

Avaliação de Reconhecimento de Padrões - Prof. Braga - 24/05/2017

Questão 01

Para a resolução desta questão, considere a base de dados de doenças do fígado, "BUPA liver disorders", que possui 5 variáveis, cujo arquivo encontra-se no Moodle e cuja sugestão de leitura em código R é apresentada a seguir.

```
> rm(list=ls())
> bupa <- as.matrix(read.csv("bupa.data", header=FALSE))
> xall<-bupa[, (1:6)] # dados de entrada
> yall<-as.matrix(2*(bupa[,7]-1.5)) # rótulos das classes como [-1,+1]
> i_cm1<-which(yall == -1) # amostras da classe -1
> i_c1<-which(yall == 1) # amostras da classe +1
> Nm1<-length(i_cm1) # tamanho da classe -1
> N1<-length(i_c1) # tamanho da classe +1
```

A resolução desta questão envolverá as seguintes etapas: análise dos dados, avaliação da coerência da rotulação e resolução do problema de classificação correspondente, conforme descrito a seguir:

- Análise dos dados:
 1. Calcular e **analisar** as correlações entre as variáveis de entrada, incluindo a variável de saída por meio da função `pairs()`;
 2. Analisar e concluir sucintamente;
- Coerência da rotulação:
 1. Utilizando um método de agrupamento (*clustering*) de dados, Kmédias, por exemplo, avalie a **coerência espacial** entre os rótulos fornecidos e os agrupamentos de dados encontrados. Os rótulos são coerentes com os agrupamentos encontrados pelo método de **clustering**? Em outras palavras, padrões de mesma classe são atribuídos ao mesmo cluster?
 2. Analisar e concluir sucintamente;
- Classificação:
 1. Utilizando um método de classificação à sua escolha, resolva o problema de classificação caracterizado por esta base de dados, separando os dados em treinamento e testes.
 2. Analisar e concluir sucintamente;