Machine Learning

Bayesian Learning

DPP: 1

Q1 Assume P(A)=0.2, P(B)=0.6,P(A U B)=0.5, Then P[A|B]=

(A) 0.2 (B) 0.3
(C) 0.6 (D) 0.5

Q2 $P(\overline{E}) = 1 - P(E)$

Match List-I with List-II

	List-I		List-II
A.	Bayer' Theorem	I.	$P(\overline{E}) = 1 - P(E)$
B.	Conditional Probability	II.	$P(E_1 \cup E_2) = P(E_1) + P(E_2)$
C.	Theorem of complementary	III.	$P(E_2/E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)}$
D.	Theorem of addition	IV.	$P(H_i/E) = \frac{P(H_i \cap E)}{P(E)}$

Choose the correct answer from the options given below:

(A) A-I, B-IV, C-III, D-II
(B) A-III, B-IV, C-II, D-I
(C) A-III, B-IV, C-I, D-II
(D) A-IV, B-III, C-I, D-II

Q3 A bike manufacturing factory has two plants P and Q. Plant P manufactures 60 percent of bikes and plant Q manufacture 40 percent. 80 percent of the bikes at plant P and 90 percent of the bikes at plant Q are rated of standard quality. A bike is chosen at random and is found to be of standard quality. What is the probability that it has come from plant P?

Q4 The chance of a defective screw in three boxes A, B, C are $\frac{1}{5}, \frac{1}{6}$and $\frac{1}{7}$ respectively. A box is selected at random and a screw drawn from it at random is found to be defective. Find the probability that it came from box A.

(A) $\frac{42}{107}$ (B) $\frac{28}{107}$
(C) 45/107 (D) $\frac{66}{107}$

Q5 Which of the following best describes Bayesian Learning?

(A) It is a type of machine learning that relies on probabilistic inference.
(B) It is a type of supervised learning that uses decision trees.
(C) It is a type of unsupervised learning that uses clustering.
(D) It is a type of reinforcement learning that uses reward systems.

Q6 In Bayesian learning, the term P(H|D) represents:

(A) The prior probability of the hypothesis.
(B) The likelihood of the hypothesis given the data.
(C) The posterior probability of the hypothesis given the data.
(D) The marginal likelihood of the hypothesis.

Answer Key

Q1	(D)
Q2	(D)
Q3	0.57
Q4	(A)
Q5	(A)
Q6	(C)

Hints & Solutions

Q1 Text Solution:

Conditional Probability formula :- P (A | B) = P(A ∩ B) / P (B)

Inclusion exclusion principle :- P (A U B) = P(A) + P(B) – P(A ∩ B)0.5 = 0.2 + 0.6 – P(A ∩ B)

P(A ∩ B) = 0.3

Given Data :- P(A) = 0.2 , P(B) = 0.6

Put all the value in conditional probability formula

P (A | B) = 0.3 / 0.6 = 0.5

Q2 Text Solution:

Theorem	Formula
A. Bayes' theorem	• It describes the probability of an event based on the prior knowledge of the conditions the might be related to the even. $\mathrm{P}\left(\mathrm{H}_{\mathrm{i}}/\mathrm{E}\right)=\frac{\mathrm{P}(\mathrm{H}_{\mathrm{i}}\cap\mathrm{E})}{\mathrm{P}(\mathrm{E})}$
B. Conditional probability	• The condition probability of event E2 is the probability that the event will occur given the knowledge that event E1 has alrady occured. • $\mathrm{P}\left(\mathrm{E}_2/\mathrm{E}_1\right)=\frac{\mathrm{P}(\mathrm{E}_1\cap\mathrm{E}_2)}{\mathrm{P}(\mathrm{E}_1)}$
C. Theorem of complementary events	• They are mutually exclusive as the two events cannot occur at the same time and tehy are also exhaustive because the sum of of their probabilities and its complement must equal 1. • $\mathrm{P}\left(\overline{\mathrm{E}}\right)+\mathrm{P}\left(\mathrm{E}\right)=1$ • i.e. $\mathrm{P}\left(\overline{\mathrm{E}}\right)=1-\mathrm{P}\left(\mathrm{E}\right)$
D. Therefore of addition	• The addition theorem refers to the happening of at least one of the events from the given two events. • $\mathrm{P}\left(\mathrm{E}_1\cup\mathrm{E}_2\right)=\mathrm{P}\left(\mathrm{E}_1\right)+\mathrm{P}\left(\mathrm{E}_2\right)$

Q3 Text Solution:

Bayes Theorem:- Let E_1, E_2,... E_n be n mutually exclusive and exhaustive events associated with a random experiment and let S be the sample space. Let A be any event which occurs together with any one of E_1, or E_2 or... or E_n, such that P(A) ≠ 0. Then

$\mathrm{P}\left(\mathrm{E}_{\mathrm{i}}|\mathrm{A}\right)\frac{\mathrm{P}(\mathrm{E}_{\mathrm{i}})\times\mathrm{P}(\mathrm{A}|\mathrm{E}_{\mathrm{i}})}{\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{P}(\mathrm{E}_{\mathrm{i}})\times\mathrm{P}(\mathrm{A}|\mathrm{E}_{\mathrm{i}})}$, i = 1, 2 , . . .n

Calculation:

$\mathrm{P}\left(\mathrm{E}_1|\mathrm{A}\right)\frac{\mathrm{P}(\mathrm{E}_1)\times\mathrm{P}(\mathrm{A}|\mathrm{E}_1)}{\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{P}(\mathrm{E}_1)\times\mathrm{P}(\mathrm{A}|\mathrm{E}_1)}$, i = 1, 2, ... n

$\mathrm{P}\left(\mathrm{E}_1\right)=\frac{60}{100}=0.6$

$\mathrm{P}\left(\mathrm{E}_2\right)=\frac{40}{100}=0.4$

$\mathrm{P}\left(\mathrm{A}/\mathrm{E}_1\right)=\frac{80}{100}=0.8$

$\mathrm{P}\left(\mathrm{A}/\mathrm{E}_2\right)=\frac{90}{100}=0.9$

As we know that according to bayes' theorem:

$\mathrm{P}\left(\mathrm{E}_{\mathrm{i}}|\mathrm{A}\right)\frac{\mathrm{P}(\mathrm{E}_{\mathrm{i}})\times\mathrm{P}(\mathrm{A}|\mathrm{E}_{\mathrm{i}})}{\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{P}(\mathrm{E}_{\mathrm{i}})\times\mathrm{P}(\mathrm{A}|\mathrm{E}_{\mathrm{i}})}$, i = 1, 2, . . . , n

$\Rightarrow\ \mathrm{P}\left(\mathrm{E}_2|\mathrm{A}\right)=\frac{\mathrm{P}(\mathrm{E}_1)\cdot\mathrm{P}(\mathrm{A}|\mathrm{E}_1)}{[\mathrm{P}(\mathrm{E}_1)\cdot\mathrm{P}(\mathrm{E}_1)+\mathrm{P}(\mathrm{E}_2)\cdot\mathrm{P}(\mathrm{A}|\mathrm{E}_2)]}$

$\Rightarrow\ \mathrm{P}\left(\mathrm{E}_1|\mathrm{A}\right)=\frac{\frac{6}{10}\cdot\frac{8}{10}}{\left[\frac{6}{10}\cdot\frac{8}{10}+\frac{9}{10}\cdot\frac{4}{10}\right]}=\frac{4}{7}$

Q4 Text Solution:

Let E_1, E_2 and E_3 denote the events of selecting box A, B, C respectively and A be the event that a screw selected at random is defective.

Then

$P(E_1) = P(E_2) = P(E_3) = 1/3$

$\mathrm{P}\left(\mathrm{A}/\mathrm{E}_1\right)=\frac{1}{5}$

$\mathrm{P}\left(\frac{\mathrm{A}}{\mathrm{E}_2}\right)=\frac{1}{6}\Rightarrow\ \mathrm{P}\left(\frac{\mathrm{A}}{\mathrm{E}_2}\right)=\frac{1}{7}$

Then, by Baye's thorem, required probability = $P(E_1/A)$

$$= \frac{\frac{1}{3}\cdot\frac{1}{5}}{\frac{1}{3}\cdot\frac{1}{5}+\frac{1}{3}\cdot\frac{1}{6}+\frac{1}{3}\cdot\frac{1}{7}} = \frac{42}{107}$$

Q5 Text Solution:

Option a) It is a type of machine learning that relies on probabilistic inference:

Correct: Bayesian learning is fundamentally about updating probabilities based on new data, hence it relies on probabilistic inference.

Option b) It is a type of supervised learning that uses decision trees:

Incorrect: While Bayesian learning can be used in supervised learning, it is not restricted to using decision trees. Bayesian methods can be applied to various models and types of learning.

Option c) It is a type of unsupervised learning that uses clustering:

Incorrect: Bayesian learning is not limited to unsupervised learning or clustering. It encompasses a broader range of learning tasks, both supervised and unsupervised, by incorporating probabilistic inference.

Option d) It is a type of reinforcement learning that uses reward systems:

Incorrect: Bayesian learning is distinct from reinforcement learning, which is specifically about learning through rewards and penalties. Bayesian methods can be applied within reinforcement learning but they are not synonymous

Q6 Text Solution:

Option a) The prior probability of the hypothesis:

- Incorrect:The prior probability, denoted as P(H), represents the initial belief about the hypothesis before observing any data. It does not account for the data D and is therefore not what P(H|D) represents.

Option b) The likelihood of the hypothesis given the data:

- Incorrect: The likelihood, denoted as P(D|H), is the probability of observing the data given the hypothesis. It represents how likely the observed data is under the assumption that the hypothesis is true. This is not what P(H|D) represents.

Option c) The posterior probability of the hypothesis given the data:

- Correct: The posterior probability, denoted as P(H|D), represents the updated belief about the hypothesis after observing the data. This is the central concept in Bayesian inference, where prior beliefs are updated with new evidence.

Option d) The marginal likelihood of the hypothesis:

- Incorrect: The marginal likelihood, denoted as P(D), is the total probability of the data under all possible hypotheses. It is used to normalize the posterior distribution but does not directly representP(H|D).