


National University of Computer and Emerging Sciences, Lahore Campus

	Course Name:	Artificial Intelligence	Course Code:	CS 401
	Program:	BS(CS)	Semester:	Spring 2018
	Duration:	180 Minutes (2 Hour)	Total Points:	40
	Paper Date:	Thursday, 17th May 2018	Weight	40%
	Section:	D, E	Page(s):	7
	Exam Type:	Final		

Student : Name: _____ Registration No. _____

Instruction/Notes: Students are allowed to use a double sided A4 size cheat sheet
 Write your final answers in the space provided on the question paper.
 If needed you might attach of your rough work/sheets at the end.

Problem 1 [Short Questions]

[4, 4, 2, 2, 4 Points]

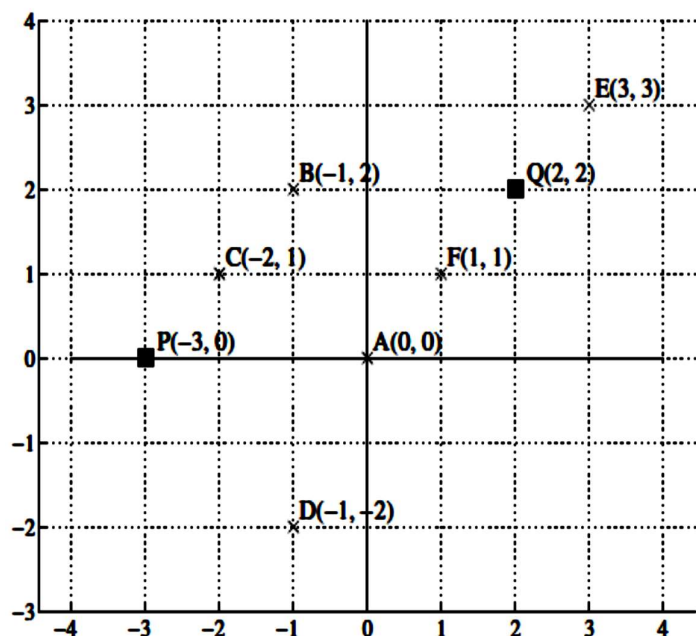
Part a)

In this question, we will use k-means clustering algorithm to cluster the points A, B, ... , F (indicated by x's in the figure below) into 2 clusters. The current cluster centers are P and Q (indicated by filled square in the figure).

Recall that k-means either requires a similarity measure or a distance function to perform the clustering. In this question you must use the following distance metric to do the clustering. Given two points, $A = (a_1 \ a_2)$ and $B = (b_1 \ b_2)$, the distance function is as follows $d(A; B) = (a_1 - b_1)^2 + (a_2 - b_2)^2$

i) Which points will get assigned to the cluster with center at P?

ii) What will be new cluster centers after the centers are updated?



Part b) To fit an ID3 based decision tree, assume we have three binary input attributes, A, B, C, and one binary outcome attribute Y. Training dataset for this problem is shown in the table below.

Which attribute (**A, B, or C**) has the highest information gain for this dataset ? Show Complete Working.

A	B	C	Y
0	1	1	True
1	1	0	True
1	0	1	False
1	1	1	False
0	1	1	True
0	0	0	True
0	1	1	False
1	0	1	False
0	1	0	True
1	1	1	True

Part c) Which of the following expressions sum up to 1?

- (a) $\sum_a P(A = a \mid B)$
- (b) $\sum_b P(A \mid B = b)$
- (c) $\sum_a \sum_b P(A = a, B = b)$
- (d) $\sum_a P(A = a) P(B = b)$

Part d) Which of the following expressions represent $P(A, B, C)$?

- (a) $P(C \mid A, B) P(A) P(B)$
- (b) $P(A \mid B) P(B \mid C) P(C)$
- (c) $P(A, B \mid C) P(C)$
- (d) $P(A \mid B, C) P(B \mid A, C) P(C \mid A, B)$

Part e) Derive a gradient descent based learning rule for a perceptron that uses ***tanh*** as activation function. Assume that the perceptron has ***D*** inputs and a bias term and we have ***N*** training examples $(X_i, Y_i) \quad i = 1, 2, \dots, N$ to learn weights of the perceptron.

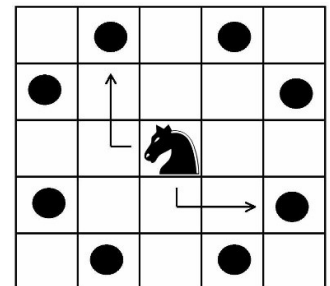
Please note that the derivative of ***tanh(x)*** is **$1 - \tanh^2(x)$**

Problem 2 [Search]

[4, 4 Points]

A knight's tour is a sequence of moves of a knight starting from a starting position on a chessboard such that the knight visits every square only once.

In this problem you are required to pose this problem as a search problem so that it might be solved using various search algorithms discussed throughout the semester.



Student Registration No _____

Part a) Describe a minimal **state** and the corresponding **successor function** needed to solve this problem using A* search algorithm

STATE:

SUCCESSOR FUNCTION:

Part b) Describe the design of a **chromosome** and that of the corresponding **fitness function** so that this problem is solved using Genetic Algorithms

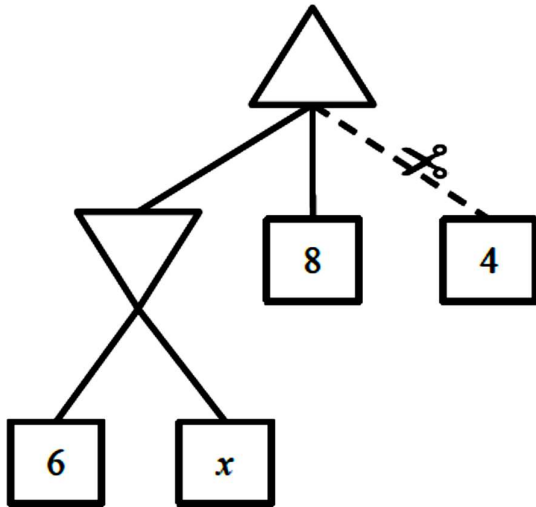
CHROMOSOME:

FITNESS FUNCTION:

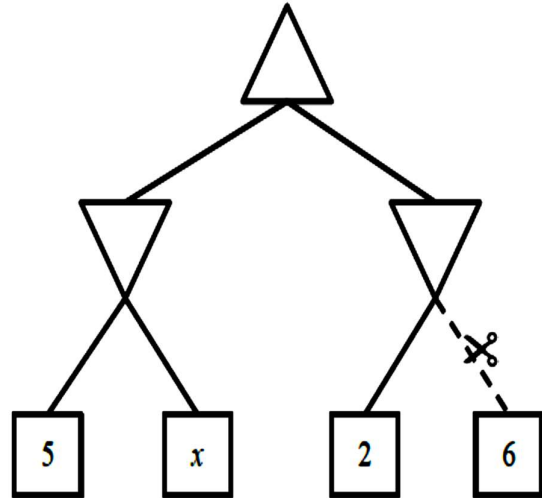
Problem 3 [Game Playing]

[1 + 1 + 1 + 1 Points]

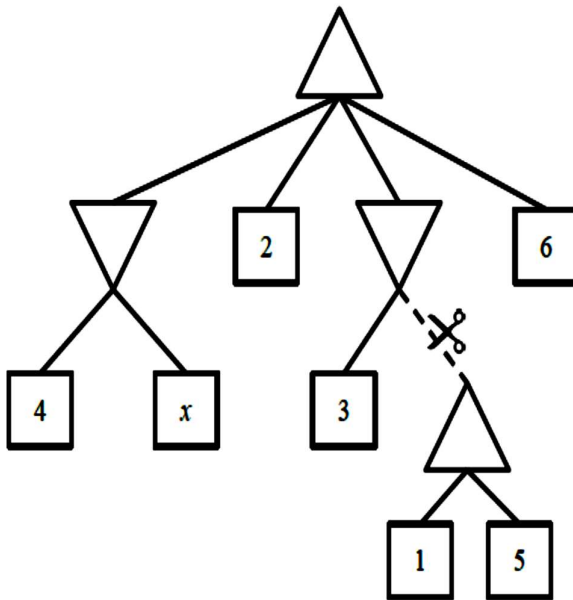
For each of the game-trees shown below, state for **which values of x** the dashed branch with the scissors will be pruned. If the pruning will not happen for any value of x write **\none**". If pruning will happen for all values of x write **\all**.



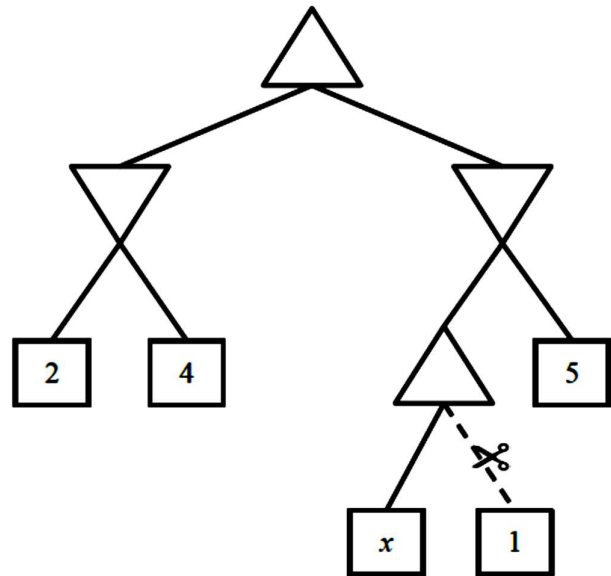
x _____



x _____



x _____



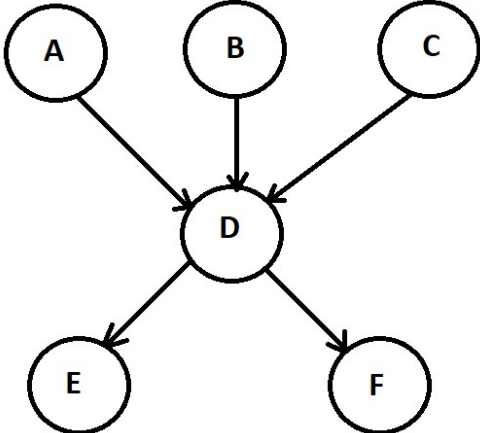
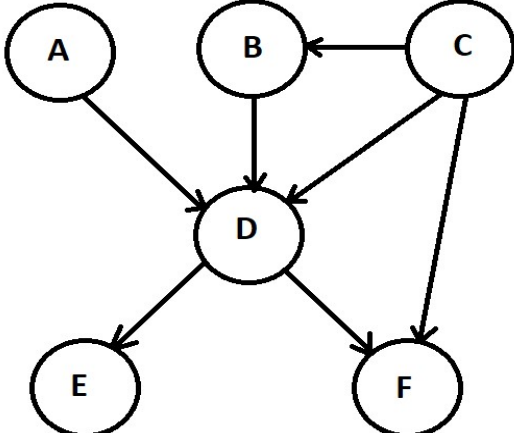
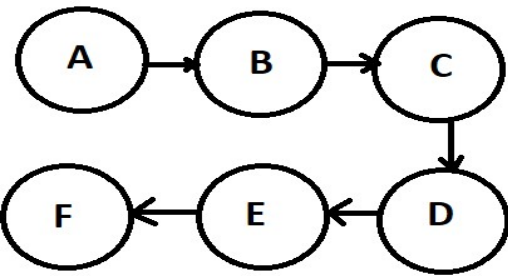
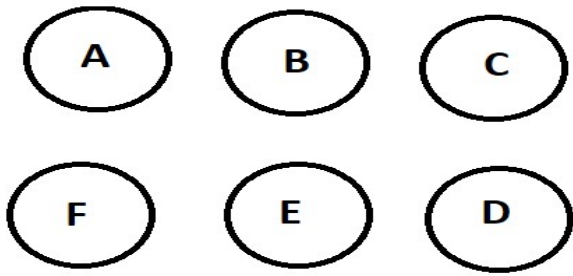
x _____

Problem 4 [Bayes Net]

[2 + 2 + 1 + 1 Points]

For each of the following Bayes Belief Network that captures the relationship between six binary valued variables (A, B, ... , F) write down the joint distribution as product of different factors according to the graph.

Also specify the minimum number of distinct probabilities values needed to completely specify the joint distribution represented by the Bayes net.

 <p>Total Probability Values _____</p> <p>$P(A, B, C, D, E, F) =$</p> <p>_____</p> <p>_____</p>	 <p>Total Probability Values _____</p> <p>$P(A, B, C, D, E, F) =$</p> <p>_____</p> <p>_____</p>
 <p>Total Probability Values _____</p> <p>$P(A, B, C, D, E, F) =$</p> <p>_____</p> <p>_____</p>	 <p>Total Probability Values _____</p> <p>$P(A, B, C, D, E, F) =$</p> <p>_____</p> <p>_____</p>

Problem 5 [Naïve Bayes]**[1 + 2 + 3 Points]**

Part a) What is our main assumptions about the relationship between various random variables when using Naïve Bayes classifier?

Part b) Use the following counts to estimate the conditional distribution for a binary variable X_1 using Laplace smoothing with $K = 10$

Counts			Probability Distribution	
C	Count X1 = 1	Count X1 = 0		
1	30	50		
2	45	35		

Part c) Use the following probability distributions involving two binary variables X_1 and X_2 to classify the following observation according to the two class Naïve Bayes classifier.

Observation: $X_1 = 1$ and $X_2 = 1$.

You must show complete working (i.e. show computation of all probabilities) to get points.

Probability Distributions				
$P(C = 1) = 0.6$	C	$P(X_1 = 1 C)$	C	$P(X_2 = 1 C)$
	1	0.7	1	0.5
	2	0.6	2	0.8