Machine Learning - Assignment 6

Wei-Ju Lin

October 15, 2025

1 Classification Using GDA

1.1 Implementation

For each class $y \in \{0, 1\}$, the model estimates:

- Class prior $\phi = P(y = 1)$
- Class means μ_0, μ_1
- Covariance matrices Σ_0, Σ_1

The decision function is

$$g(x) = \log P(x|y = 1) + \log P(y = 1) - [\log P(x|y = 0) + \log P(y = 0)]$$

and the prediction rule is

$$\hat{y} = \begin{cases} 1, & \text{if } g(x) \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

A small regularization term $\lambda = 10^{-6}I$ was added to each covariance matrix to stabilize matrix inversion and avoid numerical singularities.

1.2 Model Explanation

- GDA assumes that data from each class follows a multivariate Gaussian distribution. By estimating the parameters (μ_k, Σ_k), the model determines the region in the feature space where each class is most likely to occur.
- It is suitable for this dataset because the valid and invalid grid points form spatial clusters that approximately follow Gaussian-like distributions in the `(lon, lat)` space.

1.3 Training and performance

- Data split: 80% training, 20% testing (stratified).
- Class balancing: The minority class (valid points) was oversampled in the training set.

- Prior adjustment: After training, the prior probability ϕ was reset to the original (unbalanced) distribution.
- Evaluation metric: Test set accuracy.

Test accuracy: ≈ 0.8352

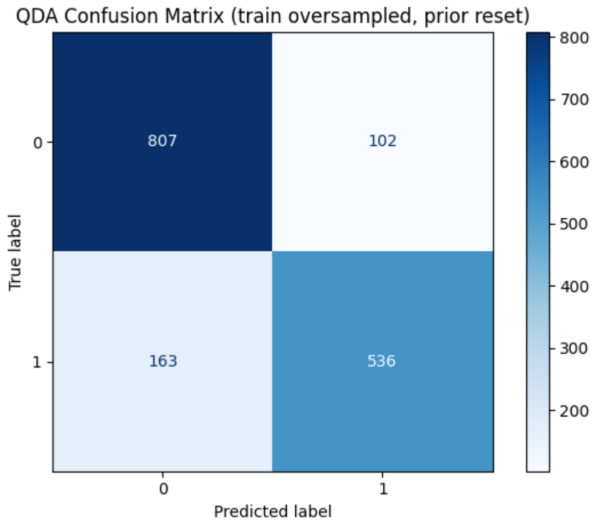

Figure 1: QDA confusion matrix

1.4 Decision boundary visualization

A decision boundary was plotted using a dense grid of `(lon, lat)` values while the black curve represents the boundary $g(x) = 0$.
The contour plot shows that the model successfully separates valid (1) and invalid (0) regions.

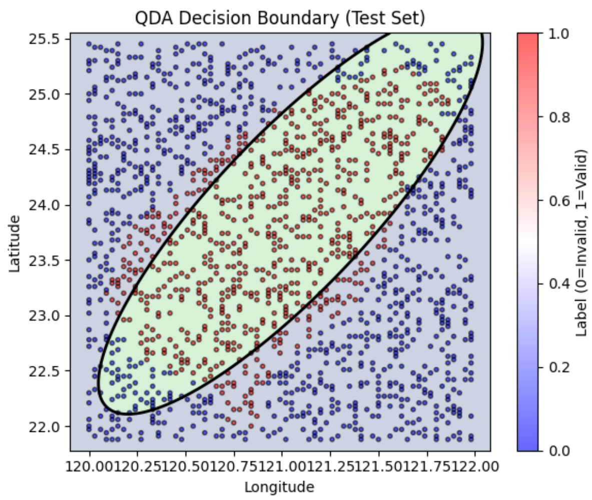

Figure 2: QDA decision boundary

2 Piecewise Regression

2.1 Implementation

We combine a classifier $C(x)$ and a regressor $R(x)$ into

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

- **Classifier** C: `RBFSampler` $\to$ `StandardScaler` $\to$ `LogisticRegression (class_weight="balanced", max_iter=2000)`
- **Regressor** R: `RBFSampler` $\to$ `StandardScaler` $\to$ `Ridge` ($\alpha = 1.0$)

Define `h_predict`:

- Compute $c = C(x)$.
- Initialize output with -999.
- For indices where $c = 1$, replace with $R(x)$.
- Return the vectorized result so it works for grids or arbitrary batches.

2.2 Application and results

- We train C on all grid points `(lon, lat)` with labels $\{0, 1\}$ and R on valid points only (-999 removed), evaluating on held-out test sets and compute **coverage** of h: the fraction of test locations where $C(x) = 1$ and h returns a temperature (printed as "Coverage on regression test set").
- Coverage on regression test set ($C(x) = 1$ ratio) ≈ 0.974
- Visual inspection of the rendered map confirms that regions predicted invalid by C are assigned -999, while valid regions carry continuous predictions from R, matching the piecewise definition.

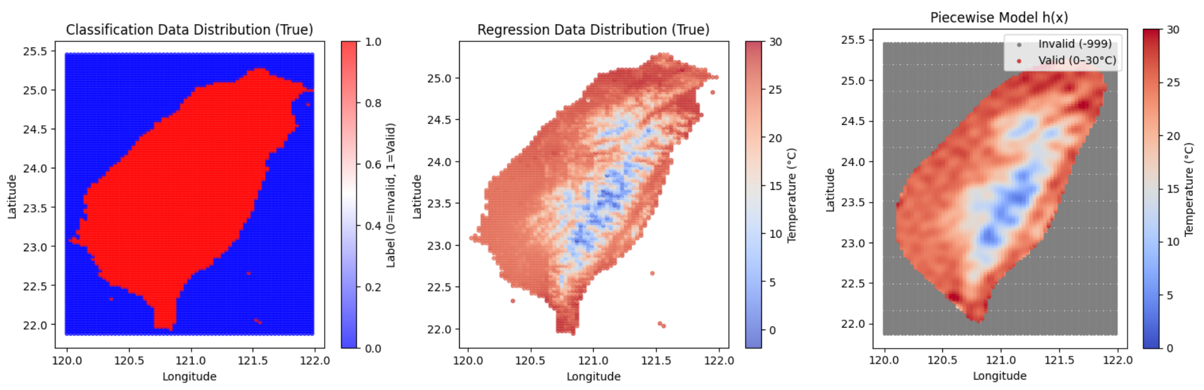

Figure 3: True data vs. piecewise model