

Task 1

Given information

$p(y = \text{Maine}) = 0.05$, $p(y = \text{Sahara}) = 0.95$, $p(x = >80 \mid y = \text{Maine}) = 0.2$, $p(x = <80 \mid y = \text{Maine}) = 0.8$, $p(x = >80 \mid y = \text{Sahara}) = 0.9$, $p(x = <80 \mid y = \text{Sahara}) = 0.1$

Part a

$$p(y = \text{Maine} \mid x = <80) = (p(x = <80 \mid y = \text{Maine}) * p(y = \text{Maine})) / p(x = <80)$$

$$= (p(x = <80 \mid y = \text{Maine}) * p(y = \text{Maine})) / (p(x = <80 \mid y = \text{Maine}) * p(y = \text{Maine}) + p(x = <80 \mid y = \text{Sahara}) * p(y = \text{Sahara}))$$

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

$$= 0.296$$

Part b

$$p(x_1 = <80 \mid x_2 = <80)$$

$$= p(x_2 = <80 \mid x_1 = <80, y = \text{Maine}) * p(y = \text{Maine} \mid x_1 = <80) + p(x_2 = <80 \mid x_1 = <80, y = \text{Sahara}) * p(y = \text{Sahara} \mid x_1 = <80)$$

$$= p(x_2 = <80 \mid y = \text{Maine}) * p(y = \text{Maine} \mid x_1 = <80) + p(x_2 = <80 \mid y = \text{Sahara}) * p(y = \text{Sahara} \mid x_1 = <80)$$

$$= 0.8 * 0.296 + 0.1 * ((p(x_1 = <80 \mid y = \text{Sahara}) * p(y = \text{Sahara})) / p(x_1 = <80))$$

$$= 0.8 * 0.296 + 0.1 * ((0.1 * 0.95) / 0.135)$$

$$= 0.30717$$

Part c

$$p(x_1 = <80, x_2 = <80, x_3 = <80)$$

$$= p(x_1 = <80, x_2 = <80, x_3 = <80 \mid y = \text{Maine}) * p(y = \text{Maine}) + p(x_1 = <80, x_2 = <80, x_3 = <80 \mid y = \text{Sahara}) * p(y = \text{Sahara})$$

$$= p(x_1 = <80 \mid y = \text{Maine}) * p(x_2 = <80 \mid y = \text{Maine}) * p(x_3 = <80 \mid y = \text{Maine}) * p(y = \text{Maine}) + p(x_1 = <80 \mid y = \text{Sahara}) * p(x_2 = <80 \mid y = \text{Sahara}) * p(x_3 = <80 \mid y = \text{Sahara}) * p(y = \text{Sahara})$$

$$= 0.8 * 0.8 * 0.8 * 0.05 + 0.1 * 0.1 * 0.1 * 0.95$$

$$= 0.02655$$

Task 2

$$p(A) + p(B) + p(C) + p(D) = 1$$

$$0.3 + 0.6 + p(C) + p(D) = 1$$

$$p(C) + p(D) = 1 - 0.9$$

$$p(C) + p(D) = 0.1$$

If this is true then P is definitely a probability function but there is no information on p(C) or p(D) so we can't be sure. However if the p(C) + p(D) do not add up to 1 then we can say it is not a probability function. Since there is no information given on the p(C) and p(D), we can say P is possibly a probability function.

Task 3

A probability density function has to be $p(x) \geq 0$ and $\int p(x) dx = 1$

Given $p(x) = 0.3$ when $0 \leq x \leq 10$

$$\begin{aligned} \int_{-\infty}^{\infty} p(x) dx &= \int_{-\infty}^0 p(x) dx + \int_0^{10} p(x) dx + \int_{10}^{\infty} p(x) dx \\ &= \int_{-\infty}^0 0.3 dx + \int_0^{10} 0.3 dx + \int_{10}^{\infty} 0.3 dx \\ &= x_1 + 0.3(10-0) + x_2 \\ &= x_1 + 3 + x_2 \end{aligned}$$

However, we do not know the values of x_1 and x_2 . If $x_1 + x_2 + 3$ equal to 0 or more then it is a density function. However if the sum is a negative number then it is not a density function. Since we do not have enough information $p(x)$ is possibly a probability density function.

Task 4

Refer to naive_bayes.m