Problem 1. Being a student in the Netherlands, you spend all your time in the cities of Amsterdam and Rotterdam. Based on your experience, the weather in Amsterdam is much nicer: the probability that it rains when you are in Amsterdam is 0.5, while the probability that it rains when in Rotterdam is 0.75. Amsterdam is where you spend most of your time: at any given moment, the probability that you are in Amsterdam is 0.8 and the probability that you are in Rotterdam is 0.2.

1. Define the random variables and the values they can take on, both with symbols and numerically.
   Solution:
   We define two random variables:

   - The variable L denotes my location. It can take values *Ams* to symbolize I am in Amsterdam, or *Rot* to symbolize I am in Rotterdam. Given the above information we have that
   $$p(L = Ams) = 0.8 \tag{1}$$
   $$p(L = Rot) = 0.2 \tag{2}$$

   - The variable R denotes the fact that it is raining where I am. As such, it can only be *True* or *False*. Again, the problem description is represented as
   $$p(R = True | L = Ams) = 0.5 \tag{3}$$
   $$p(R = True | L = Rot) = 0.75 \tag{4}$$

2. What is the probability that it does not rain when you are in Rotterdam?
   Solution:
   We are looking for $p(R = False | L = Rot)$. Since R can only take values within $\{True, False\}$, then
   $$p(R = True | L = Rot) + p(R = False | L = Rot) = 1$$
   must hold. Therefore,
   $$p(R = False | L = Rot) = 1 - p(R = True | L = Rot) = 1 - 0.75 = 0.25$$

3. What is the probability that it rains where you are?
   Solution:
   This time we are looking for $p(R = True)$. This represents the marginal probability of
   $$p(R = True) = p(R = True | L = Ams) * p(L = Ams) + p(R = True | L = Rot) * p(L = Rot)$$
   Substituting from Equation 3, Equation 1, Equation 4 and Equation 2, we have
   $$p(R = True) = 0.5 * 0.8 + 0.75 * 0.2 = 0.4 + 0.15 = 0.55 \tag{5}$$

4. You wake up on the sidewalk, after a night out which you can't remember anything about but which clearly was not such a great idea. You can't recognize your surroundings, but you must be either in Amsterdam or Rotterdam. It is raining. What is the probability that you are in Amsterdam?

   Solution:

   We would like to find $p(L = Ams | R = True)$. According to Bayes Theorem we know that

   $$p(A = a | B = b) = \frac{p(B = b | A = a) * p(A = a)}{p(B = b)} \tag{6}$$

   Substituting Equation 3, Equation 1 and Equation 5 we get

   $$\begin{aligned} p(L = Ams | R = True) &= \frac{p(R = True | L = Ams) * p(L = Ams)}{p(R = True)} \\ &= \frac{0.5 * 0.8}{0.55} = \frac{0.4}{0.55} \\ &= 0.7272727 \end{aligned}$$

Problem 2. In a particular city with a population of 500000, it estimated that 500 people have cancer. There is a blood test that 99 times out of 100 correctly diagnoses cancer in patients. It also, unfortunately, misdiagnoses 5% of people that do not have cancer.

With this information, answer the following questions:

1. What is $p(cancer)$ and $p(notcancer)$?

   Solution:

   Taking a frequentist approach, we interpret the probability of having cancer as the ratio of people having cancer in the population. Using the given information about the population we calculate $p(CANCER = True)$,

   $$p(CANCER = True) = \frac{cancerPopulation}{totalPopulation} = \frac{500}{500,000} = 0.001 \tag{7}$$

   Since the variable $CANCER$ may only have values in $\{True, False\}$, we calculate the complementary probability as

   $$p(CANCER = False) = 1 - p(CANCER = True) = 1 - 0.001 = 0.999 \tag{8}$$

2. If a patient takes the blood test and it returns positive, what is the probability the patient has cancer?

   Solution:

   We define a new variable $TEST$ to define the result of the test, which has range $\{Pos, Neg\}$. Therefore, we want to know $p(CANCER = True | TEST = Pos)$.

   To do so, first we calculate the prior probabilities $p(TEST = Pos | CANCER = True)$ and $p(TEST = Pos | CANCER = False)$ from the problem description:

   $$p(TEST = Pos | CANCER = True) = \frac{99}{100} = 0.99 \tag{9}$$

   $$p(TEST = Pos | CANCER = False) = 0.05 \tag{10}$$

Next we calculate the evidence, meaning the marginal probability of $p(TEST = Pos)$. Using Equation 9, Equation 7, Equation 10 and Equation 8

$$
\begin{aligned}
p(TEST &= Pos) \\
&= p(TEST = Pos|CANCER = True) * p(CANCER = True) \\
&+ p(TEST = Pos|CANCER = False) * p(CANCER = False) \\
&= 0.99 * 0.001 + 0.05 * 0.999 = 0.05094
\end{aligned}
\tag{11}
$$

Following Bayes Theorem in Equation 6, we have

$$
\begin{aligned}
p(CANCER &= True|TEST = Pos) \\
&= \frac{p(TEST = Pos|CANCER = True) * p(CANCER = True)}{p(TEST = Pos)} \\
&= \frac{0.99 * 0.001}{.5094} = 0.1943
\end{aligned}
\tag{12}
$$

3. What are some of the assumptions we are implicitly making when answering this question?
   Solution:
   We assume that everyone in the population (or at least a representative sample of the population) will take the test. However, the patients taking a blood test might have already presented some symptoms that hint to the fact that they might have cancer. Therefore, the ratio of patients with cancer within the test takers is not the same as the ratio in the total population.

Problem 3. For this question you will write the expression for the posterior parameter distribution for a simple data problem. Assume we observe N univariate data points $\{x_1; x_2; ....; x_N\}$. Further, we assume that they are generated by a Gaussian distribution with known variance σ^2, but unknown mean μ. Assume a prior Gaussian distribution over the unknown mean, i.e. $p(\mu) = \mathcal{N}(\mu|\mu_0, \sigma_0^2)$. When answering these questions, use $\mathcal{N}(a|b, c^2)$ to indicate a Gaussian (normal) distribution over a with mean b and variance c^2.

1. Write down the general expression for a posterior distribution, using for the parameter θ, D for the data. Indicate the prior, likelihood, evidence, and posterior.
   Solution:
   A posterior distribution is usually written as the Bayes Theorem states in Equation 6. Using θ and D we get:
   $$
   p(\theta|D) = \frac{p(D|\theta) * p(\theta)}{p(D)}
   \tag{13}
   $$
   where $p(\theta|D)$ is the **posterior**
   $p(D|\theta)$ is the **likelihood**
   $p(\theta)$ is the **prior**
   $p(D)$ is the **evidence**
   Sometimes, when $p(D)$ is not known we can rewrite it as a marginalization of a continuous variable. Therefore:
   $$
   p(\theta|D) = \frac{p(D|\theta) * p(\theta)}{\int_{\theta_0 = \theta} p(D|\theta_0) * p(\theta_0)}
   \tag{14}
   $$

2. Write the posterior for this particular example. You do not need an analytic solution.
   Solution:
   Taking Equation 14, and substituting the particular Normal distributions given in this example, we have:
   $p(\theta|D) = p(\mu|D)$
   $p(D|\theta) = p(D|\mu) = \prod_{n=1}^{N} \mathcal{N}(x_n|\mu, \sigma^2)$
   $p(\theta) = p(\mu) = \mathcal{N}(\mu|\mu_0, \sigma_0^2)$
   $p(D) = \int_\mu \mathcal{N}(\mu|\mu_0, \sigma_0^2) * \prod_{n=1}^{N} \mathcal{N}(x_n|\mu, \sigma^2)$

   $$p(\theta|D) = \frac{p(D|\theta) * p(\theta)}{\int_{\theta_0=\theta} p(D|\theta_0) * p(\theta_0)}$$
   $$= \frac{\mathcal{N}(\mu|\mu_0, \sigma_0^2) * \prod_{n=1}^{N} \mathcal{N}(x_n|\mu, \sigma^2}{\int_\mu \mathcal{N}(\mu|\mu_0, \sigma_0^2) * \prod_{n=1}^{N} \mathcal{N}(x_n|\mu, \sigma^2)} \qquad (15)$$

   As a final comment, we can guarantee that the posterior will be a Gaussian given that the prior is Gaussian as well.

Problem 4. Let A $= \begin{bmatrix} 3 & 5 \\ 2 & 3 \end{bmatrix}$ and b $= \begin{bmatrix} 9 \\ 5 \end{bmatrix}$

1. Compute Ab
   Solution:
   $Ab = \begin{bmatrix} 3 & 5 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 9 \\ 5 \end{bmatrix} = \begin{bmatrix} 3*9+5*5 \\ 2*9+3*5 \end{bmatrix} = \begin{bmatrix} 27+25 \\ 18+15 \end{bmatrix} = \begin{bmatrix} 52 \\ 33 \end{bmatrix}$

2. Compute $b^T A$
   Solution:
   $b^T A = \begin{bmatrix} 9 & 5 \end{bmatrix} * \begin{bmatrix} 3 & 5 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 9*3+5*2 & 9*5+5*3 \end{bmatrix} = \begin{bmatrix} 37 & 60 \end{bmatrix}$

3. What is the vector c for which $Ac = b$
   Solution:
   The vector c can be found by solving a linear system of equations:

   $$\begin{cases} 3c_1 + 5c_2 = 9 \\ 2c_1 + 3c_2 = 5 \end{cases} \qquad (16)$$

   we find the value for c_1 from the first sub-equation in 16

   $$c_1 = 3 - \frac{5}{3}c_2 \qquad (17)$$

   Next, we take this value and substitute in the second sub-equation of 16.

   $$2 \cdot (3 - \frac{5}{3} \cdot c_2) + 5 \cdot c_2 = 5$$
   $$6 - \frac{10}{3} \cdot c_2 + 3 \cdot c_2 = 5$$
   $$18 - 10 \cdot c_2 + 9 \cdot c_2 = 15$$
   $$c_2 = -3$$

This helps us find c_1

$$c_1 = 3 - \frac{5}{3}c_2 = 3 - \frac{5}{3} * (-3) = 3 - 5 = -2 \tag{18}$$

Finally, we state that

$$c = \begin{bmatrix} -2 & 3 \end{bmatrix} \tag{19}$$

4. What is A^{-1}?
   Solution:
   We know that the inverse of a 2x2 matrix can be found with the following formula:

$$A^{-1} = \frac{1}{det(A)} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{20}$$

   where $det(A) = | a_{11} \cdot a_{22} - a_{12} \cdot a_{21} |$

   Using Equation 20 for the given matrix A, we get:

$$A^{-1} = \frac{1}{3 \cdot 3 - 5 \cdot 2} \begin{bmatrix} 3 & -5 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} -3 & 5 \\ 2 & -3 \end{bmatrix} \tag{21}$$

5. Verify that $A^{-1}b = c$. Show that this must be the case.
   Solution:
   We substitute the values in 19 and 21

$$A^{-1}b = \begin{bmatrix} -3 & -5 \\ 2 & -3 \end{bmatrix} * \begin{bmatrix} 9 \\ 5 \end{bmatrix} = \begin{bmatrix} -3*9 + 5*5 \\ 2*9 + 5*(-3) \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \end{bmatrix} = c \tag{22}$$

   This makes sense because the inverse matrix method is equivalent to solving a linear system of equations, as we did before.

Problem 5. Find the gradient of the following functions

1. $x^2 + 2x + 3$
   Solution:
   $(x^2 + 2x + 3)' = 2x + 2$

2. $(2x^3 + 1)^2$
   Solution:

$$\begin{aligned}
(2x^3 + 1)^{2'} &= 2(2x^3 + 1)(2x^3 + 1)' \\
&= (4x^3 + 2)(3 \cdot 2x^2) \\
&= 4x^3 6x^2 + 2 \cdot 6x^2 = 24x^5 + 12x^2
\end{aligned} \tag{23}$$

Find the partial derivative of the following functions with respect to x,y, z

1. $f(x, y, z) = (x + 2y)^2 sin(xy)$
   Solution:

Partial derivative with respect to x.

$$\frac{\partial f}{\partial x} = 2(x + 2y)\frac{\partial(x + 2y)}{\partial x} * \sin(xy) + (x + 2y)^2 \frac{\partial \sin(xy)}{\partial x} \tag{24}$$
$$= 2(x + 2y)\sin(xy) + y(x + 2y)^2 \cos(xy) \tag{25}$$

Partial derivative with respect to y.

$$\frac{\partial f}{\partial y} = 2(x + 2y)\frac{\partial(x + 2y)}{\partial y} * \sin(xy) + (x + 2y)^2 \frac{\partial \sin(xy)}{\partial x} \tag{26}$$
$$= 4(x + 2y)\sin(xy) + x(x + 2y)^2 \cos(xy) \tag{27}$$

Partial derivative with respect to z.

$$\frac{\partial f}{\partial z} = 0 \tag{28}$$

2. $f(x, y, z) = 2\log(x + y^2 - z)$
   Solution:
   Partial derivative with respect to x.

$$\frac{\partial f}{\partial x} = \frac{2}{x + y^2 - z}\frac{\partial(x + y^2 - z)}{\partial x} = \frac{2}{x + y^2 - z} \tag{29}$$

Partial derivative with respect to y.

$$\frac{\partial f}{\partial y} = \frac{2}{x + y^2 - z}\frac{\partial(x + y^2 - z)}{\partial y} = \frac{2 * 2y}{x + y^2 - z} = \frac{4y}{x + y^2 - z} \tag{30}$$

Partial derivative with respect to z.

$$\frac{\partial f}{\partial z} = \frac{2 * -1}{x + y^2 - z} \tag{31}$$

3. $f(x, y, z) = \exp(x\cos(y + z))$
   Solution:
   Partial derivative with respect to x.

$$\frac{\partial f}{\partial x} = \exp(x\cos(y + z))\frac{\partial(x\cos(y + z))}{\partial x} = \exp(x\cos(y + z))\cos(y + z) \tag{32}$$

Partial derivative with respect to y.

$$\frac{\partial f}{\partial y} = \exp(x\cos(y + z))\frac{\partial(x\cos(y + z))}{\partial y} = \exp(x\cos(y + z))[x(\sin(y + z))\frac{\partial(y + z)}{\partial y}] \tag{33}$$
$$= \exp(x\cos(y + z))x * \sin(y + z) * 1 \tag{34}$$

Partial derivative with respect to z.

$$\frac{\partial f}{\partial z} = \exp(x\cos(y+z))\frac{\partial(x\cos(y+z))}{\partial z} = \exp(x\cos(y+z))[-x(\sin(y+z))\frac{\partial(y+z)}{\partial z}] \quad (35)$$

$$= \exp(x\cos(y+z))(-x*\sin(y+z)) \quad (36)$$

Problem 6. The following questions are good practice in manipulating vectors and matrices and they are very important for solving for posterior distributions. Given the following expression: $(x-\mu)^T\Sigma^{-1}(x-\mu)+(\mu-\mu_0)^TS^{-1}(\mu-\mu_0)$ where x,μ,μ_0 are vectors and σ^{-1} and S^{-1} are symmetric, invertible matrices. Answer the following questions:

- Expand the expression and gather terms.
  Solution:

$$(x-\mu)^T\Sigma^{-1}(x-\mu) + (\mu-\mu_0)^TS^{-1}(\mu-\mu_0))$$
$$= x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu - \mu^T\Sigma^{-1}x + \mu^T\Sigma^{-1}\mu$$
$$+ \mu^TS^{-1}\mu - \mu^TS^{-1}\mu_0 - \mu_0^TS^{-1}\mu + \mu_0^TS^{-1}\mu_0 \quad (37)$$

S^{-1}, Σ^{-1} are symmetric and invertible matrices with following properties:

$\mu^T\Sigma^{-1}x = x^T\Sigma^{-1}\mu$
and
$\mu_0^TS^{-1}\mu = \mu^TS^{-1}\mu_0$

Given these 2 properties we can rewrite the above as:

$$x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu - \mu^T\Sigma^{-1}x + \mu^T\Sigma^{-1}\mu + \mu^TS^{-1}\mu - \mu^TS^{-1}\mu_0 - \mu_0^TS^{-1}\mu + \mu_0^TS^{-1}\mu_0$$
$$= x^T\Sigma^{-1}x + \mu^T\Sigma^{-1}\mu - 2\mu^T\Sigma^{-1}x + \mu^TS^{-1}\mu + \mu_0^TS^{-1}x\mu_0 - 2\mu^TS^{-1}\mu_0 \quad (38)$$

- Collect all the terms that depend on μ and those that do not.
  Solution:

$$x^T\Sigma^{-1}x + \mu^T\Sigma^{-1}\mu - 2\mu^T\Sigma^{-1}x + \mu^TS^{-1}\mu + \mu_0^TS^{-1}x\mu_0 - 2\mu^TS^{-1}\mu_0$$
$$= \mu^T(\Sigma^{-1}+S^{-1})\mu - 2\mu^T(\Sigma^{-1}x+S^{-1}\mu_0) + x^T\Sigma^{-1}x + \mu_0^TS^{-1}\mu_0 \quad (39)$$

- Take the derivative with respect to μ, set to 0, and solve for μ.
  Solution:

$$\frac{\partial(\mu^T(\Sigma^{-1}+S^{-1})\mu - 2\mu^T(\Sigma^{-1}x+S^{-1}\mu_0) + x^T\Sigma^{-1}x + \mu_0^TS^{-1}\mu_0)}{\partial\mu} = 0 \quad (40)$$

Since the last two terms do not depend on μ, we can set their derivative to 0 and discard them.

$$\frac{\partial(\mu^T(\Sigma^{-1}+S^{-1})\mu - 2\mu^T(\Sigma^{-1}x+S^{-1}\mu_0))}{\partial\mu} = 0 \quad (41)$$

$$(\Sigma^{-1} + S^{-1})\mu + (\mu^T(\Sigma^{-1} + S^{-1}))^T - 2(\Sigma^{-1}x + S^{-1}\mu_0) = 0 \tag{42}$$

$$(\Sigma^{-1} + S^{-1})\mu + (\Sigma^{-1} + S^{-1})\mu - 2(\Sigma^{-1}x + S^{-1}\mu_0) = 0 \tag{43}$$

$$2(\Sigma^{-1} + S^{-1})\mu - 2(\Sigma^{-1}x + S^{-1}\mu_0) = 0 \tag{44}$$

$$2(\Sigma^{-1} + S^{-1})\mu = 2(\Sigma^{-1}x + S^{-1}\mu_0) \tag{45}$$

$$(\Sigma^{-1} + S^{-1})\mu = (\Sigma^{-1}x + S^{-1}\mu_0) \tag{46}$$

Finally, the value of μ is

$$\mu = (\Sigma^{-1} + S^{-1})^{-1}(\Sigma^{-1}x + S^{-1}\mu_0) \tag{47}$$