Exercise 1a

We are given two normal distributions:

$$P = \mathcal{N}(\mu_1, \sigma_1^2), \quad Q = \mathcal{N}(\mu_2, \sigma_2^2)$$

The goal is to calculate the Wasserstein distance W_2 between these two distributions. The formula for the squared Wasserstein distance between two distributions P and Q is given by:

$$W_2^2(P, Q) = \mathbb{E}[(X - T(X))^2]$$

where $T(x)$ is the optimal transport map that transforms $X \sim P$ to $T(X) \sim Q$.

The optimal transport map for two normal distributions is given by:

$$T(x) = \mu_2 + \sigma_2 \cdot \frac{x - \mu_1}{\sigma_1}$$

This map transforms samples from $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ to $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

Next, we calculate the difference $X - T(X)$:

$$X - T(X) = \left(1 - \frac{\sigma_2}{\sigma_1}\right) X + \left(\frac{\sigma_2}{\sigma_1}\mu_1 - \mu_2\right)$$

We introduce the following notations:

$$A = 1 - \frac{\sigma_2}{\sigma_1}, \quad B = \frac{\sigma_2}{\sigma_1}\mu_1 - \mu_2$$

Thus, the difference becomes:

$$X - T(X) = AX + B$$

We now calculate the expected squared difference:

$$\mathbb{E}[(X - T(X))^2] = \mathbb{E}[(AX + B)^2]$$

Expanding the square:

$$\mathbb{E}[(AX + B)^2] = A^2\mathbb{E}[X^2] + 2AB\mathbb{E}[X] + B^2$$

Using the following expectations for $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$:

$$\mathbb{E}[X] = \mu_1, \quad \mathbb{E}[X^2] = \sigma_1^2 + \mu_1^2$$

We get:

$$\mathbb{E}[(X - T(X))^2] = A^2(\sigma_1^2 + \mu_1^2) + 2AB\mu_1 + B^2$$

Now, we address the part you requested about the expectations. The second and third integrals we are working with represent the expected values of $(x - \mu_1)^2$

and $(x-\mu_2)^2$, respectively. These are known properties for normal distributions. Specifically:

$$\mathbb{E}_P[(x-\mu_1)^2] = \sigma_1^2$$

and

$$\mathbb{E}_P[(x-\mu_2)^2] = \sigma_2^2 + (\mu_2-\mu_1)^2$$

Thus, the second term in our calculation becomes:

$$\frac{1}{2\sigma_1^2}\left(-\sigma_1^2\right) = -\frac{1}{2}$$

And the third term becomes:

$$\frac{1}{2\sigma_2^2}\left(\sigma_2^2 + (\mu_2-\mu_1)^2\right) = \frac{\sigma_2^2 + (\mu_2-\mu_1)^2}{2\sigma_2^2}$$

By combining all these terms, we get the final result for the Wasserstein distance squared:

$$W_2^2 = (\mu_1-\mu_2)^2 + (\sigma_1-\sigma_2)^2$$

Thus, the Wasserstein distance between two normal distributions is:

$$W_2(\mathcal{N}(\mu_1,\sigma_1^2), \mathcal{N}(\mu_2,\sigma_2^2)) = \sqrt{(\mu_1-\mu_2)^2 + (\sigma_1-\sigma_2)^2}$$

Exercise 1b

Now, we derive the KL divergence between two normal distributions given by:

$$D_{\mathrm{KL}}(P \parallel Q) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right)\, dx$$

Where $p(x)$ and $q(x)$ are the PDFs of the distributions P and Q, respectively.

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right), \quad q(x) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right)$$

So:

$$D_{\mathrm{KL}}(P \parallel Q) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) \log\left(\frac{\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right)}{\frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right)}\right) dx$$

Simplifying the logarithm:

$$\log\left(\frac{p(x)}{q(x)}\right) = \log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{(x-\mu_2)^2}{2\sigma_2^2} - \frac{(x-\mu_1)^2}{2\sigma_1^2}$$

Now, we evaluate the integral term by term.
The first term:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) \log\left(\frac{\sigma_2}{\sigma_1}\right) dx = \log\left(\frac{\sigma_2}{\sigma_1}\right)$$

Since the integral of $p(x)$ is 1, we focus on the log term.
The third term can be viewed as the expectation of the numerator:

$$\frac{1}{2\sigma_2^2}\left(\sigma_2^2 + (\mu_2-\mu_1)^2\right) = \frac{1}{2} + \frac{(\mu_2-\mu_1)^2}{\sigma_2^2}$$

The second term:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) \left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) dx = -\frac{1\sigma_1^2}{2\sigma_1^2} = -\frac{1}{2}$$

Putting everything together:

$$D_{\mathrm{KL}}(P \parallel Q) = \log\left(\frac{\sigma_2}{\sigma_1}\right) - \frac{1}{2} + \frac{1}{2} + \frac{(\mu_2-\mu_1)^2}{\sigma_2^2}$$

Thus, the KL divergence between two normal distributions is:

$$D_{\mathrm{KL}}(P \parallel Q) = \log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{(\mu_2-\mu_1)^2}{\sigma_2^2}$$

Problem 2:

a) * we are trying to optimize $J(\pi) = \sum_{i,j=1}^{3} T_{ij} |b_i - a_j|^2$

Given $a = (1, 2, 3)$ & $b = (1.5, 2, -1)$

First, Sort a & b: $a = (1, 2, 3)$
$b = (-1, 1.5, 2)$

for each sorted ~~Pair~~ value of a_i Corresponding to Sorted b_j & Pair each a_i to b_j^* to minimize $|b_j - a_i|^2$

$a_1 = 1 \leftrightarrow b_1^* = -1$
$a_2 = 2 \leftrightarrow b_2^* = 1.5$
$a_3 = 3 \leftrightarrow b_3^* = 2$

the Coupling matrix T^* ~~befo~~ After Sorting is

$$\overbrace{\begin{array}{ccc} -1 & 1.5 & 2 \end{array}}^{b}$$
$$\begin{array}{c} \\ \\ \\ \end{array}\begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \Bigg\} a$$

Now returning the matrix to the Previous state before Sorting

then
$$T^* = \begin{array}{c} \overbrace{\begin{array}{ccc} 1.5 & 2 & -1 \end{array}}^{b} \\ \begin{pmatrix} 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{pmatrix} \end{array} \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \Bigg\} a$$

---

b) $J(T^*) = \sum_{i,j=1}^{3} t_{ij} |b_i - a_j|^2$

$= \frac{1}{3}[1.5 - 2]^2 + \frac{1}{3}[2-3]^2 + \frac{1}{3}[-1-1]^2$

$= \frac{1}{3} \times \frac{1}{4} + \frac{1}{3} \times 1 + \frac{1}{3} \times 4 = 1.75$