USN | | | | | | | | | | |

Internal Assessment Test 3 – Jan. 2022

Sub:	**ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING**					Sub Code:	**18CS71**	Branch:	**CSE**	
Date:	24/01/2022	Duration:	90 mins	Max Marks:	50	Sem / Sec:	7/A,B,C		OBE	
	Answer any FIVE FULL Questions							MARKS	CO	RBT
1 a)	Derive an equation for maximum a posteriori(MAP) hypothesis using Bayes theorem.							[5]	CO2	L2
1b	Explain naïve bayes classifier							[5]	CO3	L2

1 a) Derive an equation for maximum a posteriori(MAP) hypothesis using Bayes theorem.

Ans: Using Bayes theorem

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

Generally want the most probable hypothesis given the training data

Maximum a posteriori hypothesis h_{MAP}:

$$\begin{aligned} h_{MAP} &= \arg\max_{h \in H} P(h|D) \\ &= \arg\max_{h \in H} \frac{P(D|h)P(h)}{P(D)} \\ &= \arg\max_{h \in H} P(D|h)P(h) \end{aligned}$$

P(D) is constant ,it can be safely dropped

1b Explain naïve bayes classifier

Ans:

Bayesian Classifiers

- Consider each attribute and class label as random variables

- Given a record with attributes ($A_1, A_2, \ldots, A_n$)
  - Goal is to predict class C
  - Specifically, we want to find the value of C that maximizes $P(C| A_1, A_2, \ldots, A_n)$

Approach:

- compute the posterior probability $P(C | A_1, A_2, \ldots, A_n)$ for all values of C using the Bayes theorem

$$P(C | A_1 A_2 \ldots A_n) = \frac{P(A_1 A_2 \ldots A_n | C) P(C)}{P(A_1 A_2 \ldots A_n)}$$

- Choose value of C that maximizes $P(C | A_1, A_2, \ldots, A_n)$

- Equivalent to choosing value of C that maximizes $P(A_1, A_2, \ldots, A_n|C) P(C)$

- Assume independence among attributes A_i when class is given:
  - $P(A_1, A_2, \ldots, A_n |C) = P(A_1| C_j) P(A_2| C_j) \ldots P(A_n| C_j)$

  - Can estimate $P(A_i| C_j)$ for all A_i and C_j.

  - New point is classified to C_j if $P(C_j) \Pi P(A_i| C_j)$ is maximal.

2	Explain Bayesian belief networks and conditional independence with an example	[10]	CO3	L2

Definition: X is *conditionally independent* of Y given Z if the probability distribution governing X is independent of the value of Y given the value of Z; that is, if

$$(\forall x_i, y_j, z_k)\ P(X = x_i | Y = y_j, Z = z_k) = P(X = x_i | Z = z_l$$

more compactly, we write

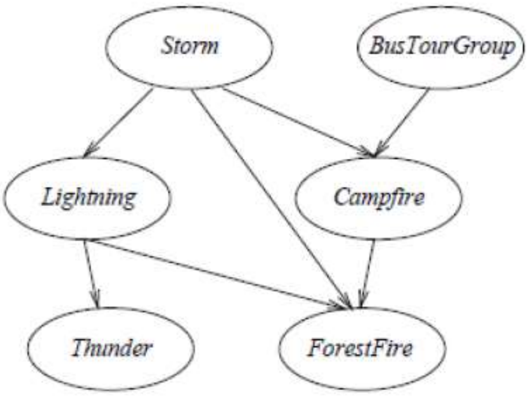
$$P(X|Y, Z) = P(X|Z)$$

Example: *Thunder* is conditionally independent of *Rain*, given *Lightning*

$$P(Thunder|Rain, Lightning) = P(Thunder|Lightning)$$

Naive Bayes uses cond. indep. to justify

$$\begin{aligned} P(X, Y|Z) &= P(X|Y, Z)P(Y|Z) \\ &= P(X|Z)P(Y|Z) \end{aligned}$$



	S,B	S,¬B	¬S,B	¬S,¬B
C	0.4	0.1	0.8	0.2
¬C	0.6	0.9	0.2	0.8

Campfire

Network represents a set of conditional independence assertions:

- Each node is asserted to be conditionally independent of its nondescendants, given its immediate predecessors.
- Directed acyclic graph

		[05]	CO3	L3

3 (a) Classify the test data {Red, SUV, Domestic} using naïve bayes classifier for the dataset given in the table below:

Color	Type	Origin	Stolen
Red	Sports	Domestic	Yes
Red	Sports	Domestic	No
Red	Sports	Domestic	Yes
Yellow	Sports	Domestic	No
Yellow	Sports	Imported	Yes
Yellow	SUV	Imported	No
Yellow	SUV	Imported	Yes
Yellow	SUV	Domestic	No
Red	SUV	Imported	No
Red	Sports	Imported	Yes

Ans: P(Yes|<Red,SUV,Domestic>) = (3/5) * (1/5) * (2/5) * (½) = 6/250= 0.024

P(No|<Red,SUV,Dom>) = (2/5) * (3/5) * (3/5) * (½) = 18/250= 0.072

P(No|D) is more. Hence |<Red,SUV,Dom> is NOT stolen

3 (b) Explain EM algorithm [5] CO2 L2

EM algorithm for k-means where k=2

- EM algorithm picks two random values as intial means $h = \langle \mu_1, \mu_2 \rangle$
- It then iteratively re-estimates h by repeating the following two steps until the procedure converges to a stationary value for h.

- Iterate the following two steps

E step: Calculate the expected value $E[z_{ij}]$ of each hidden variable z_{ij}, assuming the current hypothesis $h = \langle \mu_1, \mu_2 \rangle$ holds.

$$E[z_{ij}] = \frac{p(x = x_i | \mu = \mu_j)}{\Sigma_{n=1}^{2} p(x = x_i | \mu = \mu_n)} = \frac{e^{-\frac{1}{2\sigma^2}(x_i - \mu_j)^2}}{\Sigma_{n=1}^{2} e^{-\frac{1}{2\sigma^2}(x_i - \mu_n)^2}}$$

M step: Calculate a new maximum likelihood hypothesis $h' = \langle \mu'_1, \mu'_2 \rangle$, assuming the value taken on by each hidden variable z_{ij} is its expected value $E[z_{ij}]$ calculated above. Replace $h = \langle \mu_1, \mu_2 \rangle$ by $h' = \langle \mu'_1, \mu'_2 \rangle$.

$$\mu_j \leftarrow \frac{\Sigma_{i=1}^{m} E[z_{ij}] \; x_i}{\Sigma_{i=1}^{m} E[z_{ij}]}$$

4	What is instance based learning? Explain K-Nearest neighbour algorithm	[10]	CO3	L2

Ans:

Instance based learner is a lazy learner. It Simply stores training data (or only minor processing) and waits until it is given a test tuple. Lazy learners take less time in training but more time in predicting. Lazy method effectively uses a richer hypothesis space since it uses many local linear functions to form an implicit global approximation to the target function.

K-nearest neighbor is an instance-based learning algorithm. It assumes all instances correspond to points in the n-dimensional space. The nearest neighbors of an instance are defined in terms of the standard Euclidean distance.

Training algorithm:

- For each training example $\langle x, f(x)\rangle$, add the example to the list *training_examples*

Classification algorithm:

- Given a query instance x_q to be classified,
  - Let $x_1 \ldots x_k$ denote the k instances from *training_examples* that are nearest to x_q
  - Return

$$\hat{f}(x_q) \leftarrow \underset{v \in V}{\operatorname{argmax}} \sum_{i=1}^{k} \delta(v, f(x_i))$$

where $\delta(a, b) = 1$ if $a = b$ and where $\delta(a, b) = 0$ otherwise.

5	Explain locally weighted linear regression	[10]	CO3	L2

Ans:

Locally weighted regression constructs an explicit approximation to target function f over a local region surrounding x. It uses nearby or distance-weighted training examples to form this local approximation to f. The general regression algorithm chooses weights that minimize the squared error summed over the set D of training examples. In local regression, we normally assign different weights for the error from each training example. There are three popular methods to assign weights. They are

1. Minimize the squared error over just the k nearest neighbors:

$$E_1(x_q) \equiv \frac{1}{2} \sum_{x \in\ k\ nearest\ nbrs\ of\ x_q} (f(x) - \hat{f}(x))^2$$

2. Minimize the squared error over the entire set D of training examples, while weighting the error of each training example by some decreasing function K of its distance from x_q:

$$E_2(x_q) \equiv \frac{1}{2} \sum_{x \in D} (f(x) - \hat{f}(x))^2\ K(d(x_q, x))$$

3. Combine 1 and 2:

$$E_3(x_q) \equiv \frac{1}{2} \sum_{x \in\ k\ nearest\ nbrs\ of\ x_q} (f(x) - \hat{f}(x))^2\ K(d(x_q, x))$$

			CO2	L2

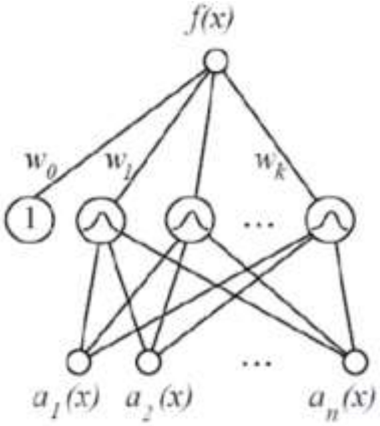
6 Write short notes on
Radial Basis Function
Case based reasoning

[5+5]

Ans: a) Learning with radial basis functions is an approach to function approximation that is closely related to distance weighted regression and also to artificial neural networks. In this approach, the learned hypothesis is a function of
the form

$$\hat{f}(x) = w_0 + \sum_{u=1}^{k} w_u K_u(d(x_u, x))$$

A typical radial basis network has the architecture shown below



b) Case based reasoning
Case-based reasoning (CBR) is an instance based learning method that uses the principle of lazy learning. In CBR, instances are typicaly represented using rich symbolic descriptions. CBR has been applied to problems such as conceptual design of mechanical devices based on a stored library of previous designs. In CADET system which is a CBR system, each instance stored in memory (e.g., a water pipe) is represented by describing both its structure and its qualitative function. New design problems are then presented by specifying the desired function and requesting the corresponding structure.

CO PO Mapping

Course Outcomes		Blooms Level	Modules covered	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3	PSO4
CO1	Appraise the theory of Artificial intelligence and Machine Learning	**L1,L2,L3**	**1,2,3, 4,5**	2	3	1	1	-	1	-	-	-	-	-	2	1	0	1	3
CO2	Illustrate the working of AI and	**L1,L2,L3**	**1,2,3, 4,5**	2	3	1	1	-	1	-	-	-	-	-	2	1	0	1	3

	ML Algorithms																		
CO3	Demonstrate the applications of AI and ML	**L1,L2,L3**	**1,23,4,5**	2	3	1	1	1	2	-	-	-	-	-	2	1	0	1	3

COGNITIVE LEVEL	REVISED BLOOMS TAXONOMY KEYWORDS
L1	List, define, tell, describe, identify, show, label, collect, examine, tabulate, quote, name, who, when, where, etc.
L2	summarize, describe, interpret, contrast, predict, associate, distinguish, estimate, differentiate, discuss, extend
L3	Apply, demonstrate, calculate, complete, illustrate, show, solve, examine, modify, relate, change, classify, experiment, discover.
L4	Analyze, separate, order, explain, connect, classify, arrange, divide, compare, select, explain, infer.
L5	Assess, decide, rank, grade, test, measure, recommend, convince, select, judge, explain, discriminate, support, conclude, compare, summarize.

PROGRAM OUTCOMES (PO), PROGRAM SPECIFIC OUTCOMES (PSO)				CORRELATION LEVELS	
PO1	Engineering knowledge	PO7	Environment and sustainability	0	No Correlation
PO2	Problem analysis	PO8	Ethics	1	Slight/Low
PO3	Design/development of solutions	PO9	Individual and team work	2	Moderate/ Medium
PO4	Conduct investigations of complex problems	PO10	Communication	3	Substantial/ High
PO5	Modern tool usage	PO11	Project management and finance		
PO6	The Engineer and society	PO12	Life-long learning		
PSO1	Develop applications using different stacks of web and programming technologies				
PSO2	Design and develop secure, parallel, distributed, networked, and digital systems				
PSO3	Apply software engineering methods to design, develop, test and manage software systems.				
PSO4	Develop intelligent applications for business and industry				