• This is a CLOSED BOOK exam. You may use one sheet (front and back) of handwritten notes.

• No calculators are permitted. You need not simplify explicit numerical expressions.

• There are a total of 40 points in the exam. Each problem specifies its point total. Plan your work accordingly.

• You must SHOW YOUR WORK to get full credit.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>

Name: ___________________________________________
Possibly Useful Formulas

Minkowski Norm
\[ \| \vec{x} - \vec{\mu} \|_p = (|x_1 - \mu_1|^p + \ldots + |x_D - \mu_D|^p)^{1/p} \]

Gaussians
\[ \mathcal{N}(\vec{x}; \vec{\mu}, \Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu})} \]

\[ \Sigma = U\Lambda U^T \]
\[ \Sigma^{-1} = U\Lambda^{-1} U^T \]
\[ U^T \Sigma U = \Lambda \]
\[ U^T U = I \]

Mahalanobis Distance and PCA
\[ d^2_{\Sigma}(\vec{x}, \vec{\mu}) = (\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu}) = \vec{y}^T \Lambda^{-1} \vec{y} \]
\[ \vec{y} = U^T (\vec{x} - \vec{\mu}) \]

Bayesian Classifier
\[ \hat{y} = \arg \max_{y \in Y} p(Y|X)(y|\vec{x}) \]
Problem 1  (10 points)

Suppose you have a dataset including the vectors

\[ \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \quad \vec{z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \]

(a) Consider the following function \( D(\vec{x}, \vec{y}) \). Is it a distance? Why or why not?

\[ D(\vec{x}, \vec{y}) = |x_1 + x_2 + x_3 - y_1 - y_2 - y_3| \]

(b) Find a diagonal matrix \( \Sigma \) such that \( d^2_{\Sigma}(\vec{x}, \vec{y}) > d^2_{\Sigma}(\vec{x}, \vec{z}) \). Express your answer in terms of the variables \( x_1, x_2, x_3, y_1, y_2, y_3, z_1, z_2, z_3 \).
Define $\Phi(z)$ as follows:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$$

Suppose $\vec{X} = [X_1, X_2]^T$ is a Gaussian random vector with mean and covariance given by

$$\vec{\mu} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$$

(a) Define $f_{\vec{X}}(\vec{x})$, to be the pdf of $\vec{X}$ evaluated at $\vec{x} = [x_1, x_2]^T$. Sketch, on the $(x_1, x_2)$ plane, the set of points such that

$$f_{\vec{X}}(\vec{x}) = \frac{1}{4\pi} e^{-\frac{1}{8}(x_2 - 2)^2}$$

(b) In terms of $\Phi(z)$, find the probability $\Pr \{ 2 < X_1 \}$. 
Problem 3  (10 points)

Suppose that a particular covariance matrix Σ has the following eigenvector matrix, $U$, and eigenvalue matrix, $Λ$:

$$U = \frac{\sqrt{2}}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad Λ = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

Let $\vec{y}(\vec{x}) = [y_1(\vec{x}) \ y_2(\vec{x})] = U^T \vec{x}$ be the principal components of a vector space $\vec{x}$.

(a) Plot the set of vectors $\vec{x}$ such that $\vec{x}^T \Sigma^{-1} \vec{x} = 1$

(b) Find the principal component representation of the following vector:

$$\vec{x} - \vec{μ} = \begin{bmatrix} 5 \\ 5 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$
Problem 4  (10 points)

Suppose that, for a particular classification problem, you have the following nine data points $\vec{x}_n$ and their labels $y_n$:

$$X = \begin{bmatrix} -1 & -1 & -1 & 0 & 0 & 1 & 1 & 1 \\ -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 \end{bmatrix}, \quad Y = [0, 1, 0, 1, 1, 1, 0, 1, 0]$$  \hspace{1cm} (1)

(a) Plot the boundaries of the nearest-neighbor classifier, for the region $-2 \leq x_1 \leq 2, -2 \leq x_2 \leq 2$.

(b) Suppose now that $f_{\vec{X}|Y}(\vec{x}|0)$ and $f_{\vec{X}|Y}(\vec{x}|1)$ are both zero-mean Gaussian pdfs, with the covariance matrices $\Sigma_0$ and $\Sigma_1$ respectfully, where

$$\Sigma_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \Sigma_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Define $\eta$ to be the odds ratio, $\eta = p_Y(0)/p_Y(1)$. Find a value of $\eta$ such that a Bayesian classifier correctly labels all of the training tokens given in Eq. (1).