Shubham Arya 1001650536

CSE 4309 – Assignment 2

Task-1

P (sensor = Maine) = 0.05, P (sensor = Sahara) = 0.95,
P (temp >= 80 | Maine) = 0.2, P (temp < 80 | Maine) = 0.8
P (temp >= 80 | Sahara) = 0.9, P (temp < 80 | Sahara) = 0.1

a) P (Maine | temp < 80) = P (Maine) P (temp < 80 | Maine) / P (temp < 80)
= P (Maine) P (temp < 80 | Maine) / [P (Maine) P (temp < 80 | Maine) + P (Sahara) P (temp < 80 | Sahara)]
= 0.05*0.8 / [0.05*0.8 + 0.95*0.1]
= 0.04 / [0.04 + 0.095]
= 0.04 / 0.135
= 0.2963

b) P (temp < 80) = P (Maine | temp < 80) * P (temp < 80 | Maine) + P (Sahara | temp < 80) * P (temp < 80 | Sahara)
= 0.2963 * 0.8 + (1 – 0.2963) * 0.1
= 0.23704 + 0.07037
= 0.30741

c) From part b, we get the probability of getting two emails indicating a daily high under 80 degrees for 2 consecutive days. Let's call this event $t_{1,2}$
The probability of first three emails indicating daily highs under 80 could be represented with the product rule like- P (t1, t2, t3) = P (t1 | t2, t3) * P (t2 | t3) * P (t3).
However, knowing the probability for the first 2 days, we could write it as:
P ($t_{1,2}$, t3) = P ($t_{1,2}$ | t3) * P (t3) = P (t3 | $t_{1,2}$) * P ($t_{1,2}$) from the bayes rule.
And P (t3 | $t_{1,2}$) = P (t3 and $t_{1,2}$) / P ($t_{1,2}$)

P (temp < 80 at either location) = 0.2963 * 0.05 + 0.7031 * 0.95 = 0.014815 + 0.667945 = 0.68276
Since each day is conditionally independent of the other day, P(t3) = 0.68276

Therefore, P ($t_{1,2}$, t3) = P (t3 | $t_{1,2}$) * P ($t_{1,2}$) = P (t3 and $t_{1,2}$) = P (t3) * P($t_{1,2}$) = 0.68276 * 0.30741 = 0.20988725

Task-2

P is possibly a probability function. Probability cannot be great than 1. With P(A) = 0.3 and P(B) = 0.6, the P(A) + P(B) = 0.9. If P(C) + P(D) is equal to 0.1, then P is a probability function, otherwise it is not.

Task-3

P is definitely not a probability function because integrating P(x) = 0.3 from the limits 0 to 10 would give us 3. This is not less than or equal to 1, so P cannot be a probability density function.

Task-4

- 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

P (F = a) = p (F = a, B = r) + p (F = a, B = b)

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

= 0.25 * 0.4 + 0.75 * 0.6

= 0.1 + 0.45

= 0.55

P (F = o) = 1 – P (F = a) = 0.45

P (B = r | F = a) = p (F = a | B = r) * p (B = r) / p (F = a) = 0.25 * 0.4 / 0.55 = 0.1818
P (B = b | F = a) = p (F = a | B = b) * p (B = b) / p (F = a) = 0.75 * 0.6 / 0.55 = 0.8181
P (B = r | F = o) = p (F = o | B = r) * p (B = r) / p (F = o) = 0.75 * 0.4 / 0.45 = 0.6667
P (B = b | F = o) = p (F = o | B = b) * p (B = b) / p (F = o) = 0.25 * 0.6 / 0.45 = 0.3333

If F = a, if the prediction is B = r, then the classifier will be 18.18% correct.
If F = a, if the prediction is B = b, then the classifier will be 81.81% correct.
If F = o, if the prediction is B = r, then the classifier will be 66.67% correct.
If F = o, if the prediction is B = b, then the classifier will be 33.33% correct.

Task-5

```
Run command for your terminal:
python3 naive_bayes.py [path of training file] [path of test file]
```

classification accuracy is 0.4483