Date / /

AI-ML IA3 QUESTION BANK ⟨MOD4⟩

1.a) DERIVE AN EQN. FOR MAP HYPOTHESIS USING BAYES THEOREM

Ans:→ **BAYES THEOREM**

* Bayes theorem provides a way to calculate the probability of a hypothesis based on its prior probability, the probabilities of observing various data given in hypothesis and the observed data itself.

* In Bayesian learning, the best hypothesis from some hypothesis space H, given the observed training data D means the most probable hypothesis, given the data D plus any initial knowledge about the prior probabilities of the various hypotheses in H.

BAYES THEOREM :

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

where $P(h)$ / $P(D)$ → prior probability

$P(h|D)$ / $P(D|h)$ → posterior probability

→ **MAP with BAYES THEOREM**

* The learner considers some set of candidate hypotheses H and it is interested in finding the most probable hypothesis $h \in H$ given the observed data D.

* Any such maximally probable hypothesis is called a MAXIMUM a POSTERIORI (MAP) HYPOTHESIS (h_{MAP})

Camlin Page
Date / /

* We determine the MAP hypotheses by using BAYES THEOREM to calculate the posterior probability of each candidate hypothesis.

* h_{MAP} is a MAP hypothesis provided:

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

$$= \underset{h \in H}{\operatorname{argmax}} \; \frac{P(D|h)\,P(h)}{P(D)}$$

$$\boxed{h_{MAP} = \underset{h \in H}{\operatorname{argmax}} \; P(D|h)\,P(h)}$$

we dropped $P(D)$ → it is a constant (independent of h)

* If every hypothesis in H is equally probable a priori
i.e. $P(h_i) = P(h_j) \;\forall\; h_i, h_j$ in H

Any hypothesis that maximizes $P(D|h)$ is called MAXIMUM LIKELIHOOD (ML) HYPOTHESIS (h_{ML})

→ $\boxed{h_{ML} = \underset{h \in H}{\operatorname{argmax}} \; P(D|h)}$ [Same & $P(h)$ is constant $\forall\; h \in H$]

↳ 4b) Explain h_{MAP} using BAYES THEOREM.

1b) CONSIDER A FOOTBALL GAME B/W TWO RIVAL TEAMS, SAY TEAM A & TEAM B. SUPPOSE T_A WINS 65% OF THE TIME + T_B WINS THE REMAINING MATCHES. AMONG THE GAMES WON BY T_A, ONLY 35% OF THEM COME FROM PLAYING AT T_B's FOOTBALL FIELD. ON THE OTHER HAND, 75% OF THE VICTORIES FOR T_B ARE OBTAINED WHILE PLAYING AT HOME. IF T_B IS TO HOST THE NEXT MATCH B/W THE TWO TEAMS, WHAT IS THE PROBABILITY THAT IT WILL EMERGE AS THE WINNER?

Ans: $P(T_A)$ = Probability that T_A wins : 0.65

$P(T_B)$ = Probability that T_B wins : $1 - P(T_A) = 0.35$

Probability that T_B hosted the match it had won

$$P(X_B|T_B) = 0.75$$

Probability that T_B hosted the match won by T_A

$$P(X_B|T_A) = 0.35$$

The given question can be solved by obtaining $P(T_B|X_B)$

⇒ probability that T_B wins the next match it hosts.

We know: BAYES THEOREM : $P(h|D) = \frac{P(D|h)\,P(h)}{P(D)}$

⇒ Using BAYES THEOREM,

$$P(T_B|X_B) = \frac{P(X_B|T_B)\,P(T_B)}{P(X_B)}$$

$$= \frac{P(X_B|T_B)\,P(T_B)}{P(X_B|T_B)\,P(T_B) + P(X_B|T_A)\,P(T_A)}$$

$$P(B) = \sum_{i=1}^{n} P(B|A_i)\,P(A_i)$$

$$= \frac{0.75 \times 0.35}{(0.75 \times 0.35) + (0.35 \times 0.65)} = 0.5357$$

$P(T_B|X_B) = 0.5357 \approx 54\%$

Camlin Page
Date / /

6a)

2a) EXPLAIN ① BRUTE FORCE MAP LEARNING ALGORITHM
② NAÏVE BAYES CLASSIFIER.

Ans ① BRUTE FORCE MAP LEARNING ALGORITHM

* We use BAYES THEOREM in CONCEPT LEARNING as it provides a principled way to calculate the posterior probability of each hypothesis given the training data.

* It acts as a basis for a straight fwd. learning algorithm that calculates the probability for each possible hypothesis ⇒ outputs MOST PROBABLE.

* Consider the concept-learning problem:
① Assume L (LEARNER) considers some finite hypothesis H defined over the instance space X, in which TASK: to learn target concept $c : X \to \{0, 1\}$

② L is given some sequence of training examples $(x_1, d_1) \ldots (x_m, d_m)$ where:
$x_i \to$ some instance $\in X$
$d_i \to$ target value of $x_i \Rightarrow d_i = c(x_i)$

③ The sequence of target values (D):
$D = (d_1 \ldots d_m)$

* Design a st.fwd. CONCEPT LEARNING ALGORITHM to output MAP hypothesis, based on BAYES THEOREM.

BRUTE FORCE MAP LEARNING ALGORITHM

① for each hypothesis h in $\in H$ calculate POSTERIOR PROBABILITY
$$P(h|D) = \frac{P(D|h)\,P(h)}{P(D)}$$

② Output h_{MAP} with highest POSTERIOR PROBABILITY
$$h_{MAP} = \underset{h \in H}{\operatorname{argmax}}\, P(h|D)$$

ASSUMPTIONS FOR BRUTE-FORCE MAP LEARNING ALGORITHM

* Let's choose $P(h)$, $P(D|h)$ to be consistent with the following assumptions

→ TRAINING DATA D → noise free $[d_i = c(x_i)]$
→ TARGET CONCEPT C is contained in HYPOTHESIS SPACE H
→ Do not have a PRIORI REASON to believe that any hypothesis is more probable than any other.

VALUES

* $P(h) = \frac{1}{|H|}$ for all $h \in H$ [all prior probabilities sum to 1]
  ↳ prior probability

* $P(D|h) = \begin{cases} 1 & \text{if } d_i = h(x_i) \text{ for all } d_i \in D \\ 0 & \text{otherwise} \end{cases}$
  ↳ probability of observing the target values D for fixed set of instances X, given a world in which hypothesis h holds.

BRUTE-FORCE MAP LEARNING ALGORITHM DERIVATION.

① Bayes theorem : $P(h|D) = \frac{P(D|h)\,P(h)}{P(D)}$

②

when h is INCONSISTENT
$P(h|D) = \frac{0 \cdot P(h)}{P(D)} = 0$
$\boxed{P(h|D) = 0}$

$VS_{H,D}$ subset of hypotheses from H consistent with D
$H' \equiv_C D$

when h is CONSISTENT
$P(h|D) = 1 \cdot \frac{\frac{1}{|H|}}{P(D)} = 1 \cdot \frac{\frac{1}{|H|}}{\frac{|VS_{H,D}|}{|H|}}$
$\boxed{P(h|D) = \frac{1}{|VS_{H,D}|}}$

∴ Under assumed $P(h)$ & $P(D|h)$

$$\boxed{P(h|D) = \begin{cases} \frac{1}{|VS_{H,D}|}, & \text{if } h \equiv_C D \\ 0, & \text{otherwise} \end{cases}}$$

(2) NAIVE BAYES CLASSIFIER

* The Naive Bayes classifier applies to learning tasks where each instance x is described by a conjunction of attribute values and where the target function $f(x)$ can take any value from some finite set V.

* A set of training examples of the target function is provided and a new instance is presented, described by : $(a_1, a_2 \ldots a_m)$ (TUPLE OF ATTR. VALUES)

* The learner L is asked to predict the target value or classification, for this NEW INSTANCE.

* Using BAYESIAN APPROACH,

[Most probable target value] $$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} P(V_j | a_1, a_2 \ldots a_n) \rightarrow (1)$$

* Using BAYES THEOREM & EQN (1)

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} \frac{P(a_1, a_2 \ldots a_n | V_j) P(V_j)}{P(a_1, a_2 \ldots a_n)}$$

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} P(a_1, a_2 \ldots a_n | V_j) P(V_j) \rightarrow (2)$$

* Assuming : $P(a_1, a_2 \ldots a_n | V_j) = \prod_i P(a_i | V_j) \rightarrow (3)$
Given target value of instance, the probability of observing the conjunction $(a_1, a_2 \ldots a_m)$ = product of the probabilities for the individual attributes.

* With eqn (2) & eqn (3), we get NAIVE BAYES CLASSIFIER:

target value output by NB classifier $\hookleftarrow$ $$V_{NB} = \underset{V_j \in V}{\operatorname{argmax}} P(V_j) \left[\prod_i P(a_i | V_j) \right]$$

2b) The following table gives data set about stolen vehicles Using NAIVE BAYES CLASSIFIER classify the new data (Red, SUV, Domestic)

N	COLOR	TYPE	ORIGIN	STOLEN?	
1	red	sports	domestic	YES	1
2	red	sports	domestic	NO	
3	red	sports	domestic	YES	2
4	yellow	sports	domestic	NO	
5	yellow	sports	imported	YES	3
6	yellow	SUV	imported	NO	
7	yellow	SUV	imported	YES	4
8	yellow	SUV	domestic	NO	
9	red	SUV	imported	NO	
10	red	sports	imported domestic	YES	5

Ans: ⇒ NEW INSTANCE = { Red, SUV, Domestic }

→ $P(\text{STOLEN} = \text{YES}) = \frac{5}{10} = 0.5$

→ $P(\text{STOLEN} = \text{NO}) = \frac{5}{10} = 0.5$

COLOR	YES	NO
Red	3/5	2/5
Yellow	2/5	3/5

TYPE	YES	NO
Sports	4/5	2/5
SUV	1/5	3/5

ORIGIN	YES	NO
Domestic	2/5	3/5
Imported	3/5	2/5

→ $P(\text{STOLEN} = \text{YES} \mid \text{NEW INSTANCE}) = P(\text{STOLEN} = \text{YES}) * P(\text{Color} = \text{Red} \mid \text{YES}) * P(\text{Type} = \text{SUV} \mid \text{YES}) * P(\text{Origin} = \text{Domestic} \mid \text{YES})$

$= \frac{5}{10} * \frac{3}{5} * \frac{1}{5} * \frac{2}{5} = 0.024$

→ $P(\text{STOLEN} = \text{NO} \mid \text{NEW INSTANCE}) = P(\text{STOLEN} = \text{NO}) * P(\text{Color} = \text{Red} \mid \text{NO}) * P(\text{Type} = \text{SUV} \mid \text{NO}) * P(\text{Origin} = \text{Domestic} \mid \text{NO})$

$= \frac{5}{10} * \frac{2}{5} * \frac{3}{5} * \frac{3}{5} = 0.072$

[not stolen]

[0.024 > 0.072]

⇒ NEW INSTANCE classified as "NO"

3a) Discuss Minimum Description Length Principle in brief.

Ans: * Inductive bias are assumptions that are made by the learning algorithm to form a hypothesis or a generalization beyond the set of training instances inorder to classify unobserved data.

* Occam's razor is a simple inductive bias that involves preference for a simpler hypothesis that best fits the data.

BAYESIAN interpretation

* MINIMUM DESCRIPTION LENGTH PRINCIPLE (MDL) is inspired from OCCAM'S RAZOR that recommends choosing the hypothesis that captures the regularities in the data and this regularity is inturn used to compress/minimize it.

COMPRESS ↑ LEARNING ↑

* The MDL is motivated by interpreting the definition of h_{MAP} in the light of basic concepts from information theory.

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

which can be equivalently expressed in terms of maximizing the $\log_2$:

$$h_{MAP} = \underset{h \in H}{\operatorname{argmax}}\ \log_2 P(D|h) + \log_2 P(h)$$

OR

$$h_{MAP} = \underset{h \in H}{\operatorname{argmin}} -\log_2 P(D|h) - \log_2 P(h) \rightarrow (1)$$

Date / /

⇒ $-\log_2 P(h)$

the description length of h under the optimal encoding for the hypothesis space H

$$L_{C_H}(h) = -\log_2 P(h)$$

C_H → optimal code for hypothesis space H.

⇒ $-\log_2 P(D|h)$

the description length of the training data D given hypothesis h, under the optimal encoding from the hypothesis space H.

$$L_{C_{D|h}}(D|h) = -\log_2 P(D|h)$$

$C_{D|h}$ → optimal code for describing data D (assuming Sender + Receiver know hypothesis h)

⇒ Rewrite Eqn (1), to show that h_{MAP} is the hypothesis h that minimizes the sum given by the description length of the hypothesis plus the description length of the data given the hypothesis

$$h_{MAP} = \underset{h \in H}{\operatorname{argmin}} \; L_{C_H}(h) + L_{C_{D|h}}(D|h) \quad (2)$$

⇒ MDL principle recommends choosing the hypothesis that minimizes the sum of these two description lengths.

$$h_{MDL} = \underset{h \in H}{\operatorname{argmin}} \; L_{C_1}(h) + L_{C_2}(D|h)$$

where C_1 & C_2 are codes to represent H, D.

⇒ If $C_1 = C_H$ & $C_2 = C_{D|h}$, then $h_{MDL} = h_{MAP}$

3b) Explain BAYESIAN BELIEF NETWORK + CONDITIONAL INDEPENDENCE with example.

Ans: ① BAYESIAN BELIEF NETWORK (BBN)

* A BBN describes the probability distribution governing a set of variables by specifying a set of CONDITIONAL INDEPENDENCE assumptions along with a set of CONDITIONAL PROBABILITIES
* It allows stating CONDITIONAL INDEPENDENCE assumptions that apply to a subset of variables represented by DIRECTED ACYCLIC GRAPHS.

→ NOTATION

* Consider an arbitary set of random variables
$\{Y_1, Y_2 \ldots Y_n\}$
where Y_i → can take on the set of possible values $V(Y_i)$

* The joint space of the set of variables Y
= cross product $V(Y_1) \times V(Y_2) \times \ldots V(Y_n)$

* ~~Each item in JOINT SPACE corresponds to one of the possible assignments of values to the tuple of variables $\langle Y_1, Y_2 \ldots Y_n \rangle$~~. The probability distribution over this JOINT SPACE is called JOINT PROBABILITY DISTRIBUTION. (J.P.D)

* J.P.D specifies the probability for each of the possible variable bindings for tuple $(Y_1 \ldots Y_n)$

* BBN describes the J.P.D for a set of variables

(2) CONDITIONAL INDEPENDENCE

* Let X, Y and Z be three discrete valued random variables. X is CONDITIONALLY INDEPENDENT of Y given Z, if the PROBABILITY DISTRIBUTION governing X is independent of the value of Y, given a value for Z that is, if:

$$\Rightarrow (\forall x_i, y_j, z_k)\ P(X=x_i \mid Y=y_j, Z=z_k) = P(X=x_i \mid Z=z_k)$$

where $x_i \in V(X)$
$y_j \in V(Y)$
$z_k \in V(Z)$

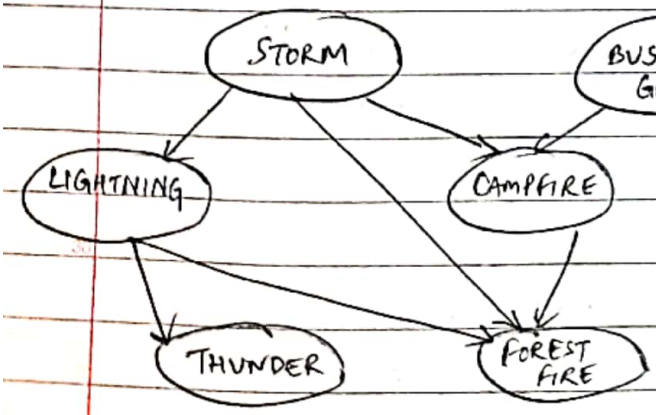
$$\Rightarrow \boxed{P(X \mid Y, Z) = P(X \mid Z)}$$

$$\Rightarrow \text{for set of variables} \quad \boxed{P(X_1 \ldots X_l \mid Y_1 \ldots Y_m, Z_1 \ldots Z_n) = P(X \ldots X_l \mid Z_1 \ldots Z_n)}$$

* Using NB CLASSIFIER for $P(A_1, A_2 \mid V)$ where A_1 is conditionally independent of A_2 given target value V:

$$P(A_1, A_2 \mid V) = P(A_1 \mid A_2, V)\ P(A_2 \mid V)$$
$$= P(A_1 \mid V)\ P(A_2 \mid V)$$

⇒ EXAMPLE



→ ASSERTION: CAMPFIRE conditionally independent of LIGHTNING, THUNDER
Immediate PARENTS: STORM, BUS TOUR GROUP.

CAMPFIRE

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

→ $\boxed{P(y_1 \ldots y_n) = \prod_{i=1}^{n} P(y_i \mid \text{PARENTS}(y_i))}$

↔ $P(\text{Campfire} = T \mid \text{Storm} = T, BG = T) = 0.4$

4a) EXPLAIN ① EM ALGORITHM
② GIBBS ALGORITHM

Ans: ② GIBBS ALGORITHM

* Although the Bayes optimal classifier obtains the best performance that can be achieved from the given training data, it can be quite costly to apply.

* An alternative, less optimal method is GIBBS ALGORITHM defined as follows:
  ① Choose a hypothesis h from H at random, according to the POSTERIOR PROBABILITY DISTRIBUTION over H.
  ② Use h to predict the classification of the next instance x.

* Given a new instance to classify, the GIBBS ALGORITHM simply applies a hypothesis drawn at random according to current posterior probability distribution.

* Suppose $p(x,y)$ is a p.d.f / p.m.f. that is difficult to sample from directly. Using conditional distributions, GIBBS SAMPLING can be done as;
  * Initialize Y^0, X^0
  * for $j = 1, 2, 3 \ldots N$ do
    sample $X^j \sim p(X \mid Y^{j-1})$
    sample $Y^j \sim p(Y \mid X^j)$
  end for

* On comparison of the errors:

$$\epsilon\left[\text{error}_{\text{GIBBS}}\right] \leq 2\epsilon\left[\text{error}_{\text{BAYES OPTIMAL}}\right]$$

→ EM ALGORITHM

* Applying EM algorithm to the existing problem; it initializes the hypothesis to $h = \langle \mu_1, \mu_2 \rangle$. (arbitary initial values)

* Then, iteratively re-estimates h by repeating the following two steps until the procedure converges to a stationary value for h.

STEP 1: Calculate the expected value $E[z_{ij}]$ of each hidden variable z_{ij}, assuming $h = \langle \mu_1, \mu_2 \rangle$

STEP 2: i) Calculate a new max. 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 in STEP ①.

ii) Then, replace $h \langle \mu_1, \mu_2 \rangle$ (OLD) with $h' \langle \mu_1', \mu_2' \rangle$ (NEW) and iterate.

---

① $$E[z_{ij}] = \frac{p(x = x_i \mid \mu = \mu_j)}{\sum_{n=1}^{2} p(x = x_i \mid \mu = \mu_n)}$$

$$= \frac{e^{-\frac{1}{2\sigma^2}(x_i - \mu_j)^2}}{\sum_{n=1}^{2} e^{-1/2\sigma^2 (x_i - \mu_n)^2}}$$

② $$\mu_j \leftarrow \frac{\sum_{i=1}^{m} E[z_{ij}]\, x_i}{\sum_{i=1}^{m} E[z_{ij}]}$$

Date / /

5a) Prove that MAXIMUM LIKELIHOOD (BAYESIAN LEARNING) can be used in any learning algorithms that are used to minimize the squared error between actual output hypothesis & predicted output hypothesis.

Ans: Consider the following problem setting;

* Learner L considers an instance space X and a hypothesis space H consisting of some REAL-VALUED FUNCTIONS defined over X.

The task of the learner is to output MAXIMUM LIKELIHOOD HYPOTHESIS given by;

$$\Rightarrow \boxed{h_{ML} = \underset{h \in H}{\operatorname{argmax}}\ \underbrace{p(D|h)}_{\text{PROBABILITY DENSITY FN.}}} \quad - (1)$$

* We assume a fixed set of training instances $\langle x_1 \ldots x_m \rangle$ with corresponding target values: $D = \langle d_1, \ldots, d_m \rangle$ such that;

$$\Rightarrow \boxed{d_i = \underbrace{f(x_i)}_{\substack{\text{NOISE-FREE} \\ \text{TARGET VALUE}}} + \underbrace{e_i}_{\text{NOISE}}} \quad - (2)$$

* Assuming the training examples are mutually independent given h, we get;

From (1), (2) $\Rightarrow$ $$\boxed{h_{ML} = \underset{h \in H}{\operatorname{argmax}} \prod_{i=1}^{m} p(d_i|h)} \quad - (3)$$

* Given noise e_i obeys NORMAL DISTRIBUTION with ZERO MEAN & SOME VARIANCE σ^2

$\Rightarrow$ d_i must also obey a NORMAL DISTRIBUTION with variance σ^2 centered around true $f(x_i)$ rather than ZERO. Then;

$$\boxed{p(d_i|h) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}(d_i - \mu)^2}} \quad - (4)$$

* From (3), (4) $\Rightarrow h_{ML} = \underset{h \in H}{\operatorname{argmax}} \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(d_i - u)^2}$

$$= \underset{h \in H}{\operatorname{argmax}} \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(d_i - h(x_i))^2}$$

* Maximizing $\ln p$ also maximizes p (MONOTONIC FN.)

$$\Rightarrow h_{ML} = \underset{h \in H}{\operatorname{argmax}} \sum_{i=1}^{m} \left(\ln \frac{1}{\sqrt{2\pi\sigma^2}} \right) - \frac{1}{2\sigma^2}(d_i - h(x_i))^2$$

($\ln \frac{1}{\sqrt{2\pi\sigma^2}}$: constant, $\to 0$)

$$\Rightarrow h_{ML} = \underset{h \in H}{\operatorname{argmax}} \sum_{i=1}^{m} -\frac{1}{2\sigma^2}(d_i - h(x_i))^2$$

* Maximizing -ve quantity = Minimizing +ve quantity

$$\Rightarrow h_{ML} = \underset{h \in H}{\operatorname{argmin}} \sum_{i=1}^{m} \frac{1}{2\sigma^2}(d_i - h(x_i))^2$$

* After discarding constants (independent of h):

$$\Rightarrow \boxed{h_{ML} = \underset{h \in H}{\operatorname{argmin}} \sum_{i=1}^{m} (d_i - h(x_i))^2}$$

$\Rightarrow$ This shows that MAXIMUM LIKELIHOOD HYPOTHESIS h_{ML} minimizes the sum of the squared errors between observed training values d_i & hypothesis predictions $h(x_i)$.

Date / /

5b) Estimate Conditional Probabilities of each attribute
⇒ {colour, legs, height, smelly} for species classes {M, H}
Using given data in table and provide probability values for NEW INSTANCE : (Colour = GREEN, Legs = 2, Height = Tall, Smelly = No)

COLOUR	LEGS	HEIGHT	SMELLY	SPECIES
White	3	Short	Yes	M
Green	2	Tall	No	M
Green	3	Short	Yes	M
White	3	Short	Yes	M
Green	2	Short	No	H
White	2	Tall	No	H
White	2	Tall	No	H
White	2	Short	Yes	H

Ans: $P(M) = \frac{4}{8} = 0.5$ $\quad P(H) = \frac{4}{8} = 0.5$

COLOUR	M	H
White	2/4	3/4
Green	2/4	1/4

LEGS	M	H
2	1/4	4/4
3	3/4	0/4

HEIGHT	M	H
Short	3/4	2/4
Tall	1/4	2/4

SMELLY	M	H
Short	3/4	1/4
Tall	1/4	3/4

→ $P(M|\text{NEW INSTANCE}) = P(M) * P(\text{COLOUR} = \text{Green}|M) * P(\text{LEGS}=2|M) * P(\text{HEIGHT} = \text{Tall}|M) * P(\text{SMELLY} = \text{No}|M)$

$= 0.5 * \frac{2}{4} * \frac{1}{4} * \frac{1}{4} * \frac{1}{4} = \frac{1}{256} = 0.0039$

→ $P(H|\text{NEW INSTANCE}) = P(H) * P(\text{COLOUR} = \text{Green}|H) * P(\text{LEGS}=2|H) * P(\text{HEIGHT} = \text{Tall}|M) * P(\text{SMELLY} = \text{No}|H)$

$= 0.5 * \frac{1}{4} * \frac{1}{4} * \frac{2}{4} * \frac{3}{4} = 0.0117$

∴ $0.0039 < 0.0117 \Rightarrow$ NEW INSTANCE ∈ Species H