Siddaganga Institute of Technology

(An Autonomous Institution affiliated to Visvesvaraya Technological University, Belgaum, Approved by AICTE, New Delhi)

B.H. Road, Tumkur, 572103, Karnataka

TEST-2 Scheme and Solutions

Course Code:	**S5CCSI01**	Course Title:	**Artificial Intelligence and Machine Learning**

I hereby certify that

- I don't have any blood relatives appearing for this paper.
- I have written down the scheme and solution myself.

	Dr. Savithramma R M
Signature of the Faculty	Name of the Faculty

	Dr. N R Sunitha
Signature of the Chairperson [BOE]	Name of the Chairperson [BOE]

2024 – 2025	√			√					**05**
Academic Year	Odd	Even	Summer	BE	B.Arch.	MCA	MBA	M.Tech.	Semester
	CIE			Degree					

Total number of pages used

Q.No.	SCHEME AND SOLUTIONS	M
1.a)	***Discuss*** the major issues in machine learning. ***Soln:*** The field of machine learning, is concerned with answering questions such as the following: 1. What algorithms exist for learning general target functions from specific training examples? In what settings will particular algorithms converge to the desired function, given sufficient training data? Which algorithms perform best for which types of problems and representations? 2. How much training data is sufficient? What general bounds can be found to relate the confidence in learned hypotheses to the amount of training experience and the character of the learner's hypothesis space? 3. When and how can prior knowledge held by the learner guide the process of generalizing from examples? Can prior knowledge be helpful even when it is only approximately correct? 4. What is the best strategy for choosing a useful next training experience, and how does the choice of this strategy alter the complexity of the learning problem? 5. What is the best way to reduce the learning task to one or more function approximation problems? Put another way, what specific functions should the system attempt to learn? Can this process itself be automated? 6. How can the learner automatically alter **its** representation to improve its ability to represent and learn the target function?	**06**

1 b)	Design candidate elimination algorithm to identify a version space.
	Soln:
	Initialize G to the set of maximally general hypotheses in H
	Initialize S to the set of maximally specific hypotheses in H
	For each training example d, do
	• If d is a positive example
	  • Remove from G any hypothesis inconsistent with d
	  • For each hypothesis s in S that is not consistent with d
	    • Remove s from S
	    • Add to S all minimal generalizations h of s such that
	      • h is consistent with d, and some member of G is more general than h
	    • Remove from S any hypothesis that is more general than another hypothesis in S
	• If d is a negative example
	  • Remove from S any hypothesis inconsistent with d
	  • For each hypothesis g in G that is not consistent with d
	    • Remove g from G
	    • Add to G all minimal specializations h of g such that
	      • h is consistent with d, and some member of S is more specific than h
	    • Remove from G any hypothesis that is less general than another hypothesis in G
		06

1 c)	Write the final version space for the below mentioned training examples using candidate elimination algorithm.	**07**

Ex	Body Temp	Gives Birth	Four Legged	Hibernates	Class label
1	Warm	Yes	Yes	Yes	Yes
2	Warm	Yes	Yes	No	Yes
3	Warm	No	No	No	No
4	Cold	No	No	No	No
5	Cold	Yes	No	No	No

Soln:

S0: {<ϕ, ϕ, ϕ, ϕ >}
G0: {<?, ?, ?, ? >}

S1:{< Warm, Yes, Yes, Yes >}
G1: {< ?, ?, ?, ? >}

S2: {< Warm, Yes, Yes, ? >}
G2: {< ?, ?, ?, ? >}

S3: {< Warm, Yes, Yes, ? >}
G3: {<?, Yes, ?, ?>, <?, ?, Yes, ?> }

S4: {< Warm, Yes, Yes, ? >}
G4: {<?, Yes, ?, ?>, <?, ?, Yes, ?>, {<Warm, Yes, ?, ?>, {<?, Yes, Yes, ?> }

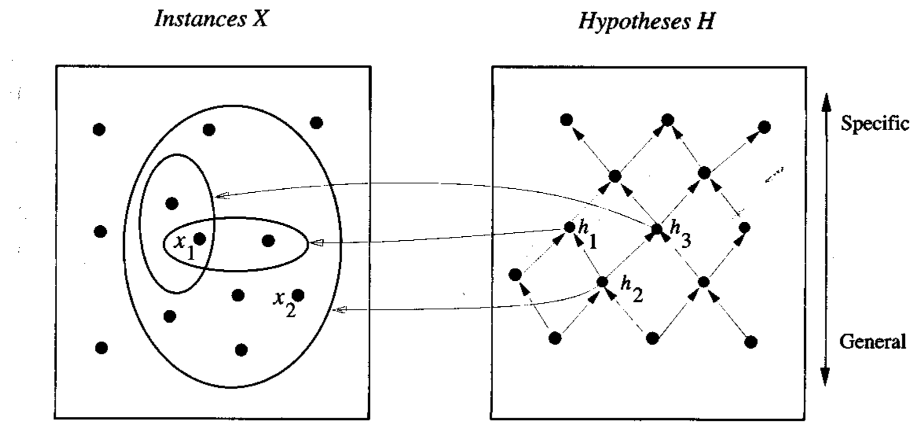
S5: {< Warm, Yes, Yes, ? >}
G5: { <?, ?, Yes, ?>, {<Warm, Yes, ?, ?>, {<?, Yes, Yes, ?> > }

Final Version Space : $VS_{H,D}$ = {< Warm, Yes, ?, ? >, < Warm, ?, Yes, ? >, < ?, Yes, Yes, ? >
Along with the boundaries S5 and G5

2 a) Define concept learning. Explain general-to-specific ordering of hypothesis with suitable example. **06**

Soln:

Concept learning: Inferring a boolean-valued function from training examples of its input and output. Any concept learning task can be described by the set of instances over which the target function is defined, the target function, the set of candidate hypotheses considered by the learner, and the set of available training examples.



x_1= <Sunny, Warm, High, Strong, Cool, Same>
x_2= <Sunny, Warm, High, Light, Warm, Same>

h_1= <Sunny, ?, ?, Strong, ?, ?>
h_2= <Sunny, ?, ?, ?, ?, ?>
h_3= <Sunny, ?, ?, ?, Cool, ?>

Explanation with any examples

2 b) Define h_{MAP} *and* h_{ML} . List the features of Bayesian Learning. **06**

Soln:

Any maximally probable hypothesis is called a ***maximum*** a ***posteriori*** (MAP) hypothesis. We can determine the MAP hypotheses by using Bayes theorem to calculate the posterior probability of each candidate hypothesis. More precisely, we will say that ***MAP*** is a MAP hypothesis provided

$$
\begin{aligned}
h_{MAP} &\equiv \underset{h\in H}{\operatorname{argmax}}\, P(h|D) \\
&= \underset{h\in H}{\operatorname{argmax}}\, \frac{P(D|h)P(h)}{P(D)} \\
&= \underset{h\in H}{\operatorname{argmax}}\, P(D|h)P(h)
\end{aligned}
$$

P(Dlh) is often called the ***likelihood*** of the data ***D*** given ***h,*** and any hypothesis that maximizes ***P(Dlh)*** is called a ***maximum likelihood*** (ML) hypothesis, h_{ML}

$$h_{ML} \equiv \underset{h\in H}{\operatorname{argmax}}\, P(D|h)$$

1. Each observed training example can incrementally decrease or increase the estimated probability that a hypothesis is correct. This provides a more flexible approach to learning than algorithms that completely eliminate a hypothesis if it is found to be inconsistent with any single example.

2. Prior knowledge can be combined with observed data to determine the final probability ~f a hypothesis. In Bayesian learning, prior knowledge is provided by asserting (1) a prior probability for each candidate hypothesis, and (2) a probability distribution over observed data for each possible hypothesis.

	3. Bayesian methods can accommodate hypotheses that make probabilistic predictions (e.g., hypotheses such as "this pneumonia patient has a 93% chance of complete recovery"). 4. New instances can be classified by combining the predictions of multiple hypotheses, weighted by their probabilities. 5. Even in cases where Bayesian methods prove computationally intractable, they can provide a standard of optimal decision making against which other practical methods can be measured.	
2 c)	Write Brute-Force MAP learning algorithm and list suitable assumptions. Derive P(D) through total probability theorem in order to specify a learning problem for the BRUTE-FORCE MAP LEARNING algorithm. **BRUTE-FORCE MAP LEARNING algorithm** **1.** **For each hypothesis h in H, calculate the posterior probability** $$P(h\|D) = \frac{P(D\|h)P(h)}{P(D)}$$ **2.** **Output the hypothesis h_{MAP} with the highest posterior probability** $$h_{MAP} = \underset{h \in H}{\operatorname{argmax}}\ P(h\|D)$$ In order specify a learning problem for the **BRUTE-FORCMEA P LEARNING** algorithm we must specify what values are to be used for ***P(h)*** and for ***P(D1h)*** (as we shall see, ***P(D)*** will be determined once we choose the other two). We may choose the probability distributions ***P(h)*** and ***P(D1h)*** in any way we wish, to describe our prior knowledge about the learning task. Here let us choose them to be consistent with the following assumptions: 1. The training data ***D*** is noise free (i.e., ***di* = *c(xi)).*** 2. The target concept ***c*** is contained in the hypothesis space H 3. We have no a priori reason to believe that any hypothesis is more probable than any other. Proof: we can derive *P(D)* from the theorem of total probability (see Table 6.1) and the fact that the hypotheses are mutually exclusive (i.e., ***(for all i !=*** *j)(P(hi ∩ hj) = 0))* $$\begin{aligned} P(D) &= \sum_{h_i \in H} P(D\|h_i)P(h_i) \\ &= \sum_{h_i \in VS_{H,D}} 1 \cdot \frac{1}{\|H\|} + \sum_{h_i \notin VS_{H,D}} 0 \cdot \frac{1}{\|H\|} \\ &= \sum_{h_i \in VS_{H,D}} 1 \cdot \frac{1}{\|H\|} \\ &= \frac{\|VS_{H,D}\|}{\|H\|} \end{aligned}$$ To summarize, Bayes theorem implies that the posterior probability *P(h /D)* under our assumed *P(h)* and *P(D/h)* is $$P(h\|D) = \begin{cases} \frac{1}{\|VS_{H,D}\|} & \text{if } h \text{ is consistent with } D \\ 0 & \text{otherwise} \end{cases}$$	**07**

3 a) The following table gives data set about the flu affecting patients recently. Using Naïve bayes classifier, estimate the prior probabilities and classify the new instance- **(N, Y, No, Y).** **07**

Sl.No.	CHILLS	RUNNY NOSE	HEADACHE	FEVER	FLU
1	Y	N	Mild	Y	No
2	Y	Y	No	N	Yes
3	Y	N	Strong	Y	Yes
4	N	Y	Mild	Y	Yes
5	N	N	No	N	No
6	N	Y	Strong	Y	Yes
7	N	Y	Strong	N	No
8	Y	Y	Mild	Y	Yes

Soln:

P(Yes) = 5/8
P(No) = 3/8

Chills	Yes	No
Y	3/5	1/3
N	2/5	2/3

Running Nose	Yes	No
Y	4/5	1/3
N	1/5	2/3

Headache	Yes	No
Mild	2/5	1/3
No	1/5	1/3
Strong	2/5	1/3

Fever	Yes	No
Y	4/5	1/3
N	1/5	2/3

Given instance X = **(N, Y, No, Y)**

$\mathbf{V_{NB}}$ **(Yes)** = P(Yes) * P(Chills = **N**/Yes) * P(Running Nose = **Y**/Yes) * *P(Headache =* ***No****/ Yes) * P(Fever =* ***Y****/ Yes)*

$$= \frac{\cancel{5}}{\cancel{8}} * \frac{\cancel{2}}{\cancel{5}} * \frac{4}{5} * \frac{1}{5} * \frac{4}{5} = 4/125$$

V_{NB} (Yes) = **0.032**

$\mathbf{V_{NB}}$ **(NO)** = P(No) * P(Chills = **N**/No) * P(Running Nose = **Y**/No) * *P(Headache =* ***No****/ No) * P(Fever =* ***Y****/ No)*

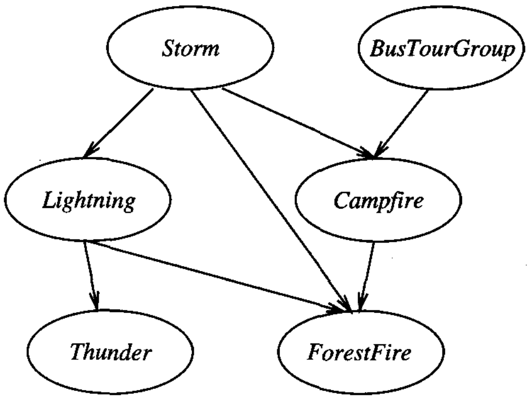
$$= \frac{\cancel{3}}{\cancel{8}} * \frac{\cancel{2}}{\cancel{3}} * \frac{1}{3} * \frac{1}{3} * \frac{1}{3} = 1/108$$

V_{NB} (NO) = **0.00925**

The given instance is classified as **"Yes"**

3 b) Illustrate the working of Bayesian Belief Network with a suitable example. **05**

Soln:



	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

A Bayesian belief network. The network on the left represents a set of conditional independence assumptions. In particular, each node is asserted to be conditionally independent of its nondescendants, given its immediate parents. Associated with each node is a conditional probability table, which specifies the conditional distribution for the variable given its immediate parents in the graph. The conditional probability table for the Campjire node is shown at the right, where Campjire is abbreviated to C, Storm abbreviated to S, and BusTourGroup abbreviated to B.

Any other example can be considered and explained