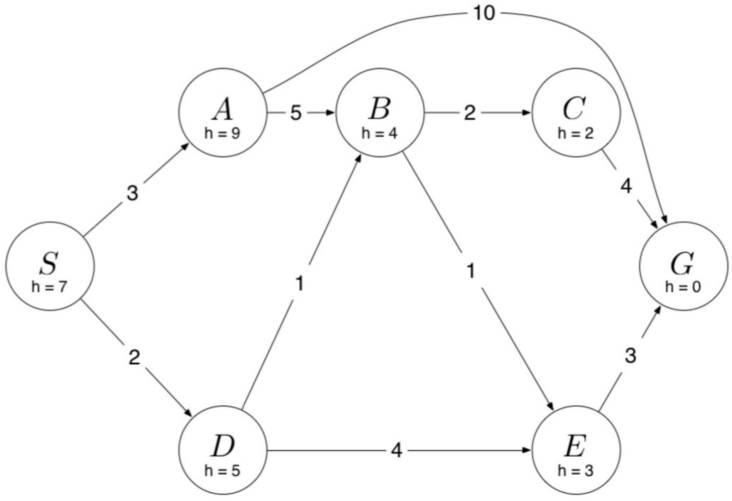
CS 480 Artificial Intelligence **Fall 2018**

Final Examination, Dec 12, 2018, 11 AM to 12:50 PM

Answer any *FIVE* questions. All questions will be equally weighted.

1. (a) Shown below is a weighted directed graph with *S* as starting vertex and *G* as the goal. In all the search strategies, the algorithm terminates when the goal vertex is removed from the fringe. Assume that the children at each node are arranged in alphabetical order. We consider a node to be visited when it leaves the fringe (open set).



(i) Write the sequence of states visited if *iterative deepening search* is used.

(ii) Write the sequence of states visited if *greedy search* is used.

(iii) Write the sequence of states visited if *breadth-first search* is used.

(iv) Write the sequence of states visited if *uniform-cost search* is used.

(v) Write the sequence of states visited if *best-first search (with h given in the figure above)* is used.

(vi) Is the function *h* admissible? Is *h* consistent?

2. (a) Consider *k*-means clustering algorithm (with $k = 2$) for the data set S = {(1,1), (2.5, 2.5), (1, 2), (0, 2), (2.5, 2.5), (3, 2), (3, 3), (4, 3)} with (1, 1) and (0.5, 2.5) as the initial cluster centers. Exhibit the clusters that result after the first and the second iteration of the algorithm.

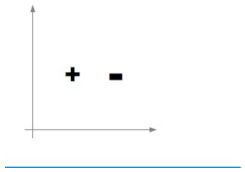
(b) Consider a finite cluster *C* of points in a plane. Thus, $C = \{(x_1, y_1) , (x_2, y_2) , \ldots, (x_n, y_n)\}$ for some real numbers $\{x_j\}$ and $\{y_j\}$, $j = 1, 2, \ldots, n$. Recall that k-means algorithm updates the center of each cluster to (x, y) where $x = (x_1 + \ldots + x_n)/n$ and $y = (y_1 + \ldots + y_n)/n$. The goal of this problem to provide a justification of this step. Suppose we define the cluster center as a point P such that sum of the distances from all the points to P is as small as possible. Show that *k*-means algorithm updates the cluster center correctly by using the average of *x* and *y* values. Specifically, create an expression $D(x, y)$ = sum of squares of distances of all points in *C* to point (x, y) and take partial derivatives of *D* with respect to *x* and *y*, set them to 0 and solve for *x* and *y*. The resulting expressions for *x* and *y* will provide the justification. Also show that that the computed value is a minimum by taking the second derivative.

3. Consider a training set *T* (of size *n*) in which each instance has just one feature that is a real number. The class label is binary (0 or 1). Assume that the distance measure *d* used by the *k*-nearest neighbor on this

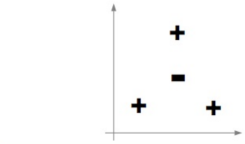
data set uses the standard measure, namely: $d(x, y) = |x - y|$. Describe an efficient preprocessing algorithm so that in the worst-case, each test instance can be classified by the k-nearest neighbor algorithm in $O(k \log n)$ time. The preprocessing algorithm takes as input the set T where each element of T is of the form (x, c). x is the value of the feature and c, the class label. What is the time complexity of the preprocessing algorithm? Also describe the test function that takes as input the feature value x of the test instance and the value of k and outputs the label. Argue that the time complexity of the test algorithm is $O(k \log n)$.

4. Apply perceptron training algorithm to train a single neuron for computing the following Boolean function of three variables (x, y, z): $f(x, y, z) = 1$ except in case $x = 1$, $y = 1$ and $z = 0$. Start with initial weights (0.05, 0.1, -0.1, 0) and $\eta = 0.1$, apply the training algorithm for each input once in the order (0, 0, 0), …, (1, 1, 1) and report the resulting weights. Determine the number of inputs that were correctly classified before and after training. (The first weight is associated with the bias variable x_0).

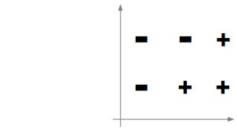
5. Each of the following plot exhibits specifies the behavior of a function of one or two inputs. Some behaviors are specified with verbal descriptions. Other behaviors are specified with graphs of desired output plotted as a function of input: points marked with "+" should produce an output of 1, while points marked with "–"should produce an output of 0. Your task is to choose which of the neural nets shown below could produce each desired behavior with suitable weights. For each desired behavior, list all possible nets that could produce it. If none could, write NONE instead. All neural net thresholds use the unit step function, and threshold (bias) inputs are not shown. If a neural net has only one input, you can choose whether that input receives x-values or y-values. You may choose the same neural net more than once, and some may not be used at all.



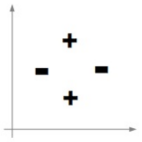
(1)



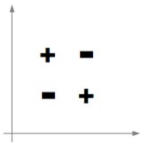
(2)



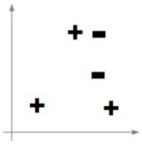
(3)



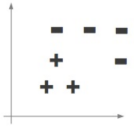
(4)



(5)



(6)



(7)

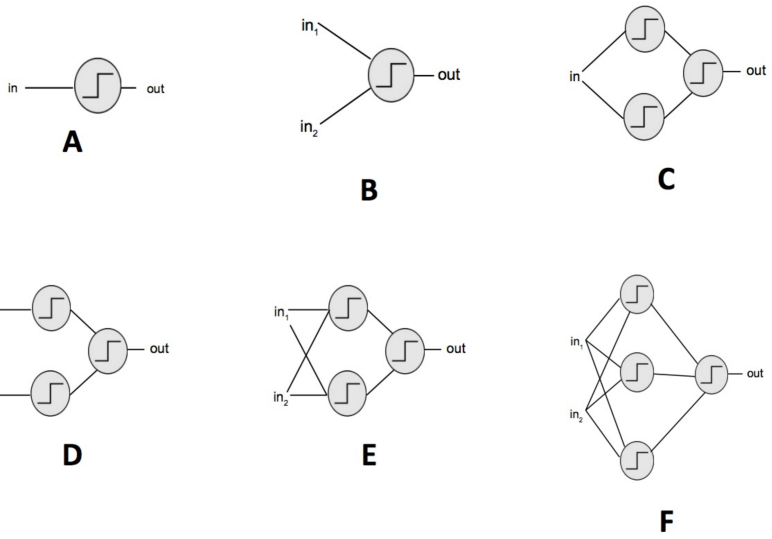
A function which takes as input a temperature in degrees Celsius; it returns one if its input is in a certain (fixed) range of degrees Fahrenheit, otherwise zero.

(8)

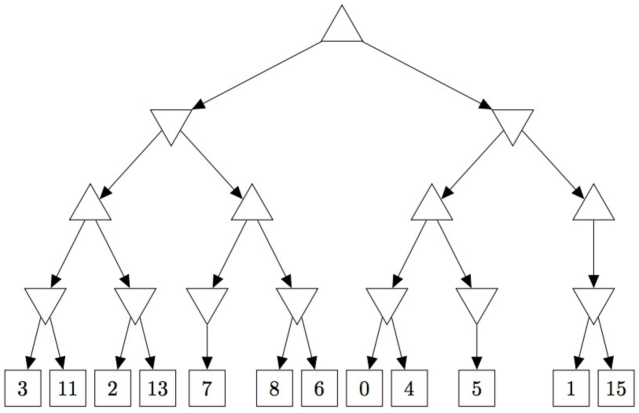
A function which returns one if its inputs are within a certain (fixed) distance of each other, otherwise zero.

(9)

The types of networks are shown below. You can choose weights as you wish.



6. Suppose the alpha-beta pruning algorithm is applied to the game tree shown below. The static evaluation given at the leaf nodes are from player 1's point of view – whose turn it is at the root of the tree.



(a) Display the alpha and beta values at each node, and mark the nodes that are pruned.
(b) Determine the value of the root node and the optimal move that player 1 should make at the root. (Call the move options 1, 2, 3, ... , from left to right.)
(c) Redraw the tree so that at each max (min) node, the children are arranged in decreasing (increasing) value. Reapply alpha-beta pruning algorithm to determine which nodes will be pruned in this new tree.
(d) Suppose Player 2 makes a random choice among moves available to him/her and Player 1 knows this. What is the optimal strategy for Player 1 in this case? Will the alpha-beta search result in any pruning? If so, identify the pruned branches.

7. (a) Shown below is a small training set that contains information about cars that have been recently stolen with three (nominal) features. Apply naïve bayes algorithm to determine the class label of the test data with feature values Color = Red, Type = SUV and Origin = Domestic. Use Laplace correction as needed to avoid 0 probability.

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

(b) Suppose an additional feature is available, namely the price of the car that follows (approximately) normal distribution with mean μ = 18K and variance σ^2 = 10K (for stolen cars) and mean μ = 10K and variance σ^2 = 6K (for non-stolen cars). Now determine the class label of the test data with features Color = Yellow, Type = SUV, Origin = Domestic and price = 12K. (A suggestion to simplify calculations: When computing the probability, conditional independence means the term to be calculated will be a product with one term that involves the density function of normal distribution. Use logarithm will not only will turn the product into sum, but will replace the error function by a quadratic function.)

8. Consider a game in which player A chooses (randomly) a string x from one of the two sets S_1 = {ABDB, CBAB, ADBD, DCAB, BDCD} and S_2 = {CADA, BDAB, AACB, DCAC, BACA}. Player B is to find out if x belongs to S_1 or S_2. B can ask questions about substring w of length 2 and A will respond with yes if w is a substring of x. For example, if x is CADA, then if w = BC, the answer is no, but if w = AD, the answer is yes. Clearly a strategy for B can be represented as a decision tree.

(a) If this problem is viewed as a machine learning problem, how many features are in the training set? How many instances are there in the training set? How many class labels are there?

(b) Calculate the entropy (information uncertainty) before the game begins.
(c) Consider two choices w = BA and w = BD. For each choice, calculate the information gain as well as the gain ratio.
(d) Which of the two strings is a better choice if the criterion used is information gain? Answer the same question if the criterion used is gain ratio.
(e) Assuming that the root node is associated with the query string BD, exhibit the two sub-problems that result (corresponding to yes and no answers).