Voyager-X Signal Classifier — Theory Summary

Sahasra Oleti - 24B4536

Overview

We consider a binary classification problem involving radio signals received from the Voyager-X spacecraft. Each signal has two features: intensity and frequency stability. Based on noisy observations, the task is to classify each signal as either a potential alien beacon (Class C_1) or background cosmic noise (Class C_0).

We develop and compare four classification methods: (1) Least Squares Classification, (2) Gaussian Discriminant Analysis with shared covariance (LDA), (3) Gaussian Discriminant Analysis with class-specific covariance (QDA), and (4) Logistic Regression. Finally, we apply Bayesian decision theory to incorporate asymmetric classification costs.

Dataset Description (Given)

The synthetic dataset consists of:

- Class C_1 (Beacons): Gaussian with mean $\boldsymbol{\mu}_1 = [2, 2]^T$ and covariance

$$\Sigma_1 = \begin{bmatrix} 0.5 & 0.2 \\ 0.2 & 0.5 \end{bmatrix}.$$

- Class C_0 (Noise): Gaussian with mean $\boldsymbol{\mu}_0 = [-2, -2]^T$ and covariance

$$\Sigma_0 = \begin{bmatrix} 1.0 & -0.3 \\ -0.3 & 1.0 \end{bmatrix}.$$

- Outliers: A small number of class C_0 samples near $[6, 6]^T$.

Part 1: Least Squares Classification

Least squares treats classification as a regression task. We encode class labels as

$$t = \begin{cases} 1, & C_1 \\ 0, & C_0. \end{cases}$$

Constructing a feature matrix Φ (including a bias term), we solve the minimization problem:

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} ||\Phi\mathbf{w} - \mathbf{t}||^2.$$

Setting the derivative to zero yields the closed-form:

$$\boxed{\mathbf{w}^* = (\Phi^\top \Phi)^{-1} \Phi^\top \mathbf{t}.}$$

Predictions are computed as:

$$y(\mathbf{x}) = \mathbf{w}^\top \phi(\mathbf{x}),$$

and classification uses

$$\boxed{y(\mathbf{x}) \geq 0.5 \Rightarrow C_1, \qquad y(\mathbf{x}) < 0.5 \Rightarrow C_0.}$$

Conclusion: Least squares provides a simple linear classifier but does not produce calibrated class probabilities and is sensitive to outliers and heteroscedastic noise.

Part 2: Gaussian Discriminant Analysis

GDA adopts a generative perspective, modeling each class conditional density as a multivariate Gaussian:

$$p(\mathbf{x}|C_k) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k), \qquad p(C_k) = \pi_k.$$

Using Bayes' theorem:

$$\boxed{p(C_k|\mathbf{x}) = \frac{p(\mathbf{x}|C_k)p(C_k)}{\sum_j p(\mathbf{x}|C_j)p(C_j)}.}$$

Case A: Shared Covariance (LDA)

If $\Sigma_0 = \Sigma_1 = \Sigma$, the discriminant comparison reduces to:

$$g(\mathbf{x}) = \mathbf{x}^\top \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) - \tfrac{1}{2}(\boldsymbol{\mu}_1^\top \Sigma^{-1} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^\top \Sigma^{-1} \boldsymbol{\mu}_0) + \ln \frac{\pi_1}{\pi_0}.$$

Decision rule:

$$g(\mathbf{x}) \geq 0 \Rightarrow C_1.$$

Conclusion: With shared covariance, the decision boundary is linear. LDA exploits Gaussian structure and is typically more robust than least squares.

Case B: Separate Covariances (QDA)

If $\Sigma_0 \neq \Sigma_1$, class likelihoods differ in shape:

$$p(\mathbf{x}|C_k) = \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^\top \Sigma_k^{-1}(\mathbf{x} - \boldsymbol{\mu}_k)\right).$$

Comparing log-posteriors yields a quadratic decision boundary:

$$g_1(\mathbf{x}) - g_0(\mathbf{x}) = 0.$$

Conclusion: QDA models different class spreads and allows curved boundaries, improving accuracy when class distributions differ, but at the risk of overfitting.

Part 3: Logistic Regression (Discriminative Model)

Logistic regression models the posterior probability directly:

$$p(C_1|\mathbf{x}) = \sigma(\mathbf{w}^\top \phi(\mathbf{x})), \qquad \sigma(z) = \frac{1}{1 + \exp(-z)}.$$

The loss is the negative log likelihood (cross-entropy):

$$\mathcal{L}(\mathbf{w}) = -\sum_{n=1}^{N} \Big[t_n \ln y_n + (1 - t_n) \ln(1 - y_n) \Big].$$

The gradient is:

$$\nabla_{\mathbf{w}} \mathcal{L} = \sum_{n=1}^{N} (y_n - t_n) \phi(\mathbf{x}_n).$$

Parameters are updated by gradient descent until convergence.

Conclusion: Logistic regression produces probabilistic predictions and often performs better than least squares when classes overlap.

Part 4: Decision-Theoretic Classification

Different misclassification types incur different losses.

Let L_{kj} denote the loss incurred from predicting class C_j when the true label is C_k.

The expected loss (risk) of choosing C_j is:

$$R(C_j|\mathbf{x}) = \sum_k L_{kj} p(C_k|\mathbf{x}).$$

The optimal decision minimizes risk:

$$\hat{C}(\mathbf{x}) = \arg\min_{C_j} R(C_j|\mathbf{x}).$$

If missing a beacon (false negative) has large cost, the classifier threshold τ is lowered:

$$p(C_1|\mathbf{x}) \geq \tau \Rightarrow C_1,$$

where $\tau < 0.5$.

Conclusion: Decision theory allows a classifier to adapt to real-world costs without retraining the model.

Summary

Least squares provides a simple baseline linear classifier. GDA models the data generatively and yields linear or quadratic decision boundaries depending on covariance assumptions. Logistic regression provides a probabilistic discriminative model with strong performance on overlapping data. Finally, Bayesian decision theory enables optimal predictions when false positives and false negatives carry asymmetric costs.