PES University, Bengaluru

(Established under Karnataka Act No. 16 of 2013)

UE19CS303

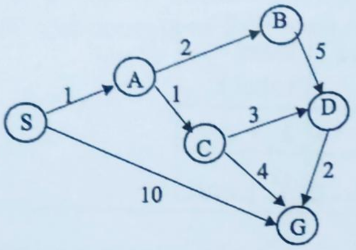
DECEMBER 2021: END SEMESTER ASSESSMENT (ESA) B TECH V SEMESTER

UE19CS303 – Machine Intelligence

Time: 3 Hrs	Answer All Questions	Max Marks: 100

1 a. Consider the graph given below. The heuristic value of all states is given in the table. Use A* algorithm to reach the goal state G from the start state S. Show all the steps. **(5)**

State	h(n)
S	5
A	3
B	4
C	2
D	6
G	0



b. Consider the following dataset of a cooking competition: **(1+2+2)**

Cook id	1	2	3	4	5	6	7	8	9
Garnishing of Dish1	C	C	I	I	I	I	I	C	I
Garnishing of Dish2	C	C	I	I	C	C	I	I	C
Average time taken(min)	1.0	6.0	5.0	4.0	7.0	3.0	8.0	7.0	5.0
Selected for next round	Yes	Yes	No	Yes	No	No	No	Yes	No

(where Garnishing of Dish 1 and Dish2 can be Completed (C) or Incomplete (I)).

Which attribute (between garnishing of dish1 and garnishing of dish2) gives the better split based on information gain?

c For the following rational agents determine the PEAS (Performance Measure, Environment, Actuator and Sensors) components. For the environment component also mention the type of environment. **(2+2+1 2)**

(i) Automated Car Driving agent

(ii) Online English Tutor agent

Is the task environment fully observable for both the above agents?

d Define the following terms: (i) Version Space (ii) Inductive Bias (iii) ROC curve (iv) Overfitting (v) Sensitivity **(1*5)**

2 a For the data given below, classify the point x=5.0 based on its 1-, 3- and 5- nearest neighbors (using Manhattan distance and majority vote) **(1*5)**

x	5.0	3.0	4.5	4.6	4.9	5.2	5.3	5.5	7.0	9.5
y	-	-	+	+	+	-	-	+	-	-

What happens if the value of k is too small or too big?

	b	Show that $\Delta w_{ji} = \eta(t_j - o_j)o_j(1-o_j)x_{ji}$ *and* $\Delta w_{ji} = \eta\delta_j x_{ji}$ where x_{ji} = the ith input to unit j, w_{ji} = the weight associated with the ith input to unit j, $net_j = x_i w_{ji} x_{ji}$ (the weighted sum of inputs for unit j), o_j = the output computed by unit j, t_j = the target output for unit j, a = the sigmoid function, outputs = the set of units in the final layer of the network, Downstream(j) = the set of units whose immediate inputs include the output of unit j, η is the learning rate and δ_j is the error term for jth unit.	8
	c	Consider the following dataset with two attributes as Attribute 1 and Attribute 2 and determine the equations of marginal planes and the optimal hyperplane.	7

	Attribute 1	Attribute 2	Class Label Y
X1	1	3	1
X2	3	1	-1

3	a	Consider a learner L on an instance space X and a hypothesis space H consisting of some class of real-valued function defined over X. L learns an unknown target function f: X→R, where R is set of real numbers. 'm' training examples are provided to the learner L and the target value of each training example is corrupted by a random noise e_i which follows a *normal distribution.* Just to bring you to a comfort zone, each training example is of the form $< x_i, d_i >$ where $d_i = f(x_i) + e_i$. Now, show that the Maximum Likelihood hypothesis h_{ML} is the one that minimizes the sum of the squared errors between the observed training values d_i and the hypothesis prediction $h(x_i)$ under certain assumptions.	6
	b	Consider the data set shown below:	3+3+1

Record	1	2	3	4	5	6	7	8	9	10
A	0	0	0	0	0	1	1	1	1	1
B	0	0	1	1	0	0	0	0	1	0
C	0	1	1	1	1	1	1	1	1	1
Class	+	-	-	-	+	+	-	-	+	+

Estimate the conditional probabilities for P(A|+), P(B|+), P(C|+), P(A|−), P(B|−), and P(C|−). Use the estimate of conditional probabilities obtained so far to predict the class label for a test sample (A= 0, B= 1, C= 0) using the Naive Bayes approach. Why m-estimate of conditional probability in Bayesian classifier is required?

	c	Hidden Markov Models are characterized by three fundamental problems. Clearly state those three fundamental problems. Say, there are two states S_1 and S_2 and four observations O_1, O_2, O_3 and O_4. The initial state probabilities of S_1 and S_2 being (0.8, 0.2). Given below are the transition and emission probabilities:	2+5

	S_1	S_2
S_1	0.99	0.01
S_2	0.1	0.9

	O_1	O_2	O_3	O_4
Given S_1	0.2	0.2	0.4	0.2
Given S_2	0.4	0.3	0.1	0.2

Compute the best hidden state sequence given observation (O_1, O_2, O_3).

4	a	The table below is the distance matrix of 6 objects:	3+2

	A	B	C	D	E	F
A	0					
B	0.12	0				
C	0.51	0.25	0			
D	0.84	0.16	0.14	0		
E	0.28	0.77	0.70	0.45	0	
F	0.34	0.61	0.93	0.20	0.67	0

Show the final result of hierarchical clustering with complete link on these 6 data objects and also by drawing a dendrogram.

b Compute the principal components of following data- 5

CLASS 1

X = 2 , 3 , 4

Y = 1 , 5 , 3

CLASS 2

X = 5 , 6 , 7

Y = 6 , 7 , 8

	c	Differentiate between the following with example: (i) Tournament Selection vs. Rank Selection (ii) Uniform Crossover vs. Masked Uniform Crossover	5
	d	Explain the working of Particle Swarm Optimization with proper expression for update of velocity and position of the particle and also state the mechanism to control the velocity of the particle in search space.	5
5	a	Suppose an RNN takes a vector input of length 2n, it has one hidden layer with n/2 neurons and one output layer with 4 neurons. If the total number of weights be 20(not including bias). Find the value of n.	2
	b	Suppose Apple has hired you as a Machine Learning Engineer. Initially you have been put onto a project where you will be using different architectures of Recurrent Neural Networks (RNNs). Consider the following tasks and justify which architecture of RNN will be most appropriate for the following: (i) Sentiment Analysis (ii) Machine Translation (iii) DNA Sequence Analysis (iv) Image Captioning (v) Predicting next character in a word	1*5
	c	Write the expressions for loss function with no regularization, with L_1 regularization and L_2 regularization. Explain briefly the concept of 'Dropout'.	3+2

		Consider the following 3 input channels as A, B, C:	8

1	0	1	0	2
1	1	3	2	1
1	1	0	1	1
2	3	2	1	3
0	2	0	1	0

I/p Channel A

1	0	0	1	0
2	0	1	2	0
3	1	1	3	0
0	3	0	3	2
1	0	3	2	1

I/p Channel B

2	0	1	2	1
3	3	1	3	2
2	1	1	1	0
3	1	3	2	0
1	1	2	1	1

I/p Channel C

And the three kernel matrices corresponding to I/p channels A, B and C are as follows respectively:

0	1	0
0	0	2
0	1	0

2	1	0
0	0	0
0	3	0

1	0	0
1	0	0
0	0	2

Bias corresponding to all 3 input channels is +1 and vertical and horizontal stride being 2. Obtain the feature map by applying the convolution operation on inputs and kernels and then apply average pool to get final representation of the input that will be sent to fully connected layer.