1. **Method Comparison Discussion**  
   In this study, we explored four approaches to the six-class Glass Identification problem in order to understand how discriminative and generative paradigms, as well as native multi-class versus one-vs-rest strategies, perform when only a few hundred samples are available. One-vs-rest logistic regression breaks the six-class task into six independent binary classifiers and labels each test sample according to the classifier with the highest predicted probability. This simplicity comes at the expense of losing any global consistency: each submodel is trained in isolation, and imbalanced class sizes can skew individual decision boundaries. By contrast, multinomial (softmax) regression jointly optimizes a single 9×6 weight matrix using cross-entropy loss, and the addition of L2 regularization helps prevent overfitting and enforces a coherent set of class separators. Linear Discriminant Analysis (LDA) takes a generative view, modeling each class as a Gaussian distribution with a shared covariance and deriving a linear discriminant that maximizes between-class variance while minimizing within-class variance; this assumption proves surprisingly robust when sample sizes are limited. Finally, the one-vs-rest logistic variant with ±1 labels differs only in loss formulation from the 0/1 version and yields very similar performance, underscoring that the choice of label encoding has less impact than the overall training strategy.
2. **Accuracy Report**  
   After stratified sampling on the 214 glass samples—reserving 10% for testing—the four methods achieved the following accuracies on the held-out test set: one-vs-rest logistic regression (0/1 labels) reached 68.42%, softmax regression with L2 regularization achieved 73.68%, LDA attained 78.95%, and one-vs-rest logistic with ±1 labels again recorded 68.42%. The superior performance of LDA suggests that the shared-covariance Gaussian assumption aligns well with the true distribution of glass features, while softmax’s unified multiclass optimization and regularization deliver the next best results. The relatively lower accuracy of both one-vs-rest variants highlights the limitations of decomposing a multi-class problem into separate binary tasks without a global consolidation of class information.

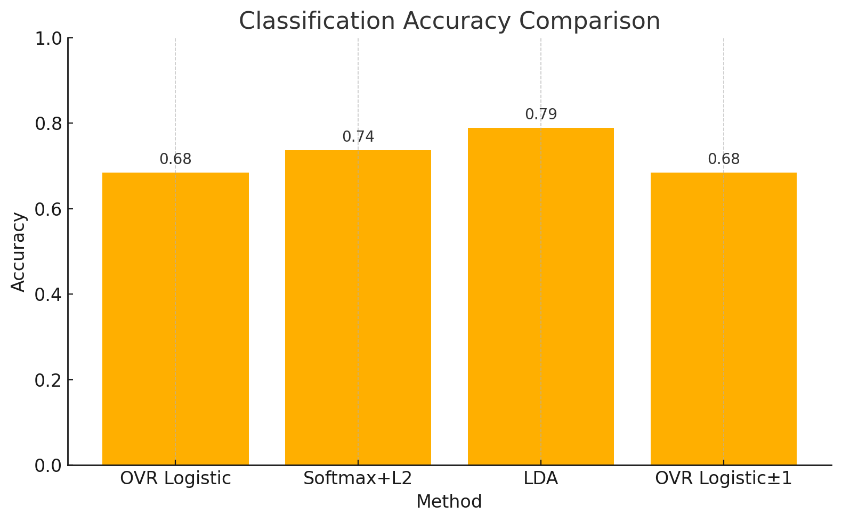


Figure 1. Classification Accurary

1. **PCA Projection and Visualization**  
   To gain geometric insight into the softmax model’s learned decision directions, we performed PCA on its 9×6 weight matrix. Treating each of the six column vectors as a point in ℝ⁹, we centered them, computed the 9×9 covariance, and projected onto the first two principal components. We then applied the same projection matrix to all training samples, yielding a two-dimensional embedding in which each point’s color reflects its true glass type. The resulting scatter plot shows distinct clusters for some classes (e.g., Types 2 and 5 lie far apart), while others (e.g., Types 3 and 4) overlap significantly, mirroring the model’s confusion patterns. Overlaying the six projected weight vectors as arrows from the origin reveals that each arrow points toward its class’s cluster centroid, confirming that the softmax model’s learned directions correspond closely to the principal axes of inter-class separation. This visualization not only validates the interpretability of the softmax parameters but also highlights which class pairs might benefit most from further feature engineering or non-linear modeling.



Figure 2. Training samples and weight vectors under PCA projection