

Universidade do Minho
Mestrado em Engenharia Biomédica
Ano Lectivo de 2011/2012
Problemas

1. Considere a seguinte imagem, onde os valores (0,20,50,99) indicam o nível de cinzento de cada pixel **(4 valores)**

```
99 99 99 99 99 99 99 99 99
20 20 20 20 20 20 20 20 20
0  0  0  0  0  0  0  0  0
0  0 50 50 50 50 50  0  0
0  0 50 50 50 50 50  0  0
0  0 50 50 50 50 50  0  0
0  0 50 50 50 50 50  0  0
0  0  0  0  0  0  0  0  0
```

- a. Qual é a entropia da fonte?
 - b. Mostre passo a passo como construir a árvore Huffman para codificar os 4 valores acima. Mostre o código para cada valor de intensidade.
 - c. Qual é o número médio de bits necessário para codificar cada pixel, usando o código de Huffman? Qual a relação com a entropia? Explique.
2. O cálculo de PageRank é baseado no algoritmo do passeio aleatório modificado para a presença de nós sem saídas.
- a. Descreva sucintamente o algoritmo do passeio aleatório. Apresente o respectivo pseudo-código.
 - b. Apresente a matriz Google para um grafo com os nós A,B,C e D; o nó A aponta para todos os outros; o B aponta para o C e D; o C aponta para o D e para o A.
 - c. Sem fazer qualquer cálculo, qual lhe parece ser o nó com maior PageRank? Porquê?
3. Suponha que temos dois sistemas de RI que devolvem para uma determinada necessidade de informação as seguintes respostas. Existem 13 documentos relevantes na colecção

Sistema 1: RRNNN RRNRN NRNRN
Sistema 2: RNNRN RRNNN NNNRR

- a. Qual a Precisão interpolada para cada sistema para 50% de Recall?

- b. Calcule a Precisão, Recall e F1 para cada sistema
 - c. Calcule o respectivo MAP e o R-Precision.
 - d. Que sistema considera melhor? Justifique a sua resposta
4. Suponha que temos um conjunto de treino com três documentos representados pelos vectores $d1 = (0, 2, 3, 1)$, $d2 = (1, 0, 4, 5)$ e $d3 = (2, 3, 0, 1)$. Os documentos 1 e 2 são da classe 1 e o documento 3 da classe 2.
- a. Considere o documento $d4 = (1, 2, 0, 0)$
 - b. Classifique-o usando o método Naive Bayes
 - c. Classifique o documento $d4 = (1, 2, 0, 0)$ usando o método 1NN¹.
 - d. Porque se diz que o método 1NN é menos estável que o 3NN ou 5NN?

¹ Nearest Neighbour (Vizinho mais próximo)