1. Generate *training data* of size 200 in the following way.

    a. Generate 10 samples m_k from the bivariate Gaussian distribution $m_k \sim N_2\left[\binom{1}{0}, \quad \mathbb{I}_2\right]$. Label these observations as "A".

    b. Generate 10 more samples m_k from the bivariate Gaussian distribution $m_k \sim N_2\left[\binom{0}{1}, \quad \mathbb{I}_2\right]$. Label these observations as "B".

    c. For each class: A and B, generate 100 observations as follows:

        i. Select m_k with probability $1/10$.

        ii. Draw a sample from $N_2[m_k, \quad \mathbb{I}_2/5]$.

2. Repeat steps c.i and c.ii of training data to generate 10,000 new *test data*. That is, you will have 5,000 new observations corresponding to each class.

3. Develop a *k*-nearest neighbor classifier using the training dataset and evaluate its performance on: (a) the training data and (b) the test data using misclassification error. Plot misclassification errors as a function of neighborhood size (*k* = 1, 5, 10, 15, 20, 25, 30, 35, 40). The X-axis corresponds to *k*-values and Y-axis is misclassification error. You may choose smaller intervals for *k*. Also, mark the misclassification rate for least-square regression in the plot.

4. Calculate true error rate (Bayes error). Which *k*-nearest neighbor model has minimum misclassification error and is closest to Bayes error.

5. Summarize main insights from the above exercise.

Note: Submit your assignment through Turnitin. Your submission should include codes used in the assignment.