

Projeto 1 – Prof. Murilo Naldi

TÉCNICAS ALTERNATIVAS DE CLASSIFICAÇÃO

Projeto 1

- ❑ Mesmos grupos do seminário
- ❑ Faça o *download* de três conjuntos de dados a sua escolha do repositório de aprendizado de máquina da UCI
<http://archive.ics.uci.edu/ml/> ou Weka
- ❑ O projeto deverá possuir **3 fases!**
 - Pré-processamento
 - Classificação
 - Comparação dos resultados

Parte 1: pré-processamento

- Explore os dados utilizando gráficos, visualizações, correlações ou qualquer outra técnica da Aula 03 que achar pertinente
 - Dica: R project (<http://www.r-project.org/>)
- Depois da análise, aplique a seleção de atributos/objetos ou normalizações que julgarem ser apropriadas

Parte 2: classificação

- Aplique árvores de decisão e a técnica discutida em seu seminário para classificar os conjuntos de dados
 - Não é necessário implementar!
 - Dica: Weka (<http://www.cs.waikato.ac.nz/ml/weka/>) ou R project
- Para isso, deve ser utilizado algum método de amostragem (*holdout*, *crossvalidation*, *leave-one-out...*) para validar a indução

Parte 3: comparação

- Compare os resultados obtidos
 - Em termos de qualidade (precisão) e
 - Se possível, em termos de tempo computacional
- Calcule **média e desvio** da amostragem
- Utilize curvas ROC e/ou testes de hipóteses para comparar as ADs com as técnicas estudadas pelo grupo
 - Dica: ferramentas como o Weka e R project para auxiliá-lo

Apresentação

- Todo o trabalho deve ser documentado em detalhes em um arquivo do tipo pdf
- A apresentação do trabalho será:
 - Dia 05/06 – Regras
 - Dia 09/06 – Vizinhos mais próximos
 - Dia 12/06 – Bayseanos
 - Dia 16/06 – RNAs

Apresentação

□ Objetivo:

- Apresentar a aplicação da técnica, cada uma das três fases em detalhes
- Mostrar que o grupo domina o processo de descoberta de conhecimento
- A nota de apresentação será dada individualmente

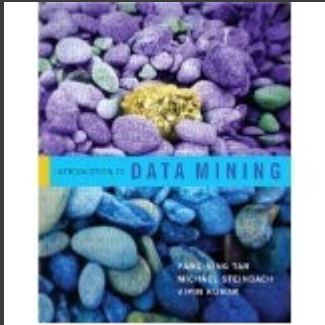
Duração

- ▮ Cada grupo deverá apresentar o tema entre 60 e 100 minutos
- ▮ Apresentações que saiam deste intervalo terão suas notas penalizadas
- ▮ Os membros do grupo devem dividir o tempo de apresentação de maneira uniforme
- ▮ Desvios no horário poderão ser penalizados

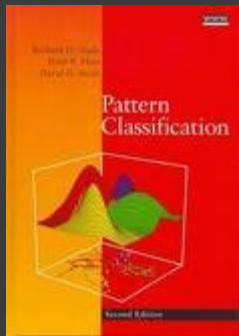
Notas

- ▮ A nota dada ao seminário será **individual** e dividida da seguinte forma:
 - Domínio conteúdo
 - Didática
 - Bom uso do tempo
- ▮ O trabalho deverá ser entregue no PVAnet (apresentação + pdf) no dia da apresentação

Referencias



STEINBACH, M., KUMAR, V. TAN, P.
Introdução ao Data Mining
(Mineração de Dados). Edição 1.
Ciência Moderna 2009. ISBN
9788573937619.

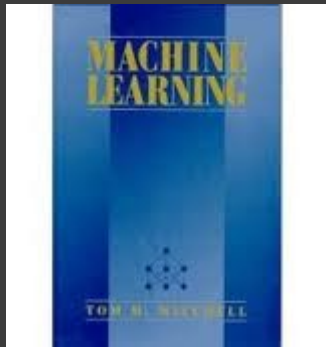


Duda, R.O., Hart, P. E. and Stork, D.
G. Pattern Classification (2nd Edition). Wiley-Interscience

Referencias



SILVA, I.N., SPATTI, D.H. e
FLAUZINO, R. A. Neurais Artificiais
para engenharia e ciências aplicadas:
curso prático. Edição 1 Artliber 2010
ISBN 9788588098534.



MITCHELL, T. Machine Learning
Edição 1 McGraw Hill 1997 ISBN
9780070428072