EE 456 Computer Assignment 2 Report

Name: Sharvari Purighalla
Student ID: 907394064
Date: October 14, 2025

Part I – Maximum Likelihood Estimation (MLE)

Overview: We are given a 2D, 3-class dataset (data1.csv) with features x1, x2, and labels {0,1,2}.
The task is to estimate the class priors, mean vectors, and covariance matrices using MLE.

Steps:

1. Read the dataset and split it into 90% training and 10% testing using stratified sampling.
2. For each class c:

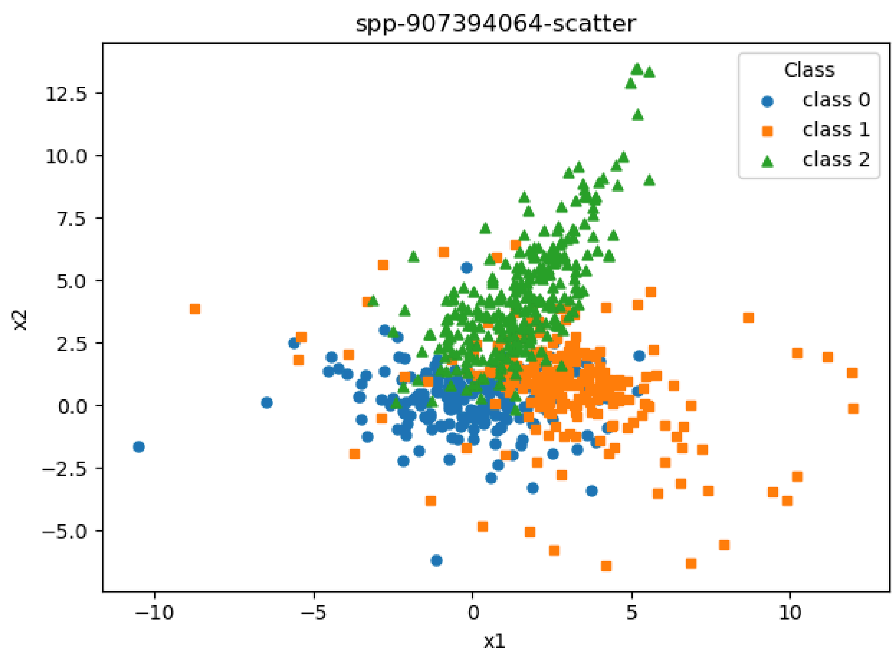
$$\pi_c = \frac{N_c}{N}, \qquad \mu_c = \frac{1}{N_c} \sum_{i:y_i=c} x_i, \qquad \Sigma_c = \frac{1}{N_c} \sum_{i:y_i=c} (x_i - \mu_c)(x_i - \mu_c)^T$$

3. Visualize training data in 2D with distinct markers/colors for each class.

Results:

Class	Prior ($\hat{\pi}_c$)	Mean ($\hat{\mu}_c$)	Covariance ($\hat{\Sigma}_c$)
0	0.2667	[-0.1286, 0.1678]	[4.50, -0.16, [-0.16, 1.50]]
1	0.3333	[2.9719, 0.8006]	[6.06, -1.43, [-1.43, 4.02]]
2	0.4000	[1.4724, 4.2966]	[2.35, 2.37, [2.37, 4.98]]

Visualization:



Part II – Bayesian Classifier

Overview: Using the estimated MLE parameters, we built a Bayesian classifier assuming each class follows a Gaussian distribution:

$$log\, p(x|c) = -\frac{d}{2}log(2\pi) - \frac{1}{2}log|\Sigma_c| - \frac{1}{2}(x - \mu_c)^T\Sigma_c^{-1}(x - \mu_c)$$

and the discriminant function:

$$g_c(x) = log\, p(x|c) + log\pi_c$$

Prediction rule:

$$\hat{y}(x) = arg\, max\, g_c(x)$$

Evaluation Metrics:

- Accuracy
  Fraction of correctly classified test samples.
- Average Negative Log-Likelihood (NLL)
  Measures model calibration:

  $$NLL = -\frac{1}{N_{test}}\sum_i log\, P(y_i|x_i)$$
- Confusion Matrix
  Compares true vs predicted classes

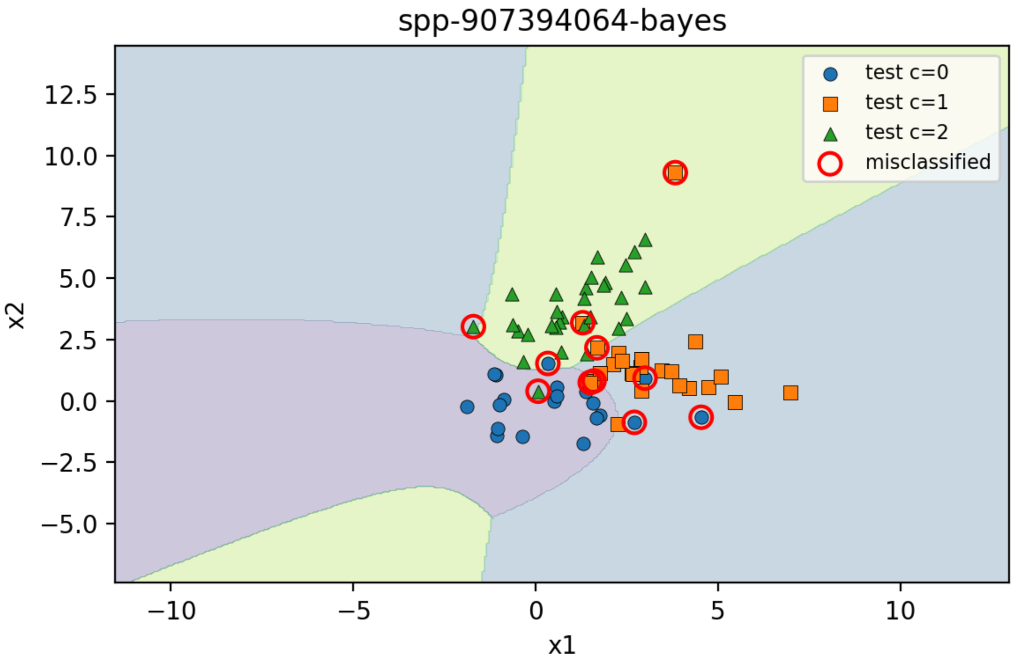
Results:

Metric	**Value**
Accuracy	0.8533 (85.3%)
Average NLL	0.5650

Confusion Matrix (rows = true, cols = predicted)

True/Pred	0	1	2
0	16	3	1
1	2	20	3
2	1	1	28

Visualization:



Part III – Perceptron Algorithm

Overview: For the binary dataset (data2.csv), we implemented the Perceptron learning rule:

$$w \leftarrow w + \eta y_i x_i, \quad b \leftarrow b + \eta y_i$$

Repeat until all points are correctly classified.

Experimental Setup:

Tried three learning rates:
$\eta \in \{0.05, 0.1, 1.0\}$

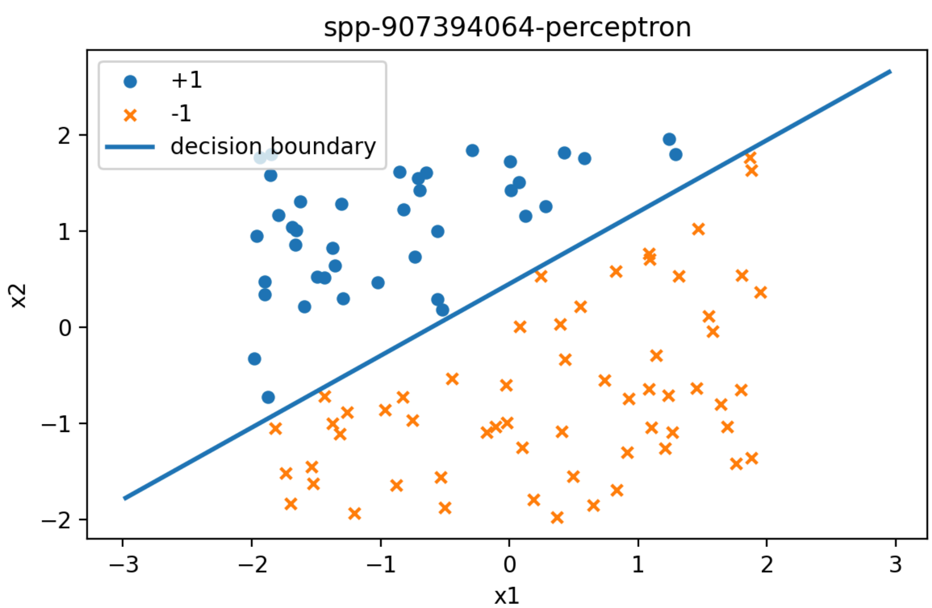
Each run starts with $w = 0$ and updates until convergence.

Results:

Learning Rate (η)	Iterations until Convergence	Training Errors
0.05	8	0
0.10	8	0
1.00	8	0

All rates converged in 8 epochs with no misclassifications, indicating the dataset is linearly separable.

Visualization:



The learned line
$w_1 x_1 + w_2 x_2 + b = 0$
Perfectly separates the two classes.