1. State the Maximum likelihood estimates for the Parameters of Gaussian Naive Bayes. Do these estimates change if we use grayscale Images Instead of RGB Images? Explain why or why not.

Ans: The Maximum Likelihood Estimates for the Parameters Involves estimating Class priors $(P(Y=k))$, the class means (μ_k) for each feature, and the class Variances (σ_k^2) for each feature.

These estimates do not directly change if we use grayscale Images Instead of RGB Images because, the estimation Process is based on distribution of the Pixel Intensities (features), regardless of whether they represent Color Channels in RGB or Grayscale Values. In both Cases, We are estimating the mean & Variance of pixel Intensities for each classes Independently.

2. The accuracy of QDA using RGB Images was lower than that of grayscale Images. what assumption does QDA make that might cause this difference in Performance?

Ans: The lower accuracy of QDA using RGB Images Compared to grayscale images Could be attributed to Several assumptions that QDA makes,

1. Assumption of Covariance Matrix:

* Challenges in accurately estimating Separate Covariance matrices for each class in high dimensional RGB feature Spaces.

2. Curse of Dimensionality:

* The Increased Complexity and Sparsity of data in high dimensional Spaces can hinder accurate parameter estimation.

3. Assumption of Quadratic Decision Boundary:

* Overfitting risks due to the assumption of Quadratic decision Boundaries particularly problematic in high-dimensional Spaces.

4. Assumption of Gaussian Distribution:

* Assumptions of Gaussian feature distributions may not accurately Capture the Complexities of RGB pixel Intensity distributions, leading to Suboptimal Performance.

3. Both LDA and Gaussian Naive Bayes Saw reduced test accuracy on grayscale Images Compared to RGB Images. why this might be the Case (is it the data, the model, or Something else)?

Ans: The reduction in test accuracy of both LDA and Gaussian Naive Bayes on grayscale Images Compared to RGB Images may be due to,

1. Loss of discriminative information in grayscale images, as they lack Color channels present in RGB Images.

2. Models' assumptions and Complexity, which may not be well suited to the characteristics of grayscale Images.

3. Reduced feature representation in grayscale Images, resulting in less expressive data for Classification.

4. Potential differences in data distribution between grayscale and RGB Images.

5. possible inadequacies in preprocessing techniques optimized for RGB Images.

6. Diminished Separability between classes in grayscale images due to reduced discriminative Information.

4. How many Parameters estimated for each model and each Image type (RGB and Grayscale)?

Ans 1. Linear Discriminant Analysis (LDA):

* Parameters estimated: For LDA, Parameters estimated include the class means (μ_K) for each feature and pooled/shared Covariance matrix (Σ) across all classes.

* Number of Parameters : 10 class priors [Prior probability for all classes]

For RGB,

For each class we estimated mean for every feature => $10 * 3072$ Class means.

Here No. of features = 3072, No. of classes = 10

for dataset we calculated 1 Shared Covariance => $3072 * 3072$ Matrix

For Grayscale,

It is same as RGB but here we have No. of features = 1024

=> Class means = $10 * 1024$, Covariance Matrix = 1024×1024 Matrix

along with we calculated weights => $10 * 1024$ for Grayscale

and $3072 * 10$ for RGB & we have same number of Intercepts in both

i.e, 10 intercepts

2. Quadratic discriminant Analysis

* parameters estimated : class means (μ_k) for each feature of each class and Covariance Matrix (Σ_k) for each class

* Number of Parameters : 10 class priors

For RGB, No. of Class means => $10 * 3072$ [Same as LDA]

No. of Covariances => 10 Covariance matrices of $3072 * 3072$

$= 10 * [3072 * 3072]$

For Grayscale, No. of class means => $10 * 1024$

No. of Covariances => $\underbrace{10\,[1024 * 1024]}_{10\,[1024,1024]}$ matrices

10 [1024, 1024] matrices.

3. Gaussian Naive Bayes:

* Parameters estimated: Parameters Include class priors $P(Y=K)$, class means (μ_K) and class Variances (σ_K^2) for each feature.

* Number of Parameters

For RGB,

Class means => $10 * 3072$ Parameters

Class Variances = $10 * 3072$ Parameters

Class priors = 10 parameters.

For Grayscale,

Class means = 1024×10 Parameters

Class Variances = 1024×10 Parameters

Class priors = 10 parameters