Computer Science and Engineering 3521:
Survey of Artificial Intelligence 1: Basic Techniques – Homework 2

Author: Danny Kan (kan.74@osu.edu)

Date of Submission: Friday, October 27, 2023

The Ohio State University
College of Engineering
Columbus, Ohio

Collaboration

I did not collaborate with any students from this course. All work found within this file is solely my own, which I attempted to the best of my knowledge and understanding of the content material.

Calculation

Something goes here...

Question 1: Bag-of-Words (BoW) Representation and Distance Computation

Question 1.1: Bag-of-Words (BoW) Construction [3 pts.]

Suppose that we are given the following dictionary:

$$S = \{\text{“}capital\text{”} : 1, \text{“}state\text{”} : 2, \text{“}team\text{”} : 3, \text{“}basketball\text{”} : 4, \text{“}hokey\text{”} : 5, \text{“}professional\text{”} : 6, \text{“}bank\text{”} : 7\}.$$

We would like to determine, that is, derive the Bag-of-Words (BoW) representation (i.e., a seven-dimensional vector) for the following three sentences without normalization:

- α = {“sacramento”, “is”, “a”, “state”, “capital”, “and”, “it”, “has”, “a”, “professional”, “basketball”, “team” },
- β = {“Columbus”, “is”, “a”, “state”, “capital”, “and”, “it”, “has”, “a”, “professional”, “hokey”, “team”, “but”, “no”, “professional”, “basketball”, “team” },
- γ = {“the”, “capital”, “bank”, “at”, “Cincinnati”, “has”, “a”, “professional”, “team” }.

When a token (i.e., a word) appears multiple times in the token sequence (i.e., a sentence), it is essential to maintain an accurate frequency count in the Bag-of-Words (BoW) representation. We do not consider tokens not present in our dictionary, that is, we do not count them in the final vector form representation.

Converting the token sequences for $|\alpha| = 12$, $|\beta| = 17$, and $|\gamma| = 9$, we get the following Bag-of-Words (BoW) representation:

- $\vec{x}_\alpha = [1, 1, 1, 1, 0, 1, 0]^T$
- $\vec{x}_\beta = [1, 1, 2, 1, 1, 2, 0]^T$
- $\vec{x}_\gamma = [1, 0, 1, 0, 0, 1, 1]^T$

We can in fact confirm that $\vec{x}_\alpha, \vec{x}_\beta, \vec{x}_\gamma \in \mathbb{R}^7$.

Question 1.2: L_1 **Distance Computation [2 pts.]**

The equation for computing the L_1 distance between two vectors, that is the magnitude of the difference vector, is given by:

$$\|\vec{x}_i - \vec{x}_j\|_1 = \sum_{d=1}^{D} |x_i[d] - x_j[d]|$$

Computing the L_1 norm between vector $\vec{x}_\alpha$ and vector $\vec{x}_\beta$, we get:

$$\begin{aligned}\|\vec{x}_\alpha - \vec{x}_\beta\|_1 &= |1-1| + |1-1| + |1-2| + |1-1| + |0-1| + |1-2| + |0-0| \\ &= |0| + |0| + |-1| + |0| + |-1| + |-1| + |0| \\ &= 0+0+1+0+1+1+0 \\ &= \boxed{3}\end{aligned}$$

Computing the L_1 norm between vector $\vec{x}_\alpha$ and vector $\vec{x}_\gamma$, we get:

$$\begin{aligned}\|\vec{x}_\alpha - \vec{x}_\gamma\|_1 &= |1-1| + |1-0| + |1-1| + |1-0| + |0-0| + |1-1| + |0-1| \\ &= |0| + |1| + |0| + |1| + |0| + |0| + |1| \\ &= 0+1+0+1+0+0+1 \\ &= \boxed{3}\end{aligned}$$

Question 1.3: L_1 **Feature Normalization [2 pts.]**

We would like to perform L_1 feature normalization on the Bag-of-Words (BoW) representation. To normalize an arbitrary vector, $\vec{x}$, we use the following equation:

$$\vec{z} = \frac{\vec{x}}{\|\vec{x}\|_1} = \begin{bmatrix} \frac{x[1]}{\|\vec{x}\|_1} \\ \vdots \\ \frac{x[D]}{\|\vec{x}\|_1} \end{bmatrix}$$

Applying L_1 normalization on vectors $\vec{x}_\alpha$, $\vec{x}_\beta$, and $\vec{x}_\gamma$, we get the following:

- $\|\vec{x}_\alpha\|_1 = \sum_{d=1}^{D} |x_\alpha[d]| = 1+1+1+1+0+1+0 = \boxed{5}$
- $\|\vec{x}_\beta\|_1 = \sum_{d=1}^{D} |x_\beta[d]| = 1+1+2+1+1+2+0 = \boxed{8}$
- $\|\vec{x}_\gamma\|_1 = \sum_{d=1}^{D} |x_\gamma[d]| = 1+0+1+0+0+1+1 = \boxed{4}$
- $\vec{z}_\alpha = \frac{\vec{x}_\alpha}{\|\vec{x}_\alpha\|_1} = \begin{bmatrix} \frac{x_\alpha[1]}{\|\vec{x}_\alpha\|_1} \\ \vdots \\ \frac{x_\alpha[7]}{\|\vec{x}_\alpha\|_1} \end{bmatrix} = \begin{bmatrix} \frac{1}{5} \\ \vdots \\ \frac{0}{5} \end{bmatrix} = [\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{0}{5}, \frac{1}{5}, \frac{0}{5}]^T = [\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, 0, \frac{1}{5}, 0]^T$
- $\vec{z}_\beta = \frac{\vec{x}_\beta}{\|\vec{x}_\beta\|_1} = \begin{bmatrix} \frac{x_\beta[1]}{\|\vec{x}_\beta\|_1} \\ \vdots \\ \frac{x_\beta[7]}{\|\vec{x}_\beta\|_1} \end{bmatrix} = \begin{bmatrix} \frac{1}{8} \\ \vdots \\ \frac{0}{8} \end{bmatrix} = [\frac{1}{8}, \frac{1}{8}, \frac{2}{8}, \frac{1}{8}, \frac{1}{8}, \frac{2}{8}, \frac{0}{8}]^T = [\frac{1}{8}, \frac{1}{8}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{4}, 0]^T$
- $\vec{z}_\gamma = \frac{\vec{x}_\gamma}{\|\vec{x}_\gamma\|_1} = \begin{bmatrix} \frac{x_\gamma[1]}{\|\vec{x}_\gamma\|_1} \\ \vdots \\ \frac{x_\gamma[7]}{\|\vec{x}_\gamma\|_1} \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ \vdots \\ \frac{1}{4} \end{bmatrix} = [\frac{1}{4}, \frac{0}{4}, \frac{1}{4}, \frac{0}{4}, \frac{0}{4}, \frac{1}{4}, \frac{1}{4}]^T = [\frac{1}{4}, 0, \frac{1}{4}, 0, 0, \frac{1}{4}, \frac{1}{4}]^T$

We can in fact verify that $\|\vec{z}_\alpha\| = 1$, $\|\vec{z}_\beta\| = 1$, and $\|\vec{z}_\gamma\| = 1$.

Computing the new L_1 norm between normalized vectors $\vec{z}_\alpha$ and $\vec{z}_\beta$, we get:

$$\begin{aligned}
\|\vec{z}_\alpha - \vec{z}_\beta\|_1 &= \sum_{d=1}^{D} |z_\alpha[d] - z_\beta[d]| \\
&= |\frac{1}{5} - \frac{1}{8}| + |\frac{1}{5} - \frac{1}{8}| + |\frac{1}{5} - \frac{1}{4}| + |\frac{1}{5} - \frac{1}{8}| + |0 - \frac{1}{8}| + |\frac{1}{5} - \frac{1}{4}| + |0 - 0| \\
&= |\frac{3}{40}| + |\frac{3}{40}| + |\frac{-1}{20}| + |\frac{3}{40}| + |\frac{-1}{8}| + |\frac{-1}{20}| + |0| \\
&= \frac{3}{40} + \frac{3}{40} + \frac{1}{20} + \frac{3}{40} + \frac{1}{8} + \frac{1}{20} + 0 \\
&= \boxed{\frac{9}{20}} = \boxed{0.45}
\end{aligned}$$

Computing the new L_1 norm between normalized vectors $\vec{z}_\alpha$ and $\vec{z}_\gamma$, we get:

$$\begin{aligned}
\|\vec{z}_\alpha - \vec{z}_\gamma\|_1 &= \sum_{d=1}^{D} |z_\alpha[d] - z_\gamma[d]| \\
&= |\frac{1}{5} - \frac{1}{4}| + |\frac{1}{5} - 0| + |\frac{1}{5} - \frac{1}{4}| + |\frac{1}{5} - 0| + |0 - 0| + |\frac{1}{5} - \frac{1}{4}| + |0 - \frac{1}{4}| \\
&= |\frac{-1}{20}| + |\frac{1}{5}| + |\frac{-1}{20}| + |\frac{1}{5}| + |0| + |\frac{-1}{20}| + |\frac{-1}{4}| \\
&= \frac{1}{20} + \frac{1}{5} + \frac{1}{20} + \frac{1}{5} + 0 + \frac{1}{20} + \frac{1}{4} \\
&= \boxed{\frac{4}{5}} = \boxed{0.80}
\end{aligned}$$

Question 2: Histogram and Kernel Density Estimation

We are given the following two data sets, A and B, consisting of eight data instances and one feature variable:

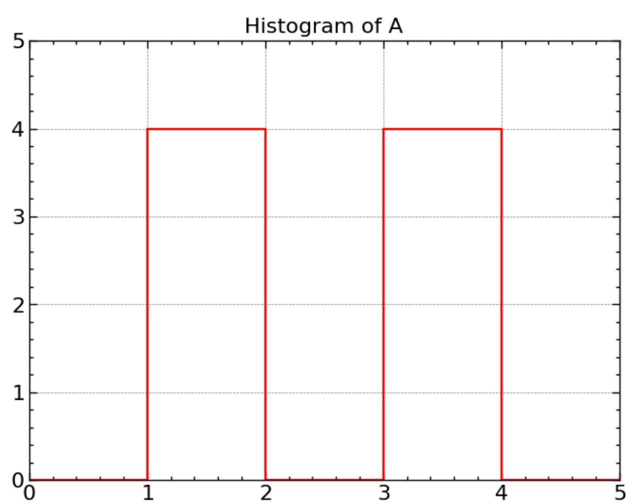

Figure 1: Histogram of A

- $\vec{x}_A = [1.2, 1.4, 1.6, 1.8, 3.2, 3.4, 3.6, 3.8]^T$

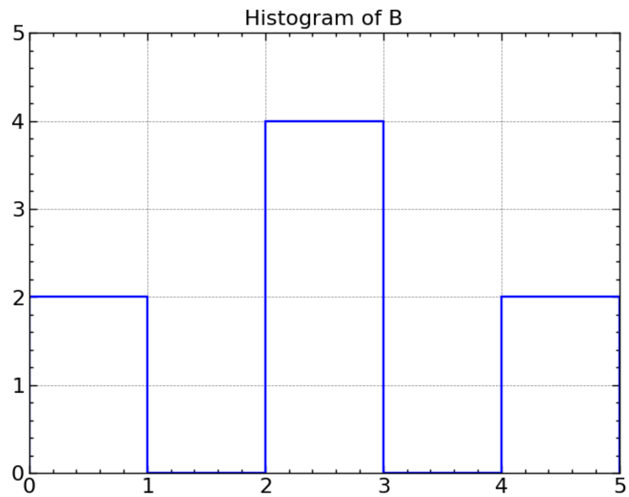

Figure 2: Histogram of B

- $\vec{x}_B = [0.6, 0.8, 2.2, 2.4, 2.6, 2.8, 4.2, 4.4]^T$

Question 2.1: First Histogram [2 pts.]

We would like to perform L_1 normalization on histograms A and B. The process is straightforward. First, we count the frequency, that is, the number of occurrences that are within our desired intervals of interest: $[0.5, 1.5)$, $[1.5, 2.5)$, $[2.5, 3.5)$, and $[3.5, 4.5)$. We get the following:

- $\vec{x}_A = [0, 4, 0, 4, 0]^T$
- $\vec{x}_B = [2, 0, 4, 0, 2]^T$

Next, to perform L_1 normalization on an arbitrary vector, $\vec{x}$, we use the following equation:

$$\vec{z} = \frac{\vec{x}}{\|\vec{x}\|_1} = \begin{bmatrix} \frac{x[1]}{\|\vec{x}\|_1} \\ \vdots \\ \frac{x[D]}{\|\vec{x}\|_1} \end{bmatrix}$$

Applying it on vectors $\vec{x}_A$ and $\vec{x}_B$, we get the following:

- $\|\vec{x}_A\|_1 = \sum_{d=1}^{D} |x_A[d]| = 0 + 4 + 0 + 4 + 0 = \boxed{8}$
- $\|\vec{x}_B\|_1 = \sum_{d=1}^{D} |x_B[d]| = 2 + 0 + 4 + 0 + 2 = \boxed{8}$
- $\vec{z}_A = \frac{\vec{x}_A}{\|\vec{x}_A\|_1} = \begin{bmatrix} \frac{x_A[1]}{\|\vec{x}_A\|_1} \\ \vdots \\ \frac{x_A[5]}{\|\vec{x}_A\|_1} \end{bmatrix} = \begin{bmatrix} \frac{0}{8} \\ \vdots \\ \frac{0}{8} \end{bmatrix} = [\frac{0}{8}, \frac{4}{8}, \frac{0}{8}, \frac{4}{8}, \frac{0}{8}]^T = [0, \frac{1}{2}, 0, \frac{1}{2}, 0]^T$
- $\vec{z}_B = \frac{\vec{x}_B}{\|\vec{x}_B\|_1} = \begin{bmatrix} \frac{x_B[1]}{\|\vec{x}_B\|_1} \\ \vdots \\ \frac{x_B[5]}{\|\vec{x}_B\|_1} \end{bmatrix} = \begin{bmatrix} \frac{2}{8} \\ \vdots \\ \frac{2}{8} \end{bmatrix} = [\frac{2}{8}, \frac{0}{8}, \frac{4}{8}, \frac{0}{8}, \frac{2}{8}]^T = [\frac{1}{4}, 0, \frac{1}{2}, 0, \frac{1}{4}]^T$

Notice that: $\|\vec{z}_A\|_1 = \sum_{d=1}^{D} |z_A[d]| = 0 + \frac{1}{2} + 0 + \frac{1}{2} + 0 = 1$ and
$\|\vec{z}_B\|_1 = \sum_{d=1}^{D} |z_B[d]| = \frac{1}{4} + 0 + \frac{1}{2} + 0 + \frac{1}{4} = 1$.

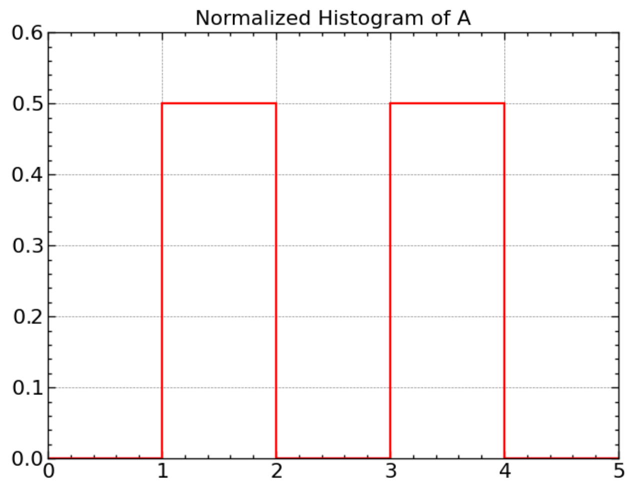

Figure 3: L_1 Normalized Histogram of A

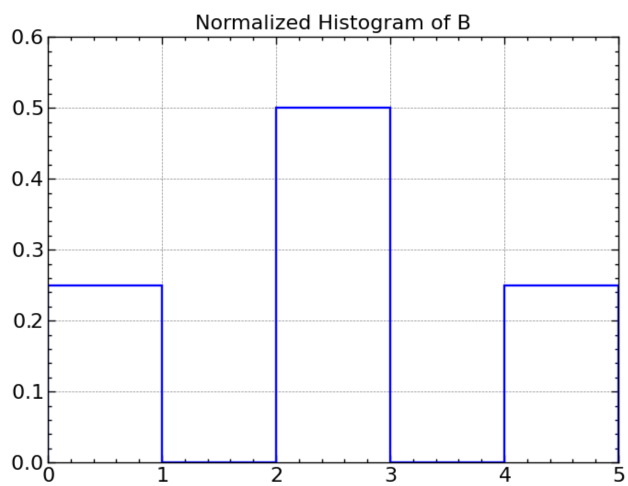

Figure 4: L_1 Normalized Histogram of B

Computing the L_1 norm between normalized vectors $\vec{z}_A$ and $\vec{z}_B$, we get:

$$
\begin{aligned}
\|\vec{z}_A - \vec{z}_B\|_1 &= \sum_{d=1}^{D} |z_A[d] - z_B[d]| \\
&= |0 - \frac{1}{4}| + |\frac{1}{2} - 0| + |0 - \frac{1}{2}| + |\frac{1}{2} - 0| + |0 - \frac{1}{4}| \\
&= |\frac{-1}{4}| + |\frac{1}{2}| + |\frac{-1}{2}| + |\frac{1}{2}| + |\frac{-1}{4}| \\
&= \frac{1}{4} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{4} \\
&= \boxed{2}
\end{aligned}
$$

That is the answer for the L_1 distance between the L_1-normalized histograms of A and B.

Question 2.2: Offset Histogram [2 pts.]

We are given the following intervals: $[0.5, 1.5)$, $[1.5, 2.5)$, $[2.5, 3.5)$, and $[3.5, 4.5)$.
The process is similar to Question 2.1, except we consider a wider bin width, δ:

- $\vec{x}_A = [2, 2, 2, 2]^T$
- $\vec{x}_B = [2, 2, 2, 2]^T$
- $\|\vec{x}_A\|_1 = \sum_{d=1}^{D} |x_A[d]| = 2 + 2 + 2 + 2 = \boxed{8}$
- $\|\vec{x}_B\|_1 = \sum_{d=1}^{D} |x_B[d]| = 2 + 2 + 2 + 2 = \boxed{8}$
- $\vec{z}_A = \frac{\vec{x}_A}{\|\vec{x}_A\|_1} = \begin{bmatrix} \frac{x_A[1]}{\|\vec{x}_A\|_1} \\ \vdots \\ \frac{x_A[4]}{\|\vec{x}_A\|_1} \end{bmatrix} = \begin{bmatrix} \frac{2}{8} \\ \vdots \\ \frac{2}{8} \end{bmatrix} = [\frac{2}{8}, \frac{2}{8}, \frac{2}{8}, \frac{2}{8}]^T = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}]^T$
- $\vec{z}_B = \frac{\vec{x}_B}{\|\vec{x}_B\|_1} = \begin{bmatrix} \frac{x_B[1]}{\|\vec{x}_B\|_1} \\ \vdots \\ \frac{x_B[4]}{\|\vec{x}_B\|_1} \end{bmatrix} = \begin{bmatrix} \frac{2}{8} \\ \vdots \\ \frac{2}{8} \end{bmatrix} = [\frac{2}{8}, \frac{2}{8}, \frac{2}{8}, \frac{2}{8}]^T = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}]^T$

Notice that: $\|\vec{z}_A\|_1 = \sum_{d=1}^{D} |z_A[d]| = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1$ and
$\|\vec{z}_B\|_1 = \sum_{d=1}^{D} |z_B[d]| = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1.$

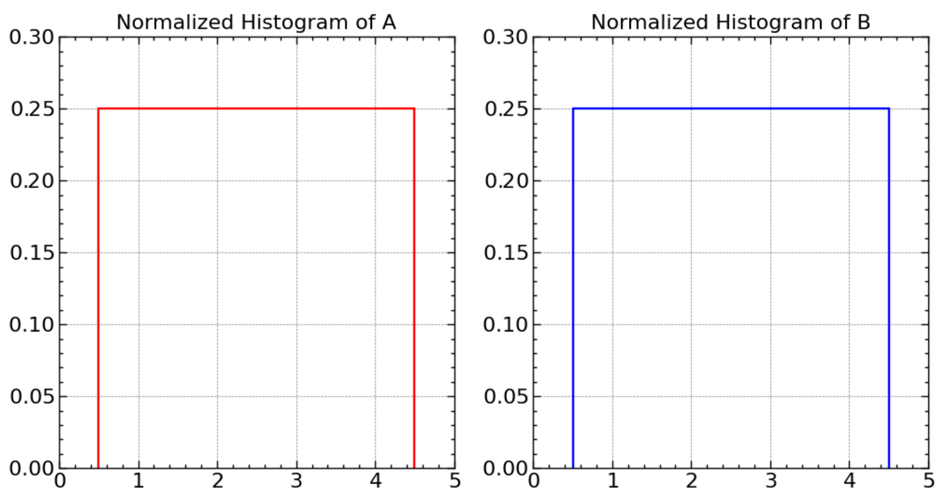

Figure 5: L_1 Normalized Histogram of A and B

Computing the L_1 norm between normalized vectors $\vec{z}_A$ and $\vec{z}_B$, we get:

$$\begin{aligned}
\|\vec{z}_A - \vec{z}_B\|_1 &= \sum_{d=1}^{D} |z_A[d] - z_B[d]| \\
&= |\frac{1}{4} - \frac{1}{4}| + |\frac{1}{4} - \frac{1}{4}| + |\frac{1}{4} - \frac{1}{4}| + |\frac{1}{4} - \frac{1}{4}| \\
&= |0| + |0| + |0| + |0| \\
&= 0 + 0 + 0 + 0 \\
&= \boxed{0}
\end{aligned}$$

That is the answer for the L_1 distance between the L_1-normalized histograms of A and B.
Notice that $\|\vec{z}_A - \vec{z}_B\|_1 = 0 \implies L_1$-normalized histograms, A and B, are identical.

Question 2.3: Narrow Bin Histogram [2 pts.]

We are given the following intervals: $[0.5, 1)$, $[1, 1.5)$, $[1.5, 2)$, $[2, 2.5)$, $[2.5, 3)$, $[3, 3.5)$, $[3.5, 4)$, and $[4, 4.5)$.
The process is similar to Question 2.1 and Question 2.2, except we consider a narrower bin width, δ:

- $\vec{x}_A = [0, 2, 2, 0, 0, 2, 2, 0]^T$
- $\vec{x}_B = [2, 0, 0, 2, 2, 0, 0, 2]^T$
- $\|\vec{x}_A\|_1 = \sum_{d=1}^{8} |x_A[d]| = 0 + 2 + 2 + 0 + 0 + 2 + 2 + 0 = \boxed{8}$
- $\|\vec{x}_B\|_1 = \sum_{d=1}^{8} |x_B[d]| = 2 + 0 + 0 + 2 + 2 + 0 + 0 + 2 = \boxed{8}$
- $\vec{z}_A = \frac{\vec{x}_A}{\|\vec{x}_A\|_1} = \begin{bmatrix} \frac{x_A[1]}{\|\vec{x}_A\|_1} \\ \vdots \\ \frac{x_A[8]}{\|\vec{x}_A\|_1} \end{bmatrix} = \begin{bmatrix} \frac{0}{8} \\ \vdots \\ \frac{0}{8} \end{bmatrix} = [\frac{0}{8}, \frac{2}{8}, \frac{2}{8}, \frac{0}{8}, \frac{0}{8}, \frac{2}{8}, \frac{2}{8}, \frac{0}{8}]^T = [0, \frac{1}{4}, \frac{1}{4}, 0, 0, \frac{1}{4}, \frac{1}{4}, 0]^T$
- $\vec{z}_B = \frac{\vec{x}_B}{\|\vec{x}_B\|_1} = \begin{bmatrix} \frac{x_B[1]}{\|\vec{x}_B\|_1} \\ \vdots \\ \frac{x_B[8]}{\|\vec{x}_B\|_1} \end{bmatrix} = \begin{bmatrix} \frac{2}{8} \\ \vdots \\ \frac{2}{8} \end{bmatrix} = [\frac{2}{8}, \frac{0}{8}, \frac{0}{8}, \frac{2}{8}, \frac{2}{8}, \frac{0}{8}, \frac{0}{8}, \frac{2}{8}]^T = [\frac{1}{4}, 0, 0, \frac{1}{4}, \frac{1}{4}, 0, 0, \frac{1}{4}]^T$

Notice that: $\|\vec{z}_A\|_1 = \sum_{d=1}^{D} |z_A[d]| = 0 + \frac{1}{4} + \frac{1}{4} + 0 + 0 + \frac{1}{4} + \frac{1}{4} + 0 = 1$ and
$\|\vec{z}_B\|_1 = \sum_{d=1}^{D} |z_B[d]| = \frac{1}{4} + 0 + 0 + \frac{1}{4} + \frac{1}{4} + 0 + 0 + \frac{1}{4} = 1$.

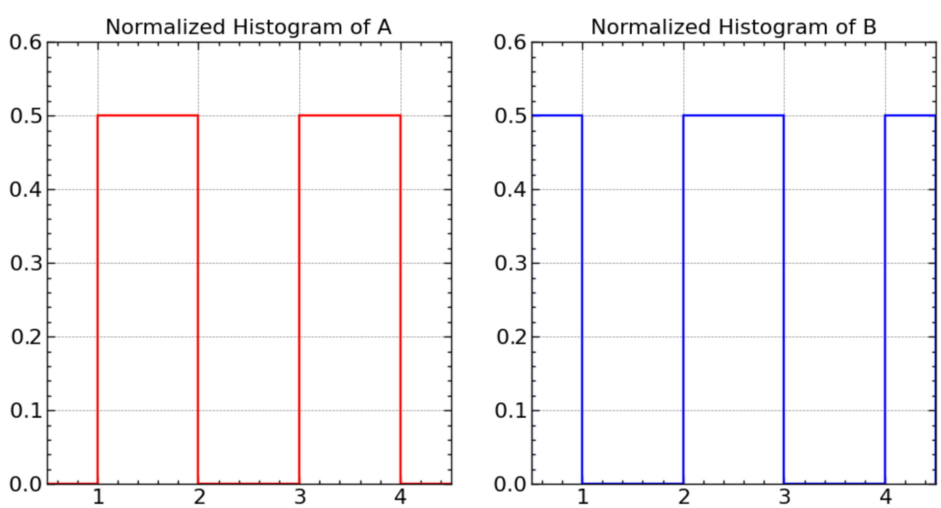

Figure 6: L_1 Normalized Histogram of A and B

Computing the L_1 norm between normalized vectors $\vec{z}_A$ and $\vec{z}_B$, we get:

$$\begin{aligned}
\|\vec{z}_A - \vec{z}_B\|_1 &= \sum_{d=1}^{D} |z_A[d] - z_B[d]| \\
&= |0 - \frac{1}{4}| + |\frac{1}{4} - 0| + |\frac{1}{4} - 0| + |0 - \frac{1}{4}| + |0 - \frac{1}{4}| + |\frac{1}{4} - 0| + |\frac{1}{4} - 0| + |0 - \frac{1}{4}| \\
&= |\frac{-1}{4}| + |\frac{1}{4}| + |\frac{1}{4}| + |\frac{-1}{4}| + |\frac{-1}{4}| + |\frac{1}{4}| + |\frac{-1}{4}| + |\frac{-1}{4}| \\
&= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} \\
&= \boxed{2}
\end{aligned}$$

That is the answer for the L_1 distance between the L_1-normalized histograms of A and B.

<u>Question 2.4:</u> Kernel Density Estimation (i.e., Parzen Window Method) [4 pts.]

Consider the following kernel function for the data set A:

$$k(u') = \begin{cases} 2 + 4 \times u', \text{ if } -0.5 \le u' \le 0 \\ 2 - 4 \times u', \text{ if } 0 < u' \le 0.5 \\ 0, \text{ otherwise} \end{cases}$$

We would like to compute the probability density function, $p(u)$, for seeing a numerical value, u, in the future. Suppose that we want to find the probabilities associated with $u = 1.5$ and $u = 2.5$.

The general form for kernel density estimation is defined as:

$$p(u) = \frac{1}{N} \sum_{n=1}^{N} k(u - u_n)$$

The calculations are self-explanatory:

$$\begin{aligned}
p(u = 1.5) = \frac{1}{8} \sum_{n=1}^{8} k(u - u_n) &= \frac{1}{8}[k(1.5 - 1.2) + k(1.5 - 1.4) + ... + k(1.5 - 3.8)] \\
&= \frac{1}{8}[k(0.3) + k(0.1) + k(-0.1) + k(-0.3) + k(-1.7) + k(-1.9) + k(-2.1) + k(-2.3)] \\
&= \frac{1}{8}[0.8 + 1.6 + 1.6 + 0.8 + 0 + 0 + 0 + 0] \\
&= \frac{1}{8}[4.8] = \frac{4.8}{8} = \boxed{0.6} \implies \boxed{60\%}
\end{aligned}$$

$$\begin{aligned}
p(u = 2.5) = \frac{1}{8} \sum_{n=1}^{8} k(u - u_n) &= \frac{1}{8}[k(2.5 - 1.2) + k(2.5 - 1.4) + ... + k(2.5 - 3.8)] \\
&= \frac{1}{8}[k(1.3) + k(1.1) + k(0.9) + k(0.7) + k(-0.7) + k(-0.9) + k(-1.1) + k(-1.3)] \\
&= \frac{1}{8}[0 + 0 + 0 + 0 + 0 + 0 + 0 + 0] \\
&= \frac{1}{8}[0] = \frac{0}{8} = \boxed{0} \implies \boxed{0\%}
\end{aligned}$$

Question 3: Covariance, Z-Score, Whitening, and Principal Component Analysis (PCA)

Question 3.1: Covariance Computation [2 pts.]

We would like to determine the covariance matrix, $\mathbf{\Sigma}$, given the following data set consisting of four two-dimensional vectors:

$$\{\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4\}$$

$$\left\{\vec{x}_1 = [20, 5]^T, \vec{x}_2 = [8, -2]^T, \vec{x}_3 = [-6, -3]^T, \vec{x}_4 = [6, 4]^T\right\}$$

$$\left\{\vec{x}_1 = \begin{bmatrix} 20 \\ 5 \end{bmatrix}, \vec{x}_2 = \begin{bmatrix} 8 \\ -2 \end{bmatrix}, \vec{x}_3 = \begin{bmatrix} -6 \\ -3 \end{bmatrix}, \vec{x}_4 = \begin{bmatrix} 6 \\ 4 \end{bmatrix}\right\}$$

We can in fact confirm that $\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4 \in \mathbb{R}^2$. $\mathbb{R}^2 = \left\{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mid x_1, x_2 \in \mathbb{R}\right\}$.
The covariance matrix is defined as:

$$\mathbf{\Sigma} = \begin{bmatrix} \mathrm{Cov}_{1,1}, \mathrm{Cov}_{1,2} \\ \mathrm{Cov}_{2,1}, \mathrm{Cov}_{2,2} \end{bmatrix} = \begin{bmatrix} \mathrm{Var}_1, \mathrm{Cov}_{1,2} \\ \mathrm{Cov}_{2,1}, \mathrm{Var}_2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2, \mathrm{Cov}_{1,2} \\ \mathrm{Cov}_{2,1}, \sigma_2^2 \end{bmatrix}$$

It is important to note that $\mathrm{Cov}_{1,2} = \mathrm{Cov}_{2,1}$.
The covariance is defined as:

$$\mathrm{Cov}_{d,d'} = \frac{1}{N} \sum_{n=1}^{N} \left[(x_n[d] - \mu[d]) \left(x_n[d'] - \mu[d']\right)\right]$$

To determine the mean vector, $\vec{\mu}_x$, we use the following equation:

$$\vec{\mu}_x = \frac{1}{N} \sum_{n=1}^{N} \vec{x}_n$$

Computing it for the specified data instances mentioned previously, we get:

$$\begin{aligned} \vec{\mu}_x = \frac{1}{N} \sum_{n=1}^{N} \vec{x}_n = \frac{1}{4} \sum_{n=1}^{4} \vec{x}_n &= \frac{1}{4}[\vec{x}_1 + \vec{x}_2 + \vec{x}_3 + \vec{x}_4] \\ &= \frac{1}{4}[\begin{bmatrix} 20 \\ 5 \end{bmatrix} + \begin{bmatrix} 8 \\ -2 \end{bmatrix} + \begin{bmatrix} -6 \\ -3 \end{bmatrix} + \begin{bmatrix} 6 \\ 4 \end{bmatrix}] \\ &= \frac{1}{4} \begin{bmatrix} 28 \\ 4 \end{bmatrix} = \begin{bmatrix} 28/4 \\ 4/4 \end{bmatrix} = \boxed{\begin{bmatrix} 7 \\ 1 \end{bmatrix}} \end{aligned}$$

We can in fact verify that $\vec{\mu}_x \in \mathbb{R}^2$. $\mathbb{R}^2 = \left\{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mid x_1, x_2 \in \mathbb{R}\right\}$.

To determine the covariance matrix, the calculations are straightforward:

$$\begin{aligned}
\mathrm{Cov}_{1,1} &= \frac{1}{4}\sum_{n=1}^{4}\left[\left(x_n[1]-\mu[1]\right)\left(x_n\left[1\right]-\mu\left[1\right]\right)\right] \\
&= \frac{1}{4}[(20-7)(20-7)+(8-7)(8-7)+(-6-7)(-6-7)+(6-7)(6-7)] \\
&= \frac{1}{4}[(20-7)^2+(8-7)^2+(-6-7)^2+(6-7)^2] \\
&= \frac{1}{4}[(13)^2+(1)^2+(-13)^2+(-1)^2] \\
&= \frac{1}{4}[169+1+169+1] \\
&= \frac{1}{4}[340] \\
&= \frac{340}{4} \\
&= \boxed{85.00}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}_{2,2} &= \frac{1}{4}\sum_{n=1}^{4}\left[\left(x_n[2]-\mu[2]\right)\left(x_n\left[2\right]-\mu\left[2\right]\right)\right] \\
&= \frac{1}{4}[(5-1)(5-1)+(-2-1)(-2-1)+(-3-1)(-3-1)+(4-1)(4-1)] \\
&= \frac{1}{4}[(5-1)^2+(-2-1)^2+(-3-1)^2+(4-1)^2] \\
&= \frac{1}{4}[(4)^2+(-3)^2+(-4)^2+(3)^2] \\
&= \frac{1}{4}[16+9+16+9] \\
&= \frac{1}{4}[50] \\
&= \frac{50}{4} \\
&= \boxed{\frac{25}{2}} = \boxed{12.50}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}_{1,2} = \mathrm{Cov}_{2,1} &= \frac{1}{4}\sum_{n=1}^{4}\left[\left(x_n[1]-\mu[1]\right)\left(x_n\left[2\right]-\mu\left[2\right]\right)\right] \\
&= \frac{1}{4}[(20-7)(5-1)+(8-7)(-2-1)+(-6-7)(-3-1)+(6-7)(4-1)] \\
&= \frac{1}{4}[(13)(4)+(1)(-3)+(-13)(-4)+(-1)(3)] \\
&= \frac{1}{4}[52+(-3)+52+(-3)] \\
&= \frac{1}{4}[52-3+52-3] \\
&= \frac{1}{4}[98] \\
&= \frac{98}{4} \\
&= \boxed{\frac{49}{2}} = \boxed{24.50}
\end{aligned}$$

Using our calculations performed previously, we get the following covariance matrix consisting of two rows and columns:

$$\mathbf{\Sigma} = \begin{bmatrix} \text{Cov}_{1,1}, \text{Cov}_{1,2} \\ \text{Cov}_{2,1}, \text{Cov}_{2,2} \end{bmatrix} = \begin{bmatrix} \text{Var}_1, \text{Cov}_{1,2} \\ \text{Cov}_{2,1}, \text{Var}_2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2, \text{Cov}_{1,2} \\ \text{Cov}_{2,1}, \sigma_2^2 \end{bmatrix} = \begin{bmatrix} 85.00, 24.50 \\ 24.50, 12.50 \end{bmatrix}$$

Question 3.2: Z-Score Normalization [2 pts.]

We would like to compute the data set consisting of four two-dimensional vectors $\{\vec{z}_1, \vec{z}_2, \vec{z}_3, \vec{z}_4\}$ by applying z-score normalization to $\left\{\vec{x}_1 = [20, 5]^T, \vec{x}_2 = [8, -2]^T, \vec{x}_3 = [-6, -3]^T, \vec{x}_4 = [6, 4]^T\right\}$.

To perform z-score normalization, we use the following equation:

$$\vec{z}_n[d] = \frac{(x_n[d] - \mu[d])}{\sigma_d}$$

The calculations are straightforward:

- $\sigma_1 = \sqrt{\sigma_1^2} = \sqrt{85.00} = \boxed{9.22}$
- $\sigma_2 = \sqrt{\sigma_2^2} = \sqrt{12.50} = \boxed{3.54}$
- $\vec{z}_1[1] = \frac{(x_1[1] - \mu[1])}{\sigma_1} = \frac{(20-7)}{9.22} = \frac{13}{9.22} = \boxed{1.41}$
- $\vec{z}_1[2] = \frac{(x_1[2] - \mu[2])}{\sigma_2} = \frac{(5-1)}{3.54} = \frac{4}{3.54} = \boxed{1.13}$
- $\vec{z}_2[1] = \frac{(x_2[1] - \mu[1])}{\sigma_1} = \frac{(8-7)}{9.22} = \frac{1}{9.22} = \boxed{0.11}$
- $\vec{z}_2[2] = \frac{(x_2[2] - \mu[2])}{\sigma_2} = \frac{(-2-1)}{3.54} = \frac{-3}{3.54} = \boxed{-0.85}$
- $\vec{z}_3[1] = \frac{(x_3[1] - \mu[1])}{\sigma_1} = \frac{(-6-7)}{9.22} = \frac{-13}{9.22} = \boxed{-1.41}$
- $\vec{z}_3[2] = \frac{(x_3[2] - \mu[2])}{\sigma_2} = \frac{(-3-1)}{3.54} = \frac{-4}{3.54} = \boxed{-1.13}$
- $\vec{z}_4[1] = \frac{(x_4[1] - \mu[1])}{\sigma_1} = \frac{(6-7)}{9.22} = \frac{-1}{9.22} = \boxed{-0.11}$
- $\vec{z}_4[2] = \frac{(x_4[2] - \mu[2])}{\sigma_2} = \frac{(4-1)}{3.54} = \frac{3}{3.54} = \boxed{0.85}$

Using our calculations performed previously, we get the following z-score normalized data set consisting of four two-dimensional vectors:

$$\{\vec{z}_1, \vec{z}_2, \vec{z}_3, \vec{z}_4\}$$

$$\left\{\vec{z}_1 = [1.41, 1.13]^T, \vec{z}_2 = [0.11, -0.85]^T, \vec{z}_3 = [-1.41, -1.13]^T, \vec{z}_4 = [-0.11, 0.85]^T\right\}$$

$$\left\{\vec{z}_1 = \begin{bmatrix} 1.41 \\ 1.13 \end{bmatrix}, \vec{z}_2 = \begin{bmatrix} 0.11 \\ -0.85 \end{bmatrix}, \vec{z}_3 = \begin{bmatrix} -1.41 \\ -1.13 \end{bmatrix}, \vec{z}_4 = \begin{bmatrix} -0.11 \\ 0.85 \end{bmatrix}\right\}$$

Question 3.3: Unbiased and Uniformly Scaled [2 pts.]

To determine the mean vector, $\vec{\mu}_z$, we use the following equation:

$$\vec{\mu}_z = \frac{1}{N}\sum_{n=1}^{N} \vec{z}_n$$

It is the same equation we used in Question 3.1.
Computing it for the z-score normalized data instances mentioned previously, we get:

$$
\begin{aligned}
\vec{\mu}_z = \frac{1}{N}\sum_{n=1}^{N}\vec{z}_n = \frac{1}{4}\sum_{n=1}^{4}\vec{z}_n &= \frac{1}{4}[\vec{z}_1 + \vec{z}_2 + \vec{z}_3 + \vec{z}_4] \\
&= \frac{1}{4}[\begin{bmatrix} 1.41 \\ 1.13 \end{bmatrix} + \begin{bmatrix} 0.11 \\ -0.85 \end{bmatrix} + \begin{bmatrix} -1.41 \\ -1.13 \end{bmatrix} + \begin{bmatrix} -0.11 \\ 0.85 \end{bmatrix}] \\
&= \frac{1}{4}\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0/4 \\ 0/4 \end{bmatrix} = \boxed{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}
\end{aligned}
$$

Notice that $\vec{\mu} = \vec{0}$, which partially confirms that our calculations have been performed correctly. When performing z-score normalization, we can reasonably expect for the new training data set to have the mean centered at the origin.

The equation for computing the standard deviation, $\sigma_z[d]$, is given by:

$$
\sigma_z[d] = \left(\frac{1}{N}\sum_{n=1}^{N}(z_n[d] - \mu_z[d])^2\right)^{1/2} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(z_n[d] - \mu_z[d])^2}
$$

Computing it for each dimension, we get:

$$
\begin{aligned}
\sigma_z[1] = \left(\frac{1}{4}\sum_{n=1}^{4}(z_n[1] - \mu_z[1])^2\right)^{1/2} &= \frac{1}{4}[(1.41-0)^2 + (0.11-0)^2 + (-1.41-0)^2 + (-0.11-0)^2] \\
&= \frac{1}{4}[(1.41)^2 + (0.11)^2 + (-1.41)^2 + (-0.11)^2] \\
&= \frac{1}{4}[1.99 + 0.01 + 1.99 + 0.01] \\
&= \frac{1}{4}[4] \\
&= \frac{4}{4} \\
&= \boxed{1}
\end{aligned}
$$

$$
\begin{aligned}
\sigma_z[2] = \left(\frac{1}{4}\sum_{n=1}^{4}(z_n[2] - \mu_z[2])^2\right)^{1/2} &= \frac{1}{4}[(1.13-0)^2 + (-0.85-0)^2 + (-1.13-0)^2 + (0.85-0)^2] \\
&= \frac{1}{4}[(1.13)^2 + (-0.85)^2 + (-1.13)^2 + (0.85)^2] \\
&= \frac{1}{4}[1.28 + 0.72 + 1.28 + 0.72] \\
&= \frac{1}{4}[4] \\
&= \frac{4}{4} \\
&= \boxed{1}
\end{aligned}
$$

Notice that $\sigma_z[1] = 1$ and $\sigma_z[2] = 1$.
This implies that we have correctly performed z-score standardization.

Question 3.4: Feature Whitening [2 pts.]

We would like to perform feature whitening given the following data set consisting of four two-dimensional vectors:

$$\{\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4\}$$

$$\left\{\vec{x}_1 = [20, -5]^T, \vec{x}_2 = [8, 2]^T, \vec{x}_3 = [-6, 3]^T, \vec{x}_4 = [6, -4]^T\right\}$$

$$\left\{\vec{x}_1 = \begin{bmatrix} 20 \\ -5 \end{bmatrix}, \vec{x}_2 = \begin{bmatrix} 8 \\ 2 \end{bmatrix}, \vec{x}_3 = \begin{bmatrix} -6 \\ 3 \end{bmatrix}, \vec{x}_4 = \begin{bmatrix} 6 \\ -4 \end{bmatrix}\right\}$$

The mean vector, $\vec{\mu}_x$, is:

$$\vec{\mu}_x = [7, -1]^T = \begin{bmatrix} 7 \\ -1 \end{bmatrix}$$

The covariance matrix, $\mathbf{\Sigma}$, is:

$$\mathbf{\Sigma} = \begin{bmatrix} 85.00 & -24.50 \\ -24.50 & 12.50 \end{bmatrix} \implies \mathbf{\Sigma}^{-1/2} = \begin{bmatrix} 0.133 & 0.096 \\ 0.096 & 0.418 \end{bmatrix}$$

The equation for feature whitening is:

$$\vec{z}_n = \mathbf{\Sigma}^{-1/2} (\vec{x}_n - \vec{\mu})$$

The calculations are straightforward:

$$\begin{aligned}
\vec{z}_1 = \mathbf{\Sigma}^{-1/2} (\vec{x}_1 - \vec{\mu}) &= \begin{bmatrix} 0.133 & 0.096 \\ 0.096 & 0.418 \end{bmatrix} [\begin{bmatrix} 20 \\ -5 \end{bmatrix} - \begin{bmatrix} 7 \\ -1 \end{bmatrix}] \\
&= \begin{bmatrix} 0.133 & 0.096 \\ 0.096 & 0.418 \end{bmatrix} [\begin{bmatrix} 20 \\ -5 \end{bmatrix} + \begin{bmatrix} -7 \\ 1 \end{bmatrix}] \\
&= \begin{bmatrix} 0.133 & 0.096 \\ 0.096 & 0.418 \end{bmatrix} \begin{bmatrix} 13 \\ -4 \end{bmatrix} \\
&= \begin{bmatrix} 0.133(13) + 0.096(-4) \\ 0.096(13) + 0.418(-4) \end{bmatrix} \\
&= \begin{bmatrix} 1.729 + (-0.384) \\ 1.248 + (-1.672) \end{bmatrix} \\
&= \begin{bmatrix} 1.729 - 0.384 \\ 1.248 - 1.672 \end{bmatrix} \\
&= \boxed{\begin{bmatrix} 1.345 \\ -0.424 \end{bmatrix}}
\end{aligned}$$

$$
\begin{aligned}
\vec{z}_2 = \mathbf{\Sigma}^{-1/2}\,(\vec{x}_2 - \vec{\mu}) &= \begin{bmatrix} 0.133 & 0.096 \\ 0.096 & 0.418 \end{bmatrix} [\begin{bmatrix} 8 \\ 2 \end{bmatrix} - \begin{bmatrix} 7 \\ -1 \end{bmatrix}] \\
&= \begin{bmatrix} 0.133 & 0.096 \\ 0.096 & 0.418 \end{bmatrix} [\begin{bmatrix} 8 \\ 2 \end{bmatrix} + \begin{bmatrix} -7 \\ 1 \end{bmatrix}] \\
&= \begin{bmatrix} 0.133 & 0.096 \\ 0.096 & 0.418 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \\
&= \begin{bmatrix} 0.133(1) + 0.096(3) \\ 0.096(1) + 0.418(3) \end{bmatrix} \\
&= \begin{bmatrix} 0.133 + 0.288 \\ 0.096 + 1.254 \end{bmatrix} \\
&= \boxed{\begin{bmatrix} 0.421 \\ 1.35 \end{bmatrix}}
\end{aligned}
$$

$$
\begin{aligned}
\vec{z}_3 = \mathbf{\Sigma}^{-1/2}\,(\vec{x}_3 - \vec{\mu}) &= \begin{bmatrix} 0.133 & 0.096 \\ 0.096 & 0.418 \end{bmatrix} [\begin{bmatrix} -6 \\ 3 \end{bmatrix} - \begin{bmatrix} 7 \\ -1 \end{bmatrix}] \\
&= \begin{bmatrix} 0.133 & 0.096 \\ 0.096 & 0.418 \end{bmatrix} [\begin{bmatrix} -6 \\ 3 \end{bmatrix} + \begin{bmatrix} -7 \\ 1 \end{bmatrix}] \\
&= \begin{bmatrix} 0.133 & 0.096 \\ 0.096 & 0.418 \end{bmatrix} \begin{bmatrix} -13 \\ 4 \end{bmatrix} \\
&= \begin{bmatrix} 0.133(-13) + 0.096(4) \\ 0.096(-13) + 0.418(4) \end{bmatrix} \\
&= \begin{bmatrix} -1.729 + 0.384 \\ -1.248 + 1.672 \end{bmatrix} \\
&= \boxed{\begin{bmatrix} -1.345 \\ 0.424 \end{bmatrix}}
\end{aligned}
$$

$$
\begin{aligned}
\vec{z}_4 = \mathbf{\Sigma}^{-1/2}\,(\vec{x}_4 - \vec{\mu}) &= \begin{bmatrix} 0.133 & 0.096 \\ 0.096 & 0.418 \end{bmatrix} [\begin{bmatrix} 6 \\ -4 \end{bmatrix} - \begin{bmatrix} 7 \\ -1 \end{bmatrix}] \\
&= \begin{bmatrix} 0.133 & 0.096 \\ 0.096 & 0.418 \end{bmatrix} [\begin{bmatrix} 6 \\ -4 \end{bmatrix} + \begin{bmatrix} -7 \\ 1 \end{bmatrix}] \\
&= \begin{bmatrix} 0.133 & 0.096 \\ 0.096 & 0.418 \end{bmatrix} \begin{bmatrix} -1 \\ -3 \end{bmatrix} \\
&= \begin{bmatrix} 0.133(-1) + 0.096(-3) \\ 0.096(-1) + 0.418(-3) \end{bmatrix} \\
&= \begin{bmatrix} -0.133 + (-0.288) \\ -0.096 + (-1.254) \end{bmatrix} \\
&= \begin{bmatrix} -0.133 - 0.288 \\ -0.096 - 1.254 \end{bmatrix} \\
&= \boxed{\begin{bmatrix} -0.421 \\ -1.35 \end{bmatrix}}
\end{aligned}
$$

Therefore, after performing feature whitening: a form of data-dependent normalization, we get the following set of four two-dimensional vectors:

$$\{\vec{z}_1, \vec{z}_2, \vec{z}_3, \vec{z}_4\}$$

$$\left\{\vec{z}_1 = [1.345, -0.424]^T, \vec{z}_2 = [0.421, 1.35]^T, \vec{z}_3 = [-1.345, 0.424]^T, \vec{z}_4 = [-0.421, -1.35]^T\right\}$$

$$\left\{\vec{z}_1 = \begin{bmatrix} 1.345 \\ -0.424 \end{bmatrix}, \vec{z}_2 = \begin{bmatrix} 0.421 \\ 1.35 \end{bmatrix}, \vec{z}_3 = \begin{bmatrix} -1.345 \\ 0.424 \end{bmatrix}, \vec{z}_4 = \begin{bmatrix} -0.421 \\ -1.35 \end{bmatrix}\right\}$$

Computing the mean vector, $\vec{\mu}$, for the z-score normalized data instances mentioned previously, we get:

$$\begin{aligned}
\vec{\mu}_z = \frac{1}{N}\sum_{n=1}^{N}\vec{z}_n = \frac{1}{4}\sum_{n=1}^{4}\vec{z}_n &= \frac{1}{4}[\vec{z}_1 + \vec{z}_2 + \vec{z}_3 + \vec{z}_4] \\
&= \frac{1}{4}[\begin{bmatrix} 1.345 \\ -0.424 \end{bmatrix} + \begin{bmatrix} 0.421 \\ 1.35 \end{bmatrix} + \begin{bmatrix} -1.345 \\ 0.424 \end{bmatrix} + \begin{bmatrix} -0.421 \\ -1.35 \end{bmatrix}] \\
&= \frac{1}{4}\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0/4 \\ 0/4 \end{bmatrix} = \boxed{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}
\end{aligned}$$

Computing the standard deviation for each dimension, we get:

$$\begin{aligned}
\sigma_z[1] = \left(\frac{1}{4}\sum_{n=1}^{4}(z_n[1] - \mu_z[1])^2\right)^{1/2} &= \frac{1}{4}[(1.345 - 0)^2 + (0.421 - 0)^2 + (-1.345 - 0)^2 + (-0.421 - 0)^2] \\
&= \frac{1}{4}[(1.345)^2 + (0.421)^2 + (-1.345)^2 + (-0.421)^2] \\
&= \frac{1}{4}[1.809 + 0.177 + 1.809 + 0.177] \\
&= \frac{1}{4}[3.972] \\
&= \frac{3.972}{4} \\
&= \boxed{0.993} \approx \boxed{1}
\end{aligned}$$

$$\begin{aligned}
\sigma_z[2] = \left(\frac{1}{4}\sum_{n=1}^{4}(z_n[2] - \mu_z[2])^2\right)^{1/2} &= \frac{1}{4}[(-0.424 - 0)^2 + (1.35 - 0)^2 + (0.424 - 0)^2 + (-1.35 - 0)^2] \\
&= \frac{1}{4}[(-0.424)^2 + (1.35)^2 + (0.424)^2 + (-1.35)^2] \\
&= \frac{1}{4}[0.180 + 1.823 + 0.180 + 1.823] \\
&= \frac{1}{4}[4.006] \\
&= \frac{4.006}{4} \\
&= \boxed{1.002} \approx \boxed{1}
\end{aligned}$$