

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

**Doutor Moura-Pires & Doutor Victor Lobo**

**Ano lectivo: 2002-2003**

**NOME:**

**Nº:**

---

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

I.1) É possível aproximar com um erro arbitrariamente pequeno qualquer função contínua e diferenciável com uma rede neuronal multicamada “feedforward” desde que:

- a) A rede tenha pelo menos 3 neurónios
- b) A rede tenha pelo menos 2 camadas de neurónios
- c) As funções de activação sejam sigmóides
- d) Nenhuma das anteriores é por si só suficiente

I.2) As fronteiras entre classes, definidas por uma árvore de decisão, são:

- a) Um hiperplano.
- b) Vários hiperplanos, sendo cada um perpendicular a um dos eixos.
- c) Uma superfície qualquer cuja forma é definida pelas funções de activação
- d) Uma superfície qualquer com um número de máximos e mínimos dependente do número de nós da rede.

I.3) Qual das seguintes afirmações é falsa

- a) Dado um número suficientemente grande de nós, existe uma árvore de decisão que define fronteiras entre classes tão perto quanto se queira das fronteiras definidas por uma dada rede neuronal.
- b) Uma das principais vantagens em usar funções de activação sigmóides em redes de perceptrões multicamadas com retropropagação (backpropagation), é o facto da derivada destas funções ser muito fácil de computar.
- c) Uma rede neuronal RBF (usando funções de base radial) pode, dado um número suficientemente alto de neurónios separar quaisquer duas classes definidas através de um conjunto de exemplos, desde que as classes sejam de facto separáveis.
- d) Uma rede neuronal SOM (Self-Organizing Map), treinada com o algoritmo standard, separará sem qualquer erro duas classes desde que o número de neurónios seja igual ou superior ao número de exemplos disponíveis para treino.

I.4) 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) O erro no conjunto de treino produz uma estimativa otimista (erro real é superior à estimativa).
- b) O erro no conjunto de validação produz uma estimativa que será geralmente mais otimista que a obtida num conjunto de teste independente.
- c) Todas as estimativas de erro obtidas a partir de dados conhecidos serão necessariamente otimistas.
- 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).

I.5) Uma das técnicas de optimização estudadas foi o "simulated annealing". Qual das seguintes afirmações é falsa ?

- a) O algoritmo de Metropolis implementa essa técnica
- b) Esta técnica é muito parecida com a de "stochastic hill-climbing", com a única diferença que em Stochastic hill-climbing nunca se aceita uma solução cujo custo seja superior à solução já obtida.
- d) Uma solução óptima para um problema só pode ser obtida se o parâmetro "temperatura" do "simulated annealing" nunca chegar a valor zero, pois neste caso o algoritmo é idêntico a "stochastic hill-climbing".
- e) Se o problema em causa for diferenciável e tiver apenas um máximo e um mínimo, a solução óptima pode sempre ser obtida usando métodos determinísticos baseados no gradiente.

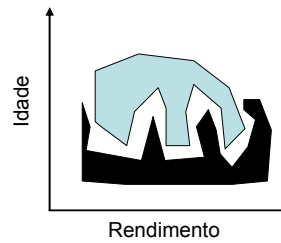
I.6) Num algoritmo genético, o operador mutação serve principalmente para:

- a) Combinar genes de diferentes soluções
- b) Atingir com menor número de iterações boas soluções
- c) Atingir soluções que incluem genes não existentes na população original
- d) Impedir o sistema de ficar preso em mínimos (ou máximos) locais

Cada um dos métodos de previsão estudado tem vantagens e desvantagens que dependem muito dos dados em questão. Normalmente os dados têm muitos atributos, tornando a sua visualização difícil. No entanto, se os dados forem bi-dimensionais, a visualização é trivial, e por vezes pode ser óbvio escolher o melhor classificador, quer em termos de taxa de erro quer em termos de facilidade de implementação. Nas questões seguintes, são apresentados gráficos com diversas distribuições de dados. Os dados (inventados) representam clientes de um banco, descritos através do seu rendimento bruto e da sua idade. Um perito dividiu os clientes que provavelmente irão aderir a uma campanha do banco (representados a escuro) e os que em princípio não estarão interessados (representados a claro). Para cada uma das situações indique qual o tipo de classificador mais apropriado.

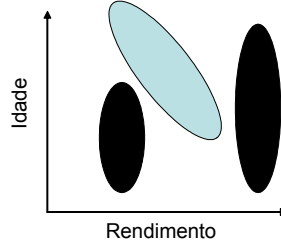
I.7)

- a) Perceptrão simples
- b) Perceptrão multicamada
- c) Árvore de decisão
- d) Rede neuronal RBF



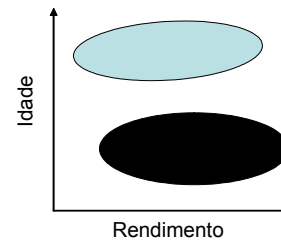
I.8)

- a) Perceptrão simples
- b) Perceptrão multicamada
- c) Árvore de decisão
- d) Rede neuronal RBF



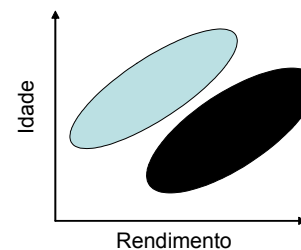
I.9)

- a) Perceptrão simples
- b) Perceptrão multicamada
- c) Árvore de decisão
- d) Rede neuronal RBF



I.10)

- a) Perceptrão simples
- b) Perceptrão multicamada
- c) Árvore de decisão
- d) Rede neuronal RBF



## II

Numa dado Instituto universitário os serviços académicos gostariam de ter a capacidade de prever se um dado aluno iria passar ou não numa cadeira, ainda antes de serem realizados os exames finais, a fim de planejar o ano seguinte. Naturalmente, a passagem do aluno depende da classificação nesse exame, mas uma previsão, ainda que sujeita a erros (de preferência quantificáveis) seria muito útil. De entre os vários métodos de previsão estudados na cadeira de NTI, a pessoa responsável pelos serviços académicos prefere que seja usada uma árvore de decisão, com um máximo de 3 níveis. Para gerar essa árvore de decisão, dispõe de alguns dados relativos a alunos de anos anteriores. Esses dados estão sumariados na tabela apresentada :

II.1 ) Obtenha uma árvore de decisão com um máximo de 3 níveis usando o método "DDT" dado nas aulas, e usando como medida de discriminação usada nos exemplos também dados nas aulas:  $f(A) = \frac{\sum C_i}{n}$  onde A é um atributo,  $C_i$  é p número de exemplos correctamente classificados pela classe mais frequente, e n é número total de exemplos pertencente ao grupo em causa.

II.2) Usando como base a árvore obtida indique quais são os atributos mais importantes e quais são os menos relevantes para prever se um dado aluno vai passar ou não.

II.3) Usando a árvore obtido na alínea 1, preveja se os seguintes alunos vão passar ou não:

Nome	Canta no coro da UNL	Entregou os trabalhos a horas	Assistiu a todas as aulas	Teve mais que 16 a Análise de dados	Teve mais que 16 a Econometria
Xerxes	S	N	N	S	S
Yuri	N	S	S	S	N
Zeferino	S	S	N	N	S

Nome	Canta no coro da UNL	Entregou os trabalhos a horas	Assistiu a todas as aulas	Teve mais que 16 a Análise de dados	Teve mais que 16 a Econometria	Passou na cadeira em causa
Abel	S	S	S	N	N	S
Belmiro	S	S	N	S	N	S
Caius	N	S	S	N	N	N
Délio	N	N	S	S	N	N
Ernesto	S	N	S	N	N	S
Felisberta	N	S	S	N	S	N
Geraldina	N	N	S	S	N	N
Hermegildo	S	N	N	S	N	S
Ildefonso	N	S	N	N	S	N
Jamila	N	S	N	S	S	N
Kermit the frog	N	S	S	S	S	S
Ludmila	S	S	N	N	N	S
Maria	S	N	S	S	S	S
Neo	S	S	S	S	S	S
Óscar	N	N	N	S	S	N

III.1) Os mapas auto-organizados de Kohonen (SOM) são muito usados para visualização de dados multidimensionais. Explique porquê, e compare os resultados obtidos com o SOM com os resultados que se obtêm projectando os dados sobre as duas primeiras componentes principais.

III.2) Explique que interações existem entre os diferentes neurónios de um SOM durante a fase de treino

III.3) Dê um exemplo de um problema prático onde um SOM pode ser usado, explicando como usaria as diversas variáveis em jogo.

Boa sorte !

