimento e de aplicação de algoritmos de aprendizagem: 30%
- Apresentação oral de 20m de um artigo/tema: 15%
- Exame: 15%

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.