

Primeira Lista de Exercícios

SI-IA

Data: 04/03/13
Prof. Flávio Varejão

Aluno:.....

1. Mineração de Dados

O processo de mineração de dados é composto por várias etapas. Alguns autores a subdividem nas seguintes etapas: entendimento do domínio da aplicação, seleção de dados, pré-processamento, transformação, escolha da tarefa de MD, escolha dos algoritmos de MD, extração de padrões, interpretação dos padrões e utilização do conhecimento. Escolha três dessas tarefas e as descreva sucintamente, especificando seu papel, suas entradas, técnicas utilizadas e os resultados esperados.

2. Conceitos

a) Apresente e discuta medidas que podem ser tomadas para reduzir a possibilidade de *overfitting* e pouca estabilidade (alta variância) dos resultados na avaliação experimental de algoritmos de aprendizado.

b) Qual medida deve ser tomada para comparar resultados de avaliações que tenham utilizado bases com diferente número de exemplos?

c) Diferencie aprendizado supervisionado de não supervisionado.

3. Pré-processamento

a) Explique o que é e para que serve a normalização de atributos. Apresente e explique uma técnica que pode ser utilizada.

b) Explique como funcionam as discretizações com particionamento por larguras iguais e com particionamento por frequências iguais.

c) Enumere as três maneiras de se reduzir o volume de dados usado em um processo de data mining. Apresente uma técnica para cada uma das maneiras enumeradas anteriormente, explicando-as.

4. Classificação

a) Explique uma maneira geral com a qual um algoritmo de árvore de decisão pode lidar com atributos contínuos.

b) Proponha uma estratégia geral para decidir se vale a pena podar um ramo de uma árvore de decisão durante a pós-poda.

c) Justifique porque é recomendável normalizar os atributos antes de aplicar o método dos k vizinhos mais próximos e também utilizar $k > 1$.

5. Meta-heurísticas

Considere um problema no qual se deseja alocar m objetos a n grupos. Leve em conta a existência de uma função FC que oferece o custo para uma dada distribuição dos m objetos nos n grupos.

a) Apresente uma estrutura de um cromossoma para esse problema.

b) Descreva um algoritmo genético simples (usando um algoritmo abstrato) para realizar essa tarefa. Não é preciso descrever como as operações do algoritmo são implementadas para o problema.

c) Mostre através de exemplos baseados no problema de alocação acima o que são e como funcionam as operações de crossover, mutação, seleção baseada no método da roleta e elitismo.

6. Meta-heurísticas

Classifique cada uma das metaheurísticas citadas abaixo em uma das seguintes categorias: (1) baseadas em soluções parciais, (2) busca local e (3) busca populacional. Escolha uma metaheurística de cada categoria e explique como funciona.

deepest descent ()

ant colony ()

<i>simulated annealing</i>	()
<i>branch and bound</i>	()
<i>grasp</i>	()
<i>a*</i>	()
<i>beam search</i>	()
<i>multistart descent</i>	()
<i>hill climbing</i>	()
<i>tabu search</i>	()
<i>genetic algorithms</i>	()
<i>simple descent</i>	()