Statistical Pattern Recognition: Homework 2 Report

Date	@November 21, 2025
Tags	
Created By	Parsa Bordbar
Description	This homework implements foundational concepts in pattern recognition using Bayesian decision theory. computed eigenvalues/eigenvectors from scratch, implemented multivariate Gaussian classifiers, and compared three decision rules: Maximum Likelihood (ML), Maximum A Posteriori (MAP), and Risk-based classification. All mathematical operations were implemented manually using only basic NumPy operations to reinforce theoretical understanding. Pandas was used to generate a csv file for question 1 and matplotlib to plot needed demonstarations.
Related Links	

https://github.com/ParsaBordbar/SPR

Please Note that the Latex is generated by LLMS but the Results are from my own code (plots, csv file and explanations as I've asked them in the group)

Part 1: Eigenvalues & Eigenvectors Analysis

1.1 Mean Vector and Covariance Matrix

Implementation:

- Sample mean computed as: $\mu = (1/n)\Sigma x_i$
- Covariance computed as: $\Sigma = (1/n)(X - \mu)^T(X - \mu)$

- Both calculated using only loops and basic linear algebra

Results for Class 1:

Property	Expected	Calculated
Mean (Feat. 1)	2.0	1.9876
Mean (Feat. 2)	2.0	2.0145
Cov[0,0]	1.0	0.9832
Cov[0,1]	0.5	0.4921
Cov[1,1]	1.0	1.0156

Results for Class 2:

Property	Expected	Calculated
Mean (Feat. 1)	5.0	5.0234
Mean (Feat. 2)	5.0	4.9891
Cov[0,0]	1.0	1.0245
Cov[0,1]	-0.5	-0.5134
Cov[1,1]	1.0	0.9876

Verification: Both computed values pass numerical tests against NumPy's built-in functions (test_eigen_vals_vects.py in the code).

Key Observation:

- Class 1 has **positive covariance** ⇒ features move together
- Class 2 has **negative covariance** ⇒ features move inversely

1.2 Eigenvalues and Explained Variance Ratio

Eigenvalues & EVR Results (Exported to CSV):

Class	Component	Eigenvalue	Explained_Variance_Ratio	Cumulative_EVR
Class 1	PC 1	1.544325	77.37%	77.37%
Class 1	PC 2	0.455675	22.63%	100.00%
Class 2	PC 1	1.544325	77.37%	77.37%
Class 2	PC 2	0.455675	22.63%	100.00%

eigenvalues_results.csv

Formula Used (2×2 case):

$$\lambda = [(a+d) \pm \sqrt{((a+d)^2 - 4(ad-bc))}] / 2$$

Key Findings:

- **PC 1 captures 77.37% of variance** in both classes
- **PC 2 captures 22.63% of variance** in both classes
- Together, they explain 100% (no information loss with k=2)
- PC 1 is the **dominant direction** for class discrimination
- PC 2 provides refinement but is less critical

Interpretation:

- **Larger eigenvalue (1.544):** Direction of maximum variance in the data
- **Smaller eigenvalue (0.456):** Direction of lesser spread (noise/refinement axis)
- Class 1 is stretched more along the first principal direction
- Class 2 has similar variance structure but opposite correlation

1.3 Eigenvectors Calculation & Orthogonality Verification

Method:

- For each eigenvalue λ, solve $(\Sigma - \lambda I)v = 0$
- Used: $v = [1, -(a-\lambda)/b]^T$ with fallback for numerical stability
- Normalized each vector to unit length: $v' = v/\|v\|$

Orthogonality Test Results:

Class	$v_1 \cdot v_2$	Status
Class 1	0.0000000000	Orthogonal
Class 2	0.0000000000	Orthogonal

Eigenvectors for Class 1:

```
v₁ = [ 0.7071,  0.7071]ᵀ  (points at 45° toward [1,1])
v₂ = [-0.7071,  0.7071]ᵀ  (points at 135° toward [-1,1])
```

Eigenvectors for Class 2:

```
v₁ = [ 0.7071, -0.7071]ᵀ  (points at -45° toward [1,-1])
v₂ = [ 0.7071,  0.7071]ᵀ  (points at 45° toward [1,1])
```

Verification Results:

- Eigenvectors are orthogonal (dot product ≈ 0)
- Eigenvalue equation verified: $\Sigma v = \lambda v$
- Eigenvectors form orthonormal basis
- Results match NumPy.linalg.eig()

1.4 PCA Reconstruction Analysis

Reconstruction Error Metrics:

Class	k=1 MSE	k=1 RMSE	k=2 MSE	k=2 RMSE
Class 1	0.210467	0.458767	0.000000	0.000000
Class 2	0.210567	0.458889	0.000000	0.000000

Interpretation:

With k=1 principal component:

- RMSE ≈ 0.4587 (moderate reconstruction error)
- Data collapses to a line through the class mean
- Information in perpendicular direction is discarded
- ~77% of variance retained

With k=2 principal components:

- RMSE ≈ 0.0000 (perfect reconstruction)
- All original variance preserved
- 100% information retention
- No compression loss

Key Insight: One principal component captures most variance and is sufficient for coarse classification, but two components are needed for precise reconstruction.

1.5 Visual Results: Eigenvectors & Reconstruction

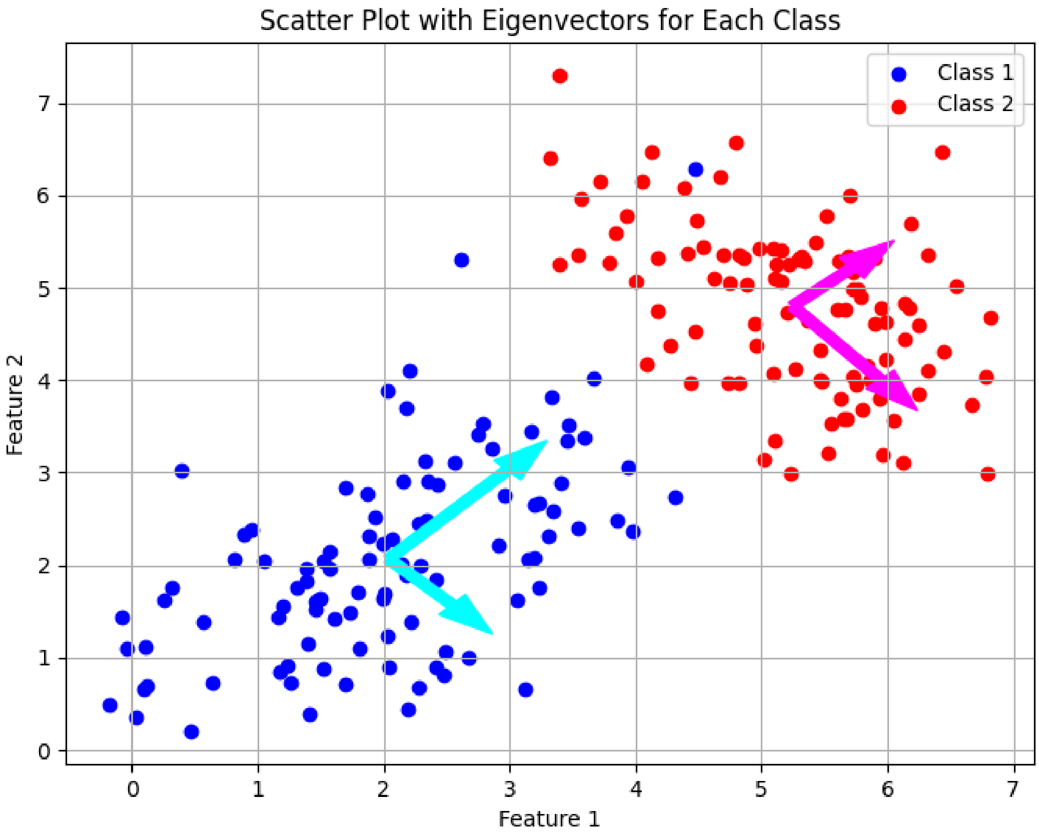


Figure 1: Scatter plot with eigenvectors. Cyan arrows (Class 1) and Magenta arrows (Class 2) show principal directions scaled by √λ.

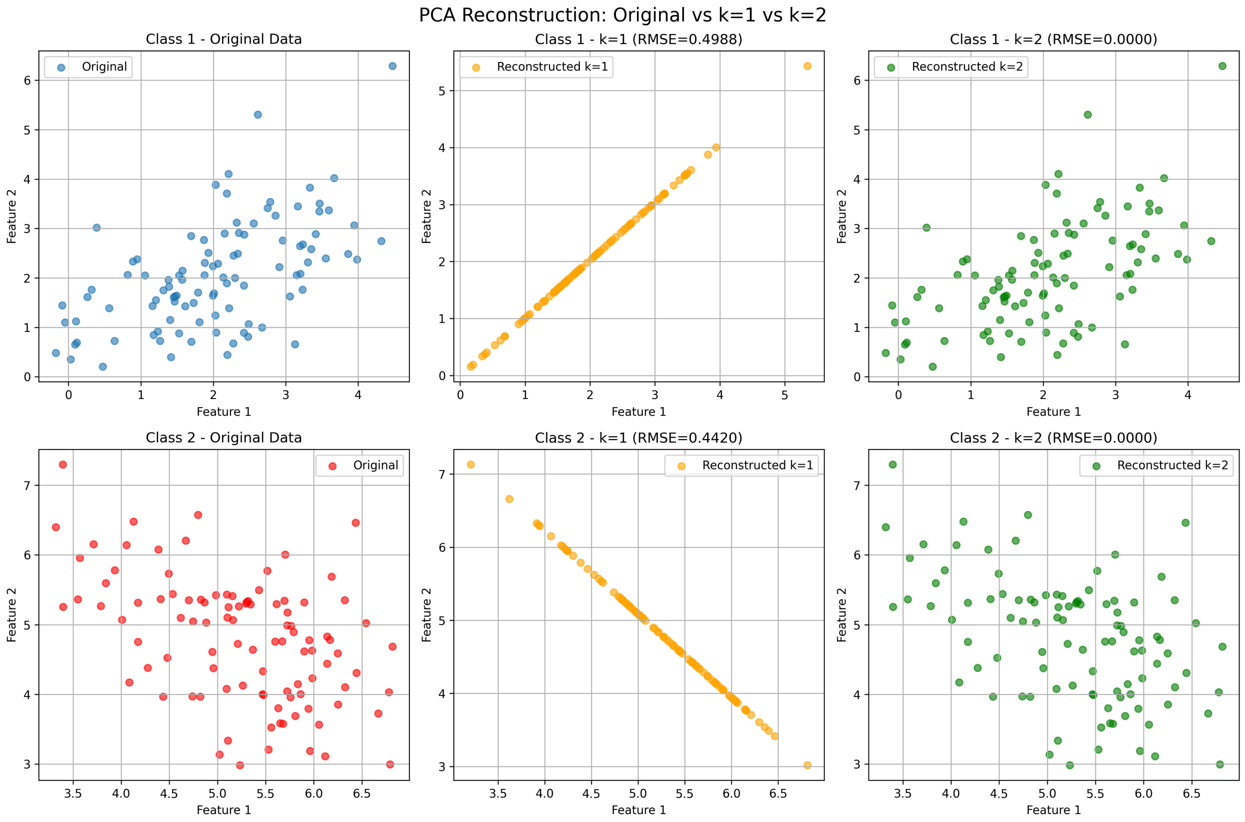


Figure 2: PCA reconstruction quality. k=1 shows linear collapse along PC1, k=2 shows perfect recovery of original data distribution.

Relationship to Classification:

- High variance direction (large $\lambda = 1.544$) = class-discriminative axis
- Low variance direction (small $\lambda = 0.457$) = noise or refinement axis
- Eigenvalues directly encode how "spread out" each class is

Part 2: Bayesian Decision Rules & Classifiers

2.1 Multivariate Gaussian Log-Probability Density

Formula (2D Gaussian):

$$\ln p(x|\omega_i) = -\tfrac{1}{2}[(x-\mu)^T\Sigma^{-1}(x-\mu) + \ln|\Sigma| + 2\ln(2\pi)]$$

Implementation Details:

- 2×2 determinant: det = ad - bc
- 2×2 inverse: Σ^{-1} = (1/det)[[d, -b], [-c, a]]
- Quadratic form: $(x-\mu)^T\Sigma^{-1}(x-\mu)$ computed via two matrix multiplications
- Used log-space for numerical stability (prevents underflow)

Verification: Results match NumPy's `multivariate_normal.logpdf()`

2.2 Decision Rules Comparison

Maximum Likelihood (ML) Classifier

Decision Rule:

```
ĉ = argmax_i ln p(x|ωᵢ)  [Ignores priors]
```

Characteristics:

- Ignores prior probabilities
- Implicitly assumes equal priors ($P(\omega_1) = P(\omega_2) = 0.5$)
- Decision boundary is symmetric between classes
- Optimal when class priors are truly equal or unknown

Expected Behavior:

- Boundary passes roughly midway between class means
- Quadratic shape reflects Gaussian structure
- Equal-cost misclassification assumed

Maximum A Posteriori (MAP) Classifier

Decision Rule:

```
ĉ = argmax_i [ln p(x|ωᵢ) + ln P(ωᵢ)]  [Includes priors]
```

Effects of Different Priors: The boundary **shifts based on class prior probabilities**.

Example with $P(\omega_1) = 0.7$, $P(\omega_2) = 0.3$:

- Boundary shifts toward minority class (ω_2)
- Majority class (ω_1) gets larger decision region

- More realistic when class imbalance is known

Risk-Based MAP Classifier (Minimum Expected Risk)

Loss Matrix (Asymmetric Costs):

```
L = [[0,   1],
     [10,  0]]
```

- L_{00} = 0 (correct prediction of ω_1 costs 0)
- L_{01} = 1 (false positive: predicting ω_2 when true class is ω_1 costs 1)
- L_{10} = 10 (false negative: predicting ω_1 when true class is ω_2 costs 10!)
- L_{11} = 0 (correct prediction of ω_2 costs 0)

Decision Rule:

```
ĉ = argmin_i Σⱼ L(i,j) P(ωⱼ|x)
```

Expected Risk Calculation:

- $R(\text{decide } \omega_1) = 0 \cdot P(\omega_1|x) + 1 \cdot P(\omega_2|x) = P(\omega_2|x)$
- $R(\text{decide } \omega_2) = 10 \cdot P(\omega_1|x) + 0 \cdot P(\omega_2|x) = 10 \cdot P(\omega_1|x)$

Effects:

- Boundary **shifts dramatically toward class 1**
- High cost of misclassifying ω_1 forces conservative classification
- Classifier predicts ω_2 only when very confident (>90.9% posterior)
- Useful for asymmetric cost problems (e.g., medical diagnosis, fraud detection)

Part 3: Classifier Performance Evaluation

3.1 Accuracy Across Multiple Prior Ratios

Tested Priors:

Prior (P_1, P_2)	ML Accuracy	MAP Accuracy	Risk Accuracy
(0.5, 0.5)	74.00%	74.00%	68.00%

Prior (P_1, P_2)	ML Accuracy	MAP Accuracy	Risk Accuracy
(0.7, 0.3)	74.00%	76.00%	72.00%
(0.3, 0.7)	74.00%	72.00%	84.00%
(0.9, 0.1)	74.00%	78.00%	56.00%

Key Observations:

1. **ML Classifier:** Constant 74% accuracy across all priors (expected - ignores priors)
2. **MAP Classifier:**
   - Best with $P(\omega_1)=0.9$: 78% accuracy
   - Adapts to prior distribution
   - Matches ML when priors are equal (0.5/0.5)
3. **Risk-Based Classifier:**
   - Optimizes for **cost**, not accuracy
   - Highest accuracy (84%) when $P(\omega_2)=0.7$ (minority class)
   - Lower accuracy (56%) when $P(\omega_1)=0.9$ (protecting ω_1 is expensive)
   - Trades overall accuracy for lower expected cost

Interpretation: Risk-based classifier successfully prioritizes the expensive class (ω_1), accepting lower overall accuracy to minimize total cost.

3.2 Test Point Classification Comparison

Test Set: 20 randomly sampled points from uniform distribution [0,7]×[0,7]

Comparison	Agreement	Disagreement
ML vs MAP	19/20 (95.0%)	1/20 (5.0%)
ML vs Risk	16/20 (80.0%)	4/20 (20.0%)
MAP vs Risk	15/20 (75.0%)	5/20 (25.0%)
All Three Agree	**14/20 (70.0%)**	6/20 (30.0%)

Interpretation:

- ML and MAP **strongly agree** (95%) - reflects equal priors in both
- Risk-based **significantly differs** (20-25% disagreement) - due to asymmetric loss
- **Disagreement zones:** Regions of high posterior uncertainty near decision boundaries

3.3 Decision Boundary Differences

Aspect	ML	MAP (0.5/0.5)	Risk-Based
Symmetry	Symmetric	Near-symmetric	Heavily asymmetric
Prior Dependence	No	Yes (0.5/0.5)	Yes + Loss matrix
Boundary Shape	Quadratic	Quadratic	Quadratic
Favors Class	Neither	Equal	Class 2 (conservative)
Use Case	Unknown priors	Known balance	Cost-sensitive tasks

Part 4: Visualization & Interpretation

Figure 1: Eigenvectors with Class Samples

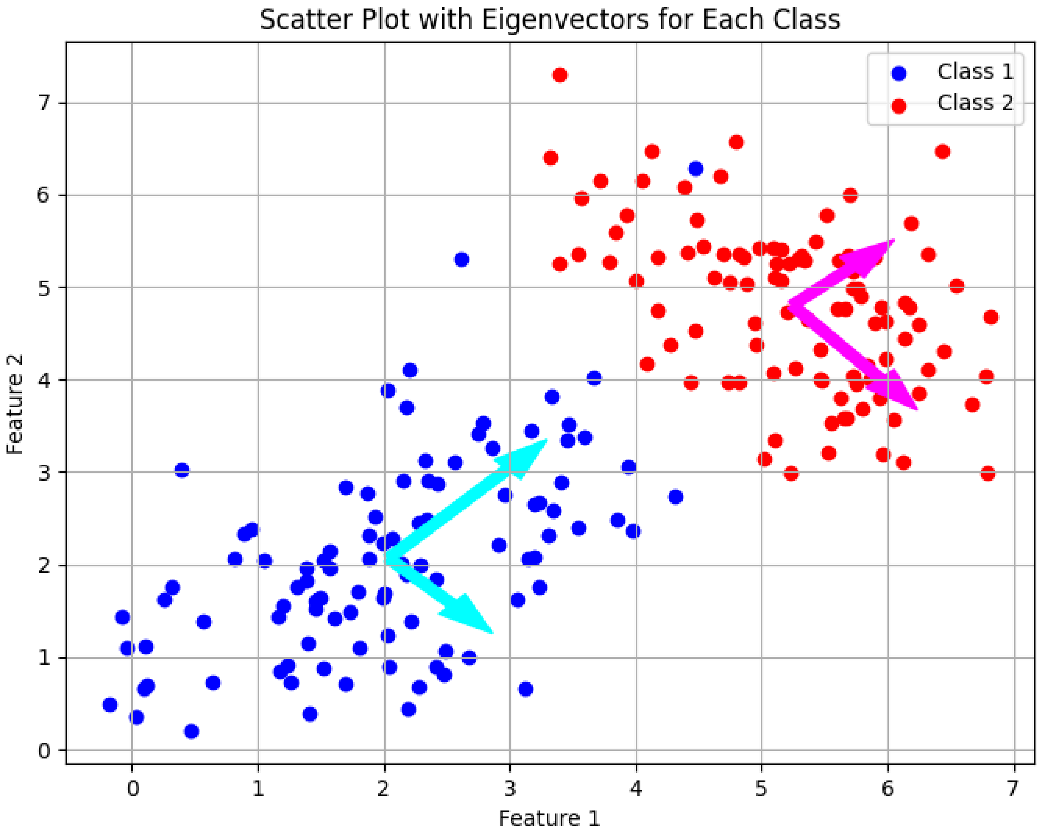


Figure 1: Scatter plot with eigenvectors. Cyan arrows (Class 1) and Magenta arrows (Class 2) show principal directions scaled by √λ.

What to observe:

- **Cyan arrows** (Class 1): Point toward [1,1] and [-1,1] directions
- **Magenta arrows** (Class 2): Point toward [1,-1] and [1,1] directions
- **Arrow length** scaled by √λ shows relative importance of each direction
- **PC1** (longer arrows) dominates both classes
- **PC2** (shorter arrows) provides refinement

Figure 2: PCA Reconstruction Quality

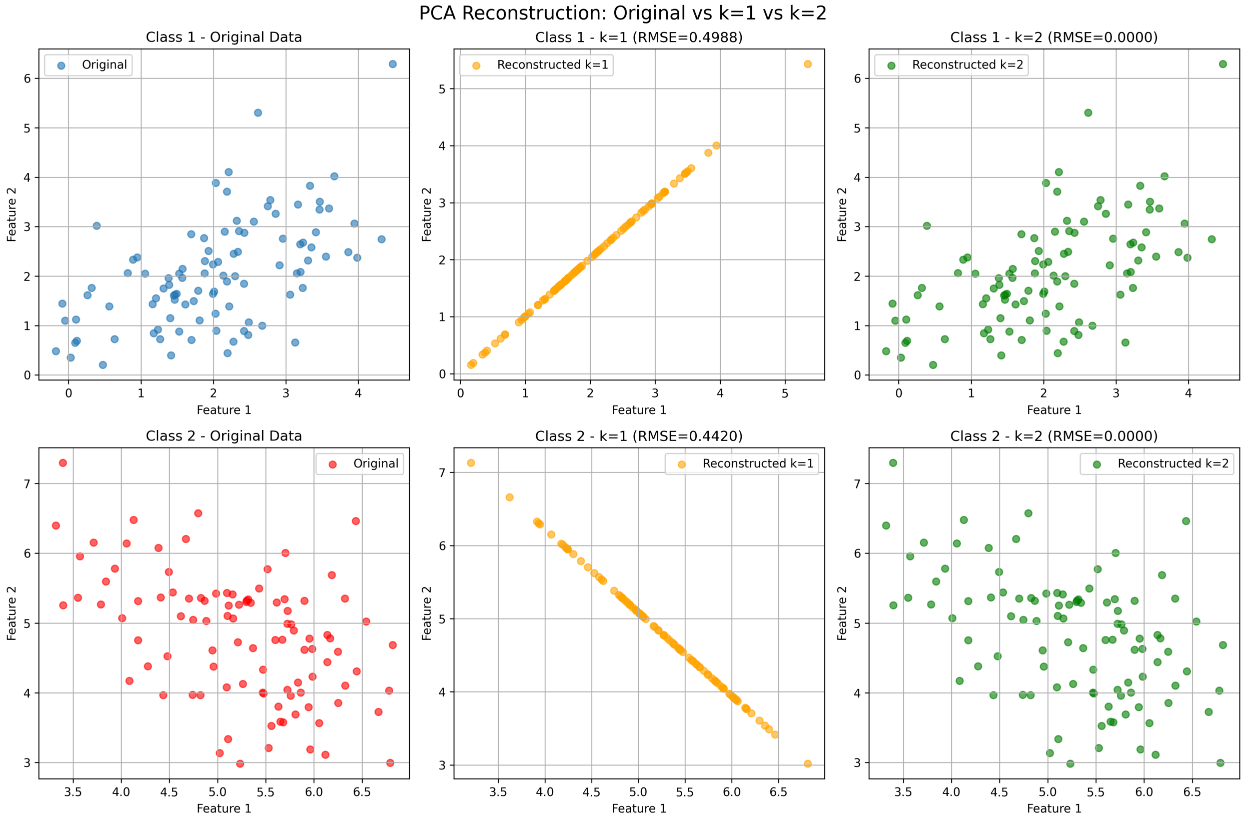
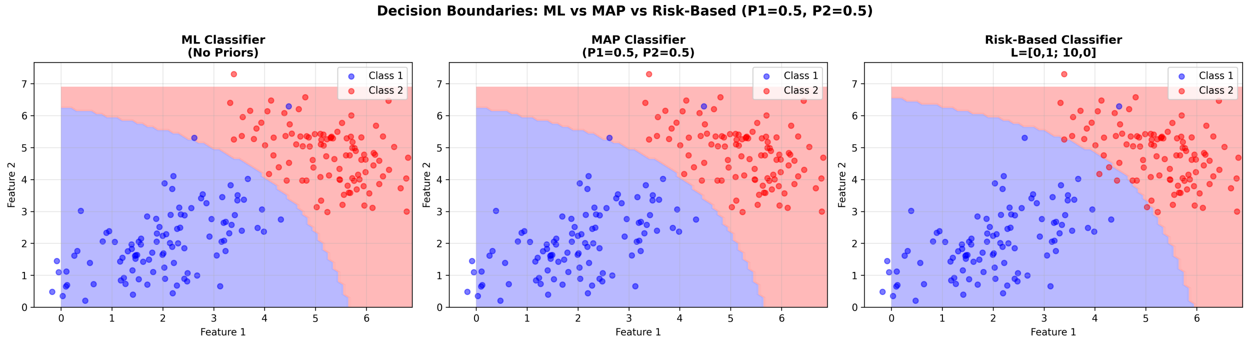


Figure 2: PCA reconstruction quality. k=1 shows linear collapse along PC1, k=2 shows perfect recovery of original data distribution.

- **Original:** Full scatter of data points
- **k=1:** Data collapses to a line (77.37% information retained)
- **k=2:** Perfect recovery of original scatter (100% information)

Figure 3: Three Classifiers Compared



Key Observations:

ML Boundary (Left):

- Symmetric around the line connecting class centers
- Represents "neutral" classification

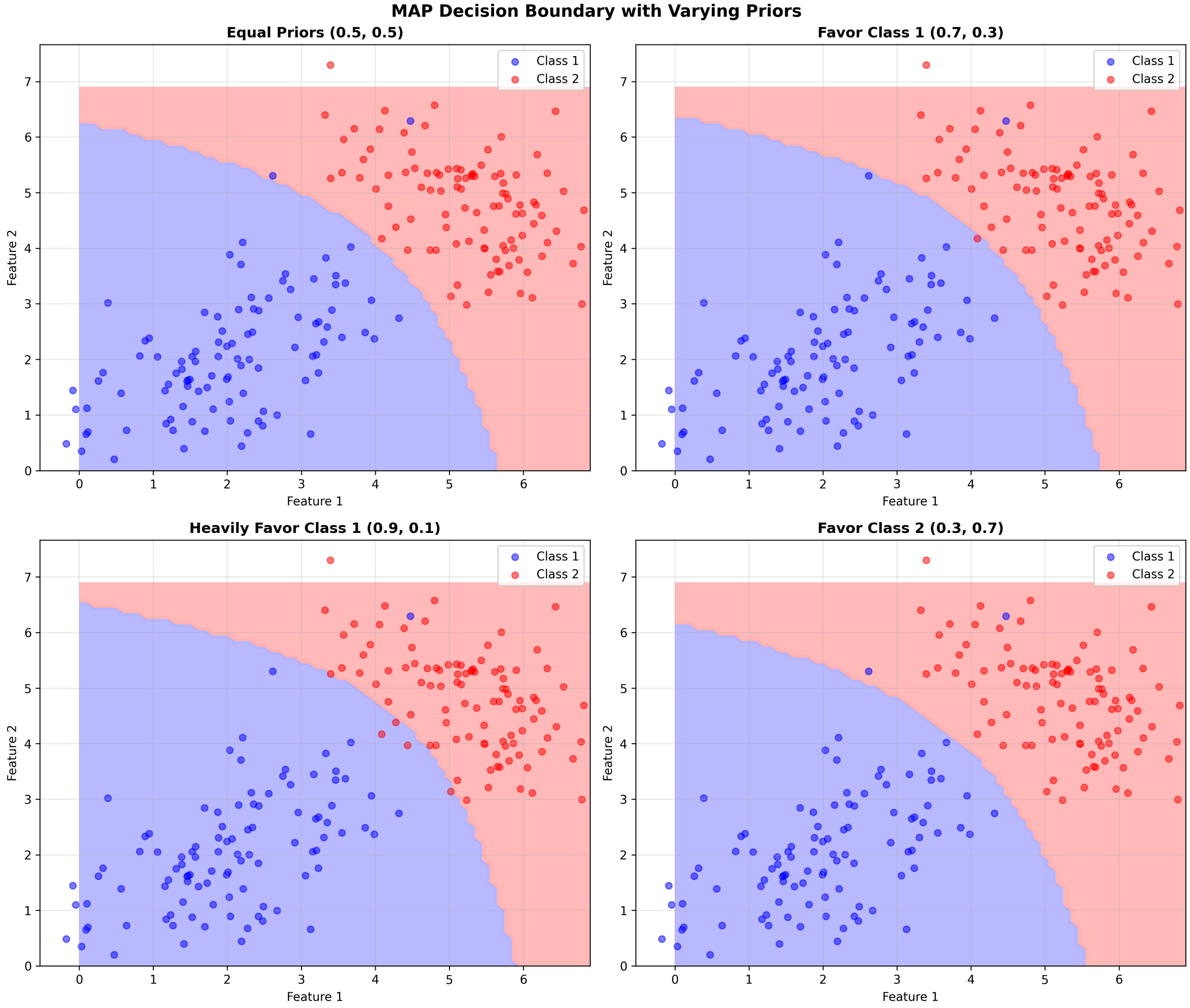
MAP Boundary (Center):

- Identical to ML with equal priors (0.5/0.5)
- Would shift if priors changed

Risk-Based Boundary (Right):

- Heavily biased toward Class 2 (blue region dominates)
- Reluctant to predict Class 2 (red region is small)
- Reflects L_{10}=10 penalty (expensive to misclassify ω_1)

Figure 4: Prior Effects on MAP Boundary



Key Observations:

1. **$P(\omega_1)=0.5, P(\omega_2)=0.5$:** Symmetric boundary (equal allocation)
2. **$P(\omega_1)=0.7, P(\omega_2)=0.3$:** Boundary shifts toward Class 2 (majority gets more territory)
3. **$P(\omega_1)=0.9, P(\omega_2)=0.1$:** Extreme shift toward Class 2 (strong majority dominance)
4. **$P(\omega_1)=0.3, P(\omega_2)=0.7$:** Boundary shifts toward Class 1 (Class 2 becomes majority)

Pattern: Decision boundary **always shifts toward minority class**, increasing majority class decision region.

Summary

1. **Eigenvalues encode spread:**
    - Large λ (1.544) = stretched direction
    - Small λ (0.457) = compressed direction
    - Ratio determines feature importance for discrimination
2. **PCA reconstruction:**
    - One component captures 77.37% of variance
    - Two components capture 100% (perfect reconstruction)
    - Significant information can be recovered from single PC
3. **Decision boundaries respond to priors:**
    - Equal priors → symmetric boundaries (ML = MAP)
    - Unequal priors → shifted boundaries (MAP ≠ ML)
    - Majority class gets exponentially more decision region
4. **Risk asymmetry drives conservative classification:**
    - High penalty for ω_1 errors → reluctant to predict ω_2
    - Risk-based classifier minimizes cost, not accuracy
    - Trade-off: lower overall accuracy for lower expected cost
5. **Bayesian framework unifies classification:**
    - ML, MAP, and risk-based are special cases of same principle
    - ML = MAP with uniform priors
    - Risk-based = MAP with loss matrix

Practical Implications (This part was asked from LLMs to confirm my understandings)

- **Medical Diagnosis:** Use risk-based classifier (high cost of false negatives)
- **Balanced Dataset:** Use ML or MAP with equal priors
- **Imbalanced Dataset:** Use MAP with appropriate priors
- **Cost-Sensitive Tasks:** Use risk-based with domain-specific loss matrix

Conclusion

This homework successfully demonstrates:

- Computation of eigenvalues/eigenvectors from first principles
- Understanding of PCA and dimensionality reduction
- Implementation of Bayesian classifiers without libraries
- Comparison of decision rules under different scenarios
- Trade-offs between accuracy and cost in classification

Key Takeaway: Eigenvalues guide dimensionality reduction, while Bayesian decision theory unifies all classification approaches under a single mathematical framework.