CSCI3230 Fundamentals of Artificial Intelligence
Written Assignment 3
Due date: 23:59:59 (GMT +08:00), December 6, 2020

I. Questions

(a) Short Answer Questions [15 marks]

i. In binary classification, the confusion matrix consists of four elements which are True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). What does each mean? [4 marks]

ii. Explain what is generalization and specialization in the logical formulation of learning. [4 marks]

iii. What is overfitting in supervised learning? Discuss the Ockham's Razor principle and how it relates to overfitting. [7 marks]

(b) Decision Tree [35 marks]

There is a smart phone evaluation database in arff format as follows:

@RELATION smart phone evaluation
@ATTRIBUTE, price, {high, med, low}
@ATTRIBUTE, performance, {high, med, low}
@ATTRIBUTE, stability, {high, med, low}
@ATTRIBUTE, size, {small, big}
@ATTRIBUTE, service, {limited, fair, good}
@ATTRIBUTE, class, {normal, good}
@DATA

high,	high,	high,	big,	fair,	good
high,	low,	high,	small,	limited,	normal
high,	low,	low,	small,	limited,	normal
med,	high,	med,	big,	fair,	normal
low,	med,	high,	small,	fair,	good
high,	low,	low,	big,	fair,	normal
low,	med,	med,	small,	limited,	good
low,	low,	high,	big,	fair,	good
low,	med,	med,	big,	good,	good
high,	med,	med,	small,	limited,	normal

i. Suppose the training set contains P "good" examples and N "normal" examples. An attribute A divides the whole data set E into subset $E_1,\ldots E_v$ according to its values, where A can have distinct values. Suppose each subset E_i has P_i "good" examples and N_i "normal" examples. Write down the definition of Entropy and the equation to compute the Information Gain for attribute A. [8 marks]

ii. What attribute will you use to construct the root node of the decision tree? Why? [12 marks]

iii. Build and draw the decision tree. [15 marks]

(c) Logic [50 marks]

i. Convert the following sentences from Horn Clause Form to Conjunctive Normal Form (CNF) and prove that Kate is happy (i.e. *Happy(Kate)*) by Resolution. Show your substitution of variables clearly.

Any student who can pass the computer exam and win a prize is happy.
Any student who enjoys study or is lucky can pass all the exams.
Kate does not enjoy study but she is lucky.
Any lucky student can win a prize.

Based on the above descriptions, we have:
$R1: Pass(x, computer) \wedge Win(x, prize) \Rightarrow Happy(x)$.
$R2: Study(y) \vee Lucky(y) \Rightarrow Pass(y, z)$.
$R3: Lucky(w) \Rightarrow Win(w, prize)$.
$F1: \neg Study(Kate)$.
$F2: Lucky(Kate)$. [25 marks]

ii. Express each of the following statements (i to vi) as First Order Logic sentences using the symbols below (together with x, $\exists$, $\neg$, $\wedge$, $\vee$, $\Rightarrow$) :

Symbol	Meaning
British(x)	x is British
Malaysian(x)	x is Malaysian
MainDish(x)	x is main dish
Like(x,y)	x likes y
Golden(x)	x looks golden
EggFriedRice(x)	x is egg fried rice
Fruity(x)	x is fruity

[25 marks]

1) *Roger* is either British or Malaysian but not both.
2) *Pizza* and *Pasta* are main dishes that look golden.
3) British people do not like any main dish which is fruity.
4) A main dish which looks golden and is not fruity can be egg fried rice.
5) If *Roger* likes fruity egg fried rice, then *Roger* is not British.

II. Assignment Submission

You **MUST** complete this assignment by using any one of the computer text editors (e.g. MS Word, WordPad, iWork Pages… etc.) and then save the document to PDF format with A4 printable page size. Scan version of the hand written work is **NOT** accepted. Hand drawn Decision Tree is **Allowed** for assignment 3 submission.

Please limit the file size of the PDF file less than 1MB.

You **MUST** submit the PDF file to the submission system on our course homepage (within CUHK network), otherwise, we will **NOT** mark your assignment.

III. Important Points

You MUST STRICTLY follow these points:

a. You MUST strictly follow the submission guidelines.
b. Remember to type your FULL NAME, STUDENT ID on the assignment.
c. Late submission will NOT be entertained according to our submission system settings.
d. Plagiarism will be seriously punished.

IV. Late Submission

According to the course homepage, late submission will lead to marks deduction.

No. of Days Late	**Marks Deduction**
1	10%
2	30%
3	60%
4 or above	100%