CS 60050 - Machine Learning
Mid-Semester Examination, Fall 2016

Total marks: 40, Total Time: 2 hours

All questions are compulsory. Please answer all parts of the same question in one place.

1. Consider the *EnjoySport* concept learning task defined in Table 1.

Table 1: Data for Question 1

Sky	Temperature	Humid	Wind	Water	Forecast	EnjoySport
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Sunny	Warm	High	Strong	Warm	Same	Yes
Rainy	Cold	High	Strong	Warm	Change	No
Sunny	Warm	High	Strong	Cool	Change	Yes

(a) Give a minimum length sequence of training examples that produces the following version space.

S :{<? *Warm* *Normal* *Strong* *Cool* ? >}
G :{<? ? *Normal* *Strong* ? ? >}

[4 marks]

(b) How many additional examples are required to make the version space converge to the following target concept?

<? ? *Normal* *Strong* ? ? >

[2 marks]

(c) Give a bound on the number of examples required to learn the target function if the learner were not trained by a teacher, but instead used an optimal query strategy. More specifically, assume each training example in the sequence is generated by allowing the learner for suggest an instance, and the trainer to provide its classification. (Note the hypothesis space in this problem contains 973 distinct hypotheses.) Assume that the entire sequence of instances must be generated by the learner, and that the target concept might be any hypothesis in H. *[4 marks]*

2. The dataset in Table 2 will be used to learn a decision tree for predicting whether a mushroom is edible or not based on its shape, color and odor.

Table 2: Data for question 2

Shape	Color	Odor	Edible
C	B	1	Yes
D	B	1	Yes
D	W	1	Yes
D	W	2	Yes
C	B	2	Yes
D	B	2	No
D	G	2	No
C	U	2	No
C	B	3	No
C	W	3	No
D	W	3	No

(a) What is the entropy $H(Edible|(Odor = 1 \quad OR \quad Odor = 3))$? *[2 marks]*

(b) Which attribute would the $ID3$ algorithm choose to use for the root of the tree (no pruning)? *[2 marks]*

(c) Draw the full decision tree that would be learned for this data (no pruning). *[2 marks]*

(d) Suppose we have the validation set given in Table 3.

Table 3: Data for question 2(d)

Shape	Color	Odor	Edible
C	B	2	No
D	B	2	No
C	W	2	Yes

What will be the training set error and validation set error of the tree? Express your answer as the number of examples that would be misclassified. *[4 marks]*

3. Suppose you have the following training set with three boolean input x, y and z, and a boolean output U.

Table 4: Data for question 3

x	y	z	U
1	0	0	0
0	1	1	0
0	0	1	0
1	0	0	1
0	0	1	1
0	1	0	1
1	1	0	1

Suppose you have to predict U using a Naive Bayes classifier.

(a) After learning is complete what would be the predicted probability $P(U = 0|x = 0, y = 1, z = 0)$? *[2 marks]*

(b) Using the probabilities obtained during the Bayes Classifier training, what would be the predicted probability $P(U = 0|x = 0)$? *[2 marks]*

In the next two parts, assume we learned a Joint Bayes Classifier. In that case ...

(c) What is $P(U = 0|x = 0, y = 1, z = 0)$? *[3 marks]*

(d) What is $P(U = 0|x = 0)$? *[3 marks]*

4. (a) In a simplified game of *Minesweeper*, there are bombs placed on a grid; you do not know where. Assume that each square (i, j) independently has a bomb ($B(i, j) = true$) with probability b. What you can observe for a given square is a reading $N(i, j)$ of the number of bombs in the adjacent squares (i.e., the eight closest squares not including the square itself). The variables $N(i, j)$ can therefore take values 0 through 8. If a square has less than eight neighbors, such as on boundaries, its N has appropriately limited domain.

   i. For a one-dimensional 4 × 1 Minesweeper grid, draw a Bayesian Network showing all the variables and arcs. *[1 mark]*

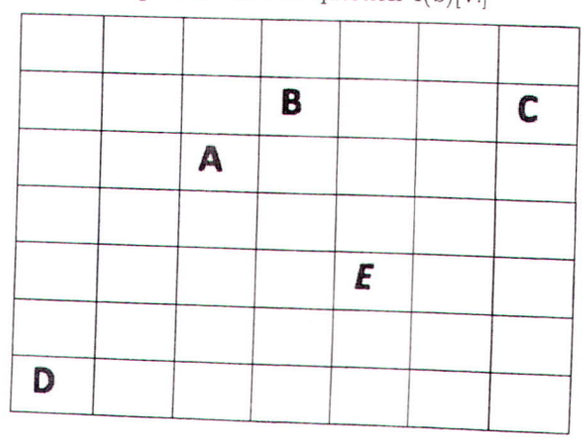
   ii. Fully specify the conditional probability tables for the variables $B(1, 1)$ and $N(1, 1)$, in terms of the bomb probability b. *[1 mark]*

   iii. What are probabilities of bombs in each of the four squares given no information? *[1 mark]*

   iv. If we observe $N(1, 2) = 1$ (for the second square), what are the probabilities of bombs in each of the squares? *[1 mark]*

   v. On the two dimensional grid in Fig. 1, assume we know the values of $N(A)$, $N(B)$, $N(C)$, and $N(D)$ and we are about to observe $N(E)$. Shade the squares whose bomb probabilities can change because of this new observation (Please try to be neat). *[1 mark]*

Figure 1: Grid for question 4(b)[v.]



(b) Draw the Bayesian networks that correspond to the following factorizations of the joint probability distribution $P(A, B, C, D)$:

   i. $P(B|A, C)P(A)P(C|D)P(D)$ *[1 mark]*

   ii. $P(A|B)P(C|B)P(B)P(D)$ *[1 mark]*

   iii. $P(D|C)P(C|B)P(B|A)P(A)$ *[1 mark]*

   iv. $P(B|A)P(A)P(C|D)P(D)$ *[1 mark]*

   v. $P(A)P(B)P(C)P(D)$ *[1 mark]*