Vijay vishnoo p.b

Roll NO: 49

data science & Machine learning

x ——— x

1. Mean $\bar{x} = 34.61 + 34.57 + 34.40 + \ldots$

$= \frac{517.95}{15} = 34.53$

Standard deviation $\sigma = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n}}$

$= \sqrt{\frac{(34.61 - 34.53)^2 + 34.57 - 34.53)^2 \ldots}{15}}$

$= \sqrt{\frac{0.0838}{15}} = \sqrt{0.005586}$

$= 0.07474$

3) in this example, the features are

f_1 = 'scoins' f_2 = "fly" f_3 = "crawl"

The class labels are

C_1 = 'Animal" C_2 = "Bird" C_3 = "Fish"

The test instance is (Slow, Rarely, No) and so we have

x_1 = 'slow' x_2 = 'Rarely' x_3 = No

Class	Skin(A_1)			Flying (A_2)			Gracol (A_3)			Total
	Hair	Feather	No	Long	Short	Neither	No	Yes	No	
Animal	2	2	1	0	0	1	4	2	3	5
Bird	1	0	3	1	2	0	1	1	3	4
Fish	1	2	0	0	0	0	3	0	3	3
Total	4	4	4	1	2	1	8	4	8	12

Step 1: We compute the following probability

$P(C_1)$ = $\frac{\text{No of records with class label Animal}}{\text{Total number of examples.}}$

= 5/12

$P(C_2)$ = $\frac{\text{No of records with class label 'Bird'}}{\text{Total Number of examples.}}$

= 4/12

$P(C_3)$ = $\frac{\text{No of records with class label Fish}}{\text{Total No. of examples}}$

= 3/12

Step 2: We Construct the following table of conditional probabilities.

Class	Features								
	Swim (F_1)			Fly (F_2)			Crawl (F_3)		
	Fast	Slow	No	Long	Short	Fast	No	Yes	No
Animal	2/5	2/5	1/5	0/5	0/5	1/5	2/5	3/5	
Bird	1/4	0/4	3/4	1/4	2/4	0/4	1/4	0/4	4/4
Fish	1/3	2/3	0/3	0/3	0/3	0/3	3/3	0/3	3/3

The conditional probabilities are created as follows

$$P(C_1 = \text{slow} / c_1) = \frac{\text{No of records with } F_1 \text{ slow } F_1 \text{ class}}{\text{No of records with class label } c_1}$$

$$= 2/5$$

Step3 : Once we calculate the following numbers

$$a_1 = P(x_1/c_1)\, P(x_2/c_1)\, P(x_3/c_1)\, P(c_1)$$

$$= (2/5) \times (1/5) \times (3/5) \times (5/12) = 0.02$$

$$a_2 = P(x_1/c_2)\, P(x_2/c_2)\, P(x_3/c_2)\, P(c_2)$$

$$= (0/4) \times (0/4) \times (3/4) \times (4/12) = 0$$

$$a_3 = P(x_1/c_3)\, P(x_2/c_3)\, P(c_3/x_3)\, P(c_3)$$

$$= (2/3) \times (0/3) \times (3/3) = 0$$

Step 4: $\max\{q_1, q_2, q_3\} 0.05$

Step 5: the maximum is q_1 as it corresponds to class label c_1 = Animal

So we assign the class label "Animal" to the test instance

"(slow, Rarely, No)"

2)

x [Academic Excellence]	~	2	distance	Rank
8	6	outstanding	31.6	3
5	6	good	1	1
7	3	good	4.47	4
6	9	outstand	2.23	2
5	7	9	9	?

$K = 3$

$$\text{distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

$$(1) = \sqrt{(5-8)^2 + (7-6)^2} = \quad \sqrt{9+1} = \sqrt{10} = 3.16$$

(2) $= \sqrt{(5-5)^2 + (7-6)^2} = \sqrt{1} = 1$

(3) $= \sqrt{(5-7)^2 + (7-3)^2} = \sqrt{2^2 + 4^2} = \sqrt{4+16} = \sqrt{20} = 4.43$

(4) $= \sqrt{(5-6)^2 + (6-9)^2} = \sqrt{1^2 + 2^2} = \sqrt{1+4} = \sqrt{5} = 2.23$

As, $k=3$: we can take the values with higher 3 rank when we study it we can know that with the 3, 2 of them stay in the classification outstanding so, student with $x=5$ and $y=7$ is classified as "out standing"

1)

class	colors		type		origin	
	Red	yellow	sports	SUV	domestic	imported
yes	3/15	2/15	4/15	1/15	2/15	3/15
NO	2/15	3/15	2/15	3/15	3/15	2/15

$\rightarrow Q_1 = P(x_1|c_1)\ P(x_2|c_1)\ P(x_3|c_1)\ P(c_1)$

probability of yes $= (3/5)\ (1/5)\ (2/5)\ (1/2) = 0.024$

Q2

probability of NO $= P(x_1|c_2)\ P(x_2|c_2)\ P(x_3|c_2)\ P(c_2)$

$= (2/5)\ (3/5)\ (3/5)\ (1/2) = 0.072$

∴ The prediction of whether a red domestic SUV car is being stolen can or not is NO.

According to Naive Bayes algorithm with the above data is predicted as 'NO'.