CSE 4309 Assignment 2

Inshaad Merchant - 1001861293

Due Date: 12 September 2024

Task 1:

Part a:

Given:

$$P(M) = 0.05, \quad P(D \mid M) = 0.8, \quad P(D \mid S) = 0.1, \quad P(S) = 0.95,$$

$$P(M \mid D) = \frac{P(D \mid M) * P(M)}{P(D)}$$

Thus,

$$P(M \mid D) = \frac{0.8 * 0.05}{(0.8 * 0.05) + (0.1 * 0.95)} = 0.296$$

Part b:

$P(\text{second email under 80} \mid \text{first email under 80}) = P(D) = (0.8*0.05)+(0.1*0.95) = 0.135$

Part c:

$$P(\text{first three under 80}) = 0.135 * 0.135 * 0.135 = 0.00246$$

Task 2:

To be a valid probability function, the following conditions must be satisfied:

1. $P(s) \geq 0$ for all $s \in S$.

2. $\sum_{s \in S} P(s) = 1$.

The sum of probabilities $P(A) + P(B) = 0.9$ satisfies the first condition, indicating that $P(C)$ and $P(D)$ could potentially satisfy the second condition if $P(C) + P(D) = 0.1$ and each is non-negative. Thus, P is POSSIBLY a probability function.

Task 3:

Given:

- $P(x) = 0.3$ for $0 \leq x \leq 10$

The function satisfies the first condition that the probability is positive. However, calculating the integral:

$$\int_0^{10} P(x)\, dx = \int_0^{10} 0.3\, dx = 3$$

Since this integral equals 3, not 1, $P(x)$ does not satisfy the second condition and is DEFINITELY NOT a probability function.

Task 4:

Marginal Probabilities:

$$\begin{aligned}
p(F = a) &= p(F = a \mid B = r) \times p(B = r) + p(F = a \mid B = b) \times p(B = b) \\
&= 0.25 \times 0.4 + 0.75 \times 0.6 = 0.55 \\
p(F = o) &= p(F = o \mid B = r) \times p(B = r) + p(F = o \mid B = b) \times p(B = b) \\
&= 0.75 \times 0.4 + 0.25 \times 0.6 = 0.45
\end{aligned}$$

Posterior Probabilities:

$$p(B = r \mid F = a) = \frac{p(F = a \mid B = r) \times p(B = r)}{p(F = a)} = \frac{0.25 \times 0.4}{0.55} = 0.182$$

$$p(B = b \mid F = a) = \frac{p(F = a \mid B = b) \times p(B = b)}{p(F = a)} = \frac{0.75 \times 0.6}{0.55} = 0.818$$

$$p(B = r \mid F = o) = \frac{p(F = o \mid B = r) \times p(B = r)}{p(F = o)} = \frac{0.75 \times 0.4}{0.45} = 0.667$$

$$p(B = b \mid F = o) = \frac{p(F = o \mid B = b) \times p(B = b)}{p(F = o)} = \frac{0.25 \times 0.6}{0.45} = 0.333$$

Classifier Accuracy:

The Bayesian classifier chooses Blue box with probability 0.818.

Task 5:

Python file is attached to this document.
classification accuracy for yeast dataset = 0.4463
classification accuracy for pendigits dataset = 0.0875
classification accuracy for satellite dataset = 0.2115