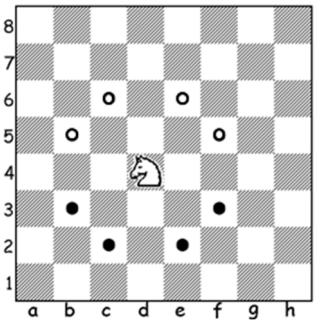
Open book exam. Duration: 2h30m. Exame da Época Normal

Note: Answer each question (1, 2, 3 and 4) in separate exam sheets (1, 2, 3 e 4).



1. [**4 points**] In a chess board, a knight moves through "L jumps", according to the figure: the horse in position *d4* may move, with a single jump, to positions *b3*, *b5*, *c2*, *c6*, *e2*, *e6*, *f3* or *f5*. To make it simple, the horse will only be allowed to make "forward" jumps: for instance, from position *d4* it may only jump to *b5*, *c6*, *e6* or *f5*.

a) Using a **depth first search** strategy, show a search tree and point out the solution found for the horse to move from position *a1* to position *f8*.

b) In this scenario, and considering the goal of moving the horse from a starting point to a destination point, which blind search strategies are complete? Justify your answer.

c) Consider the following heuristic functions, which try to estimate the number of jumps that are needed to go from the current position $i = x_i y_i$ to a goal position $f = x_f y_f$:

$h_1 = |x_f - x_i| + |y_f - y_i|$ *(Manhattan distance)*

$h_2 = \max(|x_f - x_i|, |y_f - y_i|)$ *(Chebyshev distance)*

$h_3 = \max(|x_f - x_i|, |y_f - y_i|)/2$

$h_4 = \min(|x_f - x_i|, |y_f - y_i|)$

$h_5 = \min(|x_f - x_i|, |y_f - y_i|)/2$

Which of these heuristic functions are admissible? Why? From those, which is the best one? Why?

d) Using the heuristic function identified when answering the previous question, show a search tree obtained by the **A*** search strategy, in order to move the horse from position *a1* to position *f8*. Next to each node in the tree, indicate the value of the cost function components (f=g+h). Identify the solution found.

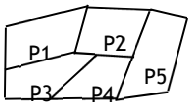
2. [**4 points**] The map coloring problem consists of assigning different colors to countries that share a border. The figure represents the map to color, which includes 5 countries (P1..P5).

There are 3 available colors: amber (A), violet (V) and black (B).

We want to employ **Genetic Algorithms** for solving this problem. Admit the existence of an initial population with 5 individuals with the following information: *colorP1, colorP2, colorP3, colorP4, colorP5*

I1: B,V,B,A,V I2: A,V,V,A,B I3: B,V,A,V,V

I4: B,B,B,B,V I5: B,B,V,B,V



a) Propose a structure for representing an individual. Explain. Exemplify by representing individual I1 from the initial population.

b) Propose a fitness function (textual description). Calculate the fitness values for each individual in the initial population.

c) In the process of selecting the individuals to use when forming the next generation, an elitist policy is employed (only for the best individual). Consider that the following random numbers have been generated (between 0 and 1): 0.7 / 0.35 / 0.15 / 0.81. Show the result of this selection process. Explain.

d) Compute the second generation of the population, explaining all your choices. Suggest a crossover strategy. The crossover probability is 75% and the following random numbers have been generated: 0.81 / 0.41 / 0.24 / 0.88. The mutation probability is 2% and only in the 15th random number a value lower than 0.02 has been obtained.

Open book exam. Duration: 2h30m. Exame da Época Normal

3. [**4 points**] In a pet shop, a robot is in charge of feeding fish; food is different for red and blue fish. The robot will learn to differentiate these two fish species by analyzing body parts, using the dataset in the table. *(Note: when answering to the following questions, take into account the existence of an unknown value.)*

Fin	Tail	Body	Fish
wide	big	fat	red
wide	small	slim	red
wide	big	fat	red
thin	small	slim	red
thin	small	slim	blue
thin	--	fat	blue
wide	small	fat	blue
thin	small	fat	blue

a) Calculate the values of the average information ("informação média") for the examples given.

b) Which attribute would you choose for the root of the decision tree, using the gain ratio criterion? We know that entropy(Fin)=0.811, entropy(Tail)=0.6 and entropy(Body)=0.95. Show all your calculations

c) Build the decision tree, considering that the error ratio in any leaf must be lower or equal to 0.4. Show every calculation and/or properly explain your answer.

4. [**8 points**] Answer **six (6)** of the following seven (7) questions (each in 5-10 lines).

a) The IDA* heuristic search algorithm uses an iterative deepening search in which the expansion limit is given by a cost instead of a depth. How should this cost limit be incremented, so as to ensure that the optimal solution will be found?

b) A student has prepared for an exam, which is indicative that she will pass (Belief=0.6); the same conclusion is supported (Belief=0.7) by the fact that she likes the respective course. On the other hand, she slept badly, which might indicate that she will fail the exam (Belief=0.3). According to the **Dempster-Shafer** model, which is the interval of belief for passing the exam?

c) In the optimization algorithm known as "simulated annealing", relate the probability of choosing a successor state with the temperature and with the difference of values between the current state and that successor state.

d) Alfa-beta cuts, when used in the minimax adversarial search algorithm, can reduce the number of examined nodes from $O(b^m)$ to $O(b^{m/2})$, where b is the branching factor and m is the maximum depth of the search tree. Explain in which circumstances this is possible.

e) Explain the compromise inherent to the usage of the "Naïve Bayes" formula.

f) "Two oranges plus four pears is six fruits." Sketch a DCG to interpret sentences like this one, where we want to validate operations over elements whose outcome is a superclass of the parcels. State which kinds of information would be needed in the knowledge base.

g) Explain the concept of *linearly separable function* and relate it with the output formula of a perceptron.