Name: Yunika Upadhayaya

Student ID: 1001631183

Assignment 2 (Written)

**Task – 1:**

Let Maine 🡪 M and Sahara 🡪 S, L 🡪 temp < 80, H 🡪 temp > 80

Given – P (M) = 5%, P (S) = 95%,

Conditions – Days are conditionally independent, and location of the sensor is also independent.

i.e., P (M) = 0.05 and P (S) = 0.95 (Prior probability of sensor placement)

For Maine:

P (temp >= 80) = 20 % = 0.2

P (temp < 80) = 80 % = 0.8 i.e., P (L | M) = 0.8

P (L | MC) = 0.1 (Here, c is compliment)

For Sahara:

P (temp >= 80) = 90 % = 0.9

P (temp < 80) = 10 % = 0.1 i.e., P (L | S) = 0.1

1. P (M | L) = P (L | M) \* P (M) / [P (M) P (L | M) + P (Mc) P (L | Mc)]

= 0.8 \* 0.05 / (0.05 \* 0.8 + 0.95 \* 0.1)

= 0.2963

1. From (a), the posterior probability of sensor placement is 29.63% for Maine and 70.37% for Sahara.

Using the law of total probability:

P (L) = 0.2963 \* 0.8 + 0.7037 \* 0.1 = 0.23704 + 0.07037 = 0.30741

1. Probability of the first three emails that indicate daily highs under 80 degrees if the sensor is in Maine = P(M) \* P (L | M) \* P (L | M) \* P (L | M) = 0.05 \* 0.8 \* 0.8 \* 0.8 = 0.025

Probability of the first three emails that indicate daily highs under 80 degrees if the sensor is in Sahara = P(S) \* P (L | S) \* P (L | S) \* P (L | S) = 0.95 \* 0.1 \* 0.1 \* 0.1 = 0.00095

Thus, Probability (first 3 emails is under 80) = 0.025 + 0.00095 = 0.02655

**Task – 2:**

Given – Function P is defined on a sample S = {A, B, C, D}

P (A) = 0.3

P (B) = 0.6

P is a probability mass function (p.m.f) since A, B, C and D are discrete points.

P (X = x) = fx(x), fx(x) >=0, ∑xfx(x) = 1

Here, fA(A) = 0.3

FB(B) = 0.6

FC(C) and FD(D) are unknown. Thus,

|  |
| --- |
| P can be probability function if P(C) + P(D) = 1 – 0.3 – 0.6 = 0.1  or P (C) + P (D) <= 1  Otherwise, it cannot be a probability density function. |

**Task – 3:**

Given – P(x) = 0.3 if x ∈ [0, 10]

For a function P(x) to be a probability density function, following conditions should be satisfied:

1. P(x) >= 0 for all x



= a + + b

= a + 3 + b

Since, we do not have the information about a and b, P(x) cannot be a valid probability density function.

|  |
| --- |
| P(x) can be a valid probability function if:  P (- ∞ < x < 0) + P (10 < x < ∞) = 1 – P (0 < x < 10)  = 1 – 0.3  = 0.7 |

**Task – 4:**

Given:

P (B = r) = 0.4

P (B = b) = 0.6

P (F = a | B = r) = 0.25

P (F = o | B = r) = 0.75

P (F = a | B = b) = 0.75

P (F = o | B = b) = 0.25

Here,

We know that if x = a:



p (x = a) = ∑ p (x | ck) \* p (ck) is the output given that (x = a) is picked.



Now,

p (F = a) = p (F = a | B = r) \* p (B = r) + p (F = a + B = b) \* p (B = b)

= 0.25 \* 0.4 + 0.75 \* 0.6

= 0.55

p (F = o) = p (F = o | B = r) \* p (B = r) + p (F = o + B = b) \* p (B = b)

= 0.75 \* 0.4 + 0.25 \* 0.6

= 0.45

Also,

p (B = r | F = a) = p (F = a | B = r) \* p (B = r) / p (F = a) = 0.25 \* 0.4 / 0.55 = 0.182

p (B = b | F = a) = p (F = a | B = b) \* p (B = b) / p (F = a) = 0.75 \* 0.6 / 0.55 = 0.82

Output – If the **box is blue**, then the **fruit is apple** 82% of the time.

p (B = r | F = o) = p (F = o | B = r) \* p (B = r) / p (F = o) = 0.75 \* 0.4 / 0.45 = 0.667%

p (B = b | F = o) = p (F = o | B = b) \* p (B = b) / p (F = o) = 0.25 \* 0.6 / 0.45 = 0.333%

Output – If the **box is red**, then **the fruit is orange** 67% of the time.

**Task – 5:**

Output produced by the training stage when given **yeast\_training.txt** as the input file:

|  |
| --- |
| Class 1, attribute 1, mean = 0.52, std = 0.10  Class 1, attribute 2, mean = 0.54, std = 0.10  Class 1, attribute 3, mean = 0.52, std = 0.07  Class 1, attribute 4, mean = 0.41, std = 0.17  Class 1, attribute 5, mean = 0.50, std = 0.01  Class 1, attribute 6, mean = 0.00, std = 0.01  Class 1, attribute 7, mean = 0.50, std = 0.05  Class 1, attribute 8, mean = 0.24, std = 0.05  Class 2, attribute 1, mean = 0.45, std = 0.11  Class 2, attribute 2, mean = 0.45, std = 0.10  Class 2, attribute 3, mean = 0.53, std = 0.06  Class 2, attribute 4, mean = 0.23, std = 0.11  Class 2, attribute 5, mean = 0.50, std = 0.04  Class 2, attribute 6, mean = 0.00, std = 0.01  Class 2, attribute 7, mean = 0.49, std = 0.06  Class 2, attribute 8, mean = 0.33, std = 0.14  Class 3, attribute 1, mean = 0.43, std = 0.10  Class 3, attribute 2, mean = 0.48, std = 0.11  Class 3, attribute 3, mean = 0.36, std = 0.06  Class 3, attribute 4, mean = 0.22, std = 0.08  Class 3, attribute 5, mean = 0.51, std = 0.05  Class 3, attribute 6, mean = 0.00, std = 0.01  Class 3, attribute 7, mean = 0.51, std = 0.04  Class 3, attribute 8, mean = 0.27, std = 0.09  Class 4, attribute 1, mean = 0.79, std = 0.07  Class 4, attribute 2, mean = 0.76, std = 0.07  Class 4, attribute 3, mean = 0.38, std = 0.06  Class 4, attribute 4, mean = 0.32, std = 0.11  Class 4, attribute 5, mean = 0.50, std = 0.01  Class 4, attribute 6, mean = 0.00, std = 0.01  Class 4, attribute 7, mean = 0.51, std = 0.07  Class 4, attribute 8, mean = 0.27, std = 0.09  Class 5, attribute 1, mean = 0.74, std = 0.16  Class 5, attribute 2, mean = 0.62, std = 0.13  Class 5, attribute 3, mean = 0.42, std = 0.08  Class 5, attribute 4, mean = 0.30, std = 0.12  Class 5, attribute 5, mean = 0.50, std = 0.01  Class 5, attribute 6, mean = 0.00, std = 0.01  Class 5, attribute 7, mean = 0.51, std = 0.06  Class 5, attribute 8, mean = 0.24, std = 0.04  Class 6, attribute 1, mean = 0.54, std = 0.14  Class 6, attribute 2, mean = 0.50, std = 0.12  Class 6, attribute 3, mean = 0.51, std = 0.05  Class 6, attribute 4, mean = 0.24, std = 0.10  Class 6, attribute 5, mean = 0.50, std = 0.01  Class 6, attribute 6, mean = 0.49, std = 0.39  Class 6, attribute 7, mean = 0.51, std = 0.03  Class 6, attribute 8, mean = 0.24, std = 0.05  Class 7, attribute 1, mean = 0.48, std = 0.11  Class 7, attribute 2, mean = 0.47, std = 0.09  Class 7, attribute 3, mean = 0.54, std = 0.06  Class 7, attribute 4, mean = 0.22, std = 0.12  Class 7, attribute 5, mean = 0.50, std = 0.04  Class 7, attribute 6, mean = 0.00, std = 0.03  Class 7, attribute 7, mean = 0.50, std = 0.06  Class 7, attribute 8, mean = 0.26, std = 0.09  Class 8, attribute 1, mean = 0.74, std = 0.11  Class 8, attribute 2, mean = 0.73, std = 0.11  Class 8, attribute 3, mean = 0.49, std = 0.05  Class 8, attribute 4, mean = 0.29, std = 0.07  Class 8, attribute 5, mean = 0.50, std = 0.01  Class 8, attribute 6, mean = 0.00, std = 0.01  Class 8, attribute 7, mean = 0.46, std = 0.08  Class 8, attribute 8, mean = 0.23, std = 0.02  Class 9, attribute 1, mean = 0.55, std = 0.14  Class 9, attribute 2, mean = 0.56, std = 0.16  Class 9, attribute 3, mean = 0.51, std = 0.07  Class 9, attribute 4, mean = 0.20, std = 0.07  Class 9, attribute 5, mean = 0.50, std = 0.01  Class 9, attribute 6, mean = 0.00, std = 0.01  Class 9, attribute 7, mean = 0.53, std = 0.05  Class 9, attribute 8, mean = 0.24, std = 0.05  Class 10, attribute 1, mean = 0.78, std = 0.06  Class 10, attribute 2, mean = 0.73, std = 0.12  Class 10, attribute 3, mean = 0.48, std = 0.11  Class 10, attribute 4, mean = 0.33, std = 0.07  Class 10, attribute 5, mean = 1.00, std = 0.01  Class 10, attribute 6, mean = 0.00, std = 0.01  Class 10, attribute 7, mean = 0.55, std = 0.02  Class 10, attribute 8, mean = 0.23, std = 0.01 |

Output produced by the last line of the test stage of the program given **yeast\_test.txt** as the input file is:

|  |
| --- |
| Classification accuracy = 0.4483 |