Problem 1. *This question is a classical question: Consider a medical test for a rare disease, where only 1% of the population is affected. Suppose the medical test is 95% accurate when a person has the disease (true positive rate) and 90% accurate when a person does not have the disease (true negative rate). Question: If a person tests positive for the disease, what is the probability that they actually have the disease?* ***Given:***

- *A: The person has the disease (1% of the population).*
- *B: The person tests positive for the disease.*
- *A': The person does not have the disease (complement of A).*
- *B': The person tests negative for the disease (complement of B).*

Using Bayes' theorem:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

Answer Key:

$P(B) = \sum_i P(B \mid A_i)$

Problem 2. *Use Bayesian Rule to answer the following question:*
A dataset is given as:

Type	***Long***	***Not Long***	***Sweet***	***Not Sweet***	***Yellow***	***Not Yellow***	***Total***
BANANA	*400*	*100*	*350*	*150*	*450*	*50*	*500*
ORANGE	*0*	*300*	*150*	*150*	*300*	*0*	*300*
OTHERS	*100*	*100*	*150*	*50*	*50*	*150*	*200*
TOTAL	*500*	*500*	*650*	*350*	*800*	*200*	*1000*

If the fruit is Long, Sweet, and Yellow, What is this fruit?

In other words:

Answer Key:

$P(BANANA \mid Long, Sweet, Yellow) = ?$
$P(ORANGE \mid Long, Sweet, Yellow) = ?$
$P(OTHERS \mid Long, Sweet, Yellow) = ?$

Problem 3. *Suppose we have the following dataset representing the joint probabilities of* ***Hours Studied (X)*** *and* ***Exam Score (Y)****:*

X	Y	Joint Probability
2	60	0.05
2	70	0.10
2	80	0.15
3	60	0.10
3	70	0.20
3	80	0.20
4	60	0.05
4	70	0.10
4	80	0.05

Table 1: Joint Probability Dataset for Variables X and Y.

Find the covariance between "Hours Studied" and "Exam Score." (Use Python to verify your answer, you need to turn in your code)

This should help you:

$$\mathrm{E}[X] = \sum_{i=1}^{n} x_i \cdot P\left(X = x_i\right)$$

$$\mathrm{E}[Y] = \sum_{j=1}^{m} y_j \cdot P\left(Y = y_j\right)$$

$$\operatorname{cov}(X, Y) = \sum_{x \in X} \sum_{y \in Y} (x - \mathrm{E}[X])(y - \mathrm{E}[Y]) \cdot P(X = x, Y = y)$$

Problem 4. *find the maximum likelihood estimates (MLE) of the mean (μ_{MLE}) and variance (σ^2_{MLE}) for the given data points $x_1 = 5$ and $x_2 = 6$ assuming a Gaussian distribution.*

For the Gaussian distribution, the probability density function (PDF) is given by:

$$P\left(x \mid \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

and the Likelihood is given by:

$$L\left(\mu, \sigma^2\right) = P\left(x_1 = 5, x_2 = 6 \mid \mu, \sigma^2\right) = \prod_{i=1}^{2} P\left(x_i \mid \mu, \sigma^2\right)$$

Problem 5. *Suppose we have two variables, X and Y, and we want to model their joint distribution using a bivariate Gaussian distribution. Let's assume the mean vector and covariance matrix are as follows:*

Mean vector: $\boldsymbol{\mu} = [3, 5]$ *Covariance matrix:* $\boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1.5 \\ 1.5 & 3 \end{bmatrix}$

Use numpy np.random.multivariate_normal to generate 1000 random points, then use the generated random points to estimate the mean vector and the Cov matrix again by using np.mean and np.cov. what is your error rate?