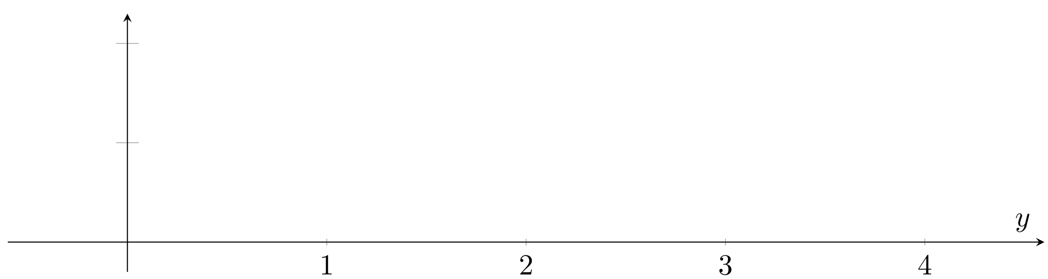
Problem 1 (Detection) *20 points*

(*This is a two-page problem with two different scenarios.*)

Scenario 1: Under H_0, Y is Gaussian $(1,1)$. Under H_1, Y is Gaussian $(3,1)$.
$\mathbb{P}[H_0] = \mathbb{P}[H_1] = 1/2$.

(a) For Scenario 1, work out the decision rule that will result in the smallest possible probability of error. Simplify your expression as much as possible.

(b) Sketch the conditional PDFs $f_{Y|H_0}(y)$ and $f_{Y|H_1}(y)$ from Scenario 1 on the following plot, along with a vertical line that clearly indicates your decision boundary.



(c) Determine the probability of error for your decision rule for Scenario 1. You may leave your answer in terms of the Φ function.

(CONTINUED ON NEXT PAGE)

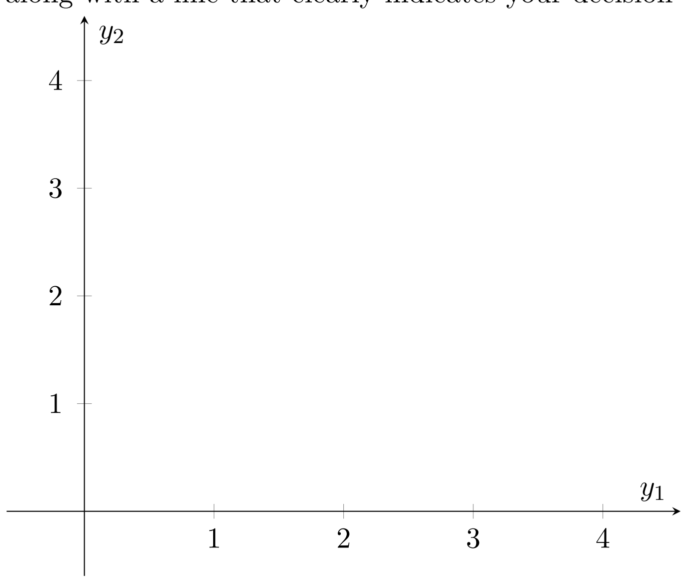
Scenario 2: Under H_0, $\underline{Y}$ is Gaussian $(\underline{\mu}_0, \mathbf{I})$. Under H_1, Y is Gaussian $(\underline{\mu}_1, \mathbf{I})$ where

$$\underline{\mu}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \ \underline{\mu}_1 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \text{ and } \mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \ \mathbb{P}[H_0] = \mathbb{P}[H_1] = 1/2.$$

It may be helpful to recall that the PDF for a Gaussian vector $\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$ with mean vector $\underline{\mu}$ and covariance matrix $\mathbf{\Sigma}$ is $f_{\underline{Y}}(\underline{y}) = \dfrac{1}{\sqrt{(2\pi)^2 \det(\mathbf{\Sigma})}} \exp\left(-\dfrac{1}{2}(\underline{y} - \underline{\mu})^\mathsf{T} \mathbf{\Sigma}^{-1} (\underline{y} - \underline{\mu}) \right)$

(d) For Scenario 2, work out the decision rule that will result in the smallest possible probability of error. Simplify your expression as much as possible.

(e) Sketch the conditional PDFs $f_{\underline{Y}|H_0}(\underline{y})$ and $f_{\underline{Y}|H_1}(\underline{y})$ from Scenario 2 as contour plots on the following plot, along with a line that clearly indicates your decision boundary.



Problem 2 (Estimation) *20 points*

(*This is a two-page problem with a single scenario.*)

Consider the following scenario: X given $Y = y$ is Exponential $(2/y)$ and Y is Uniform $(1, 5)$. Note that for parts (b), (c), and (d), it is possible to obtain an answer without integration. However, it is fine if you write your answer in terms of an integral.

(a) What is the range $R_{X,Y}$?

(b) What is the MMSE estimator of X given $Y = y$?

(c) Calculate $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

(CONTINUED ON NEXT PAGE)

(d) What is the LLSE estimator of X given $Y = y$?

(e) What is the mean-squared error of the LLSE estimator from part (d)?

Problem 3 (Statistics and Sums of Random Variables) *20 points*

The L & L Candy Company claims that the average bag contains 100 candies. After purchasing 64 bags, you calculate that your sample average is 98 and your sample variance is 16. You decide to use a significance test to determine whether you should believe the company's claim.

(a) What kind of significance test should you use?

(b) Should we reject the null hypothesis at a significance level of 0.01?

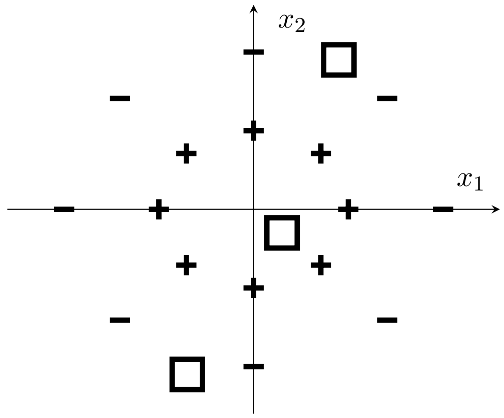
(c) Using your data, construct a confidence interval for the mean with confidence level 0.99.

(d) If you buy 32 more bags and find these additional bags have 104 candies on average, what would be your updated sample average over all $64 + 32 = 96$ purchased bags?

(e) Let $V_1, \dots, V_9$ be i.i.d. Exponential$(1/2)$ random variables. Using the Central Limit Theorem approximation, estimate the probability that $\sum_{i=1}^{9} V_i$ is less than 30. You may leave your answer in terms of the Φ function.

Problem 4 (Machine Learning) *20 points*

You are given the 16 training data points on the figure, denoted by $+$ and $-$ symbols, along with 3 test data points, denoted by squares (for which you have no labels). The axes intentionally have no tick marks, but you can assume that the scale of the x_1 and x_2 axes are the same.



(a) The *closest average classifier* has a high training error rate. Explain why this is the case.

(b) Is there a *linear classifier* of the form $D_{\text{linear}}(x_1, x_2) = \begin{cases} +1 & a_1x_1 + a_2x_2 \geq 0 \\ -1 & a_1x_1 + a_2x_2 < 0 \end{cases}$ that can attain training error rate less than $1/4$? Justify your answer.

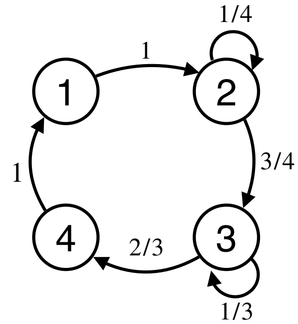
(c) Consider a *non-linear classifier* of the form $D_{\text{non-linear}}(x_1, x_2) = \begin{cases} +1 & g(x_1, x_2) \leq \beta \\ -1 & g(x_1, x_2) > \beta \end{cases}$ where the function $g(x_1, x_2)$ depends only on x_1, x_2 while the threshold β can depend on the training data. Select a good function $g(x_1, x_2)$ for this scenario. Justify your answer.

(d) Write a detailed procedure for selecting the threshold β from part (c) using the training data. Your answer should be written explicitly in terms of $(X_{\text{train},i,1}, X_{\text{train},i,2}, Y_{\text{train},i})$ for $i = 1, \ldots, n_{\text{train}}$ (but it does not need to be MATLAB/Python code or pseudo-code).

(e) Sketch the decision boundary for your trained non-linear classifier from parts (c) and (d) on the figure above. In each test data square, write the label selected by your classifier.

Problem 5 (Markov Chains) *20 points*

Consider the following discrete-time Markov chain with $\mathbb{P}[X_0 = 3] = \mathbb{P}[X_0 = 4] = 1/2$.



(a) List the communicating classes. For each communicating class, determine the period and whether it is transient or recurrent.

(b) Write down the transition matrix.

(c) Does a unique limiting state probability vector $\underline{\pi}$ exist? If so, argue why and solve for it. If not, argue why.

(d) Determine $\mathbb{P}[X_0 = 3, X_1 = 4, X_2 = 1]$.

(e) Determine $\mathbb{P}[X_1 = 1 | X_3 = 2]$.