Card Suits Classification Report

1. Data Analysis and Preprocessing

Outlier Detection

- Compared mean and median values for each feature
- Analyzed feature distributions using histograms
- Identified outliers using:
  - Z-score threshold: 3.5
  - Mean-median difference in std units
- Removed 19 outliers from dataset
- Final dataset sizes:
  - Before: 1824 samples
  - After: 1805 samples

Feature Selection

- Used plot_feature_combinations function to visualize all feature combinations
- Selected Features 2 and 4 based on:
  - Visual class separation
  - Statistical stability
  - Complementary information
- Feature characteristics:
  - Feature 2 std: 0.0089
  - Feature 4 std: 0.0010
  - Scale ratio: 8.9 (normalization not needed)

2. Classification Results

Base Performance (Full Dataset)

```
Classifier           Error Rate
Independent:         2.58%
Multivariate:        0.71%
Parzen (h=0.0005):   1.75%
1-NN:                1.81%
```

Training Set Size Impact (5 repetitions)

Reduction	Independent	Multivariate	Parzen (h=0.0005)
10%	3.19% ± 0.42%	1.07% ± 0.26%	8.39% ± 1.46%
25%	3.85% ± 1.50%	0.77% ± 0.15%	5.76% ± 0.52%
50%	2.93% ± 0.50%	0.81% ± 0.02%	3.87% ± 0.23%

Prior Probability Approaches (Reduced Dataset)

Approach	Independent	Multivariate	Parzen (h=0.0005)
Without Prior	2.12%	0.37%	14.47%
Original Prior*	2.19%	0.37%	13.30%
Targeted**	2.79%	0.80%	13.56%

*Original: Black suits (0.165), Red suits (0.085)
**Targeted: Custom class-specific priors and parts

Per-Class Error Analysis

Class	No Prior	Original	Targeted
1	0.88%	0.88%	0.44%
2	0.00%	0.00%	1.75%
3	0.00%	0.00%	0.58%
4	0.00%	0.00%	0.00%
5	0.00%	0.00%	0.00%
6	2.63%	2.63%	2.19%
7	0.00%	0.00%	0.88%
8	0.00%	0.00%	0.00%

3. Analysis of Results

Classifier Performance Patterns

1. **Multivariate Classifier:**
   - Base performance: 0.71% (full dataset)
   - Data reduction impact:
     - 50%: 0.81% ± 0.02%
     - 25%: 0.77% ± 0.15%

        - 10%: 1.07% ± 0.26%
    - Maintains strong performance even with reduced data
    - Best performance across all configurations

2. Independent Classifier:

- Base performance: 2.58% (full dataset)
- Data reduction impact:
    - 50%: 2.93% ± 0.50%
    - 25%: 3.85% ± 1.50%
    - 10%: 3.19% ± 0.42%
- Consistent but higher error rates
- Most stable with full dataset

3. Parzen Classifier (h=0.0005):

- Base performance: 1.75% (full dataset)
- Data reduction impact:
    - 50%: 3.87% ± 0.23%
    - 25%: 5.76% ± 0.52%
    - 10%: 8.39% ± 1.46%
- Strong performance with full dataset
- Significant degradation with reduced data
- Higher sensitivity to data reduction
- The visualization on how the parzen error rate changes with the window size is in the code

4. 1-NN Classifier:

- Consistent performance: 1.81%
- Comparable to Parzen's full dataset performance

4. Conclusions

1. Best Performing Configurations:

- Full Dataset:
    - Multivariate: 0.71%
    - Parzen: 1.75%
    - 1-NN: 1.81%
    - Independent: 2.58%
- Reduced Dataset:
    - Multivariate maintains performance
    - Parzen shows significant degradation
    - Independent shows slight degradation

2. Data Size Impact:

- Multivariate shows remarkable stability

- Parzen requires full dataset for optimal performance
- Independent shows moderate degradation

3. **Prior Probability Effects:**
   - Most effective with reduced datasets
   - Limited impact on full dataset performance
   - Can improve specific class performance
4. **Practical Recommendations:**
   - Use Multivariate classifier for best overall performance
   - Maintain full dataset when possible, especially for Parzen
   - Consider computational trade-offs vs. accuracy
   - No need for feature normalization
   - Use Parzen only with full dataset

The results demonstrate that while all classifiers perform reasonably well with the full dataset, the Multivariate classifier provides the most robust performance across all configurations. The Parzen classifier shows competitive performance (1.75%) with the full dataset but requires careful consideration of data size for optimal results.