國立陽明交通大學試題紙

114 學年度第一學期
博士班資格考

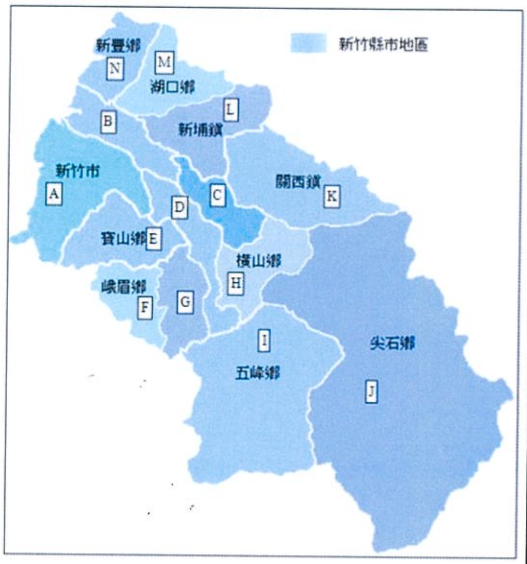
科目：人工智慧 A　　　　日期：115 年 1 月 6 日　

請 "✓" 明　　✓不可看書　　可看書

* 請將答案依題號順序寫入答案卷

答題時字跡需工整，否則不予計分。Write your answers legibly; otherwise you will get zero score.

1. [15%] Constraint Satisfaction Problem (Map-Coloring): This is the map of the Hsinchu area, with the regions represented by letters A to N. We want to assign numbers 1, 2, 3, or 4 to the regions, with adjacent regions having different numbers.

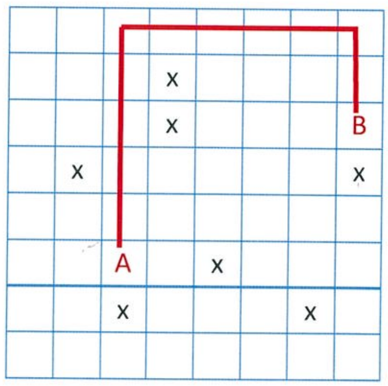


(a) Draw the constraint graph.

(b) Assume that we want to solve this problem using backtracking search with forward checking. Three commonly used heuristics are MRV (minimum remaining values), Degree Heuristic, and LCV (Least Constraining Value). Explain what these three heuristics mean in CSPs.

(c) According to the three heuristics in (b), applied in the order of MRV, Degree, and LCV, list the first three regions to be colored. You need to provide your reasoning. (To break ties, when multiple regions are equally preferred, choose according to the alphabetical order.)

2. [15%] In chess, a rook can move on a chessboard any number of squares along a straight line, vertically or horizontally, but cannot jump over other pieces. The problem is to move a rook from square A to square B in the smallest number of moves. For the figure shown to the right, the red line is a 3-step solution (the x's mark other pieces).



(a) What is the maximum branching factor from any square? Assume an 8x8 board.

(b) Assume that graph-search is used so each square is visited at most once, count the number of squares reachable from A at each different step count using breadth-first search.

(c) Describe the general criterion for node selection in A* search.

(d) Design an admissible heuristic for this problem. You need to provide your rationale (what constraint is removed) and also explain why it is admissible.

(e) Using A* search with your heuristic in (d), determine the first node to be selected from A. (If multiple nodes are equally preferred, list them all.)

3. [20%] Answer the following questions concisely.

(a) When we say a search algorithm is <u>complete</u>, what does it mean?

(b) When we say an inference rule is <u>sound</u>, what does it mean?

(c) Describe the inference rule "Resolution" in propositional logic.

(d) Why do we need evaluation functions in minimax game tree search?

國立陽明交通大學試題紙

114 學年度第一學期
博士班資格考

科目：人工智慧 B　　　　日期：115 年 1 月 6 日　第 1 頁 共 2 頁

請 "✓" 明　　✓不可看書　　可看書

* 請將答案依題號順序寫入答案卷

答題時字跡需工整，否則不予計分。Write your answers legibly; otherwise you will get zero score.

1. [50%] **Face Detection and Supervised Learning.** In recent years, machine learning algorithms have been widely applied in our daily lives. For example, smartphone camera apps use face recognition technology to automatically detect the locations of faces, adjust facial brightness, and thus take better-looking photos. Supervised learning is the key technique that enables this functionality. First, we collect a large number of face and non-face images x, each associated with a corresponding label y. If $y = 1$, the corresponding image contains a face; if $y = -1$, the image does not contain a face. Next, the training objective of supervised learning is to minimize the objective function $\min_{\theta} \sum_{i=1}^{N} L(h_\theta(x^{(i)}), y^{(i)})$, where N denotes the number of training samples. The loss function L is used to measure the discrepancy between the prediction $h_\theta(x^{(i)})$ produced by the face recognition model and the ground-truth label $y^{(i)}$. One common approach to optimizing the parameter θ is gradient descent, which iteratively computes gradients to search for a local minimum. For the binary cross-entropy loss function, the logarithm (log) is defined in the standard sense as the natural logarithm ($\log_e$).

(1) (8%) Which of the following loss functions is not differentiable with respect to the variable p?

(a) $L = \log(1 + \exp(-y \cdot p))$

(b) $L = I(-y \cdot p \leq 0)$, where I is an indicator function

(c) $L = -y \log(p)$

(d) $L = \max\{1 - y \cdot p, 0\}$

(2) (8%) What is the derivative of the function $L = -y\log(p) - (1-y)\log(1-p)$ with respect to the variable p?

(3) (8%) Let's assume $p = \frac{1}{1+e^{-\theta}}$. What is the derivative of the function $L = -y\log(p) - (1-y)\log(1-p)$ with respect to the variable p?

(4) (8%) We define the following functions as $f = \sigma(\lambda x)$ and $\sigma = \frac{1}{1+e^{-\theta}}$. Let's assume the objective function to be the Hinge Loss, i.e., $L_{Hinge} = \max\{1 - y \cdot f, 0\}$. What is the derivative of the function L_{Hinge} with respect to f ?

(5) (8%) According to the chain rule, what is the derivative of L_{Hinge} with respect to λ ?

(6) (10%) One possible choice of h_θ is neural network. Let's assume the function is a two-layer neural networks with a hinge loss as follows. Please draw the corresponding computation graph and gradients of different operation (an example is shown in the following figure).

Note:

- Two-layer neural networks: $f_\theta(x) = \theta \cdot \sigma(\mathcal{V} \cdot \phi(x))$, where $\sigma = \frac{1}{1+e^{-\theta}}$ is the logistic function. Note that $\mathcal{V}$ is a matrix.
- Hinge loss: $\max\{1 - (f_\theta(x) \cdot y) , 0\}$