## The University of Sydney

SEAT NUMBER:
STUDENT ID:
$\qquad$

SURNAME: $\qquad$
OTHER NAMES: $\qquad$
SIGNATURE:
ENROLMENT:COMP3308

## SCHOOL OF INFORMATION TECHNOLOGIES

## COMP3308/3608

INTRODUCTION TO ARTIFICIAL INTELLIGENCE EXAMINATIONS SEMESTER ONE, 2011

## TIME ALLOWED: TWO HOURS

This examination paper comprises 12 pages
INSTRUCTIONS TO CANDIDATES

Answer all questions using blue or black pen on this examination paper in the spaces provided.
The paper comprises 10 questions each with multiple parts.
Questions are not worth equal marks. The mark awarded for each part is indicated. Marks total 50.
Brief precise answers, rather than long wordy descriptions, are required.

The following material is allowed:
One double-sided A4 page of your own notes. You do not need to leave this page with your examination paper.

THIS EXAMINATION PAPER IS NOT TO BE REMOVED FROM THE EXAMINATION ROOM.

OFFICE USE ONLY

| Q1 (6) | Q2 (3) | Q3 (6) | Q4 (8) | Q5 (6) | Q6 (4) | Q7 (3) | Q8 (4) | Q9 (5) | Q10 (5) | Total <br> $(50)$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
|  |  |  |  |  |  |  |  |  |  |  |

## Question 1 [6 marks, aech part is worth 1 mark]

Consider the following graph where $s$ is the start node, $g$ is the goal node, the arcs are labelled with their respective path costs, and the heuristic (h) values are shown next to each node.


For each of the search strategies below, list the nodes that are expanded (in the order of expansion) and the solution found. Assume that among the nodes with equal priority, the last added is used first.
(a) Depth-first search
(i) Expanded nodes:
(ii) Solution found:
(b) Breadth-first search
(i) Expanded nodes:
(ii) Solution found:
(c) Uniform cost search
(i) Expanded nodes:
(ii) Solution found:
(d) Iterative deepening search
(i) Expanded nodes:
(ii) Solution found:
(e) Greedy search
(i) Expanded nodes:
(ii) Solution found:
(f) $\mathrm{A}^{*}$
(iii) Expanded nodes:
(iv) Solution found:

## Question 2 [3 marks, each part is worth 1 mark]

Answer briefly and concisely.
a) A* uses admissible heuristics. What happens if we use a non-admissible one? Is it still useful to use A* with a non-admissible heuristic?
b) What is the advantage of choosing a dominant heuristic in A* search?

c) What is the main advantage of hill climbing search over A* search?

## Question 3 [6 marks]

Consider the following game in which the static scores shown at the tip node are all from the first player's point of view. Assume that the first player is the maximazing player.

a) Compute the final backed-up values using the alpha-beta algorithm and show your answer by writing each updated value at the appropriate nodes in the tree. Assume that the nodes are examined in left-to-right order. [4 marks]
b) Circle the nodes (in the figure above) that would not need to be examined using the alphabeta algorithm. [1 mark]
c) What move should the first player choose? [1 mark]

## Question 4 [8 marks, each part is worth 1 mark]

Answer briefly and concisely.
a) $1 R$ algorithm generates a set of rules. What do they test?
b) Gain ratio is a modification of Gain used in decision trees. What is its advantage?
c) Propose two strategies for dealing with missing attributes in learning algorithms.
d) Why do we need to normalize the attribute values in the k-nearest-neighbor algorithm?
e) What is the main limitation of the perceptrons?
f) Describe an early stopping method used in the backpropagation algorithm to prevent overfitting.

g) The problem of finding a decision boundary in support vector machine can be formulated as an optimisation problem using Lagrange multipliers. What are we maximizing?
h) In linear support vector machines, we use dot products both during training and during classification of a new example. What vectors are these products of?

During training:
During classification of a new example:

## Question 5 [6 marks]

Suppose you wish to recognize good and bad items produced by your company. You are able to measure three properties of each widget (P1, P2 and P3) and express them with Boolean values. You randomly grab several items from your shipping dock and extensively test weather or not they are good, obtaining the following results:

| P1 | P2 | P3 | result |
| :--- | :--- | :--- | :--- |
| Y | Y | Y | good |
| Y | Y | N | bad |
| N | N | Y | good |
| N | Y | N | bad |
| Y | N | Y | good |
| N | N | N | good |
| N | N | N | good |

a) Build the decision tree for this set of examples, showing your calculations. Draw the decision tree after selecting an attribute to split on.
b) How will the learned decision tree classify the following new example?
$\mathrm{P} 1=\mathrm{N}, \mathrm{P} 2=\mathrm{Y}, \mathrm{P} 3=\mathrm{Y}$
You may find this table useful:

| x | y | $-(\mathrm{x} / \mathrm{y})^{*}$ <br> $\log _{2}(\mathrm{x} / \mathrm{y})$ | x | y | $-(\mathrm{x} / \mathrm{y})^{*}$ <br> $\log _{2}(\mathrm{x} / \mathrm{y}$ |
| :--- | :--- | :--- | :--- | :--- | :--- |
| 1 | 2 | 0.50 | 1 | 6 | 0.43 |
| 1 | 3 | 0.53 | 5 | 6 | 0.22 |
| 2 | 3 | 0.39 | 1 | 7 | 0.40 |
| 1 | 4 | 0.5 | 2 | 7 | 0.52 |
| 3 | 4 | 0.31 | 3 | 7 | 0.52 |
| 1 | 5 | 0.46 | 4 | 7 | 0.46 |
| 2 | 5 | 0.53 | 5 | 7 | 0.35 |
| 3 | 5 | 0.44 | 6 | 7 | 0.19 |
| 4 | 5 | 0.26 |  |  |  |

## Question 6 [4 marks]

Given the training data in the table below where credit history, debt, collateral and income are attributes and risk is the class, predict the class of the following new example using 1 R classification: credit history=unknown, debt=low, collateral=none, income $=15-35 \mathrm{~K}$.

| credit <br> history | debt | collateral | income | risk |
| :--- | :--- | :--- | :--- | :--- |
| bad | high | none | $0-15 \mathrm{k}$ | high |
| unknown | high | none | $15-35 \mathrm{k}$ | high |
| unknown | low | none | $15-35 \mathrm{k}$ | moderate |
| unknown | low | none | $0-15 \mathrm{k}$ | high |
| unknown | low | none | over 35k | low |
| unknown | low | adequate | over 35k | low |
| bad | low | none | $0-15 \mathrm{k}$ | high |
| bad | low | adequate | over 35k | moderate |
| good | low | none | over 35k | low |
| good | high | adequate | over 35k | low |
| good | high | none | $0-15 \mathrm{k}$ | high |
| good | high | none | $15-35 \mathrm{k}$ | moderate |
| good | high | none | over 35k | low |
| bad | high | none | $15-35 \mathrm{k}$ | high |

## Question 7 [3 marks]

Suppose that you need to solve the binary classification problem (class 0 : white circles vs class1: black circles) shown below using a perceptron.

a) Draw a possible decision boundary. Draw the weight vector for this boundary clearly showing its direction and its relation to the decision boundary. [1 mark]
b) Can this problem be solved with a perceptron without a bias? Why? [2 marks]

## [4 marks]

## Question 8 - For all students [4 marks]

Use the k-means algorithm to cluster the following one dimensional examples into 2 clusters: $2,5,10,12,3,20,31,11,24$. Suppose that the initial seeds are 2 and 5 . The convergence criterion is met when either there is no change between the clusters in two successive epochs or when the number of epochs has reached five. At the end of each epoch show the two clusters.

## Question 9 [5 marks, each part is worth 1 mark]

You task is to develop a computer program to rate chess board positions. You got an expert chess player to rate 100 different chessboards and then use this data to train a backpropagation neural network, using board features as the ones shown in the figure below. Mark with "yes" all of the following issues that could, in principle, limit your ability to develop the best possible chess program using this scheme. Mark with "no" all issues that could not. All issues are to be either marked with "yes" or "no".

Board Features:


1. The backpropagation network may be susceptible to overfitting of the training data, since you tested its performance on the training data instead of using cross validation.
2. The backpropagation neural network can only distinguish between boards that are completely good or completely bad.
3. The backpropagation network implements gradient descent, so your network may converge to a set of weights that is only a local minimum rather than the global minimum.
4. You should have used higher learning rate and momentum to guarantee convergence to the global minimum.
5. The topology of your neural net might not be adequate to capture the expertise of the human expert.

## Question 10 [5 marks, each part is worth 0.5 marks]

In the figure below, the circles are training examples and the squares are test examples, i.e. we are using the circles to predict the squares. Two algorithms are used: 1-Nearest Neighbour and 3-Nearest Neighbour.


We want to recover the labels for the circles given some results of classifying the squares:

| Square Point | Using 1 nearest neighbors | Using 3 nearest neighbors |
| :---: | :---: | :---: |
| 1 | - | + |
| 2 | - |  |
| 3 | + | + |
| 4 | + | - |
| 5 |  | - |

What will be the class of the following examples? Write + , - or U for cannot be determined.

1) Circle L:
2) Circle I:
3) Circle H:
4) Circle E:
5) Circle K:
6) Circle C:
7) Square 6 using 1-Nearest Neighbour:
8) Square 6 using 3-Nearest Neighbour:
9) Square 3 using 1-Nearest Neighbour?
10) Square 5 using 1-Nearest Neighbour?

