

**Novas Tecnologias de Informação**  
**Licenciatura em Estatística e Gestão de Informação**  
**Cotação: Grupo I - 1 valor cada; Grupo II-1,4, Grupo III-1,2,2**  
**ATENÇÃO: Cada pergunta de escolha múltipla errada desconta 0.4 valores**  
**Duração: 2 horas Teste A**

**Doutor Victor Lobo**

**Ano lectivo: 2003-2004 – 3ª CHMADA**

---

**I**

Escolha uma e uma só resposta para cada uma das seguintes questões

**I.1)** O problema do caixeiro viajante (TSP-Traveling Salesman Problem) tem muitíssimas aplicações, mas infelizmente é um problema de difícil resolução. Qual das afirmações seguintes é **VERDADEIRA** ?

- a) O SOM (Self Organizing Map) pode ser usado para resolver o problema do caixeiro viajante de forma óptima.
- b) O SOM (Self Organizing Map) serve para fazer aprendizagem não supervisionada e clustering, e como tal não pode ser usado para resolver o problema do caixeiro viajante.
- c) O SOM (Self Organizing Map) pode ser usado para obter uma solução para o problema do caixeiro viajante, mas o trajecto obtido não é necessariamente o mais curto.
- d) O SOM (Self Organizing Map) pode ser usado para obter uma solução para o problema do caixeiro viajante, mas o trajecto obtido poderá ser mais curto que o óptimo.

**I.2)** Assumindo que os eleitores fazem decisões racionais baseadas em dados objectivos sobre a economia e a política externa, poder-se-ia fazer um programa para, dada essa informação, prever qual dos dois candidatos à presidência dos EUA iria ganhar as eleições. Qual dos seguintes métodos de datamining, dados nas aulas, seria mais apropriado para um sistema desses?

- a) Redes neuronais auto-organizadas (Self Organizing Maps - SOM)
- b) Algoritmos genéticos
- c) Redes neuronais multicamada treinadas com retropropagação (MLP com BP)
- d) Simulated Anhealing

**I.3)** A maior parte das escolas portuguesas tem grande dificuldade em fazer os horários escolares. Na cadeira de NTI estudámos várias técnicas que podem ajudar a resolver esse problema. Qual das seguintes técnicas oferece melhores hipóteses de produzir uma boa solução para este problema?

- a) Árvores de decisão
- b) Redes neuronais multicamada treinadas com retropropagação (MLP com BP)
- c) Redes neuronais de Hopfield
- d) Algoritmos genéticos

**I.4)** Para treinar um SOM (Self Organizing Map) é necessário definir qual a função de vizinhança a usar. Qual das seguintes afirmações é **FALSA** ?

- a) Muitas vezes prefere-se uma função de vizinhança rectangular a uma função gaussiana pois é mais rápido, computacionalmente, treinar um mapa com uma função de vizinhança rectangular.
- b) A função de vizinhança deverá ser radial e monótona decrescente
- c) A função de vizinhança deverá ser radial e tender para zero quando o raio aumenta
- d) Uma função de vizinhança gaussiana normalmente dá origem a mapas mais reprodutíveis, i.e., mapas com menos variância entre eles

**I.5)** Quando se aplica uma técnica de aprendizagem automática para desenhar um classificador a partir um conjunto de dados conhecido, é vulgar dividir este em três sub-grupos: conjunto de teste, de treino, e de validação. Qual das seguintes afirmações é **FALSA** ?

- a) O conjunto de treino é normalmente o maior dos três conjuntos.
- b) O conjunto de validação serve para decidir quando se deve parar o treino
- c) O conjunto de teste não é usado na fase de desenho do classificador
- d) Há técnicas de desenho em aprendizagem automática que não necessitam do conjunto de teste.

**I.6)** Podem-se classificar a maior parte dos problemas de aprendizagem em problemas supervisionados e problemas não supervisionados. Qual das seguintes afirmações é **VERDADEIRA** ?

- a) Todos os problemas de aprendizagem não-supervisionada podem ser reformulados como problemas supervisionados
- b) Os problemas de regressão são casos particulares de problemas de aprendizagem não supervisionada.
- c) Os problemas de regressão são casos particulares de problemas de aprendizagem supervisionada.
- d) Os problemas de regressão podem ser formulados quer como problemas de aprendizagem supervisionada como de aprendizagem não supervisionada-

**I.7)** Qual das afirmações seguintes é **FALSA**:

- a) As árvores de decisão, no contexto em que foram dadas nesta cadeira, e ao contrário das redes neuronais, não necessitam de um conjunto de treino para serem construídas.
- b) Uma das vantagens das árvores de decisão é que são sempre mais simples (no sentido que exigem menos operações para chegar a um resultado) do que uma rede neuronal.
- c) As fronteiras entre classes definidas por redes neuronais podem ser sigmóides
- d) Uma das vantagens das árvores de decisão sobre outros métodos de classificação é que geralmente podem dar explicações para as suas classificações que são facilmente entendíveis por humanos.

**I.8)** Uma cadeia de hipermercados pretende abrir lojas mais pequenas orientadas para sectores de mercado mais restritas. Como base para essa segmentação de mercado, dispõe dos registos das máquinas registadores referentes às compras efectuadas pelos seus clientes. Aconselharia a utilização de:

- a) Uma árvore de decisão, em que usaria como variáveis de entrada o montante das compras que os clientes efectuam em cada tipo de produtos, e como variável de saída o segmento pretendido.
- b) Uma rede neuronal multicamada, treinada com backpropagation, em que usaria como variáveis de entrada o montante das compras que os clientes efectuam em cada tipo de produtos, e como variável de saída o segmento pretendido
- c) Uma rede neuronal SOM, em que usaria como variáveis de entrada o montante das compras que os clientes efectuam em cada tipo de produtos.
- d) Um algoritmo genético, em que usaria como variáveis de entrada o montante das compras que os clientes efectuam em cada tipo de produtos.

**I.9)** O algoritmo de “Simulated Annealing” tem sido muito popular porque:

- a) É baseado num modelo de temperaturas.
- b) Permite fazer aprendizagem supervisionada.
- c) É semelhante ao método das redes neuronais
- d) Muitas vezes consegue obter máximos ou mínimos globais de funções, ou pelo menos máximos e mínimos locais.

**I.10)** Pode-se obter uma estimativa do erro que um classificador terá ao classificar um dado novo calculando o nº de erros obtidos com esse classificador em conjuntos de dados conhecidos. Qual das afirmações é **FALSA**?:

- a) Todas as estimativas de erro obtidas a partir de dados conhecidos serão necessariamente optimistas.
- b) O erro no conjunto de validação produz uma estimativa que será geralmente mais optimista que a obtida num conjunto de teste independente.
- c) O erro no conjunto de treino produz uma estimativa optimista (erro real é superior à estimativa).
- d) O número de erros no conjunto de treino pode sempre ser reduzido a zero, desde que não haja dados contraditórios (i.e. dados que têm exactamente os mesmos valores para os atributos, mas classes distintas).

## II

Pretende-se desenhar um sistema para distinguir crianças de adultos quando estas entram num elevador. O elevador pode medir o peso da pessoa que entrou (que muitos elevadores já o fazem) e, usando um sensor de ultrasons no tecto, medir a sua altura. Mediram-se os dados realivos a um conjunto de pessoas, e obteve-se a seguinte tabela apresentada.

Por razões várias, foi decidido treinar um perceptrão com estes dados, de modo a que este consiga distinguir adultos de crianças com os dados vindos dos sensores. Para treinar o perceptrão, inicie todos os pesos sinápticos com o valor 1, e use um “bias” também de 1. Use depois a regra de aprendizagem do perceptrão simples dado nas aulas, e percorra uma vez todo o conjunto de treino. Apresente no fim os pesos sinápticos obtidos, e indique qual a taxa de erro no conjunto de treino, nesse momento.

Altura	Peso	Adulto ? (1=sim, 0=não)
167	70	1
135	40	0
170	60	0
180	85	1
153	65	1

## III

Imagine que começa a trabalhar num banco onde querem implementar um sistema de CRM (Customer Relation Management) que aconselhe o caixa sugerir (ou não) ao cliente que tem em frente a si, que faça um dado “Plano Poupança”. Para obter dados, analisou uma série de clientes e obteve os resultados apresentados na tabela.

Idade < 35	Idade > 50	Sexo	Usa gravata	Fez um plano de poupança
S	N	M	N	N
N	N	M	S	S
S	N	F	N	N
N	N	F	N	N
N	S	F	N	S
N	S	M	S	S

**III.1)** Usando o algoritmo DDT dado nas aulas, construa uma árvore de decisão para estimar se um dado cliente irá ou não aderir ao plano de poupança.

**III.2)** Num dado momento aparece um homem que aparenta ter 40 anos e não usa gravata. De acordo com a árvore de decisão obtida na alínea anterior, vale a pena sugerir-lhe que faça o plano de poupança ?

Bom trabalho !

