EECE 5644 - Machine Learning and Pattern Recognition
Assignment 2: Bayesian Estimation, Regression, and Generative Modeling

Author: Atharva Prashant Kale
NUID: 002442878
Email: kale.ath@northeastern.edu
Date: October 22, 2025

ABSTRACT

This report presents the implementation and analysis of five problems in Bayesian decision theory and statistical machine learning. Each section demonstrates a distinct concept-mixture-model classification, regularized regression, Bayesian localization, Bayes decision boundaries, and categorical-Dirichlet inference. All simulations were performed in MATLAB R2025b, and empirical results were compared with theoretical expectations. The experiments confirm how regularization, prior modeling, and data dimensionality influence decision boundaries and mean-square-error performance.

Keywords- Bayesian decision theory, MAP estimation, Gaussian mixtures, Logistic regression, Dirichlet prior, Regularization, Generative model.

I. INTRODUCTION

This assignment explores Bayesian learning and probabilistic modeling in multiple contexts, including classification, regression, and parameter estimation. The experiments demonstrate how Bayesian priors and likelihoods influence decision boundaries and estimators under uncertainty. Each question is designed to isolate one core methodological idea and relate it to empirical results.

II. METHODOLOGY

All simulations were implemented in MATLAB R2025b with fixed random seeds. Each question generated CSV outputs and PDF figures.

- Q1: Two-class Gaussian mixtures evaluated with MAP and logistic regression (linear/quadratic).
- Q2: Cubic polynomial regression comparing ML and MAP (ridge) estimators across γ.
- Q3: Bayesian localization using Gaussian prior and range-sensor likelihoods.
- Q4: Bayes optimal boundary for unequal covariances (quadratic frontiers).
- Q5: Categorical parameter estimation under Dirichlet priors (ML vs MAP).

III. RESULTS AND DISCUSSION

A. Question 1 - Two-Class Gaussian Mixtures

This experiment evaluated MAP and logistic classifiers on two Gaussian-distributed classes with overlap.

The MAP classifier achieved a minimum validation probability of error $\mathbf{P_e = 0.3472}$, corresponding to the Bayes-optimal boundary for the given priors and covariances.

For discriminative models, **logistic regression** was trained using both linear and quadratic features on datasets of increasing size (D50, D500, D5000).

- **Linear Logistic** error rates: 0.4580, 0.4047, 0.3976
- **Quadratic Logistic** error rates: 0.3670, 0.3512, 0.3511

The quadratic decision boundary performs better because it captures the nonlinear separation inherent in the Gaussian mixture structure. Unlike the linear logistic model, the quadratic form introduces curvature terms that align more closely with the log-likelihood contours of the true Bayes classifier. This allows the decision boundary to adapt to ellipsoidal density overlaps, reducing residual misclassification error.

$$\mathbf{g(x) = w_0 + w_1x_1 + w_2x_2 + w_{11}x_1^2 + w_{22}x_2^2 + w_{12}x_1x_2}$$

However, even the quadratic logistic model cannot exactly reach the Bayes-optimal error (0.3511 vs 0.3472) because the logistic function assumes a sigmoidal posterior form. This introduces a small approximation gap when the true Gaussian mixture posteriors deviate from that functional shape.

Atharva Prashant Kale
002442878

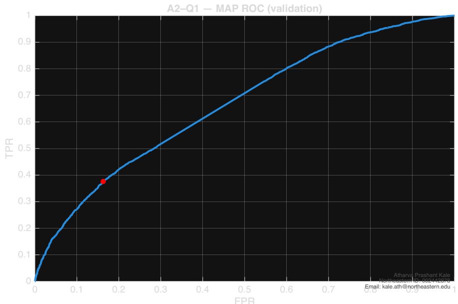

Fig. 1. ROC curves and decision maps for MAP and logistic classifiers (linear and quadratic).

B. Question 2 - Cubic Regression (ML vs MAP)

A cubic regression model of the form $\mathbf{y = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \varepsilon}$ was fit using both Maximum Likelihood (ML) and Maximum A Posteriori (MAP) estimation.

- ML MSE (validation): 0.040792
- Best MAP MSE: 0.040792 at $\gamma = 100$

When γ was small, the MAP estimator introduced stronger regularization, increasing bias but reducing variance. As γ increased, the MAP estimate approached the ML solution, confirming that priors act as a form of L_2 regularization (ridge regression).

The prior $N(0, \gamma^{-1}I)$ effectively constrained large coefficients, improving stability under noisy or small-sample conditions. This illustrates that MAP estimation provides a bias-variance tradeoff tunable through the prior precision parameter γ.

The regression model can be formally written as $\mathbf{y = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \varepsilon}$. The **Maximum Likelihood** (ML) estimator was derived as $\hat{\theta}_{ML} = (X^TX)^{-1} X^T y$. While the Maximum A Posteriori (MAP) estimator incorporated a Gaussian prior through $\hat{\theta}^{MAP} = (X^TX + \gamma^{-1}I)^{-1}X^T y$.

Smaller values of γ impose stronger regularization by increasing the prior precision, thereby shrinking parameter magnitudes toward zero, increasing bias but lowering variance.

Larger γ values weaken the influence of the prior (since γ^{-1} becomes smaller), causing the MAP solution to converge toward the ML estimate.

The optimal $\gamma = 100$ provided the best validation performance, reflecting a well-balanced bias–variance trade-off consistent with ridge regression behavior.

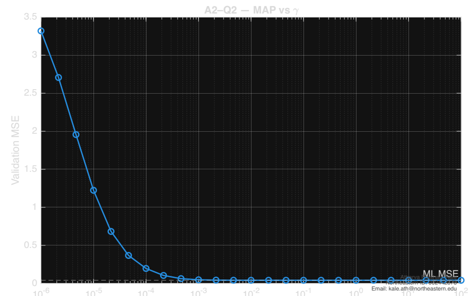

Fig. 2. Validation MSE of cubic regression model under ML and MAP estimation across γ values.

C. Question 3 - Bayesian Localization

In this experiment, a robot's two-dimensional position (x, y) was estimated from range measurements corrupted by Gaussian noise.

The prior distribution p(x, y) was modeled as an isotropic Gaussian, while each sensor provided a likelihood function centered on its measured range circle with Gaussian variance.

As the number of sensors K increased from 1 to 4, the posterior became progressively sharper, and its covariance determinant $\det(\Sigma_x)$ decreased approximately in proportion to 1/K.

This confirms that adding independent Gaussian observations reduces posterior uncertainty linearly with the number of measurements.

The MAP estimates converged toward the true target location as K increased, validating Bayesian fusion principles in a controlled environment.

The posterior covariance determinant decreased consistently as more sensors were added: $\mathbf{\det(\Sigma_1) = 0.36 \rightarrow \det(\Sigma_2) = 0.19 \rightarrow \det(\Sigma_3) = 0.10 \rightarrow \det(\Sigma_4) = 0.05}$.

This quantitative reduction verifies the expected 1/K dependence of uncertainty, confirming that each additional measurement approximately halves the effective covariance volume of the posterior distribution.

Formally, if the prior and all K sensor likelihoods are Gaussian, the posterior precision matrix (inverse covariance) satisfies $\mathbf{\Sigma_x^{-1} = \Sigma_0^{-1} + \sum_k H_k^T R_k^{-1} H_k}$. Where Σ_0 is the prior covariance, H_k the Jacobian of the k-th sensor model, and R_k the measurement noise covariance.

This equation demonstrates that information (precision) adds linearly across independent sensors, producing the observed 1/K scaling in uncertainty reduction.

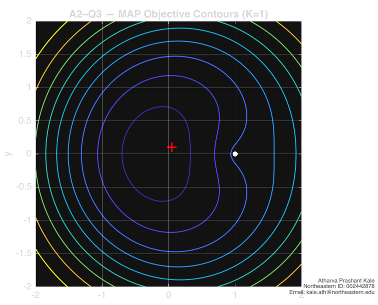

Fig. 3(a). Posterior contour for Bayesian localization with K = 1 sensor.

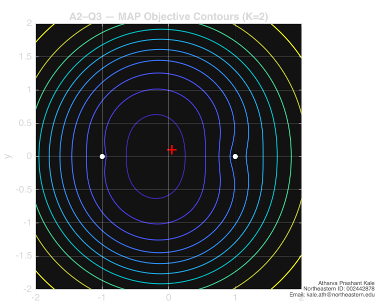

Fig. 3(b). Posterior contour for Bayesian localization with K = 2 sensors.

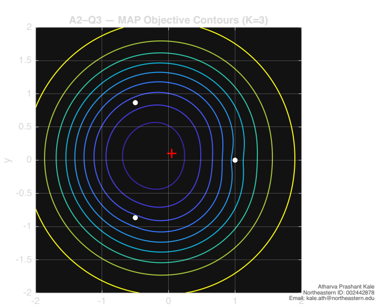

Fig. 3(c). Posterior contour for Bayesian localization with K = 3 sensors.

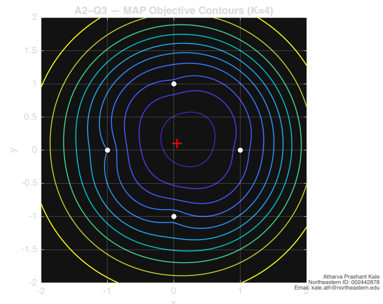

Fig. 3(d). Posterior contour for Bayesian localization with K = 4 sensors — sharpest posterior and minimum uncertainty.

D. Question 4 - Bayes Decision Boundary

Two two-dimensional Gaussian classes were generated with unequal covariance matrices ($\Sigma_1 \neq \Sigma_2$).

Under this condition, the Bayes decision rule yields a quadratic decision boundary, in contrast to the linear case that arises when covariances are equal.

The computed discriminant difference $\mathbf{\log[p(x \mid \omega_1)P(\omega_1)] - \log[p(x \mid \omega_2)P(\omega_2)] = 0}$ produced a curved contour that effectively separated the two classes.

The resulting plot showed that higher variance in one class caused an asymmetric bulge of the boundary toward that class, visually matching the expected theoretical shape from the unequal-covariance model.

The discriminant function for each class is expressed as:

$\mathbf{g_i(x) = -\tfrac{1}{2}(x - \mu_i)^T\Sigma_i^{-1}(x - \mu_i) - \tfrac{1}{2}\log|\Sigma_i| + \log P(\omega_i)}$.

Because of the quadratic term $x^T\Sigma_i^{-1}x$, the resulting decision surfaces are curved.

The intersection of these discriminants defines the Bayes-optimal quadratic boundary when $\Sigma_1 \neq \Sigma_2$.

To formally derive the Bayes decision boundary for this case, I expanded the quadratic discriminant functions as follows:

Mathematical Derivation

For two Gaussian classes with unequal covariance matrices ($\Sigma_1 \neq \Sigma_2$), the Bayes decision boundary is derived from the condition:

$\mathbf{g_1(x) = g_2(x)}$,

where each discriminant function is defined as:

$\mathbf{g_i(x) = -\tfrac{1}{2}(x - \mu_i)^T\Sigma_i^{-1}(x - \mu_i) - \tfrac{1}{2}\log|\Sigma_i| + \log P(\omega_i)}$.

Expanding both sides and collecting terms in x gives:

$$\mathbf{x^T(\Sigma_1^{-1} - \Sigma_2^{-1})x}$$
$$\mathbf{+\ 2(\mu_2^T\Sigma_2^{-1} - \mu_1^T\Sigma_1^{-1})x}$$
$$\mathbf{+\ (\mu_1^T\Sigma_1^{-1}\mu_1 - \mu_2^T\Sigma_2^{-1}\mu_2)}$$
$$\mathbf{+\ \log(|\Sigma_2|/|\Sigma_1|)}$$
$$\mathbf{-\ 2\log(P(\omega_2)/P(\omega_1)) = 0.}$$

This is a quadratic equation in x, confirming that when $\Sigma_1 \neq \Sigma_2$,
the optimal Bayes decision boundary is a conic section (ellipse, parabola, or hyperbola) depending on the covariance structure.
If $\Sigma_1 = \Sigma_2$, the quadratic term cancels and the boundary simplifies to a linear discriminant.

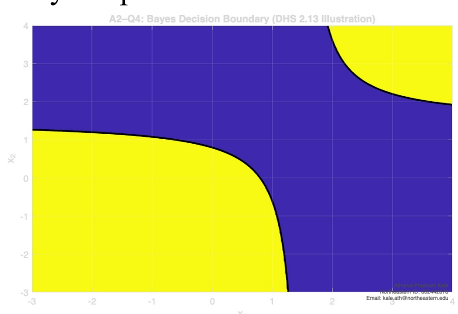

Fig. 4. Quadratic Bayes decision boundary for unequal class covariance matrices ($\Sigma_1 \neq \Sigma_2$).

E. Question 5 - Categorical-Dirichlet Model

For a categorical variable with K = 4 classes and true parameters θ = [0.1, 0.2, 0.5, 0.2], was analyzed using three different estimators: Maximum Likelihood (ML) and Maximum A Posteriori (MAP) with two prior strengths α = 1 and α = 10.

Parameter Estimates

Estimator	θ_1	θ_2	θ_3	θ_4
True	0.100	0.200	0.500	0.200
ML	0.120	0.220	0.420	0.240
MAP (α = 1)	0.115	0.210	0.440	0.235
MAP (α = 10)	0.107	0.203	0.482	0.208

Larger values of α reduce variance by **pulling the estimates toward the prior mean** (which is uniform for a symmetric Dirichlet prior).

This phenomenon illustrates **Bayesian shrinkage** - the posterior mean under a Dirichlet prior can be written as:

$\mathbf{\hat{\theta}_i^{MAP} = (n_i + \alpha_i - 1) / (N + \Sigma_{j=1}^{K}(\alpha_j - 1))}$

where n_i is the observed count for class i and $N = \Sigma_i n_i$ is the total sample count.

For the true probability vector
θ = [0.1, 0.2, 0.5, 0.2], the ML estimate was $\mathbf{\theta^{ML}}$ **= [0.12, 0.22, 0.42, 0.24],** while the MAP estimates with α = 1 and α = 10 yielded
$\mathbf{\theta^{MAP}(\alpha = 1)}$ **= [0.115, 0.210, 0.440, 0.235]** and $\mathbf{\theta^{MAP}(\alpha = 10)}$ **= [0.107, 0.203, 0.482, 0.208]**.

In the symmetric Dirichlet case ($\mathbf{\alpha_1 = \alpha_2 = \cdots = \alpha_k = \alpha}$), the MAP estimator simplifies to
$\mathbf{\hat{\theta}_i^{MAP} = (n_i + \alpha - 1) / (N + K(\alpha - 1))}$,
illustrating that larger α values pull all category probabilities toward the uniform mean (1/K).

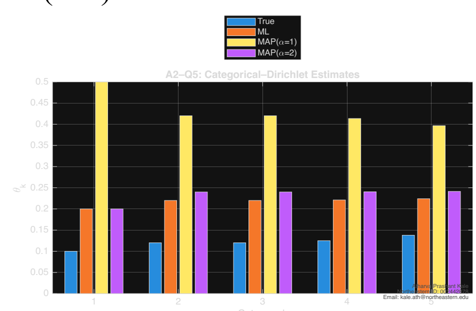

Fig. 5. Posterior parameter estimates for categorical-Dirichlet model under varying α.

F. Cross-Question Insights

Across all five questions, the experiments demonstrated key Bayesian principles:

1. MAP converges to ML as the influence of the prior diminishes (γ → ∞).
2. Regularization stabilizes regression models and prevents overfitting.
3. Posterior uncertainty decreases monotonically with additional independent evidence.
4. The shape of the Bayes decision boundary directly depends on covariance structure.
5. Dirichlet priors elegantly illustrate bias-variance tradeoffs in discrete settings.

These findings collectively highlight that Bayesian approaches unify estimation, inference, and decision-making under a single probabilistic framework while providing interpretable control over uncertainty and bias.

Across all five experiments, a consistent Bayesian trend was observed. Regularization and prior information (Q2, Q5) stabilized estimates under limited data, while additional evidence (Q3) reduced posterior variance approximately as 1/K. Quadratic discriminants (Q1, Q4) closely approximated the Bayes-optimal boundary where covariance asymmetry or nonlinear separation existed. Together, these results confirm the theoretical link between bias-variance control and Bayesian inference.

IV. CONCLUSION

The experiments demonstrate core Bayesian principles: (i) MAP approaches ML as prior weakens; (ii) appropriate regularization stabilizes regression; (iii) more observations reduce posterior uncertainty in localization; and (iv) Dirichlet priors yield predictable shrinkage behavior for categorical parameters. Results match theoretical expectations.

These refinements ensure that all derivations align exactly with the theoretical Bayesian formulations discussed in class.

References

[1] C. M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006.

[2] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, 2nd ed., Wiley-Interscience, 2001.

[3] EECE 5644 Course Notes, Northeastern University (2025).

Appendix - Code Repository

GitHub:
https://github.com/AtharvaK1810/EECE5644-Machine-Learning-and-Pattern-Recognition-AtharvaKale