Homework 2 — Pattern Recognition

TAs: Ali Eskandarian, Nima Salem

1. Eigenvalues and Eigenvectors (`eigen_vals_vects.py`)

- Compute the sample mean vector and covariance matrix for each class from scratch, using only basic linear algebra operations.
- Calculate the eigenvalues and eigenvectors of each covariance matrix, and sort the eigenvalues from largest to smallest.
- Normalize the eigenvectors and verify that they are orthogonal to each other.
- Use the eigenvectors as principal components to reconstruct the data with one or more components.
- Report and interpret the explained variance ratio of each principal component.

Exploration

Investigate how keeping only the largest eigenvalue (one principal component) versus both eigenvalues (two principal components) affects data reconstruction. Provide a simple discussion of how the eigenvalues relate to the spread and main directions of the data in each class.

2. Bayesian Decision Rules and Decision Boundaries (`bayes_decision_boundary.py`, `ml_map_risk.py`)

- Implement the multivariate Gaussian log-density and use it to build a Bayes classifier for two classes with given means and covariance matrices.
- Visualize the Bayes decision boundary for equal class priors and relate it to the distributions of the two classes.
- Implement and compare three decision rules: Maximum Likelihood (ML), Maximum A Posteriori (MAP), and a risk-based MAP classifier using a given loss matrix.
- Classify a set of test points with each decision rule and compare how often the rules make the same or different decisions.
- Discuss how changing prior probabilities or the loss matrix shifts the decision boundary and changes the classification behavior.

Expected Outcomes, Figures, and Tables

By the end of this homework, you should produce the following results and include them in your report:

1. A scatter plot of the samples for each class together with the eigenvectors drawn from the class mean. This figure should help you explain the main directions and spread of the data in each class.
2. A table listing the eigenvalues and the corresponding explained variance ratios for each class. Use this table to argue how many principal components are needed to represent most of the variation in the data.

3. A plot of the Bayes decision boundary for equal priors, overlaid with the two classes of samples. Describe in simple terms which regions of the feature space are assigned to each class and why.
4. One figure that compares the decision regions for ML, MAP, and risk-based classifiers (for example, three subplots or three colored regions on the same grid). Use this figure to point out where the methods agree and where they differ.
5. A table summarizing misclassification rates or empirical risk for different priors and for the given loss matrix. Use this table to discuss the trade-off between overall accuracy and the cost of misclassifying each class.

Summary

Summarize the key findings from all parts, focusing on how eigenvalues and eigenvectors describe the data structure, how Bayesian decision rules define decision boundaries, and how priors and risk affect classification performance.