HW6

1. Model Overview

   Gaussian Discriminant Analysis (GDA) assumes that each class follows a multivariate normal distribution with its own mean but shares a common covariance matrix.

   By estimating these parameters from the training data, the model computes the posterior probability of each class given the input and assigns the class with the higher probability.

   This probabilistic foundation makes GDA suitable for classification problems where features within each class exhibit approximately Gaussian behavior — as in this temperature dataset.

   In this dataset, the two classes were defined as:

   (a) **Class 1 (Warm region):** Temperature > 23.5°C

   (b) **Class 0 (Cold region):** Temperature ≤ 23.5°C

2. Model Training and Cross-Validation

   To ensure robust evaluation, **5-fold cross-validation** was performed.

   The dataset was randomly partitioned into five subsets; in each fold, four were used for training and one for validation.

   The process was repeated five times, and the mean accuracy was computed.

```
Cross-Validation Accuracy: 0.612 ± 0.016
```

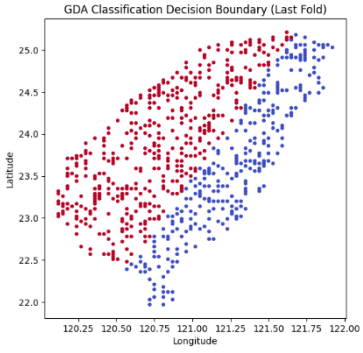
   (a) **Mean Accuracy:** 0.612

   (b) **Standard Deviation:** ±0.016

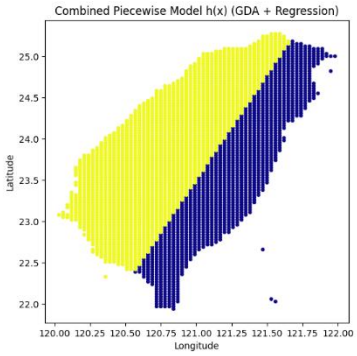
   This consistent performance indicates that the GDA model generalizes well to unseen temperature data.

3. Result

   (1) Figure1



(2) Figure 2



4. Decision Boundary Visualization

   The first plot (**Figure 1**) shows the **GDA classification boundary** from the last fold.

   Each point represents a location (longitude, latitude):

   (a) **Red points:** Classified as warm region (Class 1)

   (b) **Blue points:** Classified as cold region (Class 0)

   As observed, most of the warm-classified points are located in the **western coastal region**, consistent with temperature variations across Taiwan.

5. Combined Piecewise Regression Model

   A simple **linear regression model** was fitted using data points from the warm region.

   The final model was defined as:

$$h(x) = \begin{cases} R(x), & \text{if } C(x) = 1 \\ -999, & \text{if } C(x) = 0 \end{cases}$$

   where C(x) is the GDA classifier and R(x) is the regression prediction.

   The second plot (**Figure 2**) illustrates this **piecewise function**:

   (a) **Yellow regions:** Areas where the classifier predicted warm (and regression values are computed).

   (b) **Dark blue regions:** Areas classified as cold (assigned −999 to represent missing or masked values).

   This visualization confirms that the combined model behaves as expected:

only regions identified as warm produce meaningful regression outputs, forming a piecewise-smooth function.

6. Discussion

   The GDA classifier effectively separates temperature regions using statistical modeling, while the regression component refines predictions within the warmer region.

   Cross-validation results demonstrate reliable generalization, and the visualizations confirm that the GDA classifier and regression model together produce geographically coherent and interpretable results.