

Trabalho Prático nº3 – Codificação em binário de texto e códigos de Huffman

PARTE A

O ficheiro `datatrab` contém duas variáveis: `simbolos` e `text`. Depois de colocar o ficheiro no directório de trabalho executa os comandos

```
load datatrab
k=1
sum(text==simbolos(k))
```

1. Explique o conteúdo das variáveis do ficheiro respondendo às perguntas
 - Qual é a classe das variáveis do ficheiro? Quantos bits o MATLAB/Octave utiliza na respectiva codificação?
 - Qual é o resultado da última linha de código? O que significa?
 - Que valores pode ter a variável `k` sem haver erro na linha seguinte?
 - Qual é o conteúdo da variável `simbolos`.
 - Qual o número total de símbolos na variável `text`? Quantos caracteres diferentes existem na variável `text`? E quais são?
2. Qual é a probabilidade de cada um dos símbolos (caracteres) ocorrer na variável `text`?
3. A tabela seguinte contém duas propostas para codificar os símbolos de `text` em binário.

símbolos	código A	código B
\$	000	0
%	001	10
#	010	110
&	011	1110
”	100	11110
,	101	111110
!	110	1111110
(111	11111110

- (a) Os códigos são instâneos e não-ambíguos? Que símbolos correspondem à seguinte sequência de bits: {010101100011110}.
- (b) Para o *código B* construa a respectiva árvore binária de codificação.
- (c) Quantos bits precisa para representar **text** em binário com cada um dos códigos?

	Código A	Código B
total de bits de text		

- (d) Qual é o código mais eficiente (o que produz menor número de bits)? Justifique porquê.
- (e) Para cada caso calcule o número médio de bits por símbolo.
4. Calcule a entropia de **text**. Qual dos códigos aproxima melhor a entropia da variável **text**?
5. Construa um código de Huffman para codificar os símbolos da tabela assumindo que a probabilidade dos símbolos são as estimadas para os símbolos de **text**.